Improving information reporting in data-intensive organisations by determining individual data presentation preferences

by

Carel Mauritz Zastron

Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering in Industrial Engineering in the Faculty of Engineering at Stellenbosch University

Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. P.J. Vlok
Co-supervisor: Dr. J.L. Jooste

March 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: ....

Abstract

Improving information reporting in data-intensive organisations by determining individual data presentation preferences

C.M. Zastron

Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MEng (Industrial)
March 2016

Advancements in data capturing capabilities have allowed organisations to collect and store more data than at any other time in corporate history. However, researchers have found that a big proportion of this data is never used as managers often feel overwhelmed by the new volume of information that they need to process. An investigation into the human cognitive processes showed that the difficulty associated with interpreting information is influenced by the format in which the information is conveyed. The strain placed on an individual's cognitive processes can be reduced in two ways. Firstly, by conveying information in a format with which the receiver is familiar. Secondly, by selecting a format which supports the task that needs to be completed using the information. These are often competing factors and traditionally report designers had to decide on one of the two approaches when designing reports. This study proposes an information encoding framework which considers both the preferences of the individuals receiving the information as well as the requirements of the task that needs to be completed using the information.

A comprehensive literature review preceded the development of the framework and serves as the foundation for the steps proposed in the framework. Twenty managers from six different data-intensive industries were consulted to validate the framework. Of the managers consulted, 68% believed that some of the decisions made in their organisations are based on opinions rather than factual data, even when data is available. A mere 53% of the managers reported that they are satisfied with the way information is currently presented to them while all the managers indicated that reports tailored to their individual preferences would be of value to them. Finally, 95% of the managers consulted believed that the proposed information encoding framework could improve data communication in their organisations.

Uittreksel

Verbetering in die rapportering van inligting in data-intensiewe organisasies deur die bepaling van individuele datavoorstellingsvoorkeure

("Improving information reporting in data-intensive organisations by determining individual data presentation preferences")

C.M. Zastron

Departement Bedryfsingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MIng (Bedryfs)

Maart 2016

Verbeterings in datavaslegging laat organisasies toe om nou meer as ooit tevore in die korporatiewe geskiedenis meer data te versamel en te stoor. Navorsers het egter gevind dat 'n groot gedeelte van hierdie data nooit gebruik word nie aangesien bestuurders, deur die nuwe volumes inligting wat hulle moet prosesseer, oorweldig word. 'n Ondersoek na die menslike kognitiewe prosesse het getoon dat die uitdaging waarmee inligting geïnterpreteer word, deur die formaat waarin inligting oorgedra word, beïnvloed word.

Die lading wat op 'n persoon se kognitiewe stelsel geplaas word, kan op twee maniere verminder word. Eerstens, deur inligting in 'n formaat oor te dra waarmee die ontvanger bekend is en tweedens, deur 'n formaat te kies wat die taak ondersteun wat deur die inligting voltooi moet word. Hierdie is egter twee meedingende faktore en die ontwerpers moes tradisioneel op een van die twee besluit toe verslae ontwerp is. Hierdie studie stel 'n inligtingkoderingsraamwerk voor wat beide die voorkeure van die individue asook die vereistes van die taak wat met behulp van die inligting voltooi moet word, oorweeg.

Die ontwikkeling van die raamwerk word voorafgegaan deur 'n omvattende literatuuroorsig en dien as die grondslag vir die stappe in die voorgestelde raamwerk. Twintig bestuurders, uit ses verskillende data-intensiewe nywerhede, is geraadpleeg om die raamwerk te valideer. Agt-en-sestig persent van die bestuurders wat geraadpleeg is, glo dat sommige van die besluite in hulle organisasies op menings, eerder as feitlike data, gebaseer is - selfs wanneer daar feitlike data beskikbaar is. Slegs 53% van die bestuurders het aangedui dat hulle tevrede is met die wyse waarop data tans aan hulle voorgehou word en al die bestuurders was van mening dat verslae wat na hulle persoonlike voorkeure aangepas is, vir hulle van waarde sal wees. Laastens het 95% van die bestuurders wat geraadpleeg is, geglo dat die voorgestelde inligtingkoderingsraamwerk die kommunikasie van data in hulle organisasies kan verbeter.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations:

- Prof. P.J. Vlok for his valuable insight and dedicated guidance during this study.
- Dr. J.L. Jooste for his support during the validation of the proposed solution.
- Anglo American, for their inputs and financial support of the research.
- My family and friends for their support and their ability to take my mind off the work.
- Tania, for her uncompromising support and motivation. I could not have asked for a better companion during this journey. Anything seemed possible with her by my side.
- Our Heavenly Father who has blessed me with the abilities and determination to complete this study.

The Author
September 2015

Dedications

This thesis is dedicated to my parents Willie and Reina,
for their unwavering support and love.
Without them none of this would have been possible.

Contents

List of Figures

List of Tables

Acronyms

ACRG	Asset Care Research Group
AM	Asset Management
BSI	British Standards Institute
CEO	Chief Executive Officer
CPU	Central Processing Unit
DBMS	Database Management System
EAM	Engineering Asset Management
GDP	Gross Domestic Product
GPS	Global Positioning System
HR	Human Resources
HRM	Human Resource Management
HRIS	Human Resources Information System
HS&E	Health, Safety and Environment
IAM	Institute of Asset Management
ICU	Intensive Care Unit
IQR	Interquartile Range
IS	Information System
ISO	International Organisation for Standardisation
IT	Information Technology
JDM	Judgement and Decision Making
NLG	Natural Language Generation
PAM	Physical Asset Management
PAS	Publicly Available Specification
PDCA	Plan-Do-Check-Act
PGM	Platinum Group Metal
PMI	Project Management Institute
PMBOK	Project Management Body of Knowledge
RAM	Random-Access Memory

SMCR	Sender-Message-Channel-Receiver
SME	Subject Matter Expert
SMME	Small, Medium and Micro Enterprise
TMT	Top Management Team
TQM	Total Quality Management

Thesis Outline

Chapter 1: Introduction
Chapter 1 aims to introduce and contextualise the thesis. It discusses the background of the problem that is being investigated and summarises it in a formal problem statement. Research objectives are derived from the problem statement and the boundaries of the research are established. The design of the research and methodology used to explore and subsequently address the objectives of this study are described.

Chapter 2: Literature Review
Chapter 2 provides a comprehensive review of literature pertaining to the problem statement. The chapter commences with an overview of PAM with a particular focus on human assets, leadership, data and communication within the PAM environment. This is followed by an investigation of three popular communication models to identify the most important elements which have to be addressed in the communication process. Next, the design of information is discussed with regards to cognitive load theory and cognitive fit theory. The chapter also investigates different information presentation formats and the tasks that typically need to be completed using the data. Chapter 2 concludes with an investigation of the elements present in graphs and a discussion of popular graphs used to communicate information.

Chapter 3: Proposed Solution
Chapter 3 describes a proposed solution for the data communication problem based on the findings of the literature review in Chapter 2. The proposed solution is presented as a framework which can be used to encode information. The development and required features of the framework are discussed followed by the steps of the framework. Where possible, each step is described in terms of its purpose, theoretical grounding, an example output and the value it adds to the framework.

Chapter 4: Worked Example and Framework Validation
Chapter 4 discusses how the proposed solution was validated with the help of managers from industry. Twenty managers from six different data-intensive

organisations were consulted and tasked with providing input for the framework. The input provided by the managers was used in a worked example to illustrate how the framework would be used in industry. Thereafter, the responses of the managers to a variety of validation questions are discussed. These responses were used to validate the problem statement, the benefits and viability of the framework as well as certain assumptions made by the framework. The chapter concludes with an analysis of the primary preferences of the managers and a reflection on the implications that the selected validation process.

Chapter 5: Closure

Chapter 5 brings the study to a close. The chapter commences with an overview of the research which attempts to summarise and contextualise the findings. The chapter concludes with a discussion of the limitations encountered during the execution of the study and recommendations for future research.

Chapter 1

Introduction

This chapter aims to serve as an introduction to the study by contextualising the research problem and discussing the subsequent research process that was followed. Firstly, a background of data and the difficulties in the communication thereof are discussed. This is followed by the derivation of a formal problem statement, a research question and several research objectives. An overview of the research design and methodology followed in this thesis is then discussed briefly. Figure 1.1 outlines the structure of the thesis and is used as a roadmap when navigating through the document.

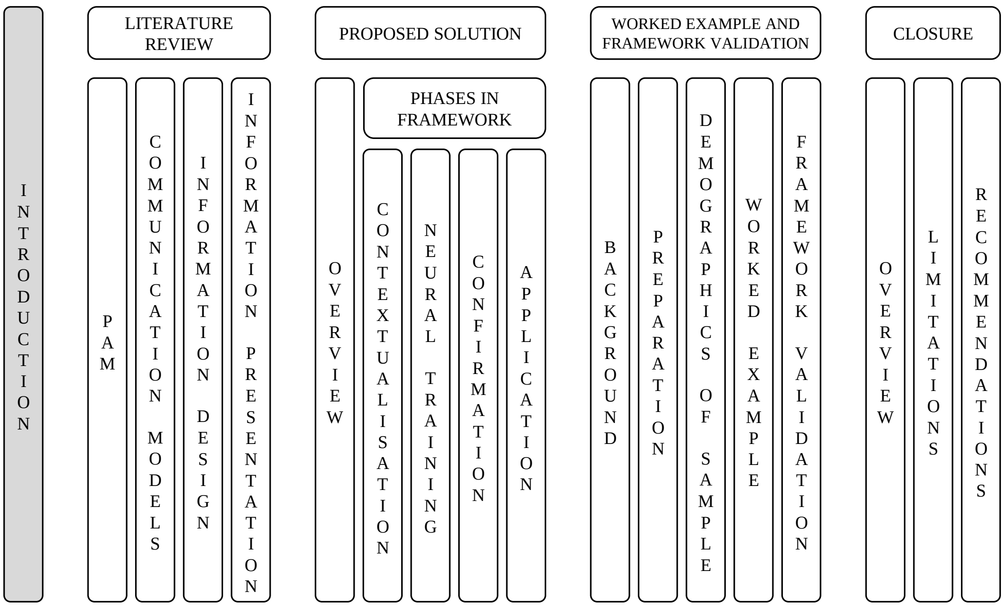


Figure 1.1: Thesis roadmap.

1.1 Background

Mayer-Schönberger and Cukier (2013) set out to determine what America's favourite pie was. When considering the supermarket sales of 30-centimetre frozen pies, apple pie was the outright winner. However, when supermarkets started selling 11-centimetre pies, apple was the fourth best selling pie. Initially the researchers found this odd, but they soon realised there is a logical explanation for it. Since a 30-centimetre pie is normally shared by the whole family, every family member has to agree on a flavour. It turns out that apple is everyone's second favourite flavour. As a 11-centimetre pie can be consumed individually, each member of the family can get their first choice. The revolution in data capturing has allowed supermarkets to see trends and patterns that were not identifiable when smaller amounts of data were available. The researchers argue that more data does not only enable users to see more of the same thing, it allows them to see new and different things. In this case, it allowed the researchers to see that America's favourite pie is not apple. Mayer-Schönberger and Cukier (2013) also noted how information has changed from being stationary and static to fluid and dynamic.

McAfee and Brynjolfsson (2012) explain that the explosion of digital data has allowed managers to measure and learn increasingly more about their businesses. The knowledge gained from all this data can be translated into improved decision making practices leading to better business performance. There is often a direct correlation between the quality of managerial decisions and the speed at which information is made available as well as the ease with which the information can be interpreted (Falschlunger, Lehner, Eisl and Losbichler, 2015).

McAfee and Brynjolfsson (2012) investigated the advantages associated with being data-driven by interviewing the executives of 330 public companies in North America and questioning them about their technology management practices. It was found that not all companies embraced data-driven decision making. However, there was a significant relationship between companies that characterised themselves as being data-driven and their performance on objective measures of both operational and financial results. Most notably, when ranked according to their use of data-driven decision making, the top third of companies in each industry were, on average, 6% more profitable and 5% more productive than their competitors.

The explosion of data is not only applicable to business environments, but also in engineering. Lin, Gao and Koronios (2006) report that embedded sensor systems in engineering assets such as aircraft, ships and processing plants all produce enormous amounts of data. Asset Management (AM) is the collective term used to describe processes related to the management of assets from inception to disposal. As a result of the increased amount of data captured by assets, AM has also become a data-intensive domain. AM is formally defined as the "*coordinated activity of an organisation to realise value from assets*" (BSI, 2014a, p. 14). This value normally refers to a combination of costs, risks, opportunities and performance benefits. Penrose (2008) identified effective and efficient AM as the single largest opportunity for business improvement in the 21st century. This is confirmed by Lin *et al.* (2006), who claim that AM is perceived as an essential business process in organisations and that it has been identified as one of the main contributors to an organisation's financial objectives. Minnaar, Basson and Vlok (2013) state that, in AM environments, organisational management should be provided with information that facilitates objective, predictable and consistent decision making, as well as the short- and long-term consequences of the possible decisions.

Despite the increase in data generation, it appears that managers and other executives are often not comfortable using this data for decision making (Lin *et al.*, 2006). Reasons cited for their discomfort include the correctness, reliability, consistency and the timeliness of the data. Hence, this greater amount of data does not always translate into better information or even into more informed decision making. Al-Hakim (2006) reports that the difficulty associated with the use of these large amounts of data leads to as much as 70% of it never being used. Koronios, Lin and Gao (2005) state that decisions are often made on the basis of judgement rather than factual data as a result of the lack of visibility and control of data in organisations. LeRoy and Snitkin (2004) emphasise that any decision which is not based on quality data is a gamble.

Many of the benefits associated with data-driven decision making are lost as a result of the data not being used. Zhu and Chen (2008) explain that decision makers often feel overloaded by the amount of business information that they need to process. Interestingly, at the same time, the managers also feel that they do not have the information they need when making decisions. More than 30 years ago, Naisbitt (1982) wrote "we are drowning in information but starved for knowledge". In the current data-driven age, however, the statement might read that we are drowning in data but starved for information.

According to Tang (2006), when managers are faced with a variety of possible outcomes for a decision, they tend to use an average of all the outcomes or even the most likely outcome when making decisions. This disregard for factual information about the more serious outcomes might result in decisions that are both inaccurate and overconfident. Kreye, Goh, Newnes and Goodwin (2012) stress that the neglect of factual data in decision making results from inadequate representations of the uncertainty in the information that is supplied to the persons responsible for making the decisions. Informed decisions can thus only be made with clear representations of all input information

(Speier and Morris, 2003; Speier, 2006). Kreye *et al.* (2012) suggest that possible outcomes of a decision and the data on which these outcomes are based are more likely to be considered thoroughly if they are represented adequately.

These statements support the findings by Platts and Hua Tan (2004) who explained that good communication and shared understanding of information are vital in a business environment which is of increasing complexity. Furthermore, Platts and Hua Tan (2004) state that the amount of information that needs to be communicated by managers is also ever increasing. A possible solution is using a diagram to convert an otherwise laborious explanation into a relationship that is both memorable and easy to identify.

Zhu and Chen (2008) mention that a variety of visualisation techniques have been created to aid people in extracting value from large information sets. Information visualisations, according to Jongsawat and Premchaiswadi (2011), can speed up perception, provide users with insight and control, and can also be used to extract valuable information from the flood of data. This, in turn, can result in a competitive advantage when making business decisions.

Graphs are widely considered to be an effective tool for presentations and business communication and the number of methods available to display data have increased drastically with advances in technology (Best, Steward and McGuire, 2008). However, with the increase of alternative data presentation formats available to designers, there are also more opportunities for the data to be displayed incorrectly or for an inappropriate format to be selected.

Cognitive load theory explains that strain is placed on an individual's working memory when interpreting information. This strain placed on an individual's working memory inversely affects task performance. Cognitive load theory provides guidelines about how the retrieval of information and learning can be improved during the execution of any task, including communicating information. The choice of presentation format can reduce the strain placed on an individual's working memory in two ways. Firstly, by encoding information in a format which inherently supports the task that needs to be completed using that information. Secondly, by encoding information in a format with which the individual is familiar. Individuals, however, are often not familiar with the most appropriate format as determined by the characteristics of the task. If the user is not familiar with the format used to convey the information, task performance will be reduced significantly.

It thus follows that designers are often tasked with deciding between encoding information in a format with which the user is familiar and a format which facilitates the task at hand. Nistal, Van Dooren, Clarebout, Elen and Verschaffel (2009) and Anderson, Potter, Matzen, Shepherd, Preston and Silva

(2011) found that the benefits of communicating information in a format with which a user is familiar often outweigh the benefits of using the most appropriate format based on the characteristics of the task. However, selecting a format based purely on either the user's familiarity with it and disregarding the task it has to facilitate, might result in poor task performance.

Nowell, Schulman and Hix (2002) found that, although the information visualisation field of research has made significant progress, studies investigating the effectiveness of visualisations have often reported conflicting or inconclusive results and suggestions. Toker, Conati, Carenini and Haraty (2012) reason that this might be due to the fact that designers focus solely on the target data set and tasks that need to be completed with little regard for the differences of the users. In addition, individual long term factors such as cognitive capabilities and experience as well as short term characteristics such as cognitive load and attention paid by individuals are often overlooked when information visualisations are created. This is particularly concerning since studies such as that of Wang Baldonado, Woodruff and Kuchinsky (2000) have found anecdotal evidence of varying personal visualisation preferences under individuals.

From the above mentioned it is clear that there are many factors to consider when designing visualisations to communicate information to managers and that the wrong choice can be detrimental to task performance. It also suggests reasons why managers might feel uncomfortable with the information presented to them and decide to disregard invaluable data which can be a competitive advantage.

1.2 Problem Statement

Advances in data capturing capabilities have enabled organisations to collect enormous amounts of data which can be converted into valuable information. The explosion of data available to managers, according to Lin *et al.* (2006), is particularly prevalent in engineering environments where automated data capturing capabilities of assets are allowing the collection and storage of more data than at any other period in corporate history.

The complex nature of data and the difficulty associated with the interpretation and communication thereof have resulted in a large proportion of the data not being used. Managers thus rely on their own opinions rather than factual data when making business decisions which could have detrimental effects (LeRoy and Snitkin, 2004; Koronios *et al.*, 2005; Al-Hakim, 2006; Lin *et al.*, 2006).

Information visualisations have been identified as a medium which can be used to communicate large amounts of data to people. However, selecting an information visualisation format in which to communicate numerical information is a complex task. This is because the choice of presentation format has to consider both the requirements of the task and the preferences of the target audience. These are often two competing factors.

The problem is thus that report designers have to consider various competing factors when reporting information and there are no clear guidelines which can be used when selecting presentation formats. Hence, the purpose of this thesis is to determine whether information encoded in a format which considers both the requirements of the task that needs to completed and the preferences of the target audience will be of value to managers in data-intensive organisations. The central empirical research question is shown below.

Is information which has been encoded in a format which considers (a) the characteristics of the task that it has to support; and (b) the preferences of the target audience, of value to managers?

The research question above can be used to formulate the following null hypothesis:

H_0:	Reporting information in a format which considers both the preferences of the target audience and the characteristics of the task that the information needs to support is not of value to managers in data-intensive organisations.

1.3 Aims and Objectives

The main aim of this thesis is to provide guidelines for the selection of presentation formats when communicating data to managers in data-intensive environments. The execution of this thesis will be guided by research objectives derived from the background and problem statement to ensure that the research conducted is aligned to the main aim. These research objectives are:

1. Gain an understanding of the role of data and leadership in AM environments.

2. Establish the fundamental principles of communication processes between two entities as described by previous researchers.

3. Investigate how various elements of visualisations influence an individual's cognitive processes.

4. Determine which tasks are typically supported by data.

5. Explore different formats in which information can be encoded as well as the types of data for which they are appropriate.

6. Develop a solution which recommends a presentation format based on the characteristics of the task to be completed and the preferences of the target audience.

7. Validate the proposed solution with regards to the problem statement.

In order for the proposed solution to be considered as a viable solution in industry, it will be have to comply with certain requirements. Incorporating the following features into the proposed solution will ensure that the requirements derived from the background and the problem statement are met.

- Practical: The proposed solution should serve as a tool which can be used in practice with clear benefits to its users. If the proposed solution is purely based on theory with no consideration for the environment in which it will be applied, it will not be adopted in practice.

- Generic: Although the proposed solution will be designed to address a problem which is prevalent in Physical Asset Management (PAM) environments, it should be implementable across a variety of industries.

- Flexible: In order to address both inter-organisational and intra-organisational differences, the proposed solution should be adaptable to specific situational needs. This requirement will also allow the proposed solution to be tailored to the organisation where it is being implemented.

- Structured: Users of the proposed solution should be able to follow a set sequence of steps which will lead to similar outputs regardless of the user.

1.4 Delimitations

The boundaries of the research have to be established to ensure that the focus remains on the intended purpose and to demarcate the generalisability of results. In this study, the proposed data encoding framework has to remain within the following boundaries:

- This study is concerned with information presentations in AM organisations which are data-intensive. Even though the framework is generic, it was designed to support managers in asset intensive industries where fact based decision making is imperative to the success of organisations.

- This study will only focus on the presentation of data which has already been processed and converted into information. It will thus not include any exploratory data presentation techniques.

- This study will only consider presentation formats which are reproducible with Microsoft Excel™. However, the framework allows for organisation-specific formats to be taken into consideration when applying the framework.

The delimitations mentioned above will be adhered to for the majority of the study, although it might be disregarded to put a certain concept into context. The approached followed during this research is discussed next.

1.5 Research Design and Methodology

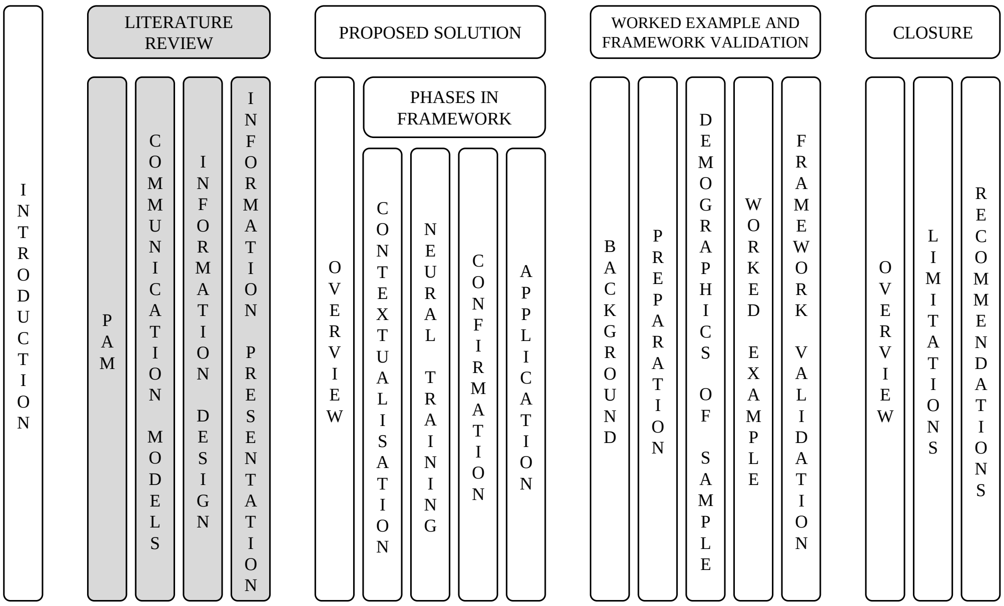
The research design provides an overview of the steps followed while conducting the research. Welman, Kruger and Mitchell (2005) explain that there are two main approaches to research, namely, quantitative and qualitative. According to Welman *et al.* (2005), quantitative research (also known as a positivist approach) focuses on natural-scientific methods and all observations and measurements should be independent of any feelings and opinions of individuals involved in the research. Furthermore, quantitative research aims to derive laws which are applicable to specific populations and requires the objective explanation of any observations and measurements. Conversely, qualitative research (or an anti-positivist approach) states that following strict natural-scientific methods is inappropriate when gathering and analysing data. While quantitative research deals with objective numerical data, the purpose of qualitative research is to evaluate the subjective responses of individuals. By combining the two approaches, the strengths of the two methods can be combined and the weaknesses of each method can be mitigated. This is referred to as a mixed-method approach and is the one used in this thesis.

Key concepts in the problem statement will be identified and investigated in a qualitative literature review. Kumar and Phrommathed (2005) explain that exploring previous research by other individuals in a field is essential to the development of a body of knowledge. A proposed solution to the problem will be conceived from the findings in the literature review. This proposed solution will aim to satisfy the objectives identified in Section 1.3. Thereafter, a sample of individuals from the target audience of the proposed solution will be consulted and asked to both provide input for the proposed solution and to evaluate it. During these consultations, the individuals will be asked to provide qualitative responses to questions in a questionnaire. The responses to the questionnaires will be analysed quantitatively and will be reported using descriptive statistics. Finally, the responses will be used to validate the proposed solution and to make recommendations for future research.

Chapter 2

Literature Review

In this chapter, key concepts that have been introduced in the background and problem statement will be explored further in order to create a body of knowledge from which a proposed solution can be derived. Hart (2001) mentions that the most important reasons for conducting a literature review include: identifying work done and progress made in the same field of research; preventing researchers from duplicating previous research; identifying and avoiding the pitfalls and errors of past research and identifying gaps in the existing body of knowledge. In addition to these reasons, the literature review also aids in placing key concepts into context to ensure a comprehensive understanding. The roadmap shown below indicates the position of the literature review in relation to the other chapters in the thesis.



The research areas explored in this literature review are broad and the relevance of certain sections might not be clear when considered in isolation. Consequently, Figure 2.1 provides a high level overview of the literature explored in this chapter.

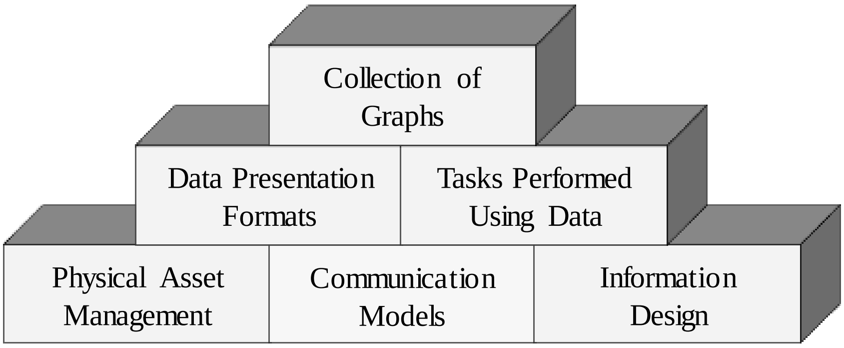


Figure 2.1: Building blocks used in literature review.

The literature review commences with an overview of PAM with a particular focus on human assets, leadership and data in this domain. This is followed by an investigation of three popular communication models to gain a better theoretical understanding of the communication process. Next, the effect that the design of information has on the cognitive processing capabilities (working memory) of individuals is discussed with reference to the cognitive load and cognitive fit theories. Once these concepts have been explored, literature regarding data presentation formats and tasks that typically have to be facilitated by data will be investigated. The chapter concludes with a discussion of popular graphs used to convey information and the appropriateness of each graph to convey different types of data.

2.1 Overview of Physical Asset Management

In an age that is characterised by a high degree of globalisation, companies need to exploit every conceivable advantage in order to stay competitive. As a result of this, there is increasing pressure on industries to decrease costs, meet higher performance standards, comply with regulations, and maximise return on assets (Ouertani, Parlikad and McFarlane, 2008). Ouertani *et al.* (2008) also state that visionary industries are trying to reduce asset maintenance costs, improve asset performance, extend the lives of assets, increase both information and decision making speeds and continuously acquire competitive advantages throughout the asset life cycle. Physical asset management addresses all of these and more aspects in order to ensure that assets are used productively.

PAM cannot be explained comprehensively before an asset is not well defined. Firstly, it is important to note that different professional domains have varying definitions for an asset. The International Accounting Standards Board (2014, p. A33) describes an asset as:

> *"... a resource controlled by the entity as a result of past events and from which future economic benefits are expected to flow to the entity."*

In the software development domain, Swanson and Curry (1989, p. 207) defined an asset to be:

> *"... a qualified entity that, through its reuse, improves quality, provides a competitive edge, and reduces software development and support costs."*

It is clear that the definition of an asset is domain dependent. The International Organisation for Standardisation (ISO) created a document to identify common practices which can be applied to a comprehensive range of assets in an extensive range of industries. The document is titled ISO 55000 and aims to explain Asset Management (AM) and AM systems (BSI, 2014a). It would thus seem reasonable to use the ISO 55000 definition of an asset for the remainder of this document.

> *"An asset is an item, thing or entity that has potential or actual value to an organisation. The value will vary between different organisations and their stakeholders, and can be tangible or intangible, financial or non-financial" (BSI, 2014a, p. 2).*

Now that an asset has been defined, PAM can be explored further. The British Standards Institute (BSI) published the Publicly Available Specification (PAS) on AM titled PAS 55. PAS 55 specifies the industry standard for AM and includes international consensus about good practices in PAM (BSI, 2008a). According to PAS 55, AM is defined as:

> *"Systematic and coordinated activities and practices through which an organisation optimally and sustainably manages its assets and asset systems, their associated performance, risks and expenditures over their life cycles for the purpose of achieving its organisational strategic plan" (BSI, 2008a, p. v).*

The Institute of Asset Management (IAM, 2014) points out that AM is not the management of assets, but rather a grouping of the knowledge and tools needed by an organisation to utilise assets to achieve its purpose. Asset Management has the ability to convert the fundamental aims of an organisation into practical guidelines when choosing, obtaining, operating and maintaining assets to reach those aims. It should be noted that AM attempts to achieve all of this while seeking the best value approach. The best value approach, in this sense, refers to the optimal mixture of cost, risk, performance and sustainability.

Implementing PAM strategies and policies might seem daunting at first, but the expected benefits make PAM implementation a lucrative proposition. According to the BSI (2014a), some of the benefits of AM adoption might include:

- improved financial performance;
- informed asset investment decisions;
- managed risk;
- improved services and outputs;
- demonstrated social responsibility;
- demonstrated compliance;
- enhanced reputation;
- improved organisational sustainability; and
- improved efficiency and effectiveness.

The potential benefits of adopting AM became evident in organisations where it was implemented successfully, but the novelty of an all round AM approach created a need for an AM guideline. PAS 55 and the ISO 55000 suite of documents were published in a response to industry's demand for an AM standard and are subsequently discussed.

2.1.1 PAS and ISO 55000

In 2004, the IAM collaborated with the BSI and, with the assistance of organisations and individuals, developed the PAS 55 document (BSI, 2008a). The main aim of PAS 55 was to assist in the life cycle management of assets with a particular focus on assets that form part of the core business. Asset intensive organisations are dependent on well-defined AM guidelines as an organisation's success is significantly influenced by the performance of its assets.

PAS 55 is divided into two parts, namely: PAS 55-1 and PAS 55-2. The first part, PAS 55-1, provides a set of requirements that define what has to be done, but there is no indication of how it must be done. PAS 55-2, in turn, provides guidelines for the successful application of PAS 55-1 without adding any additional requirements. The BSI (2008a) also stresses that AM requires an integrated approach and that the PAS 55-2 guidelines cannot be partially implemented, but that a holistic approach is required. Ratnayake (2013) explains that PAS 55 can help to reduce uncertainties involved with asset behaviour, future requirements, performance scores, associated risk and costs.

According to Van den Honert, Schoeman and Vlok (2013), PAS 55 follows the Plan-Do-Check-Act (PDCA) cycle. Qing-Ling, Shu-Min, Lian-Liang and Jun-Mo (2008) report that the PDCA cycle forms the foundation for Total Quality Management (TQM). Consequently, the use of the PDCA cycle in the PASS 55 document aids in quality assurance in the PAM environment.

The steps in the PDCA cycle follow a logical order in which a plan is conceived after identifying and analysing a problem. Once this has been completed, a solution is developed and implemented. This is followed by an evaluation of the results achieved and, finally, the results are acted upon (Johnson, 2002). If a new problem is identified in the evaluation, the PDCA cycle is initiated once again. By following these steps, the quality of work is improved in each phase. Note that its cyclical and dynamic nature means that it is a closed loop process which repeats itself after the completion of the last step (Act). Qing-Ling *et al.* (2008) highlighted that the PDCA cycle could be regarded as a summarised version of continuous and spiral improvement. Table 2.1 shows how PAS 55 links up with the steps in the PDCA cycle.

Table 2.1: PDCA cycle applied to AM.

Step	PAS 55 link to PDCA cycle
Plan	Establish the AM strategy, objectives and plans necessary to deliver results in accordance with the organisation's AM policy and the organisational strategic plan.
Do	Establish the enablers for implementing AM (e.g. asset information management system(s)) and other necessary requirements (e.g. legal requirements) and implement the AM plan(s).
Check	Monitor and measure results against AM policy, strategy objectives, legal and other requirements; record and report the results.
Act	Take actions to ensure that the AM objectives are achieved and to continually improve the AM system and AM performance.

Adapted from BSI (2008a).

Woodhouse (2014) mentions that the success of PAS 55's adoption around the world is evident as it has been translated into Spanish, French, Chinese, Russian and Portuguese. In 2009, consultation with professional bodies as well as with industry led to the submission of PAS 55 to the International Standards Organisation as the basis of a new ISO standard for AM. The standard was approved and resulted in the creation of the ISO 55000 suite of standards. ISO 55000 was developed over three years with the help of 31 participating countries and was published in January 2014. Van den Honert *et al.* (2013) and Woodhouse (2014) explain that ISO 55000 is made up of three parts, namely:

- ISO 55000 Asset management – Overview, principles and terminology
- ISO 55001 Asset management – Management systems - Requirements
- ISO 55002 Asset management – Management systems - Guidelines for the application of ISO 55001

When comparing ISO 55000 to PAS 55, it can be seen that PAS 55-1 has been split into ISO 55000 and ISO 55001. The requirements specification is contained in ISO 55001 while the important terms and definitions are explained in ISO 55000. As with PAS 55-2, ISO 55002 provides more information on the interpretation and implementation of ISO 55001. Woodhouse (2014) explains that the most significant difference between the two standards is the targeted application scope. PAS 55 focuses explicitly on physical assets with slight reference to and acknowledgement of other asset types. ISO 55001 is designed to address and apply to all types of assets even though the main focus is on the management of physical assets. The chief benefit of ISO 55000 is that it encapsulates a wide variety of industries that make use of assets within varying AM contexts.

The generalised nature of the ISO 55000 suite of documents (relative to PAS 55) has also resulted in more generic terms and definitions. For instance, ISO 55000 has reduced the lengthy description of AM in PAS 55 to:

> *"coordinated activity of an organisation to realise value from assets" (BSI, 2014a, p. 14).*

Van den Honert *et al.* (2013) mention that the ISO 55000 suite of documents provide a more comprehensive overview of AM than PAS 55 which results in a more conclusive document. PAS 55 provides a guideline in terms of what needs to be done, but there is no indication of what the minimum criteria is. In contrast to this, ISO 55001 specifies the minimum criteria that needs to be met for suggested activities while ISO 55002 offers guidelines for the interpretation an execution of those activities.

Woodhouse (2014) emphasises that the rewards associated with the successful implementation of AM cannot be ignored. The decision of ISO to recognise what needs to be done and to define the requirements of management systems to coordinate and sustain good practices was timely and that AM is considered to be a necessity in industry. ISO 55000 provides a list of generic 'must do' items that can be used by a wide spectrum of industries to ensure good AM. However, for the sake of completeness and to ensure that AM is described clearly, this thesis will refer to both ISO 55000 and PAS 55.

PAS 55 recognised that there are five broad categories of assets of which physical assets is only one. The others are human assets, information assets, financial assets and intangible assets. The BSI (2008a) points out that all of these assets have to be managed holistically in order to create maximum value. Figure 2.2 shows how the different asset categories are related and the focus of PAS 55. Although PAS 55 focuses mainly on physical assets and asset systems (assets that are interrelated), it acknowledges that all of the asset types are inextricably linked. The BSI (2008a) also emphasises that, although human factors such as leadership, motivation and culture are not explicitly handled in PAS 55, they have a significant effect on the success of an AM system and thus require due consideration. These human factors influence the owners, managers, employees, contractors and suppliers of the organisation.

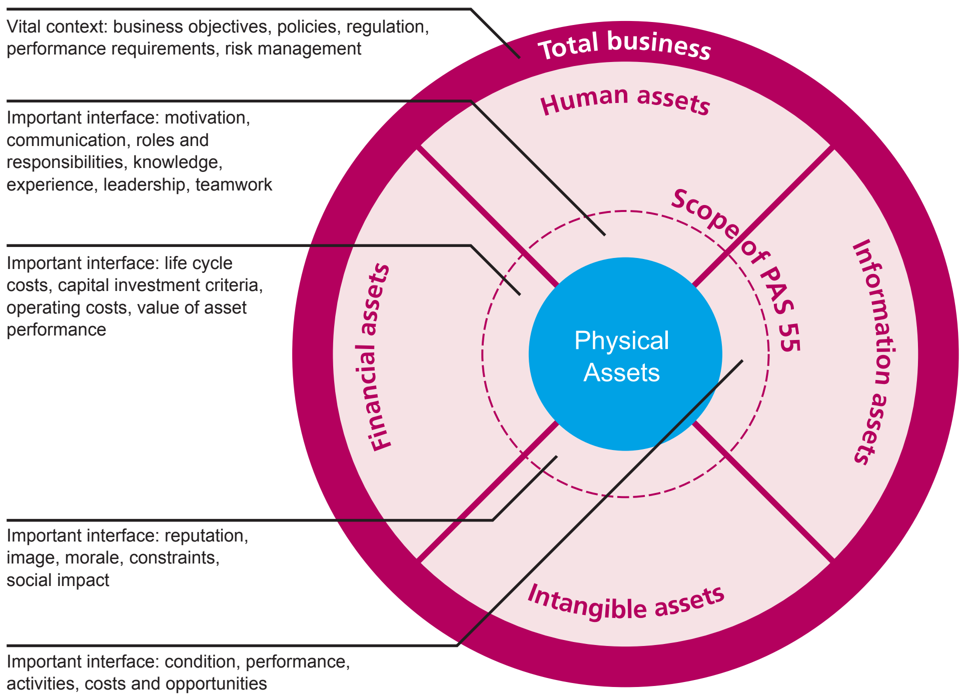


Figure 2.2: PAS 55 focus and business context in relation to other assets.

Adopted from BSI (2008a).

This leads to an examination of the foundation on which AM is built. The BSI (2014a) explains that there are four fundamental aspects on which AM is based, namely:

- **Value:** ISO 55000 emphasises that assets are not the main focus of AM, but rather the potential value that can be provided to the organisation by those assets. The value, whether tangible or intangible, is calculated according to the organisational objectives.
- **Alignment:** Effective AM enables an organisation to translate its objectives into technical and financial decisions, plans and activities. These decisions in turn, collectively support the achievement of organisational objectives.
- **Leadership:** Realisation of value depends on leadership and workplace culture. Asset Management can only be established, operated and improved within the organisation if all of the managerial levels are committed and provide leadership.
- **Assurance:** By committing to AM, organisations are assured that assets will serve the purposes that they were set out to do. Effective governance of an organisation requires assurance. Assets, AM and AM systems all require assurance.

As Figure 2.2 illustrates, human assets are one of the five assets involved in AM. Ultimately an organisation is run and managed by people and effective functioning of management is essential for an organisation to prosper. This thesis aims to improve data communication to managers and thus it is important to gain a better understanding of the different roles of people in AM environments. The next section discusses typical departments which can be found in AM organisations.

2.1.2 Human Assets within AM

With human assets being one of the five assets addressed in AM, there is an expectancy that the departments to which these human assets belong will be well researched. However, this is not the case. There is very little research available describing the different departments that can be encountered in an AM environment. A possible reason for this might be that organisations do not necessarily plan their structures around AM, but rather apply AM using their existing organisational structures. Nonetheless, Hastings (2009) found that there are certain key personnel roles within AM and they are listed below. Hastings (2009) also acknowledged that the assignment of employees to these roles depends on the size and structure of the organisation and that the positions might not require full time appointments.

- **Asset Group Manager:** Asset group managers are responsible for a group of related assets and are in charge of various asset managers. Hastings (2015) explains that AM groups (headed by an asset group manager) are stationed strategically to oversee major equipment areas in an organisation.

- **Asset Manager:** Asset managers are responsible for a distinct range of assets. According to Hastings (2009), substations are seen as asset groups in an electricity transmission organisation, but within these groups there are asset managers responsible for switchgear, transformers and secondary systems. Naturally, asset managers are expected to be competent and have an intricate knowledge of the relevant technologies and their operational context. An asset group receives guidance and leadership from asset managers and are likely to consult asset managers when considering business development in the asset manager's field of expertise. Although asset managers are likely to stem from engineering or logistic backgrounds, they need to grasp the wider business context and have a basic understanding of financial and accounting concepts and principles. Even though the ultimate detail of accounting and finances remain a specialised domain, countless decisions that are made in organisations are based on forecasts drawn from both technical and business knowledge. Hastings (2009) emphasises that asset managers play a pivotal role in providing top management with timely and accurate information on which business decisions can be based. A strong AM presence in a company will likely lead to better performance than companies without it and the asset manager plays a major role in linking technical and financial elements in an organisation.

- **Project Manager:** Project managers are involved with acquisition and development projects which form an integral part of the AM function. These project managers are required to have a high level of skill and training as well as extensive experience in the project management field. Organisations such as the Project Management Institute (PMI) and the Project Management Body of Knowledge (PMBOK) typically set out certain requirements that project managers have to fulfil in order to get certified. This ensures that employees in project manager positions are of a high standard.

- **Finance, Accounting and Costing:** These departments are often responsible for assessing costs and investigating the financial viability of new projects while managing the finances of projects that are in motion.

- **Legal department:** Hastings (2009) merely refers to the legal departments as a combination of contracting and contract managers. However, Amadi-Echendu and Amadi-Echendu (2015) recently highlighted the importance of the legal department in a PAM environment. Amadi-Echendu and Amadi-Echendu (2015) argue that the legal department is involved extensively in the processes related to the acquisition and development of land on which an organisation plans to initiate or expand its operations. The registration of an engineering asset such as an airport or processing plant on the newly acquired land also requires the input of the legal department. Finally, in an age where legal disputes regarding the sustainability and environmental impact of new business endeavours have become imperative to the feasibility of new operations, the legal department has become indispensable.

- **Engineers:** Decisions regarding an organisation's assets are frequently based on the technical knowledge provided by engineers. Engineers have technical knowledge regarding the design and operation of assets. Once the assets have been commissioned, engineers perform data analysis to calculate reliability metrics and subsequent maintenance and replacement strategies. Organisations may also use engineers as Subject Matter Experts (SMEs) on projects.

- **Logisticians:** AM is a very popular domain for logistics experts. Various logistic techniques such as logistic support analysis and level of repair analysis can be applied to AM. Hastings (2009) states that the role of logisticians can be extended to include other provinces such as configuration management, cataloguing of spare parts, identification and coding of maintainable items of equipment, determining and implementing inventory control parameters and ensuring that the systems used to distribute assets, consumables and spare parts meet the requirements.

- **General staff:** Other departments such as Human Resource Management (HRM) and administrative departments are also involved in AM, albeit more passively.

Hastings (2009) emphasises that the AM function needs personnel with business, technical, operations or service experience who will be able to integrate with finance, contract and engineering specialist. As mentioned before, these roles are not rigid and might differ from organisation to organisation. The same roles might also have varying titles in different organisations. However, it is important to note that AM affects employees from a wide spectrum of departments in organisations and that the working environments and employee characteristics in each department are often unique.

2.1.3 Leadership within AM

In Section 2.1.1 leadership was identified as one of the four fundamentals on which AM is based. The BSI (2014a) acknowledges that Top Management

Teams (TMTs) have various responsibilities that include, but are not limited to, the development of an AM policy as well as the creation of AM objectives and ensuring that they are aligned with the organisational objectives. TMTs are expected to use their authority to promote AM in an organisation. Consequently the responsibilities, accountabilities and AM objectives within an AM system are defined by top management. Alignment of the AM system with other management systems and functions within an organisation also form part of the TMT's function.

The IAM (2014) found that AM decisions are frequently limited by factors such as budgetary constraints, resources and/or regulations. It is the responsibility of all the management levels in an organisation to ensure that sufficient resources are made available to support the AM system. These resources include adequate and proficient human resources, appropriate funding and sufficient Information Technology (IT) systems.

The BSI (2014a) states that since TMTs are expected to align AM objectives with that of the business, conflict that might arise between the internal organisational culture and the performance of its AM system should be resolved by the TMT.

As mentioned in Section 2.1.1, ISO 55001 outlines the requirements of AM systems (BSI, 2014b). ISO 55001 states that TMTs shall exhibit leadership and commitment to the organisation's AM system by:

- ensuring that the AM policy, the strategic AM plan and AM objectives are created and that they can be integrated with the organisational objectives;
- ensuring that the AM system requirements are integrated into the business processes of the organisation;
- ensuring that the AM system receives adequate resources;
- ensuring that the significance of effective AM as well as the importance of conforming to the AM system requirements are communicated in the organisation;
- ensuring that the outcomes that were set for the AM system are achieved;
- motivating and guiding persons to ensure that the AM system is effective;
- encouraging the formation and collaboration of cross-functional teams within the organisation;
- creating a culture that is dedicated to continual improvement;

- accepting the responsibility of supporting other relevant manager roles;
- ensuring that there is alignment between the risk management approaches of AM and that of the organisation.

ISO 55002 (the guideline for implementing ISO 55001), emphasises that although TMTs may appoint other individuals to oversee the design, execution, operation and continual improvement of AM systems, the top management level should take ultimate ownership and accountability for AM. Top management commitment to AM can be demonstrated by:

a) ensuring that communication within the organisation reference AM principles;

b) participating in the setting of AM system objectives which can be used to measure the success of people responsible for it:
   - by prioritising the relevant objectives;
   - by ensuring that adequate resources are provided to achieve the objectives;

c) creating and maintaining a work culture that utilises collaboration to ensure that AM objectives are met;

d) seeing to it that decisions such as capital expenditures are made using a criteria that acknowledges AM principles;

e) providing support for improvement activities that are related to AM;

f) encouraging the creation and maintenance of a management-development strategy which encourages and rewards employees who spend time in positions that support AM and the operation of the AM system;

g) monitoring the performance of the AM system and adopting a continual improvement approach as well as performing corrective and preventative actions;

h) taking responsibility to ensure that AM is regarded on the same level of importance as affairs related to safety, quality, environment, etc.;

i) ensuring that risks pertaining to assets are incorporated into the organisation's risk management processes and that they are addressed appropriately;

j) seeing to it that AM and the AM system are aligned to other functions in the organisation to promote the achievement of organisational goals;

k) seeing to it that AM and the AM system are aligned to other organisational practices and management systems such as the approach taken by the organisation to manage risk.

It is evident from ISO 55001 and ISO 55002 that the responsibility of incorporating AM into organisational decisions lies primary on TMTs. Recall that, in addition to leadership, alignment is also one of the four fundamental aspects on which AM is built (Section 2.1.1). This means that all of the decisions, plans and activities should be aligned with the AM objectives and will require various parts of the organisation to collaborate (BSI, 2014a).

The AM objectives will often include the collaborative coordination, application and verification of resources as well as improving their use. Furthermore, BSI (2014a) mentions that both the development of AM plans as well as the evaluation of their effectiveness are aided by the information produced or provided by the AM system. However, as ISO 55000 explains, the Information System (IS) responsible for the collection, verification and consolidation of data related to assets is often very large and complex. This means that converting the captured data into usable information can be a complicated task. The next section provides a brief overview of data in AM environments.

2.1.4 Data within AM

Lin *et al.* (2006) report that organisations are currently generating and capturing more data than ever before in corporate history. Neely, Lin, Gao and Koronios (2006) found that in AM environments, and particularly in PAM environments, an enormous amount of data is produced during an asset's life cycle. The terms data and information are often used interchangeably in everyday terminology and thus some clarification is needed. Brous, Overtoom, Herder, Versluis and Janssen (2014, p. 126) explain the difference:

> *"Data are facts about objects, subjects or events within or without the organization. These facts generally involve the condition of the object or subject or refer to a transaction involving that object or subject. Data only becomes information once it is given context and presented in a form that people are able to understand."*

According to Neely *et al.* (2006), the data produced by engineering assets are one of two types, namely: configuration data and transaction data. Configuration data pertains to the physical attributes of the assets such as the date of acquisition, the initial cost, the value at year end and the physical location of the particular asset. Transaction data, on the other hand, is generated and collected while the asset is being operated. Transaction data can either be recorded manually by technicians during routine maintenance checks or it can be produced by sensors embedded in the assets to track when maintenance is necessary and when it is completed. Transaction data can also refer to the output of an asset such as the amount processed or the variability in output. Both configuration and transaction data can be used to support management decisions in a variety of industries. Gao, Lin and Koronios (2006) mention that a few common sources of data in PAM are:

- Inventory data: Information pertaining to the articles, goods and property owned by an organisation.
- Asset condition data: The degree to which equipment has deteriorated.
- Asset and organisational performance data: Historical information describing the availability and reliability levels of assets and the organisation or information comparing the outputs of different assets and systems in an organisation.
- Criticality data: Data about systems or subsystems which are of the greatest importance and which should be the highest priority of the organisation.
- Life cycle data: Information describing a series of changes in the life of an object (often an asset).
- Valuation data: An estimate of the value of an object an organisation owns or plans to acquire.
- Financial data: Information outlining the financial health of a business such as profits, revenues and operating income.
- Risk data: An indication of the uncertainty associated with decisions and assets.
- Reliability data: Probability that a system or an element in a system will perform its intended purpose for a set period of time.
- Technical data: Engineering or scientific information relating to the planning, development, production, creation, operation and maintenance of systems or subsystems in an organisation.
- GPS data: Information showing current and historical physical positions or locations of objects.

McAfee and Brynjolfsson (2012), however, state that even though the potential benefits of having all these data are appealing, there are very real technical and managerial challenges to using it. McAfee and Brynjolfsson (2012) explain that data does not always have the expected impact since the people making the decisions may be more interested in using their intuition than the facts provided by data. In cases where data are scarce and difficult to obtain or use, it is reasonable to allow well-placed and experienced people to make decisions. However, McAfee and Brynjolfsson (2012) express their concern about the number of organisations that rely on the opinion of the most experienced and highest placed person when making important decisions, regardless of whether there is factual data available. They conclude by saying that, while some senior executives base the majority of their decisions on factual data rather than their own intuition, most executives in the business world allow their experience and intuition to override the factual data when making decisions.

It is thus the responsibility of TMTs in AM environments to ensure that the data which is available to organisations is used to support decision making across all departments. Zaccaro, Rittman and Marks (2002) found that the biggest contributor to the success of organisational teams is the effectiveness of the leadership team. Recall that in Section 2.1.3, it was mentioned that collaboration is imperative to the successful implementation of AM at an organisation. This means that the data which is available to support decision making should be converted to information and shared among different departments as well as management levels in the organisation. ISO 55000 emphasises that communication in AM should be two-way, with leaders being able to initiate communication with subordinates as well as being open to receive information from other levels of management regarding the improvement of AM systems (BSI, 2014a). The role of communication in AM environments is discussed in the next section.

2.1.5 Communication within AM

Quirke (2007) found that, in an age where organisations are driven by a quest for value, sustainable advantage cannot depend on products, brands and services alone. Companies are required to use all assets, business processes and relationships at their disposal to stay competitive. Quirke (2007, p. 17) summarises the value of good internal communication practices as follows:

> *"In our information age, an organisation's assets include the knowledge and interrelationships of its people. Businesses take the input of information, and use the creative and intellectual assets of their people to produce value. Internal communication is one of the core processes by which business can create value."*

The need for effective communication in PAM leadership is underlined by the findings of Barrett (2006). An analysis of Chief Executive Officers (CEOs) and other executives across industries showed that good communication skills were always among the most important skills a manager must possess. Mintzberg (1973) and Ecles, Nohria and Berkley (1992) investigated the amount of time that managers spent on a variety of duties and found that communication accounted for 70 to 90 percent of their time. According to Barrett (2006), the introduction of cell phones, text messages and e-mails would have most likely caused the proportion of time spent on communication to be higher than concluded in the previous two studies. The time spent by managers on communication related activities illustrates the importance of good communication practices. However, top-down communication is not the only communication that takes place in organisations.

Bauer and Erdogan (2012) explain that there are many different directions in which communication occurs in organisations. The most common direction, according to Bauer and Erdogan (2012), is the lateral communication between two co-workers in the same department. However, information can also be communicated diagonally between different departments or between managers and their subordinates with either one initiating the communication process. Figure 2.3 illustrates some of the general communication directions.

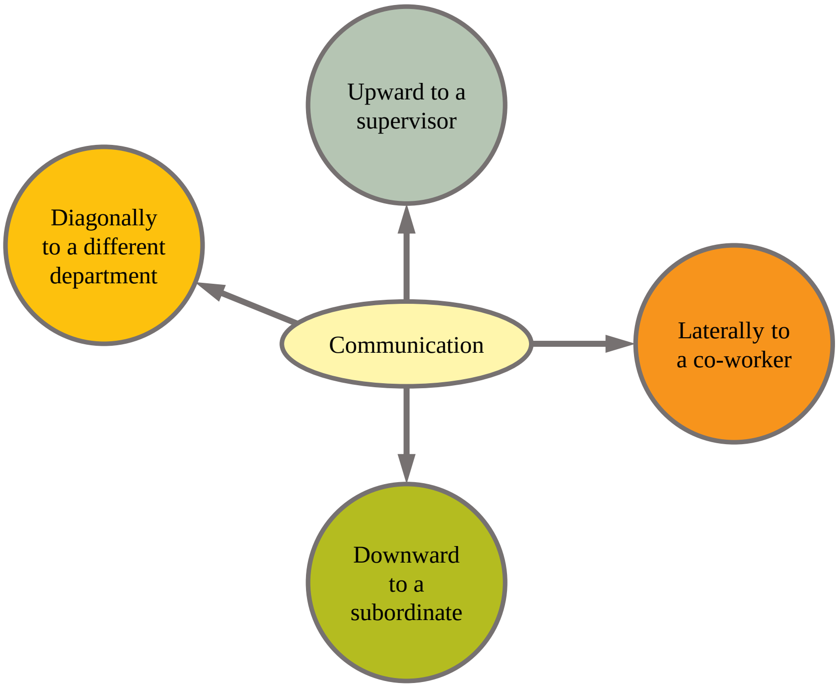


Figure 2.3: Different directions of organisational communication.

Adapted from Bauer and Erdogan (2012).

ISO 55002 states that stakeholders should be informed of AM activities that are executed by the organisation (BSI, 2014c). Furthermore, it is stated that communication with stakeholders should be carried out in a coordinated way and forms an important part of both the AM activity and the AM system of the organisation. In order to achieve this, ISO 55002 recommends that a communication plan should be created. One of the main reasons for this plan is to promote transparency and to ensure that accountability for the AM system is created by promoting active engagement with stakeholders. In other words, by sharing information across departments and stakeholders, the reason(s) for a business decision can be traced back to the person(s) responsible and the information on which the decision was based.

ISO 55002 also provides recommendations on what should be included in the communication plan. These recommendations include, but are not limited to:

- Deciding what external and/or internal knowledge should be communicated to stakeholders to ensure that they provide informed contributions or decisions, or that they give informed feedback.
- Identifying the most suitable person(s) to deliver the particular communications.
- Deciding on the most appropriate format to use for the communications.
- Establishing official processes for reporting and feedback.

In other words, the communication plan should help in establishing a formalised reporting (and feedback) process. In this formalised process an employee responsible for initiating the communication should be identified and nominated. This employee is then tasked with deciding on the internal and external knowledge which needs to be conveyed in the communications as well as the type of communication which is most appropriate for the specific message. The employee responsible for the communication process thus has a variety of decisions to make regarding the manner in which stakeholders will be informed of AM activities.

Bauer and Erdogan (2012) state that there are three types of communication, namely: verbal, written and non-verbal. Bauer and Erdogan (2012) suggest that, of these three types of communication, written communication is the most appropriate type for conveying factual information and that it can either be printed or displayed digitally. Importantly, written communication is often constructed over a long period of time and a single communication effort can contain contributions from various people. The target audience can also be a single person or a combination of people (Bauer and Erdogan, 2012). Moreover, there are a variety of data communication modalities in which the person initiating the communication can encode information such as text, tables and graphs (Prasad and Ojha, 2012). However, deciding on an appropriate format in which to communicate information is often not as simple in practice as in theory.

Kress (2004) explains that the designer of communication is faced with the task of considering the needs of a variety of stakeholders. He further indicates that the preferences and needs of the target audience as well as those of the designer will influence the ultimate choice in communication format, medium and tone of the message.

Graphs are a very common format in which data is communicated, but Toker *et al.* (2012) reveal that existing data visualisations are created around the task that needs to be completed and/or the characteristics of the datasets, while the differences among the users of the visualisations are rarely considered. These differences include both short term factors (such as the attention of the user and the cognitive load of the message) and long term characteristics (such as the user's cognitive abilities, experience and expertise). Furthermore, Conati, Carenini, Hoque, Steichen and Toker (2014) report that the effectiveness of visualisations are greatly affected by user characteristics and they suggest that visualisations can be altered to fit the needs of the user(s). This is in agreement with the findings of Prasad and Ojha (2012) who concluded that a user's comprehension of a communication effort may be influenced by a variety of factors such as the preferences and prior knowledge of the user, the situational context and his or her environment.

From this section as well as the preceding three sections, it is clear that the data generation and capturing capabilities of asset intensive industries have improved significantly in recent times. This data can often be converted into information on which business decisions can be based. Furthermore, the value of communicating this information as well as the different directions in which it can be communicated have been highlighted. The importance of well established communication principles in AM environments is evident in its inclusion in the ISO 55000 suite of documents. However, the difficulties associated with designing effective communications have been underlined as the needs of different stakeholders need to be considered. The way in which the needs of different stakeholders might vary is discussed in Section 2.1.6.

2.1.6 The Effect of Personality on Information Preferences

The demand on the capacity of humans to handle information has grown during the last decade. Becoming an information literate life-long learner is encouraged to keep up with the requirements of a fast paced society. This is something that can be learnt, but only to a certain extent. The degree of difficulty of adapting to these changing demands varies among people according to their inherent ability to meet the requirements. It should be noted that the ideal information literate citizen cannot be linked to one personality type as different traits are more suitable to different situations. Different traits have varying effects on information understanding and once the impact of each is discovered, information can be presented in a more effective way (Heinström, 2003). This is particularly important in AM environments when considering the number of different roles and departments present as described by Hastings (2009) in Section 2.1.2.

Heinström (2003) investigated the personality traits and information habits of 350 university students using a quantitative analysis. This research built on the findings of Solomon (2002). Solomon (2002) found that, although there are cases where a gap in knowledge triggers a conscious search for information, the information-seeking process is often changeable and dynamic. This information-seeking process is largely dependent on the context and also the individual performing it. The way in which information is gathered can range from planned and structured to spontaneous and flexible, depending on the individual's inner processes and needs (Heinström, 2003).

Allen and Kim (2001) found that an individual's quest for information is likely to evolve as the interaction between the context and the individual's characteristics progresses. According to Wilson (2000), variability in information seeking patterns can be explained briefly once psychological characteristics are understood. Phares (1991) stated that personality is a psychological mechanism which influences behaviour. A stable combination of personality traits are responsible for an individual's feelings, thoughts and behaviours. Heinström (2003) explored the idea that, since personality is often responsible for various characteristic reactions, personality traits could be linked to the way individuals conduct themselves when seeking information.

However, personality profiles of employees need to be determined with a practising psychologist present and the profiles are normally confidential. For these reasons, the personalities of employees will not be investigated any further, although their information requirements remain pertinent to this research. In the next section, communication models are investigated to gain a better theoretical understanding of the communication process and to identify possible opportunities for improvement in current communication practices.

2.2 Communication Models

Communication has been studied for many years and a variety of theories and models have been developed to improve the understanding thereof. 300 years

BC, Aristotle made a remark that would influence and be supported by many of the theoretical models to follow:

> *"For of the three elements in speech-making - speaker, subject, and person addressed - it is the last one, the hearer, that determines the speech's end and object."*

Aristotle's quote summarised the problem with communication: The effectiveness of communication is ultimately dependent on the listener/receiver's understanding of the initial message. Bowman and Targowski (1987) noted that the central role in human activity is occupied by language and communication and that it enables the analysis and synthesis which lead to wisdom and knowledge. However, after thousands of years, the communication process between two people is yet to be fully understood. Many communication models have been constructed in an attempt to understand the underlying elements in communication. Once these elements are understood, the sources of disruption in the messages that are sent can be identified and eliminated or at least acknowledged and controlled for. Three of the most explored models are discussed below.

2.2.1 Shannon and Weaver's Model of Communication

The model developed by Shannon and Weaver (1949) is widely considered to be part of the foundation on which the study of communication is built (Fiske, 1990). During the Second World War, Shannon and Weaver researched the channels of communication and the most efficient ways in which they could be used. The main aim of their research was to find the most efficient way to transfer information over telephone cables and radio waves. Shannon and Weaver created the model shown in Figure 2.4. The simplicity of their model attracted many derivatives, but its linearity and process centred nature has been widely criticised (Fiske, 1990). However, as mentioned before, its simplicity has contributed to its wide adoption and thus a discussion of the elements which make up the model is warranted.

The *information source* initiates the communication and decides on the message that is sent. It can thus also be seen as a decision maker as it can decide to send a newly created message or a message from a predetermined list. Once a message is chosen, the *transmitter* transforms it into a signal which is sent through the *channel* and finally the message gets delivered to the *receiver*. The *destination* represents the person who the message was intended for. In a normal conversation, the person initiating the conversation is the information source, his or her mouth is the transmitter, the sound waves represent the signal, the air through which the sound waves travel is the channel and the person listening and his ear are the destination and the receiver respectively

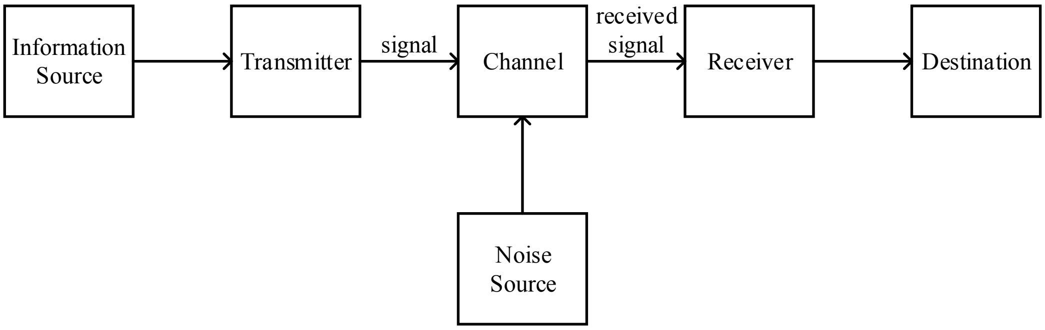


Figure 2.4: Shannon and Weaver's model of communication.

Adapted from Fiske (1990).

(Fiske, 1990). Before the noise element is investigated, the three levels of problems in the study of communication as identified by Shannon and Weaver (1963) need to be explained. These three levels are:

Level A (technical problems)	How accurately can the symbols of communication be transmitted?
Level B (semantic problems)	How precisely do the transmitted symbols convey the desired meaning?
Level C (effectiveness problems)	How effectively does the received meaning affect conduct in the desired way?

Out of all of the problems, those on Level A are the simplest to understand. It refers to the accuracy with which the actual message created by the information source arrives at the destination. The model was initially created to address these problems.

Semantic problems (Level B) are relatively easy to identify, but solving them is often a difficult task. These problems range from the different meanings of words to the varying perceptions that different people have about the same thing, for example: a 'large' piece of machinery might have a different meaning in a mining environment than in a textile factory. Shannon and Weaver were of the opinion that the meaning was contained in the message and that a better encoding effort would improve the accuracy of the semantics. Importantly, the cultures of the encoder and the decoder of the message also play an equally large role in the meaning of the message as the message itself. In other words, the degree to which the meaning of a message is conveyed depends on the listener's (destination's) understanding of the symbols used by the encoder to convey the message.

The last level of problems (Level C) is that of the effectiveness of the message. Simply put, the effectiveness of a message can be measured by the degree by which entity A has influenced entity B's response (Fiske, 1990).

Now that the three levels of problems have been defined, the last element in Figure 2.4, namely *noise*, can be discussed. Different definitions have been constructed to describe noise, but Fiske (1990) defines noise as anything that has been added to the signal between the transmission and receiving processes that was not intended by the source. Level A noises are those that appear in the channel that has been used. These Level A noises include sound that has been distorted or crackling in a telephone wire, radio signals that experience static or even random flickering dots on a television screen. However, the description of noise has been extended to include, apart from anything that was not originally transmitted by the source, anything that makes the signal originally transmitted by the source harder to interpret or decode.

An environment that hinders the successful transfer of a message from person A to person B can thus also be classified as noise. Fiske (1990) mentions that an example of such an environment might be an uncomfortable chair that distracts person B during a conversation and thus makes him or her less susceptible to the message that is being conveyed by person A.

Shannon and Weaver realised that the Level B problems could not be addressed by reducing the Level A types of noise mentioned above. A distinction was made between engineering noise (Level A) and semantic noise (Level B). Shannon and Weaver suggested that another box labelled 'semantic receiver' should be inserted between the engineering receiver and the destination. Semantic noise is described as any distortion in meaning that occurs in the transfer of information between the source and the destination. Any ambiguity can thus be classified as semantic noise as it distorts the message even if there was no engineering noise. The effect of any type of noise is thus detrimental to the communication process as it limits the amount of information that can be conveyed. It is important to note that noise can originate in the channel, the audience, the sender or even the message itself (Fiske, 1990). However, as Flensburg (2009) noted, Shannon and Weaver's model only addresses the issue of how information (the message itself) is transferred and not the content that is transferred by in the process. Shannon and Weaver's model also does not take any of the characteristics of the sender or the receiver into account. Schramm (1954) developed a model to address this problem and it is discussed in the following section.

2.2.2 Schramm's Model of Communication

Schramm (1954) adapted Shannon and Weaver's model to better address semantic noise. Schramm (1954) realised that, in order for effective communication to take place, overlapping fields of experience are required. If there is no overlapping in the fields of experience of the sender and the receiver, communication is difficult or even impossible (Lee, 1993). Schramm's model is shown in Figure 2.5. This model contains the same parts as the model developed by Shannon and Weaver, but some of the components have been renamed to apply more specifically to the communication process between two people. The *destination* component has been renamed to *receiver* while the *receiver* in Shannon and Weaver's model has been renamed to *decoder*. Finally, the *transmitter* has been renamed to *encoder*. Three new elements have also been added, namely: both the sender's and the receiver's fields of experience and a feedback loop. According to Schramm (1954), effective communication can only take place if the message falls within the area that is covered by both the sender's field of experience and the receiver's field of experience.

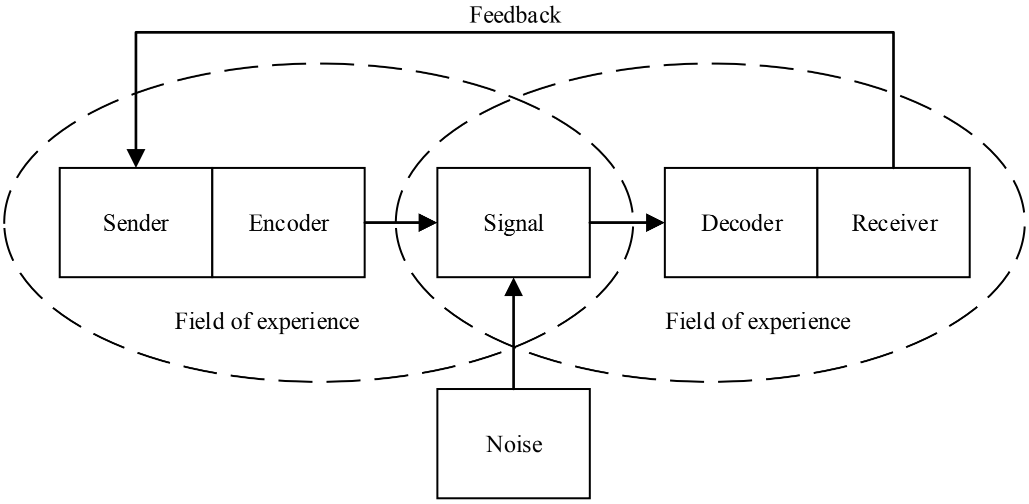


Figure 2.5: Schramm's model of communication.

Adapted from Pettersson (1989).

Pettersson (1989) argued that people do not interpret things they read, hear or see in the same way and thus the information that is derived differs according to each person's unique characteristics. A very important point raised by Pettersson (1989) is that the meaning of language symbols are not inherently understood by people, but the meaning has to be assigned to symbols by people themselves. It would thus be naive to assume that the meanings of codes and symbols are rigid throughout different societies, cultures and backgrounds.

Lee (1993) explains this concept by referring to the interaction between extensions service agents (agricultural teachers) and the American middle class. For many years cooperative extension service agents developed ways in which to communicate effectively with the American middle class. The main reason for this effective communication was the fact that many of the extension workers came from the same middle class which they were now helping and thus the overlap in the fields of experience between the middle class audience and the workers was relatively large. In the 1960s, a time in which social awareness became increasingly important, many extension workers were tasked to work with a 'disadvantaged audience'. The small overlap in the fields of experience made it difficult for the middle class extension workers to communicate with the disadvantaged receivers. This communication difficulty was reduced by employing and training individuals from the targeted disadvantaged audience to provide the important communications linkage.

Another impediment to effective communication is the tendency of people who design the message producing material for their peers and not necessarily the people who require the information. Pettersson (1989) attributes this to people being uninformed of all the communication possibilities. Pettersson (1989) mentions that, while producing information, a sender utilises representations of parts of reality which a receiver then uses to create a perception of that reality via his or her sensory impressions. The perception that is created then stimulates a response which influences the reality and/or triggers feedback to the original sender. However, there are many factors that influence the perception that is created by the receiver, for example, social status, mood, time and stage of development, experience, memory as well as some cognitive processes such as creativity (Pettersson, 1989). By considering these factors when designing communications, designers can convey messages more effectively.

Even though the model developed by Schramm (1954) improved the understanding of communication between people, it did not address the factors which determine and influence the components which make up the model.

2.2.3 Berlo's SMCR Model of Communication

The Sender-Message-Channel-Receiver (SMCR) model was created to emphasise that there are many subcomponents that influence the main elements of communication. In mathematical communication models (such as Shannon and Weaver's model in Section 2.2.1) the effectiveness of the communication process is only based on the channel that is used. Berlo's model, shown in Figure 2.6, focuses on the other components as well with specific reference to the source and the receiver. In AM, all of the components in the SMCR model are applicable when communication is initiated. The designers of reports are the *sources* while the content being communicated is the *message*. The choice of format (electronic, verbal, hard copy, etc.) used to convey the message represents the *channel*. Finally, the target audience is represented by the *receiver*. Mathematical models simplify the understanding of electronic communication, but they do not provide significant insight into the process that takes place when two humans attempt to communicate (Bowman and Targowski, 1987).

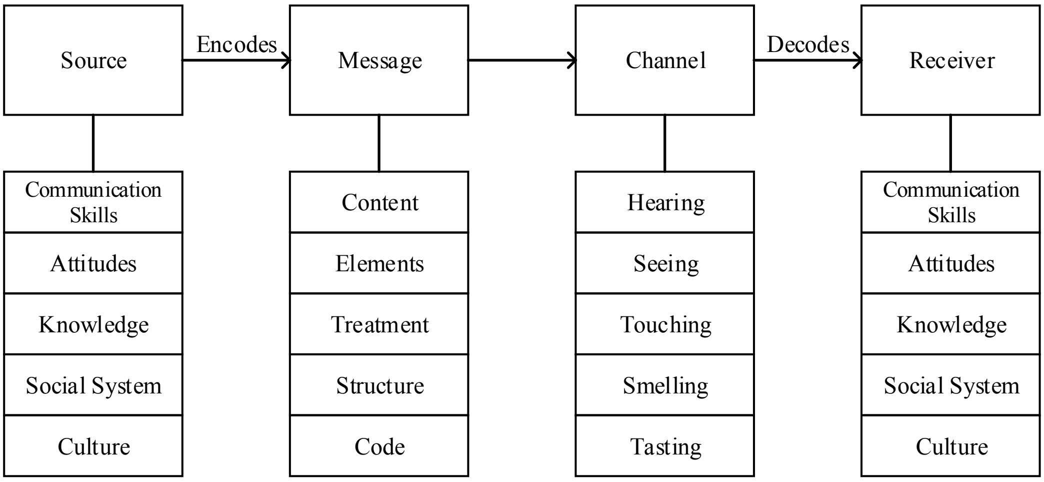


Figure 2.6: Berlo's SMCR model of communication.

Adapted from Berlo (1960).

According to Bowman and Targowski (1987), transmitting a message clearly is less important than the meaning that is conveyed in human communication. Unlike with mathematical models, a message that is only partially transmitted but fully understood is of more value than a complete message that is transmitted with little or no understanding. This is in agreement with the sentiments of Berlo (1960, p. 175) who mentions that:

> *"Communication does not consist of the transmission of meaning. Meanings are not transferable. Only messages are transmittable, and meanings are not in the message, they are in the message user."*

The aim of communication is to have a response elicited by communication with high fidelity (Berlo, 1960). In a communication environment, fidelity is described as a person getting exactly what he or she wants. A high-fidelity encoder can express the meaning of the sender or source perfectly in a message. Similarly, a high-fidelity decoder is able to decipher a message with complete accuracy. Berlo (1960) expanded Shannon and Weaver's definition of noise to include any elements of communication that reduce effectiveness. In essence, noise and fidelity are two closely related concepts. When noise is eliminated, fidelity is increased. On the other hand, fidelity is reduced by the production of noise.

As Berlo (1960) states, the effectiveness of communication can be addressed once the sources of fidelity and noise are identified. Each of the main components can be investigated to evaluate their contribution to fidelity and noise.

In the SMCR model, the source and encoder are often grouped together while the decoder and receiver components are also combined. Berlo (1960) argues that in person-to-person communication, the source encodes the message and then the receiver decodes it. The four main components of the SMCR model are thus: source, message, channel and receiver. Note that the elements in the other models can be generalised to match those of the SMCR model. These elements are discussed in more detail in the section that follows.

2.2.4 Components of Communication

All of the communication models discussed previously show the source, message, channel and receiver as the main constituents of communication. In reporting data, the source, message, channel and receiver can be replaced by the designer/reporter, data, medium and reader respectively. These elements might be better understood by means of an analogy. Suppose that the person to person(s) communication process can be represented by a person on the one side of a river attempting to send a package to a person on the other side of the river. The person sending the message is the reporter and the person receiving the package is the reader. The package represents the message that is sent and the boat and the water depict the medium that is used to convey the message.

Each of these four elements is explained in more detail below.

2.2.4.1 Medium

Bauer and Erdogan (2012) state that the accuracy with which a message is received is often determined by the medium used to initiate the communication. According to Barry and Fulmer (2004), matching the medium used to the goal of the message is the key to effective communication. The two most common formal mediums of communication in organisations, according to Bauer and Erdogan (2012), are verbal (oral) and written messages. Each medium has its own advantages and disadvantages. Table 2.2 provides guidelines which can be used to choose the appropriate communication medium based on a message's attributes.

As data reporting is based on facts describing complicated ideas or findings, this thesis focuses exclusively on written reports, regardless of whether they are hard copies or electronic.

2.2.4.2 Designer/Reporter

In the Berlo's SMCR model of communication (see Figure 2.6), the source-encoder represents the designer. Steele and Iliinsky (2011) explain that the

Table 2.2: Guidelines for when to use written and verbal communication.

Use written communication when:	Use verbal communication when:
conveying facts	conveying emotion and feeling
the message needs to become part of a permanent file	the message does not need to be permanent
there is little time urgency	there is time urgency
you do not need immediate feedback	you need immediate feedback
the ideas are complicated	the ideas are simple or can be made simple with explanations

Adapted from Bauer and Erdogan (2012).

designer identifies a goal that initiates the communication process. The type of information presentation created is based on the motivations, goals and priorities of the designer. By creating a well defined goal, subsequent design decisions are simplified and it can be used as a standard against which the design can be evaluated. Wang and Kuo (2010) state that it is vital for designers to have good understandings of the domain from which the information originates as this will allow them to provide insightful presentations of the data.

Steele and Iliinsky (2011) indicate that it is the responsibility of the designer to determine what information should be conveyed when data is reported as well as the format in which the information will be conveyed (see Section 2.4). If a designer decided to visualise information that is merely artistic, then aesthetics might be most important. However, if the report is informative or persuasive in nature, the designer needs to determine which elements in the data are of the highest importance and he or she also needs to choose the appropriate format that will highlight the relevant or important information. Penrose (2008) reveals that designers of reports have a responsibility to report accurate information as they have the ability to influence the perception of the readers by their choices of graph types, colours, scales, size and emphasis in reports.

The designers of reports thus play a crucial role in addressing the problem as outlined in the problem statement in Section 1.2. Recall that the aim of this thesis is to recommend formats in which data can be communicated to managers in data-intensive environments. The decision of which format to use is generally made by the designer of the report and thus he or she will play an important role in achieving this aim.

2.2.4.3 Data

Section 2.1.4 gives an overview of the types of data that are typically encountered and used in PAM environments. All types of data reporting are based on a dataset and it is the foundation on which data reporting is built. It thus makes sense that the data will be influential in the design of the communication and the format used to convey the information. Many software programs have been designed in such a way that almost any visualisation format can be applied to a dataset, but there is no guarantee that the visualisation will be effective or even informative for that matter. Steele and Iliinsky (2011) found that different types of data could require varying approaches, encodings (formats) and/or techniques to reveal interesting aspects in the data. Although default visualisations are often a good starting point for illustrating and making sense of data, using different reporting approaches or formats might yield new knowledge or underline different facts in the data.

The designer needs to be familiar with the data that is to be represented in order for him or her to report it effectively. Rather than forcing the data into a format that the designer finds attractive or convenient, special attention needs to be paid to consider the inherent values, relationships and structures of the data as well as the reason for the report. Steele and Iliinsky (2011) provided a list of possible considerations, but insisted that the list was by no means comprehensive. These considerations are listed below.

- The number of dimensions present in the data.
- The type of relationships among the variables (e.g. one-to-many or many-to-many).
- The degree of variability in the data.
- The data types (categorical, discrete, continuous, ordinal, etc.).

Pantelias, Flintsch, Bryant and Chen (2008) stress that even though the characteristics of the data play an integral role in the structure of the communication, data has to be reported in a way that supports decision making.

2.2.4.4 Reader

In the designer-reader-data trinity, the reader might be considered as the most influential entity. Earlier in this section it was mentioned that the success of the communication effort is ultimately determined by the reader. As Steele and Iliinsky (2011) pointed out, the designer's success is measured by the reader's success. If a designer does not take the reader into consideration when data is reported, inadequate encoding could cause the message to be received poorly or misinterpreted.

The reader will also determine the language used in the communication accompanying the visualisation. If both the designer and the reader are familiar with data that is being transferred, the communication might well be of a high complexity and the language used might be domain specific. However, as Agrawala, Li and Berthouzoz (2011) noted, the person inundated with the information might not be familiar with the field from which the data originated. As a result, the designer will be forced to simplify explanations. Steele and Iliinsky (2011) stress that simplifying explanations is not synonymous with simplifying ideas. Complex ideas can be communicated in a simplified manner by deconstructing complex concepts and then communicating those deconstructed ideas using clear and transparent language and symbols which are not domain specific. Furthermore, deconstruction can only take place when the designer has a good understanding of the data and failure to accomplish this might indicate a lack of in-depth understanding.

Although a designer is expected to have adequate knowledge of the subject that he or she is communicating, it can often have an unfavourable impact on the communication process. Steele and Iliinsky (2011) underline that too much knowledge on a subject by the designer can lead to assumptions about what the reader knows. For this reason, designers need to isolate themselves when assessing how a message will be received and interpreted by the audience. Considering the context of the reader is thus of vital importance.

Contextual considerations for the reader can be described by using a football (termed soccer in certain countries) analogy. If two players were to kick a ball to each other, player A would not simply touch the ball softly causing it to stay in the same place. Player A would exert some force on the ball to send it in the direction of player B. Other than the absolute location of player B, player A would also consider characteristics of the situation such as Player B's skill level, athleticism, speed and also direction of movement. The same principles apply to communication. The position of the designer (player A) is important, but the position of the reader (player B) is vastly more important whether they are receiving a football or interpreting a data presentation.

Nistal *et al.* (2009) argue that, although information has to be presented in a way that matches (or at least assists) the task which needs to be solved, the reader's individual characteristics and preferences play an important role in the comprehension of the information. For example, if a reader has little prior knowledge about a presentation format (table, bar graph, line graph, etc.) which is generally accepted to be the most suitable for a task, the benefits of matching the representational format to the task might be outweighed by the user's frustration and eventual disinterest caused by the lack of understanding.

Nistal *et al.* (2009) suggest that using a representational format with which the reader is familiar (although it might not be a perfect match to the task) to solve a task might yield better results than a forced match between task and presentation format.

Although the information presentation should be aligned with the needs of task, it is critical that the reader's characteristics are also considered when information is designed. Tory and Moller (2004, p. 72) summarise this concept as follows:

> *"We believe that more attention should be paid to users who must view and manipulate the data because how humans perceive, think about, and interact with images will affect their understanding of information presented visually."*

The next section discusses how humans acquire and process information and how tailoring the presentation of information to an individual's needs can affect his or her understanding thereof.

2.3 Information Design

Albers and Mazur (2014) describe high-quality information design as communication that transfers information in a way that is appropriate and relevant to a reader's situational context. All of the focus has to be placed on the reader to ensure that the reader's requirement which initiated the search is satisfied and that the right information is obtained. According to Kelton, Pennington and Tuttle (2010), different formats in which information can be disseminated to users are referred to as the information's presentation formats. Engle (2002) emphasised that the way in which information is presented can influence the cognitive load that is experienced by a person's working memory system which affects performance. The concepts of cognitive load and working memory as well as the relationship between them are subsequently discussed.

2.3.1 Cognitive Load Theory and Working Memory

Pioneering research by Sweller (1988) found that the load placed on a person's cognitive processes influences the problem-solving capabilities of the individual. Anderson *et al.* (2011) define cognition as the process involved with the acquisition and understanding of knowledge. The human brain uses working memory to execute these cognitive processes (Engle, 2002). Paas and Van Merriënboer (1994) described cognitive load as a multidimensional concept that describes the load that is imposed on a cognitive system of an individual when performing a certain task. Sweller, Van Merriënboer and Paas (1998) state that cognitive load theory aims to provide suggestions which could aid in the presentation of information to optimise intellectual performance.

In order to explain the concept of working memory, Baddeley (1992) uses the analogy of computer architecture. The Central Processing Unit (CPU) of a computer represents working memory while short-term memory is represented by temporary data buffers or Random-Access Memory (RAM). Finally, the sensory perceptions and reactions resulting from it are represented by the input/output communications. If the cognitive processes that are required in problem-solving are greater than an individual's cognitive processing capacity, performance will be negatively affected. In the computer architecture analogy, this negative effect on the performance results in a computer freezing or lagging. Of course, this analogy is only intended to aid in the understanding of working memory and is by no means comprehensive enough to represent the complex construct of working memory fully.

Chandler and Sweller (1991) distinguished between three different types of cognitive load, namely: germane, intrinsic, and extraneous. Each of these loads have a distinct effect on the decision making process. Seufert, Jänen and Brünken (2007) explained that the sum of the three types of cognitive loads determines the overall load that is placed on the person's working memory. Anderson *et al.* (2011) summarised these findings in Figure 2.7. Notice how the *Task Performance* curve is inversely affected by the *Cognitive Load* curve. The load-sub types remain constant in the decision making process. A description of the three types of cognitive loads and their impact on information presentation is given below.

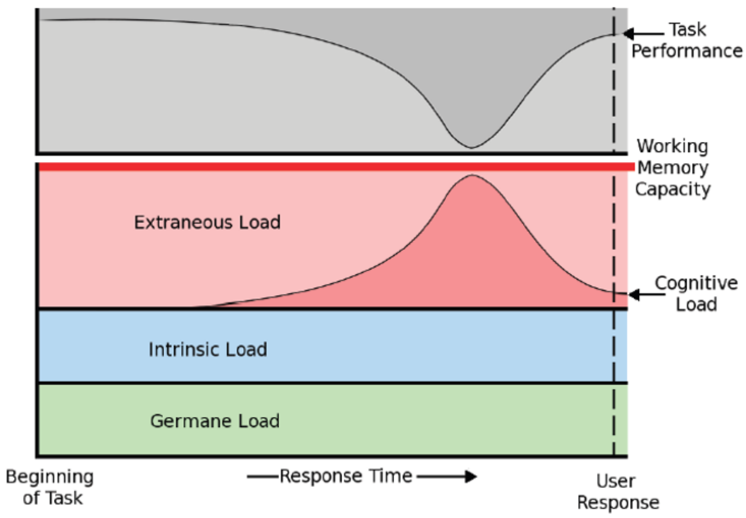


Figure 2.7: Combination of germane, intrinsic and extraneous load to form working memory capacity.

Adopted from Anderson *et al.* (2011).

Germane cognitive load is concerned with the way in which knowledge is acquired (Sweller, 2010). It is described by Anderson *et al.* (2011) as the working memory that is used to learn a new cognitive schema. This schema can be described as the internal representations which are constructed when a person learns how to interpret a new presentation format. Once a new schema is developed, it can be used multiple times in tasks which are similar. As a result of this, the contribution of the germane cognitive load to the overall cognitive load is high until the schema is established in which case the germane load becomes minimal.

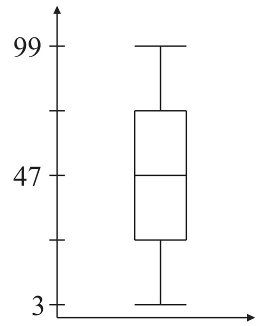
In terms of information presentation, this can be described by the construction of mental representations of information by using, for example, a pie graph. Pie graphs can be used to illustrate different proportions. The germane load is described as the effort needed to create mental representations of the proportions displayed by the pie graph which can then be used in problem solving tasks and decision making. The first time a person is tasked with using a pie graph, the germane load is high because of the person's unfamiliarity with pie graphs. However, once the person has become comfortable with using a pie graph, the germane load involved in decoding information conveyed in pie graphs is minimal.

Intrinsic cognitive load is referred to by Anderson *et al.* (2011) as the demands that are placed on a person's working memory by the underlying complexity of the information that is being transferred. Sweller (2010) explains that the intrinsic load of a task is determined by the task itself and the knowledge of the person. The intrinsic load of a task can thus only be altered by changing either the task or the knowledge of the person and is not affected by the presentation of the information. There is an intrinsic cognitive load involved in adding two numbers or learning the meaning of a chemical symbol. In information presentations, there is an intrinsic cognitive load involved in finding the highest number in a set, regardless of whether the information is presented in a graph, a table or text.

Extraneous cognitive load is a measure of the cognitive load that is placed on a user as a result of a task's design (Paas, Renkl and Sweller, 2003). Seufert *et al.* (2007) have shown that the way in which information is presented affects this type of load. Anderson *et al.* (2011) used Figure 2.8 to illustrate two ways of describing data. The numerical description is on the left while the right is a visual description. It is clear that the visual summary provided by the box plot requires less working memory to interpret than the numerical display. However, it is important to notice that the box plot on the right only requires less working memory if the schema of interpreting a box plot has been established. In other words, the box plot is only easier to read if a person has already learnt how to interpret a box plot.

3,	5,	28,	78,	72,	40,	52,	37
76,	6,	26,	68,	96,	70,	66,	75
34,	33,	20,	74,	36,	85,	99,	51
99,	33,	18,	38,	14,	18,	37,	53
25,	8,	69,	85,	25,	65,	30,	28
12,	87,	59,	54,	6,	30,	16,	59
97,	66,	23,	84,	87,	76,	36,	15
97,	87,	93,	12,	70,	56,	94,	97

Table of Data



Box Plot

Figure 2.8: Two ways of describing data.

Adopted from Anderson *et al.* (2011).

Ginns (2006) explained that tasks with high intrinsic and/or germane loads are generally perceived to be difficult by the user. Designers of information should ensure that difficult tasks are presented in a format which places a low extraneous load on the user to promote understanding. Recall that the overall cognitive load placed on the working memory is the sum of the germane, intrinsic and extraneous loads. Figure 2.7 shows that the task performance effectively becomes zero or insignificant if the overall cognitive load exceeds the capacity of the working memory. Since the intrinsic load is not influenced by the presentation of information, a high extraneous load results in little working memory being available for the germane load. If the working memory that is available for the germane load is insufficient, there might be a reduction in the deeper meaningful cognitive processes of the person interpreting the information. This, in turn, could result in the person not obtaining a deep understanding of the information.

Huang, Eades and Hong (2009) discovered that the overall cognitive load is influenced by certain factors. However, instead of assigning the factors to the three load types of Chandler and Sweller (1991), they were labelled causal factors and categorised into one of six different types of complexities. These causal factors are believed to influence the overall cognitive load which, in turn, influences the assessment factors such as response time and response accuracy. This concept is illustrated in Figure 2.9.

The six causal factors and short descriptions of each are given below.

1. Domain complexity: The data formats, contents, tasks and visualisation requirements are often related to the domain in which they are created

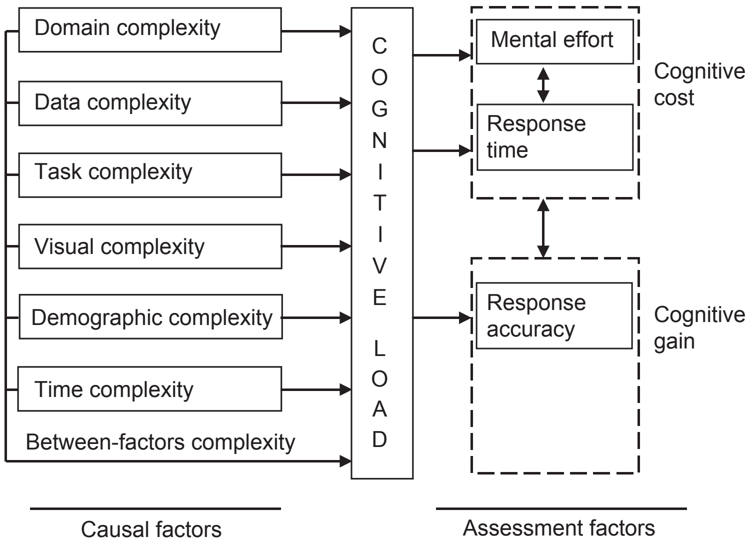


Figure 2.9: The construct of cognitive load for visualisation understanding.

Adopted from Huang *et al.* (2009).

and used. Kashihara, Oppermann, Rashev, Simm *et al.* (2000) found that, although exploring concepts of an unknown domain might be an effective way of learning, it can also result in cognitive overload. Huang *et al.* (2009) state that this factor serves as the basis of the overall complexity as it is responsible for the specific visualisation constraints. For example, the visualisation requirements for sociologists and biologists will differ as the information required by each is unique to their domain. In a PAM environment, data related to certain asset performance and maintenance metrics might be difficult for the other departments (such as the legal and accounting departments) to interpret. On the other hand, these metrics should be common in the engineering department.

2. Data complexity: This refers to the characteristics of the data such as, but not limited to, the number of observations, dimensions present, relationships between variables and attributes of the objects. Lin *et al.* (2006) and Koronios and Lin (2006) summarise the differences in data characteristics between an Engineering Asset Management (EAM) environment and other typical business environments. These differences are shown in Table 2.3. Aggarwal (2013) pointed out that advances in hardware technology have enabled the collection of data by using Global Positioning Systems (GPS) and other sensory equipment. These new data types which are being recorded require novel interpretation methods. The large diversity of data types in the AM environment means that the cognitive load associated with the presentation of data is normally higher than that of other typical business environments. This makes the task of the choice of presentation all that more important in AM.

Table 2.3: Comparison of data characteristics between AM and other business environments.

Typical Business Environment	EAM Environment
Self-descriptive	Non self-descriptive
Static	Dynamic
Intrinsic quality	Intrinsic/extrinsic quality
Discrete value with fewer or no constraints	Continuous value with constraints (e.g. within a range), precision value
Current	Temporal
Transactional data	Time-series streaming data
Often structured	Often unstructured
Easy to audit	Difficult to be audited
Can be cleansed using existing tools	Difficult to be cleansed using existing tools
Similar data types	Diversity of data types

Adapted from Koronios and Lin (2006).

3. Task complexity: Huang *et al.* (2009) reason that tasks have certain demands that impose a load on the person(s) executing the task. These demands include the number of objects and how the objects interact. Kalyuga (2011) described the task complexity as a measure of the load that has to be processed by the working memory as a result of the interconnectedness of the information related to the essential elements in a task. Task complexity should not be confused with task difficulty. Task complexity refers to an objective measure which is independent of the subject completing the task. Difficulty refers to a viewer's perception of the task based on other factors in addition to task complexity.

   Tuovinen and Sweller (1999) explained that learning the order of words is an example of this. Word order can only be learned when all of the words are considered simultaneously. In AM, the same could be said about deciding between projects. The task complexity of deciding between three projects will be higher than deciding between two projects regardless of the person(s) responsible for the decision or the presentation of the information of each alternative.

4. Visual complexity: Visual elements as well as their spatial distributions influence visual complexity. The visual representations of the elements, how well the spatial relationships and intrinsic structural links match,

the extent of conformity between spatial relationships and human visual perception conventions and the required cognitive processes required by the task all determine the visual complexity (Huang *et al.*, 2009). Consequently, visualisations that contain fewer elements or that are based on general aesthetics are not necessarily of lower visual complexity. However, Heylighen (1999) and Harper, Michailidou and Stevens (2009) mention that the selection of objects, colours, patterns, range of objects and surface styles all contribute to the overall cognitive load.

Bacim, Ragan, Scerbo, Polys, Setareh and Jones (2013) investigated the relationship among several factors such as visual complexity, task scope, display fidelity and the spatial ability of the user. The study found that visual complexity and task scope had a direct impact on user response speeds with higher levels of either factor resulting in slower user performance. Three different complexity levels were used in the study and they are illustrated in Figure 2.10.

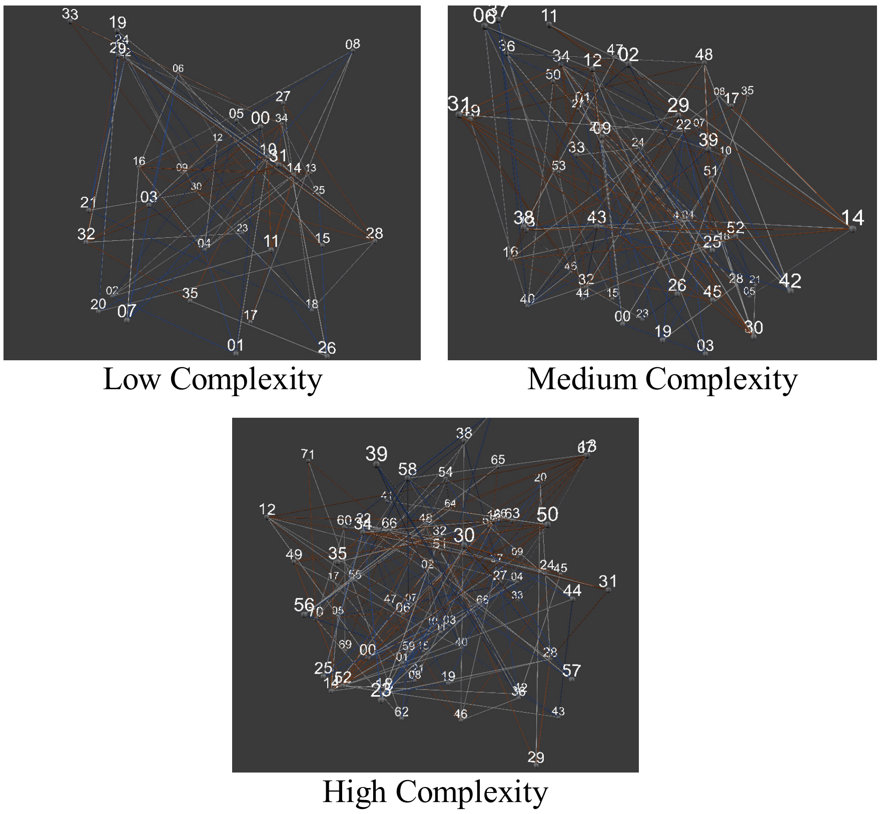


Figure 2.10: Examples of the three levels of visual complexity.

Adopted from Bacim *et al.* (2013).

In a PAM environment, the choice of visualisation used to convey information could have a significant impact on the visual complexity. Even if the right choice is made with respect to the type of visualisation to be used, colours and patterns used to distinguish between different data sets or elements could increase the visual complexity. As employees in PAM environments are generally exposed to large amounts of data which they have to process in short amounts of times, illustrations which are too visually complex might be disregarded.

5. Demographic complexity: Demographic complexity refers to the characteristics of the individual interpreting the information and includes a wide variety of factors such as motivation, age, gender, cognitive capabilities, knowledge of the domain as well as mental status. These factors contribute to the cognitive load experienced by the subject when completing a task. For example, a user with sufficient domain knowledge will require less effort to understand social networks. Recall that task complexity was described as an objective measure of the size of the load placed on the user and that the difficulty of the task is dependent on the user. This perceived difficulty of the task, as experienced by the user, is determined by the user's demographic or personal characteristics.

   Engineers with years of maintenance experience and a clear understanding of the inner workings of the specific AM system will interpret condition monitoring and failure statistics with relative ease compared to a young accountant with no experience or natural aptitude for the information.

   Moreover, as Oblinger (2003) notes, younger professionals do not consider computers and other pieces of electronic hardware as new technology as they are accustomed to computers. Consequently, the cognitive load associated with using advanced electronic condition monitoring systems might not be as large for younger employees as for older employees who have always used the same traditional systems.

6. Time complexity: The amount of time required to understand visualisations depends on the situation. Time complexity is also influenced by the user's perception of the time available. A study by Kerick and Allender (2004) investigated the effect that various task demands had on the shooting performances of soldiers. The study concluded that time stress had the most significant effect on the soldiers' perceptions of the workload and subsequent shooting performances. If the time that is available for the interpretation of information is short, lengthy reports (even if they are insightful) will add to the unnecessary cognitive load experienced by the user.

Huang *et al.* (2009) underline that cognitive load is not only impacted by the factors individually, but also by the interaction between the factors.

The overall cognitive load affects performance in three ways, namely: mental effort, response accuracy and response time (see Figure 2.9). Mental effort and response time are considered to be cognitive costs. Generally, cognitive costs should be as low as possible. Response accuracy, on the other hand, is considered as the cognitive gain and a higher amount is better. As with the causal factors, there is also interaction between the assessment factors. An example of this is when more time is used (higher cognitive cost) to achieve better performance (higher cognitive gain). The only way to improve performance in a set task while not increasing the cognitive costs is by exerting more mental effort.

Huang *et al.* (2009) proposed another model which aims to explain the interaction between cognitive load and the factors used to assess it under the limits of memory. Figure 2.11 shows the theoretical model of task performance, mental effort and cognitive load. The model shows that performance is dependent on both the memory that is being used and the memory that is demanded by the task. There are five regions of varying cognitive loads in the model and each of the regions are subsequently discussed.

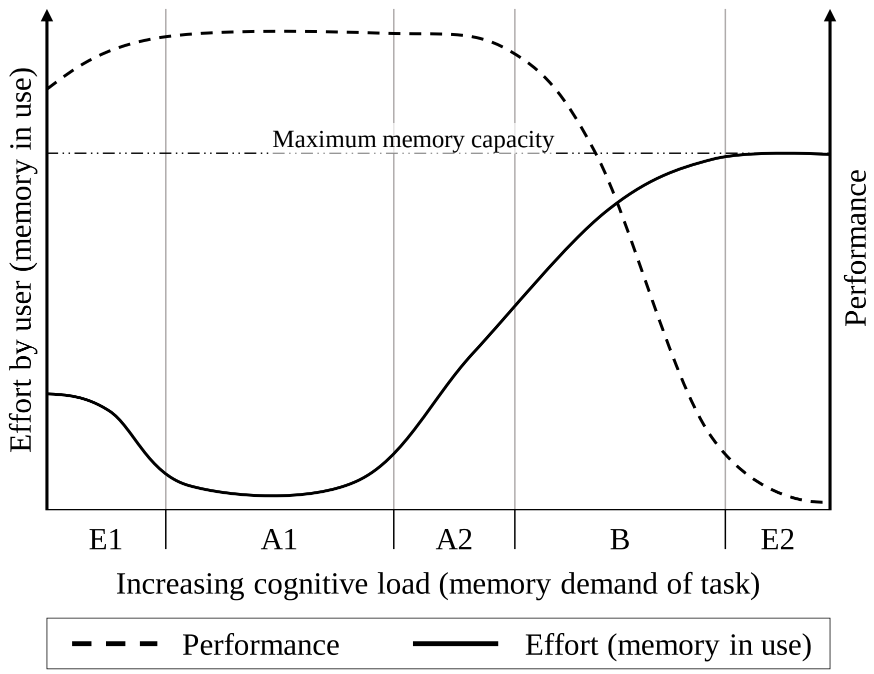


Figure 2.11: Mental effort, task performance and cognitive load.

Adapted from Huang *et al.* (2009).

1. In Region *E1*, boredom and disregard affect performance negatively as there is a consistently low cognitive load. The increased memory that is being used is as a result of the viewer attempting to remain focused. In PAM, this can happen when irrelevant or self-evident information is being emphasised in data reporting. Managers and other employees alike lose interest in the information being conveyed as they do not find it intellectually stimulating.

2. In Region *A1*, there is an increased demand on memory, but the demand is met by the viewer without devoting any extra effort. The resulting performance remains at an optimum level and the memory used is well within the viewer's memory capacity.

3. In Region *A2*, the performance remains at an optimum high, but it can only be maintained if the viewer exerts more effort than in Region A1 (though still well within the viewer's memory capacity). If the effort which is required by the viewer's memory approaches the maximum memory capacity, a shift from Region A2 to Region B occurs. All reporting in a PAM environment should aim to take place in regions A1 or A2 as this is where optimal performance takes place.

4. In Region *B*, the memory effort needed by the viewer approaches the viewer's maximum capacity which leads to a decrease in task performance. Any increase in effort in this region severely decreases performance. Once the effort needed reaches the maximum capacity, Region E2 is entered.

5. In Region *E2*, the demands placed on the viewer exceeds the capabilities of the viewer's memory to such an extent that performance becomes insignificantly small. If the overall cognitive load of the message exceeds capabilities of user, no information can be transferred.

It should be noted that the maximum memory capacity is dependent on the individual performing the task. Figure 2.11 draws attention to the fact that a low cognitive load does not necessarily lead to better performance. As a result of little cognitive stimulation, a viewer might become bored and thus disengaged. On the other hand, if complex information produces a low cognitive load it implies that a visualisation is not being understood effectively. As the cognitive load induced is made up of different factors, it is difficult to isolate and measure the effect of each individual factor (and the combined effect). Huang *et al.* (2009) state that any cognitive load that is not related to the task that needs to be accomplished (such as the cognitive load that results from visual complexity) should be minimised. This is in agreement with the findings of Anderson *et al.* (2011). Notice the similarity in the shape of the performance curves in Figure 2.7 and Figure 2.11.

It is clear that a cognitive load with a magnitude that approaches the working memory capacity sharply decreases the task performance. The cognitive load of the message that is being reported should thus be in region A1 or A2 (in Figure 2.11) of the user's cognitive capacity. This can be achieved by tailoring the presentation of information to fit the task that has to be executed. The concept of cognitive fit is discussed in the next section.

2.3.2 Cognitive Fit

Building on the research conducted by Sweller (1988), Vessey (1991) found that humans create a mental representation of the scenario when facing a problem. The representation occurs in the human working memory and it can be seen as a subset of the total problem space. The mental representation that is formed comprises of characteristics of both the problem representation and the task that needs to be executed (collectively referred to as the problem solving-elements). Some cognitive processes are applied to the information in the representation of the problem and some are required to solve the problem. The interactions between these processes result in the mental representation of the problem.

If the types of information that are emphasised in the problem-solving elements match, the mental representation of the problem also emphasises the same information. As a result, there will be consistency in the processes that are used to interpret the information and to complete the task which, in turn, facilitates problem solving. The initial mental representation that is created will thus not need to be transformed to accommodate the processes needed to interpret the information and those needed to execute the task. Cognitive fit occurs when information is provided in a format that leads to the creation of a mental representation that does not require any additional transformations. Vessey (1991) also noted that a mismatch in the problem-solving elements would lead to decreased performance in problem solving or information processing. Figure 2.12 shows the problem-solving model developed by Vessey (1991).

Shaft and Vessey (2006) extended the model created by Vessey (1991) to address the argument that the mental representation for a task solution is influenced by both internal and external representations as well as the interactions between the two representations. Shaft and Vessey (2006) simplify and explain this concept by using the task of understanding a software program as an example. The software itself is seen as the external problem representation while the user's prior knowledge of the software domain represents the internal representation of the domain. Finally, the task that needs to be completed by using the software is considered to be the problem-solving task.

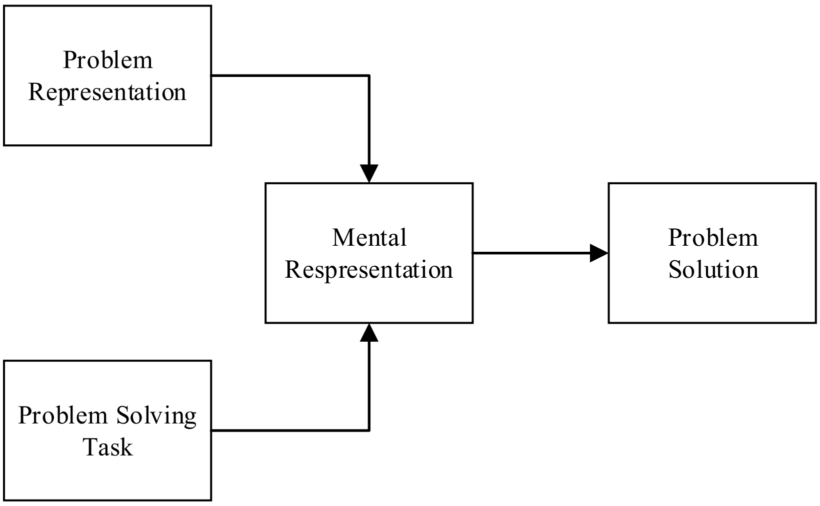


Figure 2.12: General problem-solving model.

Adapted from Vessey (1991).

Expanding on the work done by Shaft and Vessey (2006), Kelton *et al.* (2010) developed a framework which can be used to detail the impact of information presentation on individual Judgement and Decision Making (JDM). The framework, shown in Figure 2.13, suggests that decision making is influenced in two primary ways by the information presentation. As was already concluded by Vessey (1991), the information presentation format influences the mental representation created by the user. Another way in which the information presentation format might affect the decision making process is by influencing specific characteristics of the user such as active involvement (in the case of users identifying relevance and importance of information) and task knowledge.

Mostyn (2011) explains that the main aim of cognitive load theory is to provide guidelines which promote information retrieval and learning. There are two main ways of accomplishing this. Firstly, by reducing germane cognitive load by using standardised or easily recognised visualisations. Secondly, the extraneous load can be improved (reduced) by making use of appropriate formats and designs for information visualisations based on the task that needs to be executed. Reducing the extraneous load in this way is termed cognitive fit (Lehner, Falschlunger, Losbichler and Eisl, 2015). By fitting the visualisation to the task at hand, the problem representation will not have to be transformed to match the task and neither will the decision processes need to be transformed to match the problem representation. This will reduce the cognitive effort needed by the problem solver which, in turn, improves decision making.

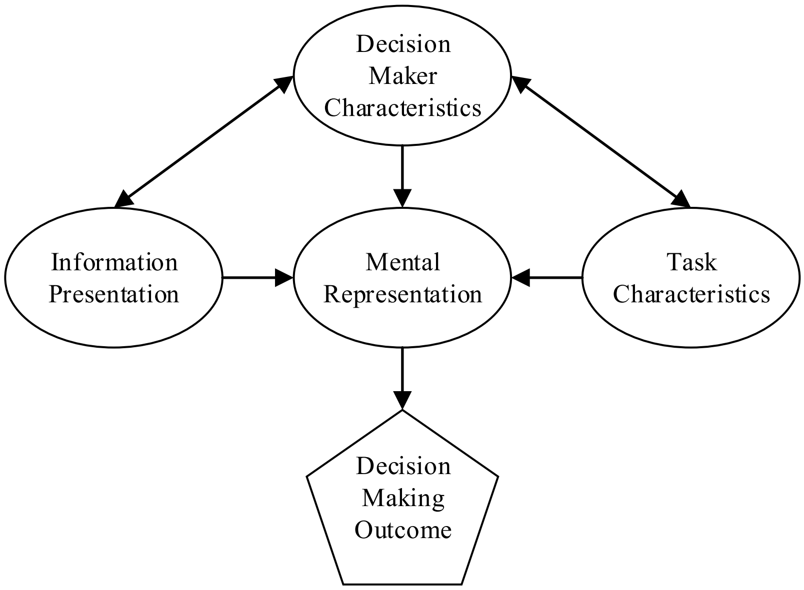


Figure 2.13: Decision-making framework for information presentation effects.

Adapted from Kelton *et al.* (2010).

Literature emphasises that the process involved in designing information should concentrate on the receiver of the information. Various factors need to be taken into account in order to decide on the most appropriate presentation format for the information. The next section discusses the main presentation formats used to convey data and information to the intended audience.

2.4 Data Presentation

The preceding sections have mainly focused on how to communicate effectively and how the individuals in the target audience affect communications. This section will explore the three most common presentation alternatives that are available to designers of reports, namely: text, tables and graphics.

Statistical data analysis commences with the exploration of data and concludes with the presentation of results (Koschat, 2005). Data presentations can thus be divided into two categories: exploration and presentation. As Chen, Härdle and Unwin (2007) mention, there is a fundamental difference in form and practice between the two categories and Steele and Iliinsky (2011) point out that there are tools and approaches that might only be appropriate for the one or the other.

Exploratory graphics are used when looking for results in a large dataset. Steele and Iliinsky (2011) state that, because of the advantages associated with visualisations, translating data into a visual representation can aid in the identification of features such as curves, trends, lines, or anomalous outliers. Chen *et al.* (2007) emphasise that exploratory graphics should be fast and informative and that accuracy is less important than speed. The main use of exploratory graphics is to generate ideas and to find information. Koschat (2005) explains that the main aims of data exploration include gaining an improved understanding of the datasets, examining the integrity of the data, finding any unusual features which can aid a formal analysis and preparing the data for a more extensive analysis. According to Steele and Iliinsky (2011), data exploration is often conducted at a high level of granularity and generally forms part of the data analysis phase.

Presentation graphics (also referred to as explanatory graphics) are generally static and used to summarise information. Chen *et al.* (2007) highlight that explanatory displays are required to be of high quality and should be accompanied by clear definitions and explanations of the variables and the shape of the graphic. Another distinction made by Steele and Iliinsky (2011) is the fact that the features of the data are known to the designer of the graphic at the outset. As a result, the graphic can be designed in such a way as to emphasise the appropriate areas or outliers. As the name suggests, presentation graphics normally form part of the presentation phase and aim to communicate the results of the analysis to the target audience (Koschat, 2005).

This thesis aims to improve information reporting to relevant stakeholders during the presentation phase. For this reason, the rest of the research will only focus on data that has already been processed into numerical information that needs to be reported to stakeholders. As mentioned before, designers of reports have three common formats which can be used in the presentation of information to stakeholders, namely: text, tables and graphs. However, as Porat, Oron-Gilad and Meyer (2009) highlight, the correct choice between table, text and graph is not as apparent as one would think as each method has advantages and disadvantages associated with it.

Research by Gillan, Wickens, Hollands and Carswell (1998) led to the creation of a flowchart (shown in Figure 2.14) which can be used to decide on a presentation format based on the amount of data, the uses of the data as well as the value of visualising the data. Gillan *et al.* (1998) maintained that a small number of data points with a simple relation, such as one value being significantly greater than the rest of the values in the small dataset, does not justify the use of a table or a graphic as the increase in cognitive costs (loads) outweighs the potential benefits. However, it is emphasised that presenting larger datasets, especially those with underlying trends, requires more consideration.

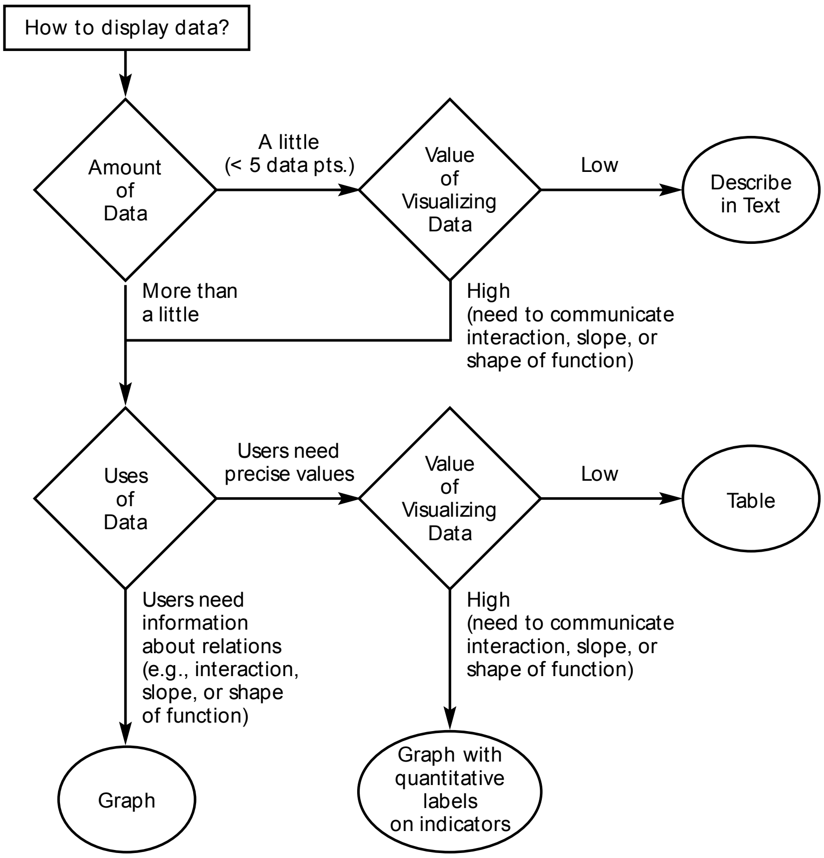


Figure 2.14: Flowchart showing the steps involved in deciding how to present data. Adopted from Gillan *et al.* (1998).

Gillan *et al.* (1998), similarly to various other authors, highlight that the design of information (including the choice of format used to encode it) should support the reader's requirements. However, Gillan *et al.* (1998) also acknowledge that the choice of presentation format should consider the requirement(s) of the task that needs to be executed. The choice between text, table and graph is a crucial one and Franzblau and Chung (2012) explain that important information can be misinterpreted or even ignored if not presented in the appropriate format. Each of the presentation formats are discussed next.

2.4.1 Text

Data can be conveyed by merely including the numbers or results in the main body of the text. Law, Freer, Hunter, Logie, Mcintosh and Quinn (2005) compared how well textual summaries and trend graphs of physiological data supported decision making capabilities of staff in a neonatal Intensive Care Unit (ICU). Data that were communicated consisted of both continuous physiological variables such as heart rate as well as discrete variables such as descriptions of staff actions. The study was conducted using 40 neonatal ICU doctors and nurses with varying degrees of experience. Participants were presented with electronic information (either graphic trends of the patient data or textual summaries) and asked to choose an appropriate action from a predetermined list with 18 possibilities. Two human experts generated the textual summaries which included descriptions of the changing physiological metrics and any of the medical interventions which were deemed relevant, but they did not provide any interpretations of the data. Law *et al.* (2005) concluded that the textual summaries resulted in better performances by the participants than the set of trend graphs and suggested further research into computer generated descriptive summaries.

Van Der Meulen, Logie, Freer, Sykes, McIntosh and Hunter (2010) conducted a study similar to that of Law *et al.* (2005) in an attempt to validate their findings. A mixture of 35 staff members including senior doctors, junior doctors, senior nurses and junior nurses at a neonatal ICU participated in the study. Participants were once again exposed to anonymised data that were presented as textual summaries, multiple line graphs and computerised Natural Language Generation (NLG). After interpreting one of the three data representations, participants were asked to select an appropriate response from a predetermined list.

The results supported the findings of Law *et al.* (2005) with participants performing the best when using the human generated text (summaries) for decision making. The decisions resulting from the use of NLG were at least as good as those resulting from the graphical displays when a direct comparison was made. Van Der Meulen *et al.* (2010) suggested that using textual summaries and NLG might reduce visual complexity and mental workload which could improve decision-making performance, especially for inexperienced staff. This recommendation supports the proposition of Huang *et al.* (2009) (refer to Section 2.3.1) which states that users with good knowledge of the domain, such as senior staff members, will experience a reduced cognitive load. On the other hand, the increased cognitive load placed on junior staff members by their inexperience of the domain might lead to a reduction in memory that is available to interpret complex visualisations. For the purpose of this thesis, it is assumed that the designers of the information will not always be experts who can generate data summaries.

A simple example where using text might be appropriate in PAM environments is in the expected revenue to be generated by a single project. There is no value added when as single value is communicated using a table or a graphic instead of using a sentence to communicate data. Figure 2.11 in Section 2.3.2 shows that employees often lose interest when simple information is communicated in a complex format. As there are already a large number of graphics used in AM reports of data-intensive industries, unnecessarily encoding simple information in complex formats will add to the overall cognitive load.

In addition, the flowchart shown in Figure 2.14 indicates that it is only appropriate to describe data in text when there are less than five data points. For these reasons, only tables and graphs will be considered as viable alternatives for presenting data in this thesis.

2.4.2 Tables

Academics are divided when it comes to choosing between using tables or graphs.

Gelman (2011) explains that tables are used to communicate raw data and can often contain a lot of data. A conventional table consist of multiple rows of information that are divided into columns with column headings in the first row. Row headings are optional. Gelman (2011) notes that tables are not created to be read as narratives and thus clarity is not necessarily their main priority. Tables, however, do need to be highly accurate and should be used to summarise the most important results.

A study by Chan (2001) investigated the performance of 40 part-time Master of Science Information Systems students in predicting the operating performance of listed companies in Hong Kong. The participants were divided into four groups to test whether tables or graphs performed better under high and nominal information loads. There were no significant differences between tables and graphs, although lower information loads resulted in better performances by participants for both formats. Chan (2001) concluded that there was no real advantage of graphs over tables when displaying financial information and that conveying the right amount of information in data communication was more important.

Hawley, Zikmund-Fisher, Ubel, Jancovic, Lucas and Fagerlin (2008) explored the impact that various graphical presentation formats had on health-related treatment choices made by patients. Participants in the study perceived data conveyed in tabular format to be the most trustworthy and scientific. Tables were also found to be the best choice for conveying specific numeric information such as the exact number of patients that were expected to experience side effects.

Porat *et al.* (2009) pointed out that tables are equally effective as graphs or even marginally better for tasks which can be performed by looking at single data points. However, graphs have a definite advantage over tables when tasks require configurations of data points to be considered. Schonlau and Peters

(2012) summarised studies in which certain types of graphs and tables were compared across a variety of tasks. When estimating the absolute size of proportions, tables outperformed graphs in terms of accuracy, but took longer to read. Tables also outperformed graphs when the maximum of a row had to be identified.

Koschat (2005) reports that the benefits of using tables to display information are not acknowledged as enthusiastically by the statistical community. The only indication of tables being an appropriate format to display information in is in the frequency of their use. According to Koschat (2005), the space devoted to tables in statistical reports and research papers often exceeds that which is devoted graphs. However, the lack of interest shown in tables by the broader statistical community is evident in their disregard as a topic in statistical education, little mention in statistical practice and lack of research of their use in the field of statistics.

Following from the literature on tables as a presentation format, a possible application of tables in a PAM environment would be listing the GPS coordinates of various assets. Coordinates specify the position of an object on earth as a set of numbers, typically in terms of a longitude and a latitude. The longitude and the latitude of a coordinate are given as degrees ranging from 0 to 180 and -90 to 90 respectively. Since the entire earth is represented by such a small range of numbers, the exact longitudinal and latitudinal values of a position - down to the decimal - are of extreme importance. Since a table is superior to a graph when displaying exact values, the ideal format for the communication or coordinates would be tables.

2.4.3 Graphics

As mentioned before, tables and graphs are often seen as two competing means of presenting data and information. Chan (2001) noted that tables were traditionally used to present numbers, but with the emergence of technologies supporting information and graphics, graphs have become an inexpensive alternative. Fishwick (2008) found that good visualisations facilitate readability and enhance understanding. Tory and Moller (2004) referred to the process by which representing information visually assists humans with data analysis as Cognitive Support (see Cognitive Load in Section 2.3.1).

Card, Mackinlay and Shneiderman (1999) created a table which summarised the main methods in which visualisations support cognition. The table was then modified by Tory and Moller (2004) to generalise the ideas to all types of visualisations. Table 2.4 summarises the generalised advantages of using graphics.

Table 2.4: Cognitive support provided by the use of visualisations.

Method	Description
Increased Resources	
Parallel processing	Parallel processing by the visual system can increase the bandwidth of information extraction from data.
Offload work to the perceptual system	With an appropriate visualisation, some tasks can be done using simple perceptual operations.
External memory	Visualisations are external data representations that can reduce demands on human memory.
Increased storage and accessibility	Visualisations can store large amounts of information in an easily accessible form.
Reduced search	
Grouping	Visualisations can group related information for easy search and access.
High data density	Visualisations can represent a large quantity of data in a small space.
Structure	Imposing structure on data and tasks can reduce task complexity.
Enhanced Recognition	
Recognised instead of recall	Recognising information presented visually can be easier than recalling information.
Abstraction and aggregation	Selective omission and aggregation of data can allow higher level patterns to be recognised.
Perceptual monitoring	Using pre-attentive visual characteristics allows monitoring of a large number of potential events.
Manipulable Medium	Visualisations can allow interactive exploration through manipulation of parameter values.
Organisation	Manipulating the structural organisation of data can allow different patterns to be recognised.

Adapted from Tory and Moller (2004).

Franzblau and Chung (2012) point out that visual representations of data are valuable additions to any paper or presentation as the way in which they depict trends and key points improve understanding. Two of the various advantages of visual representations is their ability to illustrate relationships and trends that would not be visible if the data was presented in tabular format. This can be invaluable in PAM environments when making sense of large sets of data with no apparent patterns. For example, by plotting the outputs of two mines for twelve months on the same graph (using a presentation format which allows this), a comparison between the outputs of the mines for each month can be made and the trend in outputs for each mine can also be investigated. Any anomalies identified when plotting the outputs can then be isolated and investigated further.

The previous sections have repeatedly found that the presentation formats used in reporting information should support the needs of the reader or the target audience. In this section, the three most popular presentation formats have been discussed, but it has also been shown that data or information should only be communicated in the textual body of a report when fewer than five data points are present. It has also been mentioned that the choice of presentation format depends largely on the dataset and the task that the data needs to facilitate. In the next section, common tasks that need to be completed using data are discussed.

2.5 Tasks Performed Using Data

The choice of presentation format depends on a variety of factors. Section 2.3 concludes that the target audience of the communication plays a pivotal part in the decision. Furthermore, Figure 2.14 shows that the characteristics of the data in the dataset also influence the choice of presentation. However, this section will show that the task that needs to be performed using the data is also instrumental to the final choice of presentation format.

Sant'Anna and Douglas (2015) discuss a framework which can be used to classify data into categories such as qualitative or quantitative, nominal or ordinal, discrete or continuous and metric or non-metric to name a few. Shmueli, Patel and Bruce (2011) on the other hand classify data into categories such as numerical or text, continuous, integer or categorical and ordered (ordinal) or unordered (nominal). These frameworks might enable a designer to classify the different data types, but it does not aid the data reporting process. Even if the data were classified or categorised perfectly, Hartley, Cabanac, Kozak and Hubert (2014) explain that it still does not ensure that choosing the optimal reporting format thereof is straightforward.

Studies by Hollands and Spence (1998), Law *et al.* (2005), Few (2007), Van Der Meulen *et al.* (2010) and Schonlau and Peters (2012) all show how different representational formats have varying strengths and weaknesses. Recently, in a study by Kozak and Hartley (2013), both a table and a figure were successfully used to illustrate different features in the same dataset. This shows that, although the data type is important, it is the decision that needs to be made using the data that should ultimately determine the reporting format that is used. Braithwaite and Goldstone (2013, p. 1) summarise this concept as follows:

> *"Humans employ external representations of information, such as graphs, diagrams, tables, and equations, to assist in performing a wide variety of tasks in a wide variety of domains. Often, the same information may be represented using more than one representational format."*

Simon (1978) labelled two representations which contain the same information and which can be reconstructed perfectly using the other as *informationally equivalent*.

Even though there is ample research suggesting that the format used to present data should support the task at hand, very little research has attempted to create a formalised list of typical tasks which need to be supported by data. Amar, Eagan and Stasko (2005) proposed ten low level visual analytical tasks, all of which are relevant to graphics created from datasets. These ten analytical tasks and general descriptions of each are illustrated in Table 2.5.

Yang, Li and Zhou (2014) tested 500 users' task-oriented visual insights when presented with a variety of graphics. Users were asked to provide free-text descriptions of the graphics which were analysed (and coded) by three investigators. The investigators discovered two groups of insights: basic and comparative. The basic group consisted of four insights, namely: read value, identify extrema, characterise distribution and describe correlation. There were also four insights in the comparative group, namely: compare values, compare extrema, compare distribution and compare correlation. Table 2.6 provides short descriptions of each of the insights.

Fausset, Rogers and Fisk (2008) differentiate between local tasks and global tasks. Local tasks are described as tasks that can be performed by observing a data representation without having to do any calculations. Global tasks, on the other hand, require using values derived from calculations performed on data that is presented. The authors reiterated that tasks which require users to derive values that are not explicitly displayed in a data representation affect performance negatively by placing a higher demand on the user's working

Table 2.5: Ten low level visual analytical tasks.

Task	General description
Retrieve Value	Find attributes given a set of specific cases.
Filter	Find elements in the data which satisfy a concrete list conditions.
Compute Derived Value	Calculate an overall numeric representation of a number of data cases.
Find Extremum	Identify data points within a data range which possess and extreme value of an attribute.
Sort	Rank a set of data cases using an ordinal metric.
Determine Range	Determine the span of an attribute of interest within a set.
Characterise Distribution	Characterise the distribution of an attribute of interest over the set.
Find Anomalies	Identify values in a dataset which do not illustrate the expected relationship i.e. statistical outliers.
Cluster	Find groupings of similar data values.
Correlate	Identify possible useful relationships between two variables or attributes.

Adapted from Amar *et al.* (2005).

memory. Fausset *et al.* (2008) propose that this can be avoided by using information presentations and formats which facilitate task performance. The task requirements used in their study were:

- Judgement of change over time
- Comparing trends
- Identification of trends
- Discrete data comparison
- Comparing two points between data series
- Exact point value extraction
- Identifying maximum or minimum point values
- Estimating proportion of the whole
- Comparing more than one component

Table 2.6: Basic and comparative types of insights.

Types	Description
Basic Insights	
Read Value	Determine the value or range of one or more data points.
Identify Extrema	Determine the identity of the variable(s) which possess extreme values of the variable being measured.
Characterise Distribution	Describe the variation in the values across the range of a variable being measured.
Describe Correlation	Describe the relationships between the variable being measured and categorical variable.
Comparative Insights	
Compare Values	Contrast the values of the variable being measured at some data points in the display. Can also imply contrasting variables of two or more data sets at a specific point in time.
Compare Extrema	Compare two or more values identified as extrema.
Compare Distribution	Compare the variation in the values across the range for two or more categorical variables.
Compare Correlation	Compare the relationship between the variable being measured with relation to one categorical variable to the relationship of the variable being measured to another categorical variable.

Adapted from Yang *et al.* (2014).

Fausset *et al.* (2008) used a graphic similar to Figure 2.15 to illustrate that a graphical format can be appropriate for a certain task but not for another. For example, there is a match between graph and task when a user is asked to use Figure 2.15 to determine which stock was traded the most at 1:00pm (Stock D). On the other hand, a graph to task mismatch occurs when Figure 2.15 needs to be used to determine the value/price of Stock A at 1:00pm as it cannot be answered using the information in the graph.

The example used by Fausset *et al.* (2008) along with the findings of Kozak and Hartley (2013) in the beginning of this section show that some graphs might be better for communicating information for a certain task, but not for another. The next section will discuss some of the basic elements that make up these graphs as well as how data is presented in them.

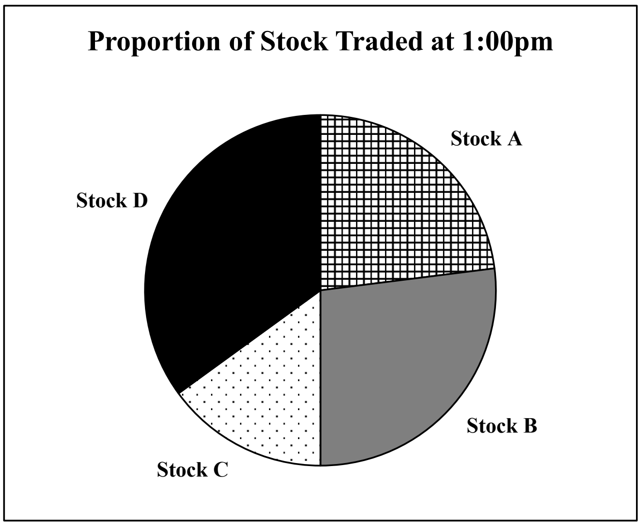


Figure 2.15: Pie graph showing proportions of stock sold at 1:00pm.

Adapted from Fausset *et al.* (2008).

2.6 Anatomy of Graphs

There are a variety of graphics that can be used to communicate information, but before they can be explored, a brief overview of the basic elements that make up graphs is required. Figure 2.16 shows the basic structure of most of the graphs in Section 2.7, although there are some exceptions such as pie graphs and radar graphs.

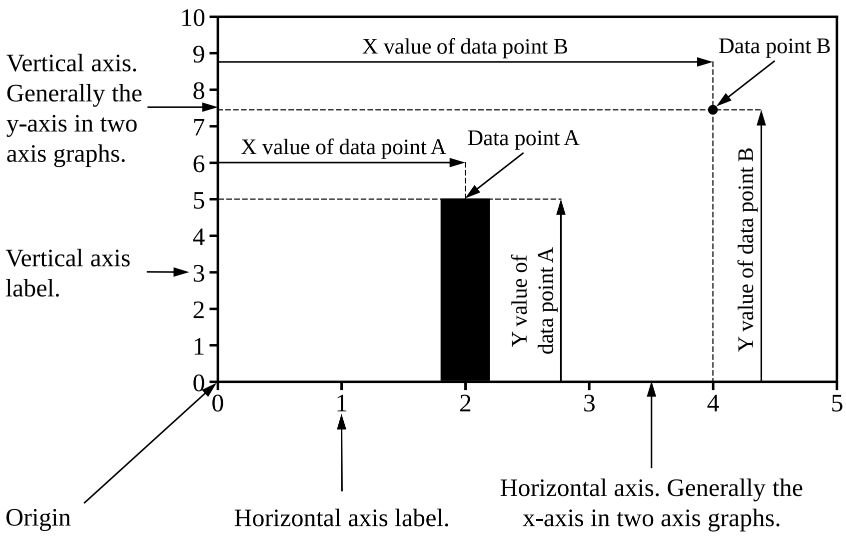


Figure 2.16: Basic elements in graphs.

Adapted from Harris (1999).

Two dimensional graphs typically consist of a vertical axis and a horizontal axis. The vertical axis is normally referred to as the y-axis while the horizontal axis is referred to as the x-axis. According to Harris (1999), axes serve various purposes such as serving as the origin from which coordinates are measured, indicating the orientation of the graph, serving as a mechanism to display tick marks and scales and framing the graph. Generally, the horizontal axis is used to display independent variables while the vertical axis is used to show the values of the dependent variables.

Independent variables are controlled by the researcher or the person conducting the investigation. Any changes to this variable will likely affect changes in the dependent variable(s). Dependent variables are the variables that are expected to change when the independent variable changes in an experiment or investigation (Spatz, 2007). In other words, the independent variable serves as the input while the corresponding dependent variable is the output for that particular input. The types of independent variables used in this thesis include: nominal, ordinal, interval and continuous variables.

2.6.1 Independent Axis

Even though this thesis distinguishes between presentation formats for different tasks rather than for different data types, some presentation formats are only appropriate for certain data types. Section 2.5 mentions a few popular ways of classifying data into different categories. In this thesis, the data will be classified into either nominal, ordinal, interval or continuous. The latter is the only data type which is not categorical, but the continuous data will be divided into equispaced intervals or groups and thus be converted into interval data in this thesis.

According to Shmueli *et al.* (2011), nominal variables do not have any natural or intrinsic ordering. Examples of these are colours of cars such as blue, green and red or names of continents such as Africa, Asia and Europe. Furthermore, Agresti and Kateri (2011) explain that the ordering of these variables is irrelevant and does not influence the statistical analysis thereof.

Ordinal variables are similar to nominal variables, but they do possess a natural ordering (Shmueli *et al.*, 2011). Examples of ordinal variables include the sizes clothes such as small, medium and large or the social classes such as lower, middle and upper or even political position such as liberal, moderate and conservative. However, even though there might be a natural ordering, the distances between the different categories are unknown. Agresti and Kateri (2011) explain that a person who has been classified as moderate is more conservative than a person classified as liberal, but the difference in the amount of conservatism cannot be described using a numerical value. Ordinal values are normally listed or reported according to their natural ordering.

Finally, interval variables are ordered and also have a numerical distance between values. Examples of interval variables are the blood pressure level of a person or the annual salary of an employee. Interval variables can also refer to grouped categories, preferably of equal size. Months in a year are grouped categories of interval variables and the size of each category is equal to the number of days in that particular month. Figure 2.17 gives examples of each of the scales (independent variables) discussed in this section.

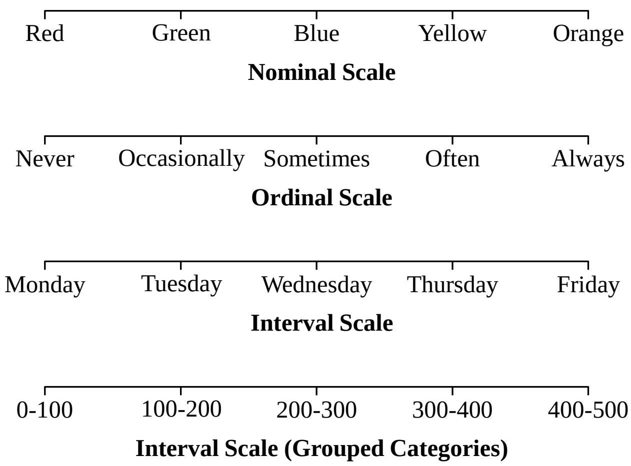


Figure 2.17: Types of independent scales (labels) used in graphs.

Adapted from Harris (1999).

In Figure 2.17, the order in which the colours making up the nominal scale are presented does not matter and presenting them in the reverse order will not influence the result at all. However, the order of the independent variables shown on the ordinal scale needs to be presented in their natural order. Even though trends should not be indicated on graphs with ordinal independent variables, presenting the variables in their natural order can give the reader a rough indication of the distribution of the dependent variables. Finally, interval scales and grouped categories should always be presented in their natural order as it enables users to identify trends and distributions. If the days of the week are presented in order of the size of their dependent variables, it might lead the user to believe that the dependent variable is decreasing or increasing which could be an inaccurate assumption.

For the remainder of this thesis, data will be classified into one of two groups. Data types that are nominal or ordinal will be grouped together as they have similar presentation guidelines. The only condition that applies to ordinal data but not to nominal data is the requirement for ordinal data to be presented or displayed in their natural order. The second group will comprise of interval and continuous data. As mentioned before, continuous data will be divided into grouped categories (bins) of the same size.

The designer can use his or her own discretion when deciding on the sizes of the bins when grouping continuous data into categories, although it is recommended that sensible intervals are chosen when possible. For example, if data is recorded daily for a year, it would be sensible to present it in weekly, monthly or quarterly intervals. However, if there is no clear way in which the continuous data should be divided into intervals, Sturges' rule can be used as a guideline. Sturge's rule is a formula which is traditionally used to determine the number of bins that should be used for histograms based on the number of data points in the sample or dataset (Scott, 2009).

According to Scott (2009), the following formula can be used to calculate the number of bins, k:

$$k = \lceil 1 + \log_2(n) \rceil \tag{2.6.1}$$

where n represents the number of data points in the dataset and its logarithm is taken to the base 2. The $\lceil\ \rceil$ operator indicates that the enclosed value should be rounded up to the nearest integer. If the minimum and maximum values in a dataset are denoted by a and b respectively, equally spaced bins can be constructed by dividing the interval (a,b) into k amount of equally spaced intervals. The size of each bin, h, can be calculated using the following formula:

$$h = \frac{b-a}{k} \tag{2.6.2}$$

For the dataset with range (a,b), the interval for the n^{th} bin is calculated using:

$$\mathrm{bin}_n = (a + h(n-1)\,,\, a + hn) \tag{2.6.3}$$

Once the format of the independent variables has been decided on, data points can be used to indicate their corresponding values in the dataset that is being communicated.

2.6.2 Data Points

A data point represents the value of the dependent variable at a specific independent variable. In Figure 2.16, the two data points (A and B) are indicated using two different conventions. Data point A is shown using a vertical bar (box) that originates at the zero base line (x-axis) and extends to the corresponding y-value of that data point. The height of the bar represents the y-value of the data point and the centre of the bar shows the corresponding x-value. Data point B is indicated using a single marking and the coordinates of its centre represent the x and y-values. If a data point is indicated using a bar, the bar should start at the axis representing the independent variable and extend perpendicular to it. In Figure 2.16, the orientation of the bar representing data point A shows that the independent variable is depicted on the horizontal axis.

An investigation into some of the most frequently used reporting graphics follows in the next section.

2.7 Collection of Graphics Used to Communicate Information

Work by Huang, Chung and Cheng (2008), Adelheid and Pexman (2010) and Harger and Crossno (2010) mention some of the most common graph types which can be used to convey information that can also be tabulated. These graphs will be discussed individually in this section with a focus on the typical application of each. The suitability of each graph and that of its basic variations to communicate nominal, ordinal and interval as well as single and multiple datasets are indicated in Table 2.10 in Section 2.8.

Recall that Section 2.6 discusses independent and dependent variables. In a PAM environment, a typical independent variable could be a range of mines and the dependent variables could refer to the ore processed by each mine. Table 2.7 contains fictional operational data of four mines. The rest of this section describes some the most popular graphs used to convey information. Note that, where possible, the graphs are based on the data in Table 2.7 to demonstrate their possible applications in a PAM environment.

2.7.1 Bar Graphs

Also referred to as: *horizontal bar, horizontal column or rotated column graphs.*

The term bar graph is used loosely in literature and consequently a clear definition of bar graphs is needed before they can be discussed in more detail. Authors such as Petkosek and Moroney (2004), Abu Doush, Pontelli, Simon, Son and Ma (2009), Okan, Garcia-Retamero, Galesic and Cokely (2012) and Borkin, Vo, Bylinskii, Isola, Sunkavalli, Oliva and Pfister (2013) use the term bar graph to describe graphs which are referred to as column graphs by authors

Table 2.7: Fictional operational data for demonstrative purposes.

	Tons produced per mine (Year End 2014).			
Month	Mine A	Mine B	Mine C	Mine D
January	5 460	8 075	9 734	3 151
February	5 208	8 127	7 520	3 027
March	4 713	7 864	6 783	3 182
April	5 201	1 500	10 954	3 186
May	4 038	1 300	13 066	3 001
June	4 090	7 953	11 443	3 101
July	4 306	8 085	9 464	3 029
August	4 175	8 037	7 702	3 170
September	5 055	8 065	13 413	3 028
October	4 435	8 027	9 127	3 123
November	4 289	7 891	8 580	3 119
December	4 013	7 940	8 718	3 123
Total	54 983	82 864	116 504	37 240
Average	4 582	6 905	9 709	3 103
Standard Deviation	500	2 464	2 033	63

such as Harris (1999), Fischer, Dewulf and Hill (2005) and Oakland (2008).

In order to understand this, consider that there are two groups of people: Group A and Group B. Graphs with vertical bars (see Figure 2.22) are referred to as column graphs by Group A and as bar graphs by Group B. However, graphs with horizontal bars (see Figure 2.18) are referred to as bar graphs by Group A and as horizontal bar graphs by Group B. Group B generally also acknowledges that graphs with vertical bars can be referred to as column graphs. Table 2.8 summarises these two groups.

Table 2.8: Bar graph and column graph naming conventions.

	Vertical Bars	Horizontal Bars
Group A	Column Graph	Bar Graph
Group B	Bar Graph Column Graph	Horizontal Bar Graph

The degree to which the two names used to describe graphs with horizontal bars (i.e. column graph and bar graph) are used interchangeably in literature is illustrated in the work done by Abu Doush *et al.* (2009). In the article, Abu Doush *et al.* (2009) discuss Microsoft Excel™ graph construction principles. However, the graph with vertical bars is labelled as being a bar graph in the article, but is constructed using the column graph function in Microsoft Excel™.

Even though the recommendations made by Abu Doush *et al.* (2009) can be used when using the column graph function instead of the bar graph function, it highlights an even bigger problem. The careless use of the term bar graph complicates the analysis of research where remarks are made about bar graphs without providing an example of a bar graph. In other words, it is impossible to deduce whether the authors of the article used the naming convention of Group A or Group B in Table 2.8 when referring to bar graphs.

Abu Doush *et al.* (2009) report that in a survey conducted by Forrester, 87.1% of the 2 300 companies which participated used Microsoft ™ applications. In light of the popularity of Microsoft Excel™ as a graphing tool, the naming conventions of Group A will be used to describe graphs with bars.

Harris (1999) explains that bar graphs use a series of horizontal rectangles to display quantitative information. The length of a bar which represents the value 50 would be twice as long as a bar representing a value of 25. Bar graphs are often used to illustrate and/or compare values of multiple entities at a specific point in time. Data points in a dataset are depicted as rectangles of which the ends are located at the values they represent (see Section 2.6.2). The data series is thus represented by the complete set of bars. The horizontal axis of a bar graph normally has a quantitative or continuous scale which usually starts at zero and extends further than the highest data point in the data set. Annesley (2010) mentions that the independent variables on the vertical axis are generally discrete values meaning that they could be nominal (e.g. eye colour) or grouped intervals (e.g. heights 140-160 centimetres, 160-180 centimetres, 180-200 centimetres).

Figure 2.18 shows a bar graph indicating the tonnage processed by four mines for the year 2014 based on the information in Table 2.7. As mentioned before, the vertical axis is made up of discrete values (names of mines in this case) while the horizontal axis comprises of a continuous variable (tons processed). The horizontal bars originate at the zero value on the horizontal axis and end in line with the corresponding value of the data points they represent on the horizontal axis.

A major advantage of bar and column graphs, according to Abu Doush *et al.* (2009), is that they simultaneously convey absolute information (i.e.

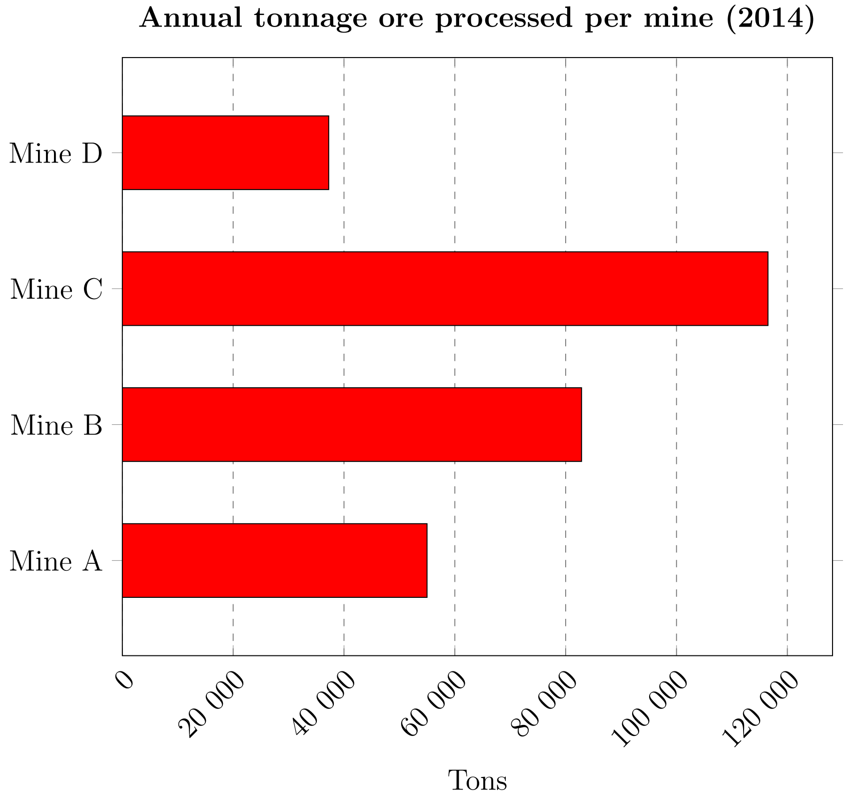


Figure 2.18: An example of a bar graph.

length/height of bar) and relative information (i.e. differences in heights/widths across bars). Schonlau and Peters (2012) and Joshi, Venkatesh and Joshi (2014) agree that these two graphs are ideal when comparing the size of two categories. Figure 2.18 clearly shows that Mine C processed the most ore in 2014 while Mine D processed the least. However, as Fausset *et al.* (2008) and Schonlau and Peters (2012) mention, bar and column graphs are not ideal when determining which sums of proportions are greater. For example, it is difficult to determine whether the total combined tonnage processed by Mine A and Mine B in Figure 2.18 is more or less than the combined tonnage processed by Mine C and Mine D.

Fischer *et al.* (2005) mention that there is a *distance effect* and a *size effect* when processing numbers which is very relevant to bar and column graphs. The distance effect suggests that it is more difficult to determine which of two numbers is the greater or smaller number if the difference between them is small. For example, comparing the numbers 2 and 3 takes longer than comparing the numbers 2 and 9. The size effect implies that when the difference (distance) between two numbers is held constant, discriminating between two small numbers, such as 10 and 15 (difference of 5), is easier than discriminating between two large numbers, such as 90 and 95 (difference 5). It is thus important to realise that bar and column graphs will be more effective when the sizes of the numbers being plotted are small and the differences are large.

Although bar graphs and column graphs seem to have very little differences other than the orientation of the bars, Fischer *et al.* (2005) discovered that human performance varies when using the two graphs. Contrary to their initial expectation, vertically oriented bars (column graphs) produced faster decision times. This is in line with the findings of Feldman-Stewart, Kocovski, McConnell, Brundage and Mackillop (2000) who discovered that column graphs had a lower mean percentage of errors and a lower average median response time than bar graphs.

Two variations of bar graphs and short descriptions of each are provided in Appendix A.

2.7.2 Box and Whisker Plots

Also referred to as: *box diagrams, box graphs, box plots and box symbols.*

Box and whisker plots are based on robust statistics according to Massart, Smeyers-Verbeke, Capron and Schlesier (2005). Robust statistics comprise of methods which are more resistant to the presence of outliers than traditional statistics which focus on normally distributed variables. As Pfannkuch (2006) and Hansen, Chen, Johnson, Kaufman and Hagen (2014) explain, the five-number summary (maximum, upper quartile, median, lower quartile and minimum) provided by the box and whisker plot enables a simple, yet effective, analysis and comparison of distributions. In effect, the box and whisker plot divides an ordered dataset into four segments each containing approximately 25% of the values in the dataset. However, before the box and whisker plot can be explained in more detail, it might be sensible to introduce its components first.

Massart *et al.* (2005) provide a brief, yet adequate description of the components in a box and whisker plot. The median is used as the measure of central tendency (also referred to as Q2 or sometimes the second quartile). If a dataset contains an odd number of entries, the median is defined as the observation in the middle of a ranked series of data points. If there is an even number of data points, the median is the average between the two data points in the middle of the ranked dataset. After the median has been identified, the median of the values remaining below Q2 is identified and labelled as the first quartile (or Q1) and the median of the values remaining above Q2 is labelled as the third quartile (or Q3). If there is an odd number of values in the dataset, the median is included in the remaining values when calculating Q1 and Q3. However, if the number of values is even, the newly calculated median is not used when calculating Q1 and Q3. Examples of how to calculate Q1, Q2 and Q3 in datasets with an odd and an even number of entries are shown in Figure 2.19.

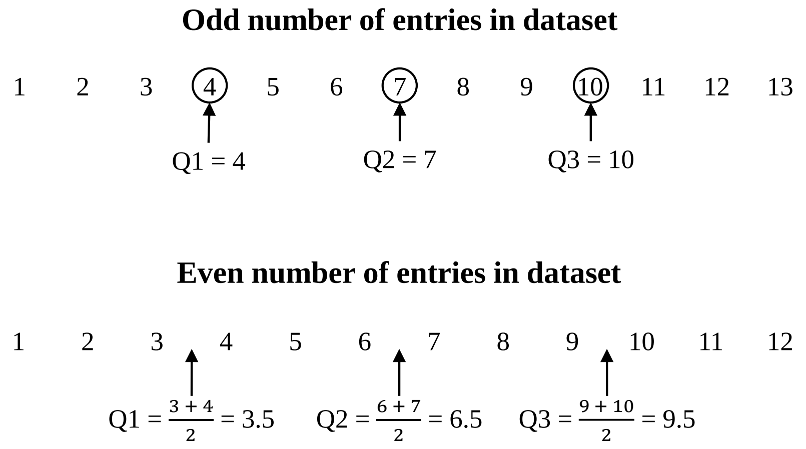


Figure 2.19: Calculating the tree quartiles of box and whisker plots.

Another common component of box and whisker plots is the Interquartile Range (IQR) which describes the way the data is dispersed. Potter, Hagen, Kerren and Dannenmann (2006) explain that the IQR represents the range between the upper quartile and the lower quartile and shows the middle 50% of the values in the dataset. The IQR is calculated using the following formula:

$$\text{IQR} = \text{Q3} - \text{Q2} \tag{2.7.1}$$

Lastly, the values of the fence (also referred to as the extreme value limits) are calculated. The fence extends 1.5 times the height of the IQR above Q3 and below Q1. Although it does not feature on the box and whisker plot, it is used to identify outliers which fall outside the permitted range. Data points which fall outside the allowable range are indicated using alternative markings such as crosses or dots. Figure 2.20 provides the basic anatomy of a box and whisker plot.

Hubert and Vandervieren (2008) provide the following step-by-step explanation of the how to construct a box and whisker plot using the components described above:

1. A horizontal line is constructed at the height corresponding to the value of the median (Q2) of the sample and its width determines the width of the box.

2. Two lines of the same length are drawn parallel to the median, one above it and the other below. The line above the median is drawn at the height

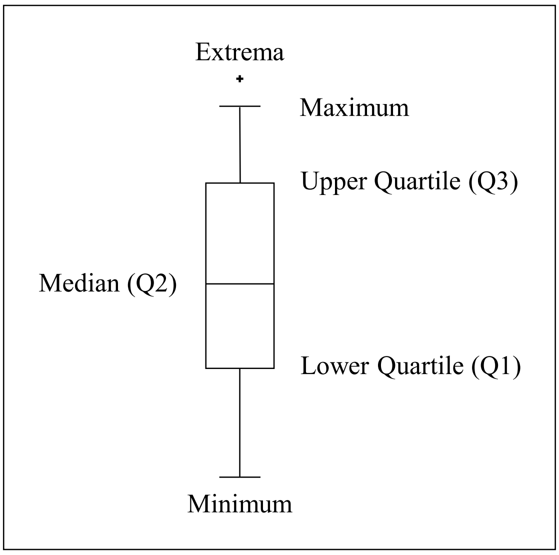


Figure 2.20: The anatomy of a box and whisker plot.

Adapted from Hansen *et al.* (2014).

of the third quartile (Q3), while line below the median is drawn at the height of the first quartile (Q1). Two parallel vertical lines connect the ends of the lines at Q1 and Q3 to form a box. Note that the height of the box is equal to the IQR.

3. All data points that fall outside the fence are identified. The fence is described as the maximum and minimum limits for values not to be shown as outliers and can be calculated using Equations 2.7.2 and 2.7.3.

$$\text{Fence}_{\text{min}} = \text{Q1} - 1.5\text{IQR} \tag{2.7.2}$$

$$\text{Fence}_{\text{max}} = \text{Q3} + 1.5\text{IQR} \tag{2.7.3}$$

Note that the fence is only used to identify values that should be indicated as outliers, but it is not shown on the box and whisker plot.

4. A whisker is drawn from the bottom of the box at Q1 to the lowest data point which is not an outlier (i.e. lowest data point larger or equal to $\text{Fence}_{\text{min}}$). Another whisker is drawn from the top of the box at Q3 to the highest data point that is not an outlier (i.e. highest data point smaller or equal to $\text{Fence}_{\text{max}}$).

5. All points outside the fence are indicated using individual markings.

As mentioned before, one of the main advantages of box and whisker plots is that they facilitate simple comparisons of distributions of datasets. Consider the distributions of the monthly production outputs of Mine B and Mine

C in Figure 2.21. When considering the summary data/statistics shown in Table 2.7, little can be deduced regarding the distributions of the outputs of Mine B and Mine C. The only statistic that comments on the distribution of the output is the standard deviation, which is similar for both mines.

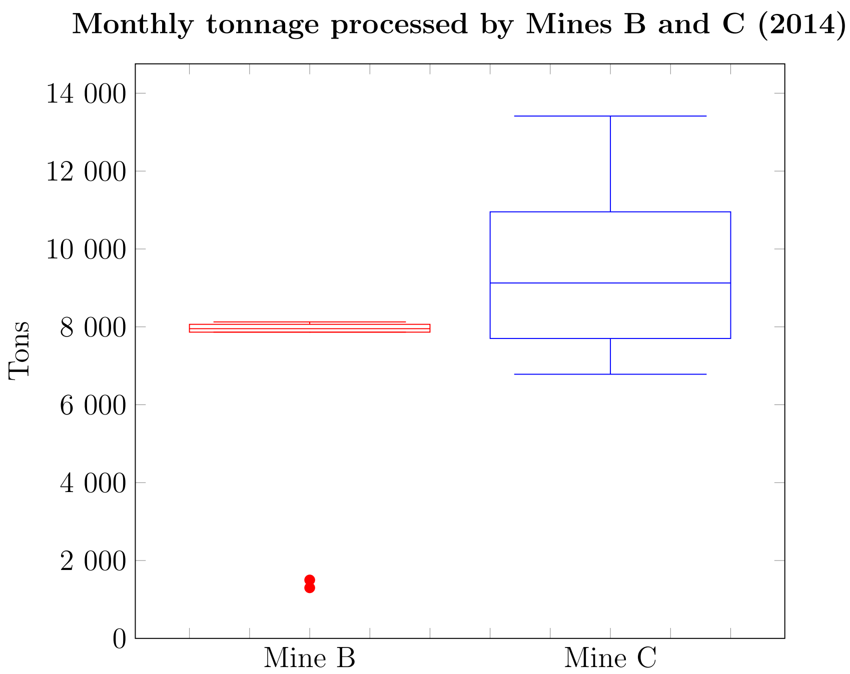


Figure 2.21: An example of a box and whisker plot.

When, however, considering the box and whisker plots of the outputs in Figure 2.21, it becomes clear that there is little variation in the monthly production outputs of Mine B other than the two outliers (indicated by the dots). Mine C, on the other hand, has a large variation with an approximate IQR of 2 500 tons. In a PAM environment, a typical application of the box and whisker plot would be to identify anomalies (like the two outliers of Mine B) and to see which processes are stable and which could benefit from process quality improvement programs (for example Mine C with high production variation).

A disadvantage of box and whisker plots, however, is that they do not indicate any independent variables, but only the distribution of the dependent variables. Although Figure 2.21 shows that Mine B had two outliers, the two months in which these low amounts of ore were processed cannot be retrieved from the graph.

2.7.3 Column Graphs

Also referred to as: *bar, rotated bar and vertical bar graphs.*

Column graphs (also referred to as bar graphs, see Section 2.7.1) consist of vertical rectangular bars (columns) each representing a data element in a dataset. The height of a column indicates the value of the data element that is being represented by the column (Harris, 1999). As with bar graphs, the height of a column which represents the value 50 would be twice as high as a column representing a value of 25. Abu Doush *et al.* (2009) and Adelheid and Pexman (2010) mention that column graphs are ideal for illustrating differences between groups or independent variables, although they are sometimes also used to illustrate change over time.

Figure 2.22 shows a column graph indicating the tonnage processed by four mines for the year 2014. Notice how column graphs consist of the same components as bar graphs (see Figure 2.18), but with different orientations. The independent variables are located on the horizontal axis while the continuous dependent variables are located on the vertical axis. In Figure 2.18, the horizontal axis shows the names of mines while the vertical axis shows the number of tons processed. The vertical bars originate at the zero value on the vertical axis and end in line with the corresponding value of the data points they represent on the vertical axis.

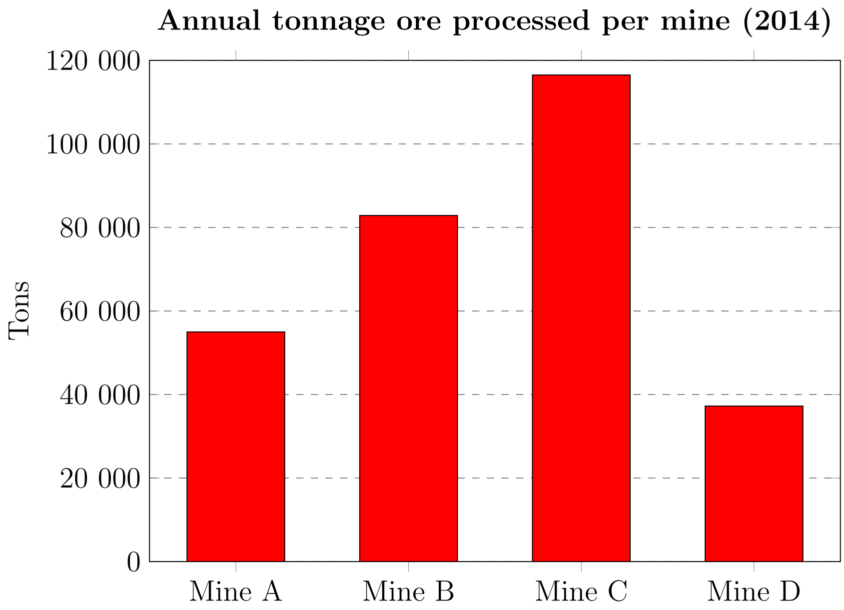


Figure 2.22: An example of a column graph.

If the range of the values of the data elements being displayed is very large, Hager, Scheiber and Corbin (1997) suggest that designers break the dependent axis using a double slash (//). Furthermore, this technique is not limited to column graphs, but can be used in any two dimensional graph with a quantitative dependent axis. However, this is not encouraged as it can give a false impression of the distribution of the variable(s) being represented.

Two variations of column graphs and short descriptions of each are provided in Appendix A.

2.7.4 Histograms

Also referred to as: *frequency diagrams and frequency polygons.*

According to Steinberg (2010), histograms are normally used to convey continuous data. Histograms are similar to column graphs in their construction, but the columns are connected (whereas the columns in a column graph are separated). The x-axis is normally populated with quantitative data such as age and distance which are then grouped into bins (intervals) (Adelheid and Pexman, 2010). Equations 2.6.1, 2.6.2 and 2.6.3 in Section 2.6.1 can be used to determine the number, widths and intervals of bins. Jongen, Smit and Janssen (2008) state that histograms can facilitate simple analysis of data in AM. In a histogram, the heights of the bars indicate the observed frequencies of the data points in the corresponding bins on the x-axis. Adelheid and Pexman (2010) and Steinberg (2010) emphasise that it is imperative that the intervals on the x-axis are arranged from smallest to largest to illustrate continuity. Ideally, the intervals on the x-axis should include all possible data points in the population.

It is difficult to illustrate the value of histograms using the data in Table 2.7. Consider the following range of ages for 40 pieces of equipment at a mine (in years rounded down):

10, 6, 16, 8, 1, 20, 2, 9, 17, 4, 12, 18, 7, 2, 16, 4, 19, 17, 3, 11, 20, 4, 0, 17, 4, 9, 4, 20, 1, 20, 15, 10, 3, 17, 5, 2, 15, 1, 13, 16.

The minimum age is zero years and the maximum is 20 years with an average age of 9.9 years. If it is the company's policy to replace equipment after 20 years, the manager might feel comfortable with an average equipment age of 9.9 years. However, a closer inspection reveals that the distribution is not uniform and that there are in fact 13 pieces of equipment (26%) that are 16 years or older. Figure 2.23 shows an example of the equipment ages illustrated in a histogram.

Figure 2.23 highlights that there is a large proportion of the equipment approaching end of life status. The use of a histogram can aid managers in

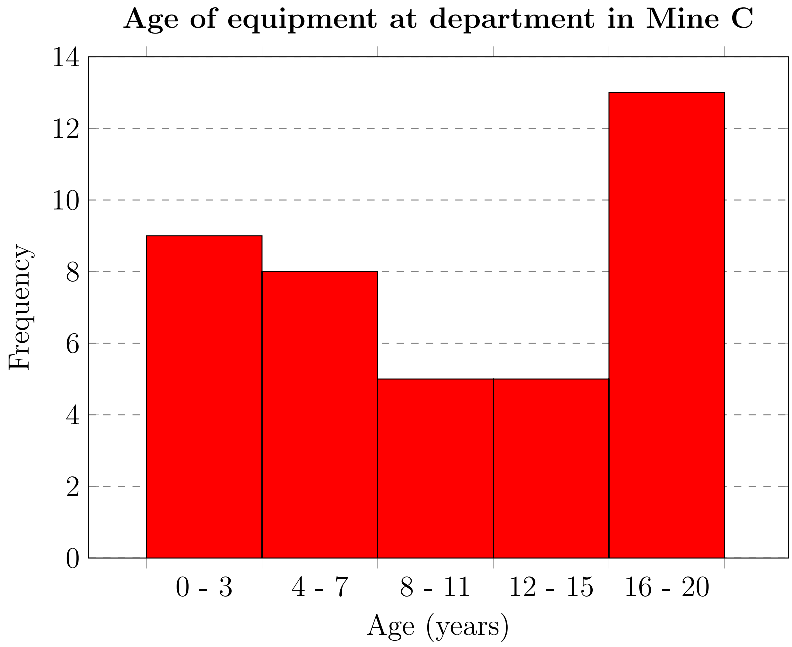


Figure 2.23: An example of a histogram.

noticing irregularities in the data.

2.7.5 Line Graphs

Also referred to as: *curve graphs.*

Line graphs connect a number of individual data points over a change in the independent variable using line segments. Fausset *et al.* (2008) explain that the independent variables are placed on the x-axis (horizontal axis) while the magnitudes of the corresponding dependent variables are shown on the y-axis (vertical axis). Fausset *et al.* (2008), Abu Doush *et al.* (2009) and Adelheid and Pexman (2010) all agree that line graphs are generally the best format to depict trends or change over time. However, as Peebles and Ali (2009) mention, line graphs can only be used to show changes if the data are ordered in some way (i.e. ordinal or interval data). However, by definition, ordinal data do not have constant numeric intervals between data points and thus any trend that is indicated might be misleading. For this reason, only the trends of interval data will be illustrated in this thesis. Figure 2.24 shows an example of a line graph in a PAM environment.

The markings (dots) in Figure 2.24 indicate the amount of tons (dependent variable) produced each month in 2014 (independent variable) at Mine B. Note that the markings are not compulsory as the line passes through the

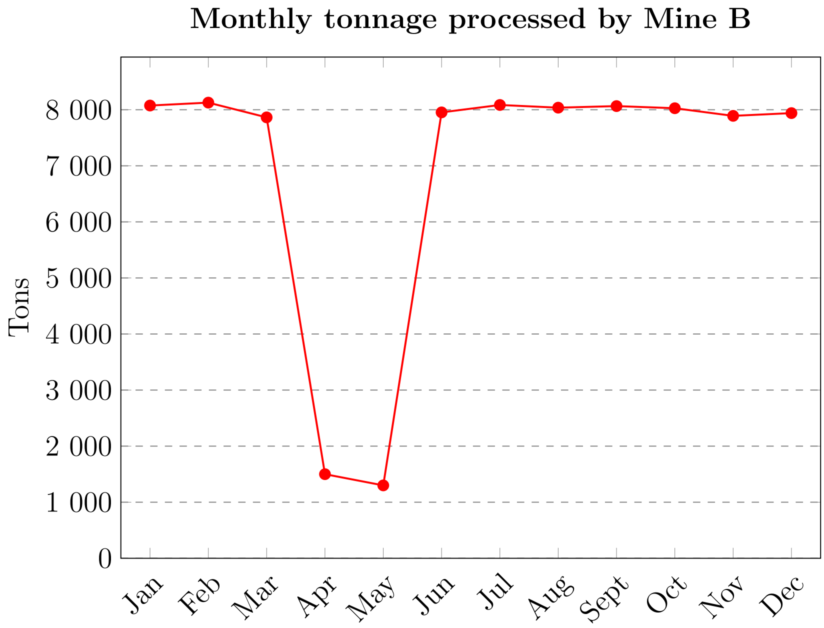


Figure 2.24: An example of a line graph.

exact data point regardless of whether the markings are present or not.

A variation of the line graph is the *area graph*. According to Harris (1999), a line graph showing one data series can be converted into an area graph by filling the area between the data curve and the horizontal axis. As with line graphs, area graphs are not as effective when used to show specific values as they are when used to convey trends or relationships, place emphasis on specific information or to show parts of the whole. When area graphs are used to show two or more data series, there are two formats which can be used: grouped area graphs and stacked area graphs. An example of a stacked area graph is shown in Figure 2.25.

The main difference between the two types of area graphs is that the bottoms of all the data series plotted on the grouped area graph start at the horizontal axis and the top of each filled area represents the actual value of that data series. A possible problem with this format is filling the areas when the lines used to indicate the various data series cross. This generally happens when a data element of one data series is larger than a data element of another data series at a particular value of the independent variable and then smaller than the data element of the other data series at the next value of the independent variable. Contrarily, with stacked area graphs, each data series is plotted on top of the previous data series and the actual value of the data series at a certain point is represented by the vertical height of the filled area. Only the bottom series touches the horizontal axis and the top of the upper

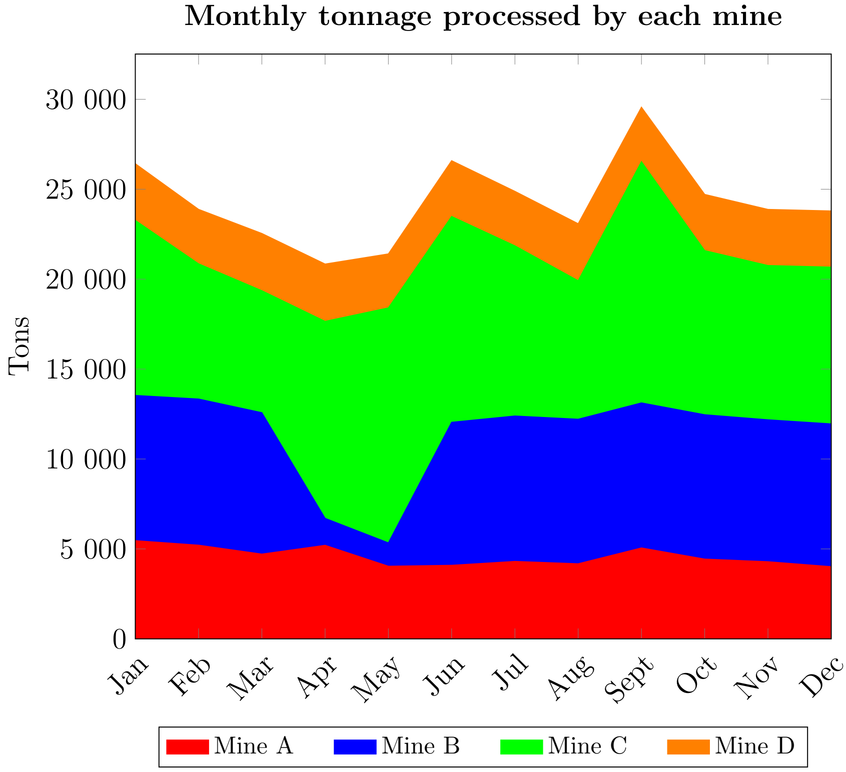


Figure 2.25: An example of a stacked area graph.

data series shows the grand total of all the data series plotted. As a result, a stacked area graph illustrates both the individual data elements and the sum of all the data elements for each particular independent variable on the x-axis.

In Figure 2.25, the top left point of the upper data series (representing Mine D), shows the total tons produced of all of the mines for January 2014. The top left point is located at a y-value of 26 420 which is the sum of 5 460, 8 075, 9 734 and 3 151 tons produced in January by mines A, B, C and D, respectively (see Table 2.7).

2.7.6 Pie Graphs

Also referred to as: *cake charts, divided circles, circular percentage charts, sector charts, circle diagrams, sectograms, circle graphs or segmented charts.*

In its most basic form, a pie graph consists of a circle which is divided into a number of wedge-shaped sections originating at the centre point of the circle and extends to the circumference of the circle. Wedges are thus made up out of two straight lines with lengths equal to the radius of the circle and an arc (which forms part of the circumference of the circle). Harris (1999) clarifies that the area of each of the sections is the same proportion of the complete circle as the data element being represented by it is to the sum of all of the elements in the data set. According to Harris (1999), the proportion of the data element can be determined in three ways. These three ways are:

1. The angle of the wedge (out of the 360 degrees in a circle).
2. The area of wedge (in comparison to the total area of the circle).
3. The length of the arc (in comparison to the circumference of the circle).

Even though Joshi *et al.* (2014) stress the popularity of pie graphs, Siirtola (2014) states that pie graphs have produced conflicting experimental results with some practitioners criticising them heavily while others defend their use. In Table 2.9, Few (2013) summarises how well pie graphs perform in the five effectiveness criteria he uses to evaluate graphs.

In Figure 2.26, the tonnage produced by each of the four mines in Table 2.7 is represented by a wedge. Each wedge has an area with the same proportion to the area of the full pie graph as the proportion that the data element it represents has to the sum of the data elements in the dataset. For example, note that Mine C produced 116 504 tons of the total 291 591 tons produced in 2014 which equates to 39.95% of the total production. The area of the wedge representing Mine C in Figure 2.26 thus covers 39.95% of the area of the pie graph. In accordance with the comments of Harris (1999), the angle formed by the wedge representing Mine C and the length of its arc are 39.95% of the total degrees (360) and the total circumference of the pie graph respectively. The same is true for Mine A, B and D with proportions of 18.86%, 28.42% and 12.77% respectively.

Hollands and Spence (1998) found that pie graphs outperformed column graphs when judging part-to-whole relationships (when measuring response time and accuracy). Hollands and Spence (1998) concluded that column graphs were generally outperformed by pie graphs when judging proportions and even more so when the number of components (data elements) is increased.

Schonlau and Peters (2008) investigated the effect of adding a third dimension to pie graphs (i.e. three dimensional wedges, but with a constant thickness throughout). It was found that response accuracy was significantly worse for three dimensional graphs than for two dimensional graphs. The conclusion of Schonlau and Peters (2008) was confirmed by Stewart, Cipolla and Best (2009) who tested the response accuracies for pie graphs using easy, moderate and difficult tasks. It was found that response accuracies when using three dimensional graphs were lower for all three tasks when compared to two dimensional

Table 2.9: Summary of advantages and disadvantages of using pie graphs.

Criteria	Answer	Comments
Nature of relationship indicated clearly?	Yes	The ability of a pie graph to indicate part-to-whole relationships is clearly its primary strength.
Quantities represented accurately?	No	The three visual attributes used when encoding values (the area of the wedge, the angle formed by the wedge and the length of the wedge's arc on the circumference) results in visual redundancy. Schonlau and Peters (2012) and Few (2013) agree that humans are not capable of decoding angles, areas and curved distances accurately.
Quantities can be compared accurately?	No	Values cannot be compared if they cannot be perceived accurately. However, pie graphs do facilitate the comparison of proportions.
Shows the ranking of variables?	No	Even though it is a good practice to arrange the wedges from largest to smallest (or vice versa), identifying subtle differences between two slices is not simple. Even though it is simple to see that a wedge which covers 50% of the area is larger than a wedge covering 20% of the area, it is more difficult to find and rank three wedges covering 20%, 21% and 22% of the area (especially if they are not adjacent and/or ranked).
Indicates the planned application of the information?	Partially	Encoding information using pie graphs encourages the comparison of data elements with each other, but the operation is not supported effectively by the graph.

Adapted from Few (2007).

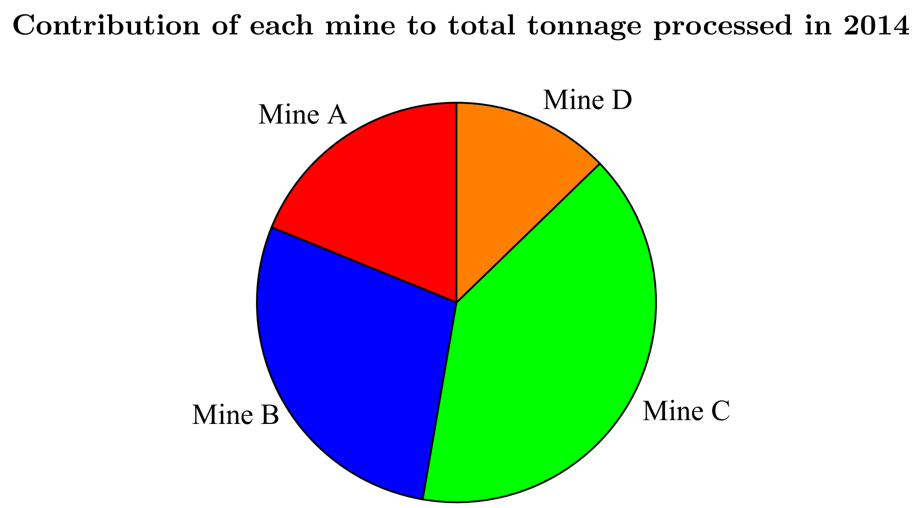


Figure 2.26: An example of a pie graph.

graphs.

An extensive review of pie graphs was done by Few (2007) to determine whether pie graphs have any practical advantages. The only advantage of using a pie graph that Few (2007) reported with confidence was based on a study by Spence and Lewandowsky (1991). Spence and Lewandowsky (1991) investigated the relative effectiveness of conveying part-to-whole relationship information using pie graphs, column graphs and tables. Pie graphs had an advantage over the other two formats when the task required participants to compare the relative sizes of two sets of summed parts. An example of this task using Figure 2.26 would be to compare the combined output of Mine A and Mine B to the combined output of Mine C and Mine D. Few (2007) states that it is easier to sum the areas of the wedges than it is to imagine the combined lengths of bars. Refer to Figures 2.18 and 2.22 for a bar graph and a column graph which is based on the same information as Figure 2.26.

Very recently, Kozak, Hartley, Wnuk and Tartanus (2015) argued that, although there might be advantages when considering individual pie graphs, multiple pie graphs should be avoided. It was found that multiple pie graphs are difficult to analyse and interpret and even more so when they are adjacent.

2.7.7 Radar Graphs

Also referred to as: *circle diagrams, Kiviat graphs, polygons, spider graphs, star graphs and competency reports.*

Harris (1999) and Saary (2008) describe a radar graph as a comparative tool which consists of a series of spokes originating at a central point. Each spoke represents a different discrete independent variable with a minimum value at the origin and a maximum value at the end of each spoke (or vice versa if a smaller value is better than a large value). Goldberg and Helfman (2011) explain that radial graphs encode quantitative values on axes which all originate at a centre point and radiate outward. In effect, radar graphs are line graphs turned into circular visualisations, but without the restriction that the independent variables have to be ordinal (as with line graphs). The only limitation on the number of spokes is that there has to be more than two and not too many as to affect readability negatively.

Figure 2.27 shows an example of a radar graph which conveys hypothetical information about two potential employees. A possible application of such information is when a manager needs to decide which one of the two employees should be promoted. Each employee received a rating out of ten in five different categories that are considered important for the management position. Note that all of the spokes do not have to have the same scale (although it is the case in this example). The scores of each employee for each of the different categories are marked on the spokes and then connected using straight lines. Both employees had the same score for *communication skills* with Employee A getting higher scores than Employee B in the *experience* and *team player* categories and while Employee B outperformed Employee A in *training* and *motivation*.

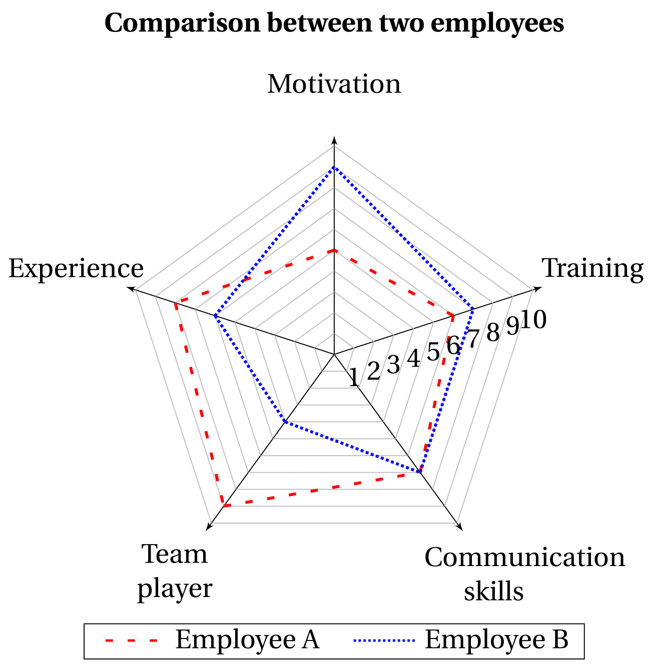


Figure 2.27: An example of a radar graph.

Harris (1999) explains that one of the advantages of radar graphs is that they can compare data series with many variables. Not only can individual data points of different datasets be compared, but also the areas enclosed by the polygons formed by the data points. By overlaying two or more datasets on the same radar graph, the shapes formed by datasets can be compared and any differences can be investigated further.

Another advantage of using radar graphs is their ability to highlight deficiencies or strengths in performances between two groups. Areas where one group's dataset balloons out or caves in can be investigated to find possible root causes of variations in performances.

As with all graph types, there are certain disadvantages associated with the use of radar graphs. Subtle differences are difficult to identify. If the scale chosen on each axis/spoke is too large, then the differences between two sets of datasets will become imperceptible. However, Harris (1999) still recommends that the scales of the axes should be similar. Finally, radar diagrams are most effective when two datasets are compared to each other or when a dataset is compared to the norm. Plotting too many datasets on the same radar graph can confuse readers and can complicate comparisons.

2.7.8 Point Graphs

Also referred to as: *dot plots, scatter plots, scatter graphs and scattergrams.*

Adelheid and Pexman (2010) describe point graphs as representations of single events as a function of two quantitative variables which are indicated along the x-axis and y-axis. Harris (1999) mentions that point graphs are frequently used to explore the relationships and correlations between two or more data sets. The patterns formed by the plotted data points are then analysed to find relationships about the datasets graphed. Trend lines may be used to approximate relationships once they are identified. Note that the relationships may be linear, non-linear or even random (no relationship).

According to Louit, Pascual and Jardine (2009), plotting cumulative failures of an asset (or an asset system) against time is the simplest method for testing a trend. The plots in Figure 2.28 labelled A, B, C and D show an increasing trend, no trend, two clearly different periods and a non-monotonic trend respectively. By analysing these trends, predictions and assumptions about the datasets they represent can be made.

The cumulative failures of Machine A are plotted against its age (in hours) in Figure 2.29. Notice how the pattern formed by the plotted data in Figure 2.29 resembles that of plot A in Figure 2.28. There is thus an increasing

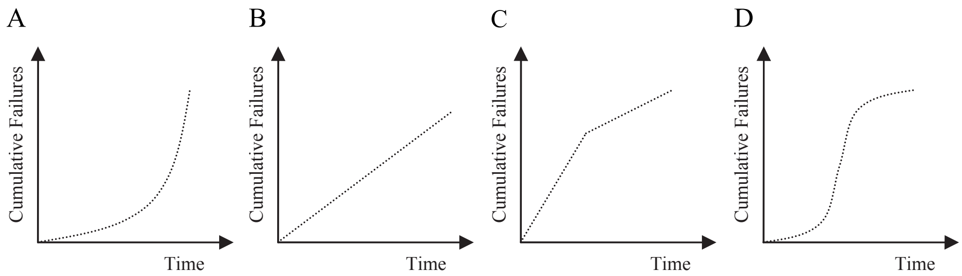


Figure 2.28: Examples of cumulative failures vs. time plots.

Adapted from Louit *et al.* (2009).

trend in the data which shows that the times between failures has decreased as Machine A got older. This information can be used to show management that the times between failures of Machine A are decreasing and that Machine A should be serviced or replaced.

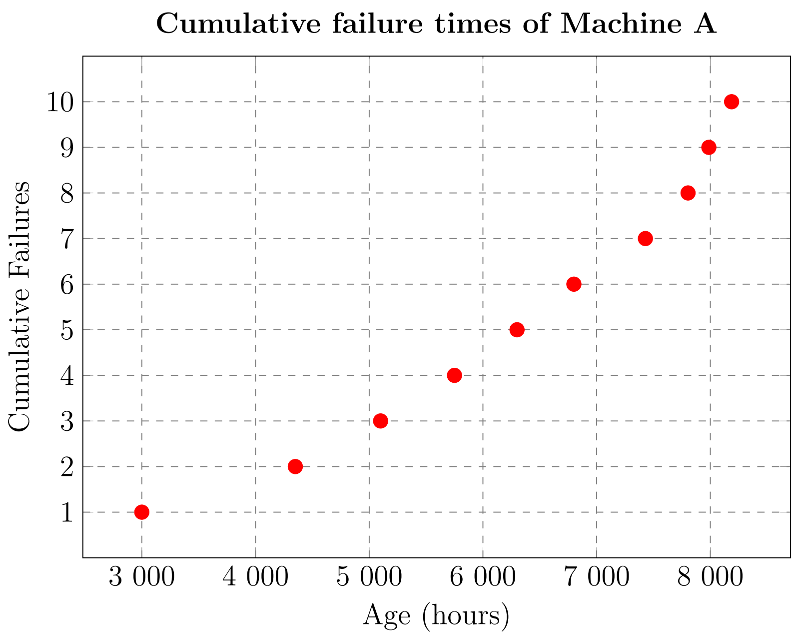


Figure 2.29: An example of a scatter plot.

There is no restriction on the types of axes used in point graphs. In other words, both the horizontal and the vertical axes can be either nominal, ordinal, continuous or interval. This property makes point graphs appealing when indicating events without set independent variables.

2.8 Chapter Summary

This chapter commences with a discussion of PAM and its importance in today's industrial climate. Special focus is placed on human assets, leadership, data and communication within PAM environments to contextualise the problem that is being addressed in this thesis. Three communication models are discussed and from these models it is found that the receiver of the message determines the success of the entire communication process. The cognitive load or strain that is placed on an individual's working memory is discussed and it is found that the task performance is influenced by the total cognitive load of a task. This cognitive load can be reduced by using communication formats with which the user is familiar or by choosing a format which matches the requirements of the task being attempted.

Three different information presentation formats used to communicate processed data are discussed, namely: text, tables and graphics. It is found that using text is only appropriate when communicating small amounts of data and thus this research only focuses on tables and graphics. Although some formats are only appropriate for certain types of data and numbers of datasets, the choice of format should be determined by the task that needs to be completed using the data rather than the characteristics of the dataset. The chapter concludes with a discussion of popular graphics used to communicate data. These formats are listed in Table 2.10 and their suitability for communicating different data types as well as the number of datasets they can convey are also indicated.

The following five objectives identified in Section 1.3 were addressed and achieved in this chapter:

1. Gain an understanding of the role of data and leadership in AM environments.
2. Establish the fundamental principles of communication processes between two entities as described by previous researchers.
3. Investigate how various elements of visualisations influence an individual's cognitive processes.
4. Determine which tasks are typically supported by data.
5. Explore different formats in which information can be encoded as well as the types of data for which they are appropriate.

In Chapter 3 a proposed solution to the problem statement is developed based on the body of knowledge discussed in this chapter.

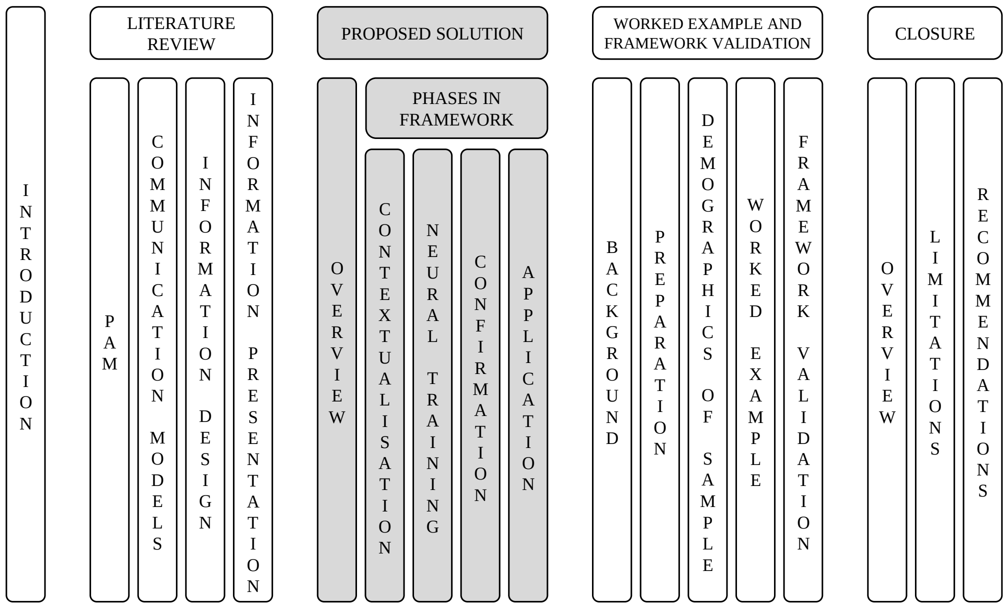
Table 2.10: Collection of graphs and their suitability to communicate different data types and numbers of datasets.

Format	Nominal or Ordinal	Interval or Continuous	Single Dataset	Multiple Datasets
Table	Yes	Yes	Yes	Yes
Point Graph (Dot Plot)	Yes	Yes	Yes	Yes
Radar Graph	Yes	Yes	Yes	Yes
Bar Graph (Simple)	Yes	Yes	Yes	No
Bar Graph (Grouped)	Yes	Yes	No	Yes
Bar Graph (Stacked)	Yes	Yes	No	Yes
Column Graph (Simple)	Yes	Yes	Yes	No
Column Graph (Grouped)	Yes	Yes	No	Yes
Column Graph (Stacked)	Yes	Yes	No	Yes
Pie Graph	Yes	Yes	Yes	No
Pie Graph (Multiple)	Yes	Yes	No	Yes
Box and Whisker	No	Yes	Yes	Yes
Histogram	No	Yes	Yes	No
Histogram (Multiple)	No	Yes	No	Yes
Line Graph	No	Yes	Yes	Yes
Line (Area) Graph	No	Yes	Yes	Yes

Chapter 3

Proposed Solution

The preceding chapter shows that data is ever more prevalent in PAM environments as it is captured continuously by automated systems. Data mining and analysis techniques have enabled the conversion of data into information which can be used in both everyday and long term decision making, but the presentation formats selected to convey the information is often suboptimal. Chapter 2 established that information presentations should be tailored to the needs of the target audience while still considering the task which needs to be completed using the information.



This chapter uses the literature discussed in Chapter 2 as the basis to propose a framework which can be used to encode information in a manner which is appealing to the targeted audience, or at least the majority thereof.

3.1 Overview

The literature in Chapter 2 underlines the importance of basing PAM decisions on objective and factual information. Although data analysis techniques have enabled organisations to convert raw data into usable (and often invaluable) information, the decision making process is often dictated by the opinions of experienced and powerful individuals who do not necessarily consider all the facts and information available to them. Even when individuals are receptive to the information, it is often encoded in a manner which does not match the needs of the target audience. This can lead to the individuals being frustrated and the information being disregarded.

All of the communication models discussed in Chapter 2 indicate that the success of the communication effort is determined by the receiver of the message. If the message is encoded in a way that is inappropriate for the needs of the receiver, the communication has failed, regardless of the success of the preceding or subsequent steps in the process. Cognitive load theory in Section 2.3.1 shows that if the mental effort required by a task exceeds the mental capabilities of an individual, the individual's performance in that task will decrease significantly. However, if the information is presented in a format which supports the task which needs to be executed using the information (termed *cognitive fit*), the task performance of the individual can be improved substantially.

The aim is thus to develop a solution which can be used to ensure that information is encoded in a way which matches the needs and preferences of the majority of the targeted audience while, at the same time, considering the task that needs to be executed with the support of the information. Instead of measuring traditional task performance metrics (such as response time or response accuracy), the information presentation preferences of the receiver will be determined for different tasks. Successful development and application of the solution will guide employees in deciding on appropriate information presentation formats when reporting numerical information. Although the proposed solution is intended to address information reporting problems in asset intensive industries, the structured steps can be applied to various domains where information reporting is seen as a problem.

This section discusses the development and features of an information encoding framework and will serve as an introduction to the comprehensive exploration of the model in the subsequent sections.

3.1.1 Development of Framework

As mentioned before, the proposed solution is a framework which aims to serve as a guideline when deciding on information presentation formats. The three communication models in Section 2.2 are considered to be adequately comprehensive to create a basic understanding of human communication. Section 2.2 discusses the four generalised components of communication (source, message, channel and receiver) and it is discovered that for communication to be successful, it has to be tailored to the needs of the reader or receiver (target audience). Figure 3.1 is a graphical representation which summarises how the proposed framework aims to tailor the presentation of information to the needs of the target audience.

Notice that the framework has four main phases, namely: contextualising the framework, training the framework, confirming the preferences stored in the framework and finally, applying the framework. Three of the four phases are compulsory, while the confirmation phase is optional. As mentioned before, the framework will be applicable to most data-intensive industries, but it still has to be adapted to consider the organisation's unique attributes. The contextualisation phase ensures that the recommendations made by the framework are applicable to the organisation where it is being implemented. Neural training, the second phase, is generic and aims to determine the data presentation preferences of the employees and to store or update them in a repository (typically a database). Confirming the preferences is optional and aims to ensure that the preferences stored in the database are in fact the correct ones.

The purpose of the first three phases (contextualising, neural training and confirmation) is to ensure that the employee preferences which are collected are relevant, reproducible, easily obtainable and correct. Once the first three phases have been completed, the output will be used repeatedly without further collection of information or preferences unless new employees join the workforce. In this case, only the information and preferences of those new employees will be collected. It is advised that the first three steps be reviewed every one to two years.

Finally, the application phase contains the steps that are recommended when the framework is used to aid in the selection of appropriate information presentation formats. It is important to note that the application phase uses the repository created and populated in the neural training phase as its source. This means that the quality of the ultimate output of the framework depends on the quality of the information stored in the repository.

The steps in Figure 3.1 are numbered and serve as a guideline for the sequence in which they should be performed. Since the confirmation phase

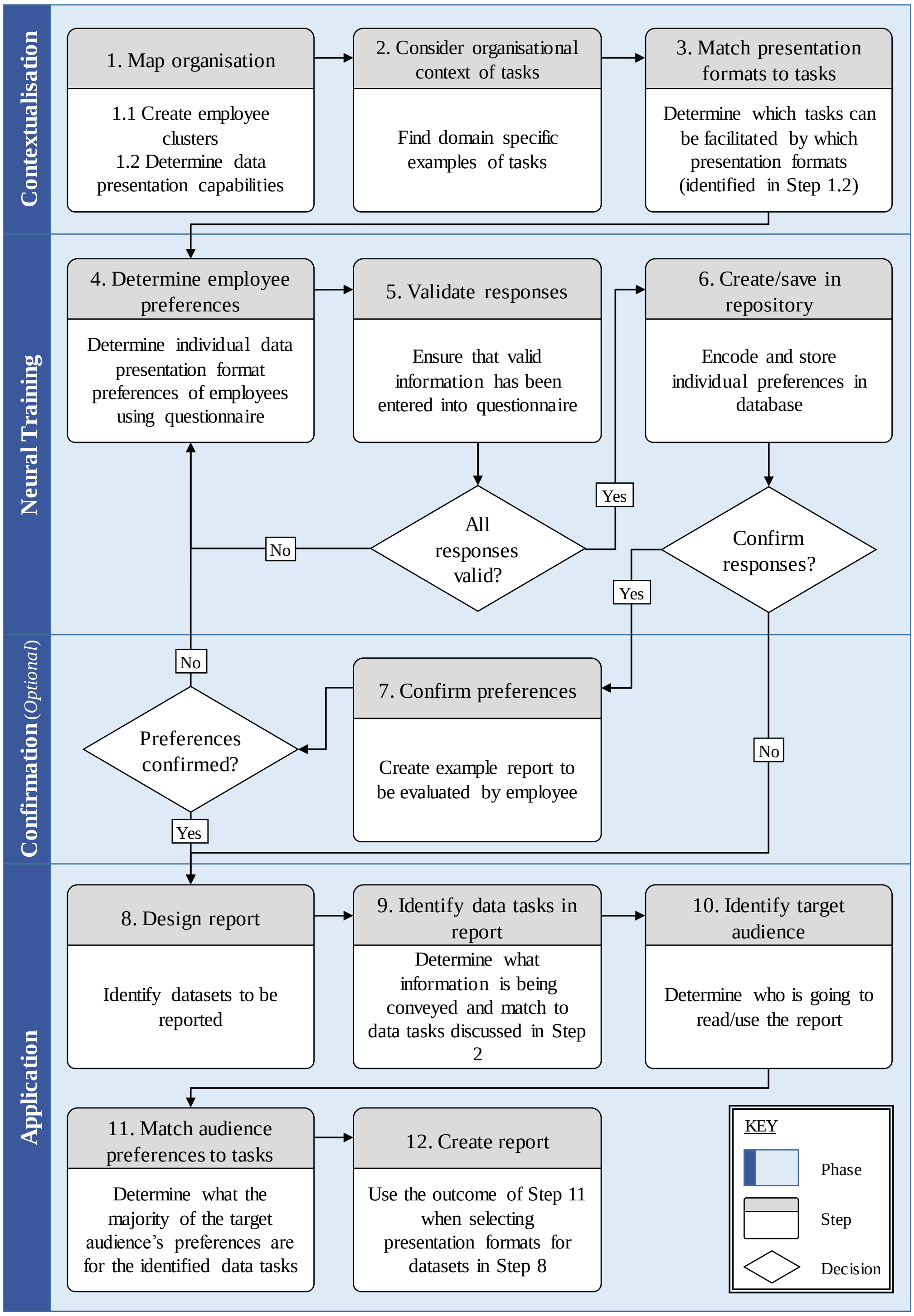


Figure 3.1: Proposed information encoding framework.

is optional, there might be instances where the user of the framework ignores step seven and moves directly from step six in the neural training phase to step eight in the application phase. As mentioned before, once the first three phases of the framework have been completed, only the steps in the application phase will be repeated (i.e. steps 8 - 12).

3.1.2 Required Framework Features

Chapter 1 describes some of the biggest problems faced by data-intensive industries. If the framework is to be considered a viable solution to these problems, it needs to meet the following requirements identified in Section 1.3:

- Practical: The framework should serve as a tool which can be used in practice with clear benefits to its users. If the framework is purely based on theory with no consideration for the environment in which it will be applied, it will not be adopted in practice.
- Generic: Although the framework will be designed to address a problem which is popular in PAM environments, it should be implementable across a variety of industries.
- Flexible: In order to address both inter-organisational and intra-organisational differences, the framework should be adaptable to specific situational needs. This requirement will also allow the framework to be tailored to the organisation where it is being implemented.
- Structured: Users of the framework should be able to follow a set sequence of steps which will lead to a similar output regardless of the user.

Section 2.1.3 mentions that managers in PAM environments have a variety of responsibilities and are faced with a plethora of decisions on a daily basis. Furthermore, even the ground level employees are required to perform at a very high level in order for asset intensive industries to remain competitive. This framework aims to decrease the effort needed by the whole range of employees to interpret reports without placing additional strain on the designers of the reports.

It is necessary to realise that the framework will not predict the preferences of employees based on certain characteristics such as age, education and personality. The framework will receive information about both the datasets to be reported and the target audience as inputs and will recommend presentation formats which best satisfy the needs of the majority of the target audience. In order to accomplish this, the framework needs to be populated with preference data which can then be used repeatedly to produce outputs which satisfy the criteria.

All four phases of the framework and their subsequent steps are discussed in more detail in Sections 3.2, 3.3, 3.4 and 3.5. The majority of the steps in the first two phases (contextualisation and neural training) are discussed using the following structure:

1. Purpose: The reason why this step is important.
2. Theoretical grounding: How this step is supported by literature.
3. Artefact/Example: A typical output of this step populated with hypothetical information.
4. Value proposition: How does this step contribute to the framework.

The purpose and the value proposition of the third phase (confirmation) are also indicated, but the theoretical grounding and artefact/example are replaced by a summary. In the application phase, the purpose of each step is followed by a short summary and an example output of the step.

Note: Data tasks refer to tasks that need to be executed using data.

3.2 Phase 1: Contextualising the Framework

As mentioned in Section 3.1.2, the framework will be adapted to the organisation in which it is being implemented. This will be done by mapping the organisation, considering the organisational context of the data tasks and finding presentation formats which the organisation is capable of creating to facilitate those tasks. Each step is explained in more detail below.

3.2.1 Step 1: Map Organisation

Since the framework aims to find the most appropriate presentation formats to use when presenting information to the organisation's employees, it seems appropriate to start by reviewing both the employee clusters and the information presentation capabilities of the organisation.

3.2.1.1 Step 1.1: Create Employee Clusters

Purpose: To identify attributes which can be used to create employee clusters. This step will enable designers to ensure that only the preferences of the target audience are used when creating reports. It will serve as the basis for Step 10 (see Section 3.5.3).

Theoretical grounding: Literature in Section 2.1.6 shows that information presentation preferences differ among people. Recall that Allen and Kim (2001) found that an individual's quest for information is likely to evolve as the interaction between the context and the individual's characteristics progresses. Although demographic characteristics cannot be used to predict the preferences of individuals, if groups of people are exposed to similar information on a regular basis, they will likely have had to perform similar tasks with the data. Employees working in different levels in an organisation (from executive managers to lower level employees) are rarely exposed to the same reports. Even employees that are on the same level, but who work in different departments, rarely see and use the same information.

Schramm's model of communication in Section 2.2.2 best explains the reason for this step. Schramm (1954) concluded that communication can only be effective if it falls within the fields of experience of both the sender and the receiver. This step aims to find characteristics which can be used to narrow down the number of employees whose fields of experience need to be taken into consideration when reporting information.

By identifying clusters of employees who are typically exposed to similar reports, designers of new reports will be capable of isolating the presentation preferences of the target audience from those of the rest of the employees. This will ensure that each report is only tailored to the preferences of its target audience and not that of the whole population of employees. Because of the variability in preferences among people, reducing the number of people whose needs the report has to satisfy will reduce the preferences that need to be considered.

Artefact/Example: This step does not have a specific artefact which can be reused in future, but an example of how attributes can be used to isolate clusters of employees is shown in Table 3.1. The department, managerial level in the organisation and security clearance level of nine employees are shown.

Even though the attributes used in identifying clusters might be static, the clusters themselves can be dynamic. In Table 3.1, all of the employees in the *Human resources (HR)* department have been highlighted to form a cluster. However, by collecting multiple attributes of employees (such as level in organisation, department and security clearance), report designers will be able to isolate, for example, the middle managers in the HR department with security clearance of two or higher. Using the employees in Table 3.1 as an example, only *Employee C* and *Employee D* satisfy this criteria. As a result, only the presentation preferences of those two employees need to be considered when designing the report. By using dynamic filtering of multiple attributes,

Table 3.1: Example of creating clusters of employees based on attributes.

Employee Name	Department	Level	Clearance Level
Employee A	Asset management	Top	3
Employee B	Legal department	Top	3
Employee C	Human resources (HR)	Middle	3
Employee D	Human resources (HR)	Middle	3
Employee E	Project management	Middle	2
Employee F	Asset management	Middle	2
Employee G	Human resources (HR)	Middle	2
Employee H	Logistic department	Low	2
Employee I	Human resources (HR)	Low	1

clusters can be manipulated to suit the needs of the designer. In order to allow designers maximum freedom when isolating clusters, the attributes measured should be organisation-specific to enable the isolation of very small groups of employees. This information can also be supplemented by attributes of employees that are procedurally collected by the HR department.

Value proposition: This step enables the identification and isolation of preferences of the target audience by the framework. As a result of this step, the presentation formats recommended by the framework will only consider the needs of the individuals who a report is created for.

3.2.1.2 Step 1.2: Determine Data Presentation Capabilities

Purpose: To ensure that all of the preferences tested by the framework can be created by the designer of the report. Additionally, to ensure that data presentation formats used by the organisation, which are not mentioned in Section 2.7, are added to the framework.

Theoretical grounding: The purpose of the framework is to guide report designers when deciding on a data presentation format which satisfies the needs of the majority of the target audience. However, before an ideal format can be selected, a list of all the formats that the organisation's reporting tool or software is capable of generating needs to be created. Section 2.5 explains that different representational formats have varying strengths and weaknesses which make them more suitable for some tasks than others. Nevertheless, it is essential that the organisation is able to reproduce all of the formats that are presented as options to the users when determining their presentation preferences in Section 3.3.1. Furthermore, it is equally important that any presentation formats not mentioned in the framework, but which can be produced by the organisation, are included as presentation options in the framework.

All of the graphs (with the exception of box and whisker graphs) discussed in Section 2.7 can be created using templates in Microsoft Excel™ - a popular data analysis and presentation tool. Although there is not a template for it in Microsoft Excel™, box and whisker graphs can be created using alternative steps. The list of graphs in Section 2.7 can be used as a starting point, but it should not be considered to be an exhaustive list or all-inclusive. Once a new list is created, it can be used to test the preferences of the employees in Section 3.3.1.

Recall that, in Section 2.5, it was explained that representations containing the same information are labelled as informationally equivalent. As the framework is intended to aid in the selection of a presentation format out of a variety of alternative formats which convey approximately the same information, it has been decided that neither single nor multiple pie graphs will be included as alternatives in the generic version of the framework. Table 2.9 in Section 2.7.6 outlines the advantages and disadvantages of pie graphs. Pie graphs only show the value of each data element as a proportion of the whole without providing the user with the option of retrieving the actual values of the data elements. In other words, the information of a single data element is only provided in relation to the other data elements in the set and thus it would be unreasonable to provide it as an alternative to the other formats in the framework. Although not all of the other formats in the framework are informationally equivalent, they provide substantially more information than pie graphs.

Artefact/Example: The artefact for this step is a check list similar to the one illustrated in Figure 3.2. All of the presentation formats used in the generic framework are listed on the left with empty lines which the user can fill with presentation formats not listed in the framework. On the right of each presentation format listed, the user is required to indicate whether the organisation is capable of reproducing that particular format or not.

Figure 3.2 has been filled in for illustrative purposes only. It shows that a hypothetical organisation is not capable of reproducing box and whisker plots, line (area) graphs and the point graphs (dot plots) discussed in the generic list of formats in the framework. Furthermore, the hypothetical organisation is able to produce heat maps, bubble plots and pictographs - all of which are formats not discussed in the framework.

Presentation format	Reproducible?
Bar Graph (Grouped)	Yes
Bar Graph (Simple)	Yes
Bar Graph (Stacked)	Yes
Box and Whisker	No
Column Graph (Grouped)	Yes
Column Graph (Simple)	Yes
Column Graph (Stacked)	Yes
Histogram	Yes
Histogram (Multiple)	Yes
Line (Area) Graph	No
Line Graph	Yes
Point Graph (Dot Plot)	No
Radar Graph	Yes
Table	Yes
Heat maps	Yes
Bubble plots	Yes
Pictographs	Yes

Figure 3.2: Example of a check list which can be used to determine all reproducible presentation formats.

Value proposition: This step verifies that the list of presentation format options being tested in Step 4 (see Section 3.3.1) can be reproduced by the organisation. In addition, this step also ensures that the list of presentation formats contains all formats which do not form part of the generic framework, but which can be produced by the organisation.

3.2.2 Step 2: Consider Organisational Context of Tasks

Purpose: To find a list of domain specific tasks with which the employees will be familiar when asked to provide their presentation format preferences.

Theoretical grounding: Since the framework will differentiate between different tasks rather than data types, a list of all the possible tasks that can be completed using data (referred to as data tasks in this thesis) needs to be created. Section 2.5 in Chapter 2 mentions three sets of visual tasks as described by Amar *et al.* (2005), Fausset *et al.* (2008) and Yang *et al.* (2014) (from this point on they will be referred to as the three primary task lists). In order for the framework to be as comprehensive as possible, the three primary task lists have been consolidated into one task list which will be used in the framework. The new task list and the way it incorporates the other lists in Section 2.5 is shown in Table 3.2.

Table 3.2: Data tasks used in this thesis.

Combined Task Name	Amar *et al.* (2005)	Fausset *et al.* (2008)	Yang *et al.* (2014)
Characterise or Compare Distribution	Determine Range Characterise Distribution		Characterise Distribution Compare Distribution
Compare Discrete Value		Discrete data comparison Comparing two points between data series	Compare Values
Compute Derived Value	Compute Derived Value	Comparing more than one component	
Find or Compare Extreme Values	Find Extremum	Identifying maximum or minimum point values	Identify Extrema Compare Extrema
Identify or Compare Trend	Correlate	Identification of trends	Describe Correlation Compare Correlation
Not used in final list			
Retrieve Exact Value	Retrieve Value	Exact point value extraction	Read Value

Similar tasks among and within the three primary task lists were grouped together. If a group of tasks contained tasks from two or more of the three primary task lists, the group was assigned a new name (shown in the first column of Table 3.2). Altogether, six new task names were created, but only five will be used in the framework.

Recall that in Section 2.4.2, it was found that tables are the best format when conveying specific information or single data points. Moreover, graphs containing large amounts of data points rarely contain the values of data points as the plotting space is limited. For these reasons and to keep the questionnaire used in Step 5 as short and concise as possible, it was decided not to test the preferences for the task of retrieving exact values.

The tasks not included due to only one appearance in the primary task lists include the 'cluster', 'filter', 'find anomalies' and 'sort' tasks mentioned by Amar *et al.* (2005) as well as the 'estimating proportion of the whole' task contained in the list described by Fausset *et al.* (2008).

Clustering involves identifying groups of datasets with similar characteristics. These characteristics can refer to various attributes such as trends, proportions, distributions, as well as descriptive statistics such as means. Since clustering is facilitated by other tasks, it will not be tested individually in this study.

Filtering and sorting data elements are too broad for the purposes of this study. Both tasks are accomplished using a criteria which, in turn, makes use of other tasks. Identifying anomalies can also be facilitated using other tasks such as *characterise distribution*, *find extreme values* and *identify trend*.

Finally, the task of judging proportions is excluded on the basis of consistency in the selection process since it was only discussed by Fausset *et al.* (2008) and does not appear in the list of Amar *et al.* (2005) or Yang *et al.* (2014). Fausset *et al.* (2008) concluded that pie graphs are the best format to convey proportions which is in agreement with the findings of Spence and Lewandowsky (1991), Hollands and Spence (1998) and Few (2007). As with the task of retrieving exact values, it was decided not to test the presentation preferences for judging proportions.

The data tasks in Table 3.2 are applicable to both single datasets and multiple datasets. In the case of the first task, a single dataset would require a user to characterise the distribution whereas multiple datasets require users to both characterise and compare them. The same is true for extreme values where a single dataset would merely require a user to identify extreme values, while multiple datasets might require both the identification and comparison of multiple extreme values. Recall that research by Porat *et al.* (2009) and Schonlau and Peters (2012) discussed in Section 2.4.2 show that tables are often preferred to graphics when presenting single datasets while graphics are the format of choice when presenting multiple datasets. If this is true when a table is compared to a graphic, there might be different preferences when there are more graphics presented as alternatives to individuals. For this reason, the presentation preferences for both single datasets and multiple datasets are collected.

Short summaries of the selected five data tasks in Table 3.2 are given below.

- **Characterise or Compare Distribution**
  In a single dataset, users are required to identify and characterise the variation in the variables and to determine the range of the dependent variable. In the case of multiple datasets, in addition to characterising the distributions, the characterisations of all the distributions need to be compared. Characterising the distribution(s) includes determining the span of the distribution and for this reason those two tasks described by Amar *et al.* (2005) have been combined in this thesis.

- **Compare Discrete Values**
  This task requires the user to contrast two or more discrete values explicitly. These values can be intra-dataset or inter-dataset. An intra-dataset comparison often requires a user to compare two or more dependent variables of the same dataset at different values in the independent variable. For example, comparing the output of Mine A in January to the output of Mine A in February in Table 2.7. On the other hand, inter-dataset comparisons typically require a user to compare two or more dependent values of different datasets at the same independent value. For example comparing the output of Mine A in January to the output of Mine B in January in Table 2.7.

- **Compute Derived Value**
  There are instances where the value(s) of interest is not displayed and thus needs to be calculated using other quantities that are displayed. In Section 2.5, Lee, Plaisant, Parr, Fekete and Henry (2006) explain that this might include computing attributes such as the average, the median and the count.

- **Find or Compare Extreme Values**
  If there is any variability in a dataset, there will be some values that are higher than others and conversely also values that are lower than others. This task involves identifying extreme values in the dataset. For the purpose of this research this task will only focus on explicit information and not attributes that need to be computed such as gradients of trends.

- **Identify or Compare Trend**
  Robertson, Fernandez, Fisher, Lee and Stasko (2008) describe a trend as a general tendency which can be observed. They also state that plotting a variable's change over time (or some other ordinal independent variable) is the most common method used to identify trends. Trends can be increasing, decreasing, reversing (changing from increasing to decreasing or vice versa) or cyclic (noisy data). When multiple variables are plotted over a continuous variable, users can easily identify trends which don't fit the general pattern (counter-trends).

  In this study, the identification of trends and correlations are combined as their definitions are closely related. The relationship between variables is described as a correlation in the task lists of Amar *et al.* (2005) and Yang *et al.* (2014) while it is labelled as a trend by Fausset *et al.* (2008).

All of the data tasks in Table 3.2 can be applied to single or multiple datasets. However, some are only applicable to continuous or interval data and others to both interval or continuous data and nominal or ordinal data. Since the number of datasets and the data type being displayed both affect the task, a complete list of the combination of tasks is given in Table 3.3. The subtasks have been codified and these codes will be used in the rest of the framework when referring to a specific combination of task, number of datasets and data type.

Artefact/Example: The artefact for this step is a list containing domain specific examples of each of the 16 tasks. An example of domain specific examples of the five tasks that can be applied to single datasets of data type interval or continuous is shown in Table 3.4. The tasks also need to be accompanied by a hypothetical scenario or background to make the information more digestible and to provide context for the tasks. For the sake of consistency, the example shown in Table 3.4 is once again chosen to depict a mining scenario. Note that the task codes in Table 3.4 correspond with those in Table 3.3.

Table 3.4 only provides examples for five out of the 16 tasks described in Table 3.3. In reality, domain specific examples of tasks would have to be created for the remaining eleven tasks as well before the next step in the framework can be initiated.

Value proposition: This step ensures that the data tasks are understandable to the employees when the information presentation preferences of indi-

Table 3.3: Task codes for all combinations of task names, data types and number of datasets used in framework.

Task Name	Data Type	Number of Datasets	Task Code
Characterise or Compare Distribution	Interval/Continuous	Single	TC1
	Interval/Continuous	Multiple	TC2
Compare Discrete Value	Nominal/Ordinal	Single	TC3
	Nominal/Ordinal	Multiple	TC4
	Interval/Continuous	Single	TC5
	Interval/Continuous	Multiple	TC6
Compute Derived Value	Nominal/Ordinal	Single	TC7
	Nominal/Ordinal	Multiple	TC8
	Interval/Continuous	Single	TC9
	Interval/Continuous	Multiple	TC10
Find Extreme Values	Nominal/Ordinal	Single	TC11
	Nominal/Ordinal	Multiple	TC12
	Interval/Continuous	Single	TC13
	Interval/Continuous	Multiple	TC14
Identify Trend	Interval/Continuous	Single	TC15
	Interval/Continuous	Multiple	TC16

Table 3.4: An example of domain specific tasks.

Task Code	Example
Background	The graphs show the tons of ore processed by a mine in each month of 2014.
TC1	What is the approximate range of the outputs (difference between highest and lowest output)?
TC5	In which month was more ore processed in 2014, May or July?
TC9	Approximately how many tons of ore was processed in the first quarter of 2014 (Jan, Feb and Mar)?
TC13	In which month of 2014 was the least ore processed?
TC15	Was the trend in outputs for 2014 increasing, decreasing, reversing or noise?

viduals are collected. Without it, some employees might not interpret a task correctly when indicating their preferences for that task.

3.2.3 Step 3: Match Presentation Formats to Tasks

Purpose: To determine the structure of the questionnaire in Section 3.3.1.

Theoretical grounding: After all the possible data presentation formats have been established, they need to be matched to the tasks that they can support. In other words, each task needs to be assigned a number of presentation formats which can be used to support the execution of that specific task. Recall that, in Sections 2.3 and 2.5, it was mentioned that data can be conveyed using a variety of formats. This step aims to find every possible presentation format which can support a task regardless of how appealing that format might appear to the designer. The framework focusses on the preferences of the reader and, as such, the appeal of each format for a task is ultimately determined by him or her.

Artefact/Example: Table 3.5 shows how presentation formats have been assigned to tasks that they can be used for.

Table 3.5: Table showing all possible data presentation formats for each task.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	TOTAL
TC1	✓	✓	✓	✓			✓			✓	✓		✓	✓	**9**
TC2	✓	✓	✓		✓	✓		✓	✓	✓		✓	✓	✓	**11**
TC3	✓	✓	✓	✓			✓								**5**
TC4	✓	✓	✓		✓			✓							**5**
TC5	✓	✓	✓	✓			✓				✓		✓	✓	**8**
TC6	✓	✓	✓		✓			✓				✓	✓		**7**
TC7	✓	✓	✓	✓			✓								**5**
TC8	✓	✓	✓		✓			✓							**5**
TC9	✓	✓	✓	✓			✓				✓		✓	✓	**8**
TC10	✓	✓	✓		✓			✓				✓	✓		**7**
TC11	✓	✓	✓	✓			✓								**5**
TC12	✓	✓	✓		✓			✓							**5**
TC13	✓	✓	✓	✓			✓				✓		✓	✓	**8**
TC14	✓	✓	✓		✓			✓				✓	✓		**7**
TC15	✓	✓	✓	✓			✓				✓		✓	✓	**8**
TC16	✓	✓	✓		✓	✓		✓	✓			✓	✓	✓	**10**

Notice that the task codes discussed in Table 3.3 in Section 3.2.2 form the row headings with the possible presentation formats forming the column headings. If additional presentation formats have been identified in Section 3.2.1.2, columns will be added to the table to accommodate them. On the other hand, if there are any presentation formats in Section 2.7 that cannot be reproduced by the organisation using the framework, the columns showing those tasks will be removed. Together, the tasks in the rows and the presentation formats in the columns form a matrix.

If a presentation format can support the execution of a task, a check mark (✓) is placed at the intersection of that task's row with the relevant presentation format's column. If not, the block is filled with grey to show that it is not a viable option. The sum of the presentation format possibilities for each task is calculated and indicated in the far right column of the matrix labelled *TOTAL*. This shows the number of possibilities that will be tested for each task in Section 3.3.1.

Value proposition: Once this step is completed, it will serve as a directory of all the presentation formats at a report designer's disposal. It will also determine the layout of the questionnaire in the next step.

3.3 Phase 2: Neural Training

After the framework has been adapted to consider the characteristics of the organisation where it is being implemented, it needs to be populated with information regarding the employees and their preferences. The first step in this phase is the collection of employee preferences followed by the validation and storage thereof.

3.3.1 Step 4: Determine Employee Preferences

Purpose: To collect descriptive information of individuals which can be used to form clusters as well as to determine the information presentation preferences of those individuals.

Theoretical grounding: In Section 2.1.6, Solomon (2002) indicated that the information-seeking process is largely dependent on the individual performing it. Heinström (2003) supplemented these findings by stating that each individual's way of gathering information is largely dependent on that individual's inner processes and needs and will vary among different individuals.

In Section 2.5 it was concluded that the same information can be communicated using a variety of formats which are informationally equivalent and that the optimal format depends on both the task at hand and the individual performing it.

The importance of encoding information in a format which considers the preferences of the end user (reader) of the information was highlighted in Section 2.2.4.4. Nistal *et al.* (2009) maintained that encoding information in a format that matches the preferences of the user outweighs the benefit of following strict rules to find the most suitable representational format for the task. This finding accorded well with those of Anderson *et al.* (2011) in Section 2.3.1. Anderson *et al.* (2011) found that using a format with which a user is familiar will lead to better performance even if a more suitable format that the user is unfamiliar with, exists.

Therefore, using a representational format which the end user of the information prefers over one which has been chosen based on the best theoretical match between the task and the representational format might yield an improved understanding of the information being conveyed.

Artefact/Example: The framework aims to recommend presentation formats based on the preferences of the target audience. As mentioned before, these preferences will not be predicted using certain attributes of the employees, but rather by testing the individual preferences of the employees using a questionnaire. Employees will be required to indicate their data presentation preferences as well as some personal information on a questionnaire. All of the preceding steps are executed to ensure that the questionnaire is relevant to both the organisation and the employees. The questionnaire will consist of two parts, namely: employee information and presentation preferences.

As was mentioned in Section 3.2.1.1, attributes of employees need to be collected so that clusters can be created. The employee information which is recorded does not refer to demographic information, but rather information that can be used to identify similar groups of employees within the organisation. It should be noted that most of this information might already be stored in the organisation's Human Resources Information System (HRIS). However, if it is not possible to retrieve these attributes pertaining to the employees from the HRIS or if any additional categories need to be created, adding these to the questionnaire would be the simplest way of collecting additional information. As was mentioned in Section 3.2.1.1, typical attributes which can be measured include employee department, level in organisation and security clearance (level). These attributes will be used to isolate the information presentation preferences of the target audience from those of the other employees in the database when creating a report.

The main benefit of the framework is its ability to provide report designers with a recommendation of the format in which information should be encoded. However, the employee preferences that are retrieved by the framework need to be captured before they can be used as a guideline. Employees will be required to indicate their preferences for each of the data tasks shown in Table 3.5 on a questionnaire.

The format of the questionnaire will be determined by the organisation's version of Table 3.5 after the contextualisation phase of the framework has been completed. The number of information presentation alternatives for each task code is shown in the last column of Table 3.5 labelled *TOTAL*. For example, the first task code (*TC1*) will have nine options for the employees to choose from. All of the information presentation formats that have a check mark in the row of Task Code *TC1* in the organisation's version of Table 3.5 will be presented as an option in the questionnaire for task code TC1.

Employees will be asked to indicate their first or primary preferences for each of the tasks. In addition to this, they will also be asked to indicate a secondary preference. Recall the example given by Mayer-Schönberger and Cukier (2013) in Chapter 1 of this research. Although apple pie was the fourth most popular pie when considering the sales of small pies (consumed by individuals), it was found that apple pie is the outright winner when considering the sales of large pies shared by families. Mayer-Schönberger and Cukier (2013) concluded that apple pie is everyone's second favourite choice of pie and when the preferences of a group had to be considered, apple pie was the best solution. The purpose of determining the secondary preference is to ensure that the best solution can be found when combining the preferences of a group of employees.

Finally, employees will be asked to indicate whether there are any presentation formats which they would like to avoid if possible. Box and whisker plots have been described as providing users with simple, yet effective, analysis and comparison of distributions. However, as Anderson *et al.* (2011) explain, if employees have not been exposed to these plots previously, they will not be able to benefit from any of the features that make box and whisker plots so attractive. This means that if a report designer uses a box and whisker plot to convey information to an employee who is not familiar with the format, the communication process will have failed. Indicating that a presentation format should be avoided should only be used to indicate options that the respondent does not understand and not to indicate that the option is not the employee's first or second choice.

If a presentation format is not indicated to be a user's primary or secondary preference or that he or she would like it to be avoided, it indicates that the user has a neutral sentiment towards the use of that format for a particular task code.

Individuals will be required to indicate one primary preference, one secondary preference and an unrestricted amount of data presentation formats which they would like to avoid for each task code. The options are summarised in Table 3.6.

Table 3.6: Description of responses used in questionnaire.

Response	Symbol	Description	Number of uses
Primary preference	P	Indicates the respondent's first choice. It is compulsory for all of the task codes and can only be indicated once.	1
Secondary preference	S	Indicates the respondent's second choice. This response is also compulsory and cannot be omitted by the respondent. Each task code should have exactly one secondary choice.	1
Avoid	A	Indicates presentation formats that should be avoided if possible. After the primary and secondary preferences have been indicated, any of the remaining can be marked as avoid.	Unlimited

An example of how the possible presentation formats can be presented to employees is shown in Appendix B.

Value proposition: This step aims to collect the data that will serve as the source of all recommendations made by the framework. Without it, the framework would not be able to provide any type of output.

3.3.2 Step 5: Validate Responses

Purpose: This step has been included to act as a fail-safe to ensure that quality data is entered into the repository.

Summary: As was explained in the previous steps, the questionnaire will comprise of two sections. The first section will gather personal or descriptive information about the employees while the second section will collect their data presentation preferences. Although employees will be guided through the questionnaire, there are still opportunities for the wrong data to be recorded.

Potential problems which may occur as well as possible solutions to restrict the effects of those problems are subsequently discussed.

Problems in section collecting personal information:

In the section of the questionnaire where employees are required to indicate descriptive information which will be used to create clusters, spelling mistakes can occur or wrong information can be filled in mistakenly. If designers try to form clusters using information which is encoded incorrectly, the preferences of those employees with erroneous information will be ignored by the framework. This can be avoided if the section collecting personal information is designed so that employees have to choose a single option out of a list showing all the possible options. For example, if an employee has to indicate his or her department, a list containing all of the departments in the organisation can be provided in the questionnaire with boxes next to them that can be marked if the employee belongs to that department. This will eliminate spelling mistakes and will ensure that the responses provided by users are usable and comparable.

There are two possible disadvantages to this approach. Firstly, the designer of the questionnaire might not consider all the possible departments to which an employee can belong. Secondly, the list of options might be inappropriately long. For example, consider a situation where the framework is implemented at all of the departments in a government. Including a list of all possible departments would be impractical and place unnecessary additional strain on an employee completing the questionnaire. If the information can be recorded electronically, cascaded drop down lists can be used to dissect large lists into several smaller options which follow on from each other. A typical use of cascaded drop down lists is on websites of second hand car dealers. Instead of providing customers with a long list of all the available cars, customers are asked to answer certain questions to narrow down the number of choices until only a few cars remain. For example, customers might first be asked to fill in the manufacturer, then the year, then the model to find the car they are looking for. By using these three attributes of the car, a very long list of possible cars can be reduced to a shortened list only containing cars similar to the one the customer is looking for. Using the same approach, a designer can drastically reduce the list of all the departments in an organisation by merely providing a few identifying attributes.

If the wrong information is filled in deliberately (for whatever reason), the questionnaire cannot be designed to prevent it. One solution would be for someone with enough knowledge of the employee to be present while the section pertaining to the employee's personal information is filled in. Another, more ideal solution would be to link the Database Management System (DBMS) in which the employee preferences are stored (see Step 6) to the organisation's HRIS. This will prevent employees from entering wrong information and will ensure that the information stored in the preference database is in agreement with the official information stored by the HR department. Even if the employee preferences are not stored in a database, it would still be beneficial to obtain and use as much of the information from HR as possible.

Problems in section aiming to collect employee preferences:

The second opportunity for wrong data to be recorded is in the section where employee preferences for the various task codes are determined. Table 3.6 describes how the primary and secondary preferences for each task should be indicated as well as how an employee can show that a presentation format should be avoided. The following four rules summarise instructions for each task code:

1. A primary preference is compulsory and should be indicated one time only.

2. A secondary preference is compulsory and should be indicated one time only.

3. A presentation format not indicated as a primary or secondary preference can either be left blank or marked to be avoided.

4. A particular presentation format cannot have more than one choice assigned to it. In other words, each presentation format should be indicated as either a Primary preference (P), Secondary preference (S), Avoid (A) or left blank.

Should any of the four rules be broken by an employee, the data stored in the repository will not be accurate. If more than one primary choice has been indicated for a task or if it has been omitted altogether, the framework will not provide the designer with an appropriate recommendation when reporting data to that respective employee. Although more than one primary preference improves the odds of a designer choosing a presentation format which the employee prefers, the chosen format will ultimately still be determined by the designer and not the receiver of the information. In the case where no primary preference has been indicated, the designer will be forced to use the secondary preference. However, the secondary preference might indicate a format which the employee will be least dissatisfied with if the primary preference cannot be used.

The same is true if more than one secondary preference has been indicated or if it has been omitted altogether. If a single primary preference has been indicated, having more than two or no secondary preferences will not affect the recommendations made by the framework if it is applied to a single employee (as only primary preferences are considered). Nevertheless, the wrong amount of secondary preferences will influence the recommendations of the framework if it is applied to a cluster of employees as the results will be biased.

Finally, the last possible problematic outcome is when a presentation format is indicated as being more than one of either P, S or A. Although there might be means to narrow down and identify the possible mistake(s), it would be in violation of the framework's requirement to be structured (see Section 3.1.2) as different report designers might use different reasoning and logic to identify the mistake(s). Consequently, if a task contains a presentation format which has been indicated as being more than one of either P, S or A, the responses for the entire task will be disregarded.

By collecting the preferences electronically, the four rules can be set as conditions that need to be satisfied before a response can be saved. For example, if an employee has indicated one primary preference and tries to indicate another for the same task code, an error message can warn an employee that there can only be one primary preference per task code. Notice that, by collecting the information in both sections of the questionnaire electronically with predefined conditions for each section, the collection of invalid or faulty data can be reduced substantially.

Failure by any employee to adhere to the four rules mentioned above will require that employee to undergo Step 4 an additional time (or until all rules have been adhered to). However, it should be noted that only the tasks that were problematic would have to be readdressed. It has also been explained that this step will be redundant if the information can be collected electronically with conditions governing the possible user responses. Once all of the responses to the different task codes have been deemed valid, the information should be stored in a repository.

The combination of primary and secondary preferences as well as the formats which the employee would like to avoid should receive a status of *unconfirmed* upon entry into the repository in the next step of the framework.

Value proposition: Including this step will ensure that the data entered into the repository is in the correct format and that it satisfies all the constraints. This step will also prevent errors in the application phase of the framework.

3.3.3 Step 6: Create or Save in Repository

Purpose: To store the information in a database that allows easy retrieval of the desired information.

Theoretical grounding: This step is included in the framework to simplify the manipulation and retrieval of information as well as to improve data security. Neither data security nor databases lie within the scope of this research. However, a short discussion regarding databases is provided to underline the possible benefits of using databases to store the information. Ramakrishnan and Gehrke (2000) describe a DBMS as follows:

> *"...software designed to assist in maintaining and utilizing large collections of data, and the need for such systems, as well as their use, is growing rapidly."*

Ramakrishnan and Gehrke (2000) briefly highlight some advantages of using a DBMS to manage data, the following four are highly relevant to the framework:

- Efficient data access: Data is stored and retrieved efficiently by means of a variety of sophisticated techniques. If data is stored on external devices or shared by facilities that aren't linked, this step is especially important.

- Data integrity and security: If a DBMS is always used to enter and access the data, integrity constraints can be enforced on the data. Integrity constraints refer to certain conditions that the new data has to satisfy before it can be entered or stored. This property is invaluable for the purposes of the framework even if Step 5 has been executed.

  Firstly, it will ensure that the correct data has been filled in by the employee and that no typographical errors occur when the information is transferred from the questionnaires to the repository. If an employee states that he or she is in a certain department which does not exist, the DBMS will be able to highlight this fact if it is connected to the HRIS of the organisation. The DBMS will also ensure that consistent spelling is used for each attribute used to form clusters. Finally, if an employee has not followed the criteria for indicating their primary and secondary preferences and the presentation formats which they would like to avoid (see Table 3.6), the DBMS would be able to prevent the wrong data from being entered and saved. Most of benefits regarding data integrity can also be realised if the questionnaire is administered electronically with certain conditions that need to be satisfied before an employee can proceed to the next step (see Section 3.3.2).

- Data administration: If the data is shared among various users, significant improvements can be realised when the data is administered centrally. By assigning the responsibility of organising and storing data to a selected group of professionals, redundancy can be minimised and retrieval of information can be made more efficient.

- Concurrent access and crash recovery: A DBMS is designed to allow multiple users access to the data simultaneously. This means that there will be no restrictions or user conflicts when there is more than one user retrieving information at a certain point in time. A DBMS also saves progress and a backup of all the data to minimise the effects of system failures.

Another relevant advantage of a DBMS is its support of the use of multiple criteria when querying or filtering information. This means that the stored information can be subjected to various criteria to find and isolate a specific group(s) of employees. For example, the preferences of employees who are either in top management positions in the AM department or middle management positions in the legal department can be isolated using a single query. This decreases the need to manually isolate and combine preferences of different groups of employees.

If the organisation where the framework is implemented decides not to use a database to store the information, it is recommended that, at the very least, the information be stored in electronic format on a server that is backed up regularly. The application used to store the data should also enable the user to form clusters, even if multiple filters are required repeatedly.

Artefact/Example: Once the questionnaire discussed in Section 3.3.1 has been completed by an employee, the data needs to be stored in a database. Recall that one of the required features of the framework is that it should be practical. In other words, it should be easy to use with clear benefits to the users or report designers. By storing the information collected by the questionnaires in a database, the preferences can be manipulated (filtered, isolated and tallied) to suit the needs of the report designer. The information in the database can then be subjected to multiple criteria to isolate the preferences of the target audience.

Furthermore, the preferences will be transformed into numbers in the database. Figure 3.3 shows how the preferences of one fictional employee have been transformed into numerical scores.

In Figure 3.3a, the template in Table 3.5 has been populated with the Primary (P) and Secondary (S) preferences of the respondent as well as the

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC1	A			P			S							
TC2	A		A		S			P						A
TC3	S	P												
TC4	P	S												
TC5				S							P			A
TC6			A		S								P	
TC7	P		S											
TC8	P							S						
TC9	A			P							S			
TC10	S		A		A									
TC11		S					P							
TC12	P							S						
TC13	S						A				A		P	
TC14			S					P						
TC15	A												P	S
TC16	A							A	A				P	S

a)

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC1	-2			2			1							
TC2	-2		-2		1			2						-2
TC3	1	2												
TC4	2	1												
TC5				1							2			-2
TC6			-2		1								2	
TC7	2		1											
TC8	2							1						
TC9	-2			2							1			
TC10	1		-2		-2									
TC11		1					2							
TC12	2							1						
TC13	1						-2				-2		2	
TC14			1					2						
TC15	-2												2	1
TC16	-2							-2	-2				2	1

b)

Figure 3.3: Transforming preferences into scores.

formats which the respondent has indicated should be Avoided (A) if possible. This information will be retrieved directly from the questionnaire in Section 3.3.1. Figure 3.3b shows how the preferences have been transformed into numbers using the following weightings:

Primary preference (P)	= 2 points
Secondary preference (S)	= 1 point
Avoid (A)	= -2 points

By storing the information in numerical format, simple arithmetic procedures (such as additions and subtractions) are facilitated. This is discussed in more detail in Section 3.5.4.

Value proposition: Storing employee attributes and data presentation preferences in a database firstly ensures that the information is safeguarded. Secondly, it allows the isolation of complex clusters of employees by means of queries. Finally, it allows arithmetic procedures to be performed on the data once it is encoded.

3.4 Phase 3: Confirmation

The confirmation phase only has one step and, although highly recommended, the execution of the step is optional.

3.4.1 Step 7: Confirm Preferences

Purpose: To ensure that the preferences stored in the repository for an employee are, in fact, the actual preferences of that specific employee.

Summary: After an employee has indicated his or her preferences, the information is entered into a repository for future use. There is, however, a possibility that the information entered into the repository is not a true reflection of the employee's actual preferences. The two main reasons why the wrong employee preferences might be stored are that an error occurred when the preferences were transferred from the questionnaire into the repository or that the wrong preferences were indicated by the employee.

A possible solution to the erroneous transfer of data would be to collect the preferences electronically. The questionnaire can be linked to the repository and as soon as all of the responses for a task have been saved they can be entered into the repository automatically. If the use of electronic questionnaires is not possible, the erroneous transfer of information can be reduced by having a policy stating that two or more people have to confirm that the information has been transferred correctly before it is made available to report designers.

The other reason why the wrong preferences could be stored in the database is because they have been indicated incorrectly by the employee completing the questionnaire. Regardless of whether the employee misunderstood the question or whether he or she marked a presentation format incorrectly, the fact remains that the wrong preferences would have been stored in the repository leading to incorrect recommendations when used by a report designer.

Artefact/Example: An employee's preferences, as indicated on the questionnaire, could be confirmed by creating a sample report containing all 16 task codes as well as the primary and secondary preferences they indicated for each task code. The examples used in the task codes of the sample report should ideally be different to those used in Step 4 while still having the same requirements for each task code. Each task code can be followed by the primary and secondary preferences previously indicated by the employee. Finally, all of the alternative presentation formats can be supplied after the primary and secondary preferences. This will enable the employee to compare his or her two chosen primary and secondary presentation formats to other possibilities.

Recall that an employee's combination of preferences for each task code is assigned a status of *unconfirmed* upon entry into the database (see Section 3.3.2). If the stored preferences for a task code are confirmed to be correct, that particular task code's status can be changed to *confirmed* in the repository. However, should an employee indicate that his or her primary and secondary preferences are not those displayed on the sample report, the task code will be marked as *dismissed*.

The next step in the framework is dependent on whether any of the statuses of the 16 task codes have been changed from *unconfirmed* to *dismissed*. If any of the preferences for any of the task codes have been marked as *dismissed*, Step 4 (Section 3.3.1) has to be repeated, but only for the task codes marked as *dismissed*. In other words, none of the preferences of the task codes with a status of *confirmed* need to be collected again and neither will the status of those tasks be changed when Step 4 is repeated. Once all of the responses have been marked as *confirmed*, the application phase can commence.

Value proposition: This step ensures that the recommendations made by the framework in Step 12 are, in fact, based on the actual preferences of the employee(s).

3.5 Phase 4: Applying the Framework

In Section 3.1.1 it was mentioned that the framework has four phases. The first three phases are designed to collect, store and confirm information that is specific to the organisation where it is being implemented. The fourth phase is generic and will be used in the same way regardless of the organisation where it is being implemented.

In order for the framework to be used as a tool when reporting data, certain general steps need to be followed to ensure that the desired results are achieved. Each of the recommended steps which should be used when applying the framework are explained below.

3.5.1 Step 8: Design Report

Purpose: To identify the target audience of the report as well as to determine all of the numerical information that needs to be presented.

Summary: It should be emphasised that the framework does not recommend what information should be included when reporting data. It merely makes a recommendation of which format should be used to convey that information. Designers should therefore first design the report and identify the information that needs to be presented or communicated. Once all of the datasets have been identified, the rest of the framework can be used to determine the ideal presentation formats for each of those pieces of information.

Example Output: Table 3.7 shows a summary of five fictional datasets that will be used for illustrative purposes in the rest of this phase. Each dataset has been assigned a number to simplify the identification and referencing of a dataset. A name or description is recommended to give an indication of the contents. Furthermore, mentioning the period and source of the data would be beneficial as it could be used to locate the origin of the data if the designer has queries. Finally, the summary should include an indication of whether the datasets are single or multiple as well as their data types.

Table 3.7: An example of a dataset summary.

No.	Dataset Name	Period	Source	Datasets	Data Type
1	Crusher Outputs	Feb 2014	Mine A	Multiple	Interval
2	Incident Report	Feb 2014	Mine A	Single	Nominal
3	Incident Report	Feb 2014	Mine B	Single	Nominal
4	Downtime Report	YTD 2014	Mine M	Multiple	Continuous
5	Ore Processed	YTD 2014	Various	Multiple	Nominal

A summary of all the datasets similar to the one shown in Table 3.7 can simplify the identification of data tasks in the next step significantly.

3.5.2 Step 9: Identify Data Tasks in Report

Purpose: To identify the data tasks which need to be completed by employees when interpreting the datasets listed in Section 3.5.1.

Summary: After the summary of all the datasets to be reported has been compiled (see Table 3.7), the main reasons for the inclusion of each piece of information on the list included in the report should be identified. Each reason should be stated in terms of the tasks described in Section 3.2.2 and listed in Table 3.3.

Example Output: The output of this step is merely an additional two columns added to Table 3.7 showing the name and code of the main data task that needs to be executed using each dataset.

For example, if the reason why dataset number 1 (*Crusher Outputs*) in Table 3.7 is included in the report is to show the trend in the outputs of crushers it should be indicated in a new column. Table 3.3 can facilitate the assignment of the particular task code to the task based on the data task, number of datasets and data type. Note that the only task code in Table 3.3 which is a combination of the *identify trend* data task using *multiple* datasets of *interval* data type is task code *TC16*. Each task code has a unique combination of task name, data type and number of datasets so there will always be a definite answer for each combination.

Table 3.8 shows an example of how a task code has been assigned to each dataset. Note that Table 3.8 has been created by removing the *Period* and *Source* columns from Table 3.7 (for representative purposes) and by adding an additional column to show the task name and task code of each dataset.

This will show the designer which preferences need to be retrieved from the repository in the next step. A summary similar to the one shown in Table 3.8 can be set up to identify the tasks that are being used in the report. Note that the summary shown here is merely for illustrative purposes.

Table 3.8: An example of a dataset summary with task codes assigned to datasets.

No.	Dataset Name	Datasets	Data Type	Task Name	Task Code
1	Crusher Outputs	Multiple	Interval	Identify Trend	TC16
2	Incident Report	Single	Nominal	Compare Discrete Value	TC3
3	Incident Report	Single	Nominal	Compare Discrete Value	TC3
4	Downtime Report	Multiple	Continuous	Compare Distribution	TC2
5	Ore Processed	Multiple	Nominal	Compare Discrete Value	TC4

If a report had to be created to convey the information contained in the five datasets shown in Table 3.8, the designer would need to retrieve the presentation preferences for task codes TC2, TC3, TC4 and TC16. However, only the preferences of the employees who the report is intended for need to be retrieved from the repository. Section 3.5.3 explains how this can be done.

3.5.3 Step 10: Identify Target Audience

Purpose: To identify and isolate the target audience according to the attributes discussed in Section 3.2.1.1.

Summary: It has been mentioned repeatedly that the optimal way in which different people receive information is not always the same. By identifying the group of employees for which the report is intended, the number of unique preferences which need to be accommodated by the designer can be reduced. This will not only assist the designer when deciding on information presentation formats, but it will also serve as a reference when deciding on a register (degree of formality and choice of vocabulary) when encoding information. The audience should be identified in terms of the attributes discussed in Section 3.2.1.1.

Example Output: A typical output of this step is a criteria which can be used to narrow the entire employee population down to only the target audience. Recall that in Table 3.1, the target audience was identified by using multiple criteria. In that particular example, the target audience had to work in the HR department, be middle managers and have a security clearance level of three. Clusters can also be formed by requiring employees to satisfy some of the criteria, but not all of them. For example, the target audience might be employees who satisfy any of the following criteria:

1. Employees in the AM department.
2. Employees in the Project Management department with a clearance level higher than one.
3. Employees in the HR department who are middle managers and with a clearance level of two or three.

Lastly, employees who do not form part of the groups identified by the criteria mentioned above can be added individually. This approach can also be followed if the target audience comprises of a relatively small number of employees from a variety of groups or if the target audience is a single employee. Section 3.5.4 discusses how the designer can determine which presentation formats should be used once the target audience has been identified.

3.5.4 Step 11: Match Audience Preferences to Tasks

Purpose: To determine the optimal presentation format for the datasets being reported based on the preferences of the target audience.

Summary: In this step, the framework will be used to match the preferences of the target audience identified in Section 3.5.3 to the task codes identified in Section 3.5.2. Only the preferences of the employees in the target audience will be retrieved from the database. In addition, only the preferences for the task codes identified in Section 3.5.2 need to be retrieved. The preferred presentation format for a task code can be obtained by identifying the cell with the highest score or value in the row representing that task code. As an example, consider Tables 3.9 and 3.10 which show the encoded questionnaire responses of a single employee and the responses of multiple employees respectively. In both tables the cells with the highest values have been shaded green to indicate the recommended formats.

Table 3.9: Example of selected presentation formats of an individual employee.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC2	-2		-2		1			2						-2
TC3	1	2												
TC4	2	1												
TC16	-2							-2	-2				2	1

For an individual employee (see Table 3.9), the preferred formats are selected by using the primary preferences of that individual (encoded using the number two). When creating a report for the employee whose preferences are shown in Table 3.9, the designer would use a grouped column graph, a point graph, a table and a line graph to report information for task codes TC2, TC3, TC4 and TC16 respectively.

Table 3.10: Example of selected presentation formats of multiple employees.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC2	5		-8		3			8		8		7	10	-2
TC3	8	6	-10	7			21							
TC4	14	1	-12					20						
TC16	3		2					6				11	18	1

When the report needs to be designed for multiple employees, the preference scores for each combination of task code and presentation format of the whole target audience are summed and the highest score is selected. For example, the first score of five in the top left corner of Table 3.10 is the sum of all the Primary (2), Secondary (1) and Avoid (-2) responses indicated by the employees in the target audience for using tables to complete task code TC2. Table 3.10 shows that using a line graph to complete task two was the most popular combined preference of the target audience. In addition, a simple column graph, a grouped column graph and a line graph were indicated as being the optimal formats in which to convey information to the target audience for task codes TC3, TC4 and TC16 respectively.

As the combined preferences of a group of employees are determined by simply summing the encoded individual responses, there is a possibility that more than one cell in a row can have the same value. For example, the combined preference scores for grouped column graph and a box and whisker plot in for task two are both eight in Table 3.10. This has no effect on the recommended format that should be used for task code TC2 in Table 3.10 as a line graph has a higher score than both formats. However, if more than one format shares the highest score for a single task, the following rules could be used to decide which format should be used:

1. The presentation format with the smallest number of *-2*'s (avoid) responses should be chosen.
2. If there is still a tie, the format with the most *1*'s (secondary preferences) should be used.
3. If there is still a tie, it means that the number of P, S and A responses for the tied formats are the same. The designer can then use his or her own discretion in deciding on one of the tied formats.

Example Output: The most preferred presentation format for each task code, as indicated in Tables 3.9 and 3.10 (depending on whether the target audience is an individual or multiple employees), can be indicated next to each dataset in an additional column in Table 3.8. Table 3.11 shows the recommended presentation format for each task and summarises the output of the framework.

Table 3.11: Condensed example of the framework output showing the ideal presentation format for each task.

			Presentation Format	
No.	Dataset Name	Task Code	Individual Employee	Multiple Employees
1	Crusher Outputs	TC16	Line Graph	Line Graph
2	Incident Report	TC3	Point Graph	Column Graph (Simple)
3	Incident Report	TC3	Point Graph	Column Graph (Simple)
4	Downtime Report	TC2	Column Graph (Grouped)	Line Graph
5	Ore Processed	TC4	Table	Column Graph (Grouped)

Once again, some columns (*Period*, *Source*, *Datasets*, *Data Type* and *Task Name*) have been removed for representative purposes. Step 12 explains how Table 3.11 can be used when creating the report.

3.5.5 Step 12: Create Report

Once the ideal presentation format has been determined for each task in the report, the designer can create information presentations with the assurance that the majority of the target audience's preferences will be satisfied.

In conclusion, Table 3.12 explains the meaning of each of the column headings of the framework's ultimate output.

Table 3.12: Names and descriptions of ultimate framework outputs.

Attribute	Description
No.:	Dataset number which simplifies referencing a particular dataset.
Dataset Name:	A name which is logical and gives the designer an idea of its contents.
Period:	The time period during which the data was collected.
Source:	The origin of the dataset which can be used to locate its creator if the designer has any queries.
Datasets:	An indication of whether the dataset consists of a single dataset or multiple datasets.
Data Type:	An indication of whether the data type of the dataset is nominal, ordinal, interval or continuous.
Task Name:	The name of the main task that needs to be facilitated by the data. The possible task names are illustrated in Table 3.2.
Task Code:	The code of the main data task which needs to be completed by employees when interpreting the dataset. A complete list of the possible data tasks is shown in Table 3.3.
Presentation Format:	The presentation format which should be used when representing that dataset. This recommendation is based on the number of datasets, the data type of the dataset, the data task which needs to be completed and the preferences of the target audience. All the steps in the framework are executed to get the optimal match between the presentation format and the dataset.

3.6 Chapter Summary

This chapter describes a proposed solution to the problem statement. The proposed solution is conceived in the form of an information encoding framework which is based on the literature discussed in Chapter 2. The chapter commences with an overview of the framework which describes the development of the four phases in the framework as well as the features of the framework. Section 1.3 explains that the proposed solution (framework) is required to be practical, generic, flexible and structured.

As mentioned before, the framework comprises of four main phases, namely: contextualisation, neural training, confirmation and application. Each one of these four phases serves a different purpose and contains one or more steps. All four phases as well as their steps are summarised in Figure 3.1. The first three phases are only executed once to collect the information on which the framework's recommendations will be based. Once the source information has been collected, it will be used repeatedly. However, the fourth phase will be executed every time that the framework is used to aid in deciding on presentation formats for reports.

Literature in Chapter 2 has shown that an individual's task performance (when acquiring information) can be improved by altering the germane and extraneous cognitive loads associated with the presentation of information. What makes the proposed framework unique is that it attempts to recommend a presentation format which reduces both the germane and extraneous cognitive loads placed on individuals. This chapter addressed and achieved the sixth objective identified in Section 1.3:

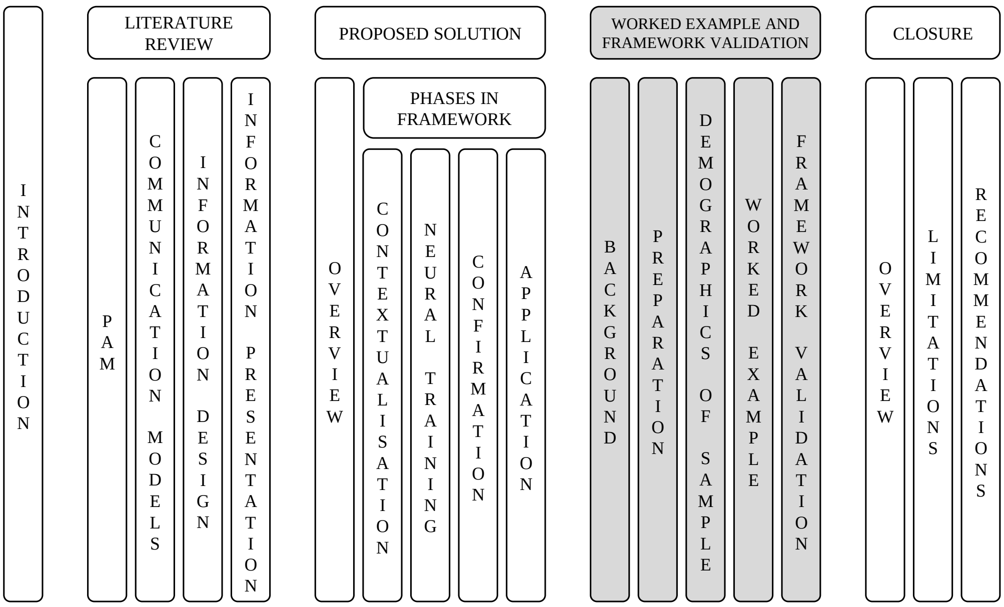
6. Develop a solution which recommends a presentation format based on the characteristics of the task to be completed and the preferences of the target audience.

The next chapter aims to validate the information encoding framework by consulting managers from six different data-intensive organisations.

Chapter 4

Worked Example and Framework Validation

Even though the information encoding framework was derived from literature to address the problems outlined in the problem statement, there is no guarantee that it is a viable solution in practice. Therefore, individuals from industry were consulted to provide both input for the framework and responses to validation questions. This chapter commences with a background of the organisations and individuals who took part in the validation and a description of the structure of the consultations. Thereafter, the demographics of the individuals consulted are reported, followed by a worked example and the responses to the validation questions. The chapter concludes with an analysis of the primary preferences indicated by the individuals consulted and a reflection on how the design of the validation might have influenced the results.



4.1 Background of Individuals and Organisations Used in Validation

In this chapter, the proposed information encoding framework discussed in Chapter 3 is applied and validated. This thesis was conducted with the support of the Asset Care Research Group (ACRG) at the University of Stellenbosch, South Africa. One of the main aims of the ACRG is to identify and propose scientific solutions for problems experienced by the PAM industry. Six organisations with which the ACRG has a good working relationship were approached and requested to both provide input for the framework and to evaluate the framework as a possible solution to the problems outlined in the problem statement. These organisations were:

- Anglo American Platinum: Anglo American is one of the largest mining companies in the world operating in five different continents. Anglo American Platinum is the division responsible for the mining, processing and marketing of Platinum Group Metals (PGMs) and is a global leader.

- Appletiser: A producer of sparkling fruit juice which is based in South Africa and the products are exported to more than 20 countries worldwide. Appletiser was acquired by Coca-Cola in 2014.

- eAsset Management: A global consulting firm that specialises in AM, change management, training and technology which aims to deliver sustainable business benefits.

- Namdep Diamond Corporation: Namdeb is a Namibian organisation who performs land-based prospecting and mining for diamonds. Namdeb is owned in equal shares by the Namibian government and the De Beers Group - the world's leading diamond company. The contribution made by Namdeb to the Namibian Gross Domestic Product (GDP) is larger than the combined contribution of all the other mining activities in Namibia.

- Pragma: A consulting company based in South Africa which focuses on improving physical asset performance in a sustainable way. Pragma delivers asset care services, AM improvement projects, training and consulting interventions.

- TechnoServe: A non-profit organisation focussed on enterprise development and promotes the growth of Small, Medium and Micro Enterprises (SMMEs) in Africa, Asia and South America. TechnoServe accomplishes this by means of programmes that are based on global best practices and run by experienced staff with extensive private sector an consulting experience.

Feedback of 20 employees in low, medium and high management levels in these six organisations was collected. It was confirmed that all 20 of the employees occupied managerial positions and hence they will be referred to as managers for the remainder of this chapter. Prior to the engagement with the organisations, it was agreed that the responses from the managers would remain confidential to preserve their anonymity and as such the responses were pooled together and are reported collectively.

The problem that the framework attempts to address was described to the managers followed by a discussion of the purpose of each step in the framework. Once the managers were familiarised with the framework, they were asked to complete a generic version of Step 4 of the framework. After the managers indicated their preferences which were used as input for the framework, the managers were asked to assess the validity of the problem that the framework attempts to address. In addition to the problem, managers were also asked to assess the framework as a whole and some of its assumptions. In order to ensure consistent and comparable assessments, managers were asked to provide responses to a set number of questions. However, it should be noted that it was not a survey, but rather content validation by means of a questionnaire to test the practicality and implementability of the proposed solution.

The target group for the content validation was a select number of managers from data-intensive industries who would typically benefit from the framework. More emphasis was placed on the quality of managers than the quantity of managers used in the validation process. Consequently, the small sample size did not affect the results as the aim was to obtain insights from industry rather than to test and report the statistical significance of responses. However, descriptive statistics such as proportions and distributions will still be reported to summarise the responses.

4.2 Preparation of Consultations

As mentioned before, inputs from individuals in industry were used to validate the proposed solution (framework) discussed in Chapter 3. Twenty managers from the six organisations listed in Section 4.1 were consulted and asked to provide their inputs regarding various aspects of the framework. For practical reasons, the 20 managers could only be consulted once. Consequently, a predetermined set of questions had to be formulated to ensure that the responses collected from the various managers were comparable. These questions and their responses are discussed in the subsequent sections.

Factors such as limited contact time with managers and the spatial requirements of presenting a large amount of alternative presentation formats to managers at once led to a decision to collect the responses in paper format. Making use of paper instead of an electronic format to collect responses was the most resource effective option for the small sample of managers consulted and its once-off use as no software programs had to be developed. Another advantage of using paper instead of an electronic format is that no additional equipment or skill is needed by a manager to provide his or her responses.

Section 3.3.2 outlines some of the problems that can occur when paper is used to collect information. These problems are very relevant to the proposed information encoding framework as most of the inputs needed by it are subject to some conditions. Consequently, the information that would later be used to create sample clusters was collected by asking the managers to choose a single response from a list of options. Although providing users with a list of all the possible options should reduce the number of incorrect responses, it does not remove the possibility thereof entirely. Unfortunately, no control measures could be put in place to ensure that the preferences recorded for each task complied with the criteria discussed in Table 3.6 in Section 3.3.1. The information presentation preferences that were collected and the handling of incorrect responses are discussed in more detail in Section 4.4.2.1.

The responses that were collected from the managers were divided into three parts, namely: demographic information, information presentation preferences and validation responses. Information in the first part was collected to characterise the managers in the sample and to specify and limit the generalisability of the results. The second part is discussed in the form of a worked example. In this worked example, the steps in the framework are followed using the preferences collected to illustrate how the framework would be applied to a hypothetical scenario. Finally, the validity of the problem statement and the framework as assessed by the managers are discussed.

4.3 Demographic Information of Managers

All of the 20 managers who were consulted were asked to indicate some demographic and descriptive information. The main purpose of this information was to detail the characteristics of the sample, but some of the information will also be used to form clusters in the worked example discussed in Section 4.4. Managers were not asked to indicate their names or any information that would compromise their anonymity. In order to minimise the intrusiveness of the consultations, managers were also allowed to not provide an answer to a demographic question if they felt that it was inappropriate. Each demographic question as well as a summary of the responses is discussed next.

4.3.1 Gender

The sample of managers consisted of 15 (75%) males and 5 (25%) females. The sample was a sample of convenience merely determined by the availability of managers to provide input and thus no attempt was made to specify the ideal proportions of males and females. It should be noted that the gender of the managers was only recorded to characterise the sample and the information is not used by the framework in any way.

4.3.2 Age

Out of the 20 managers that were consulted, only 18 provided their age. Recall that all of the demographic questions were optional, although the managers were encouraged to provide as much information as possible. The average age was 34 years and reported ages ranged from 27 years to 45 years. The histogram in Figure 4.1 shows the distribution of the ages in five-year intervals.

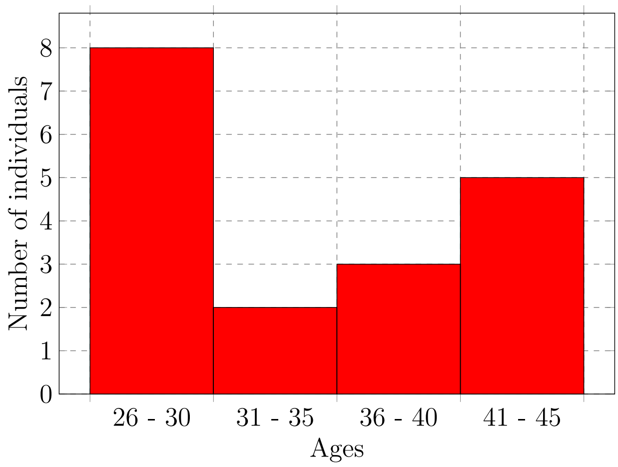


Figure 4.1: Distribution manager of ages in sample.

The bins in Figure 4.1 were chosen to be of equal size in convenient intervals and that is why the first bin accommodates managers who are 26 years old even though the youngest reported age is 27 years. Notice that even though the average age of the managers is 34 years, only two out of the 18 reported ages fall in the 31-35 years bin. Out of the 18 managers who indicated their age, eight (44%) were 30 years old or younger. Once again, the ages of the managers are not used as inputs for the framework and are only reported to detail the sample of managers consulted.

4.3.3 Time at Company and Time in Industry

Managers were asked to indicate the number of years spent in their current industry as well as the number of years spent at their current organisation. All 20 of the managers provided responses for both questions. The average number of years spent in industry is 13 years with a maximum and minimum of 38 years and three years respectively. The managers have been employees of their current organisations for an average of eight years with a maximum of 23 years and a minimum of one year. In Figure 4.2, a box and whisker graph shows the distributions of the number of years that the managers have worked in the industry and at their current organisation.

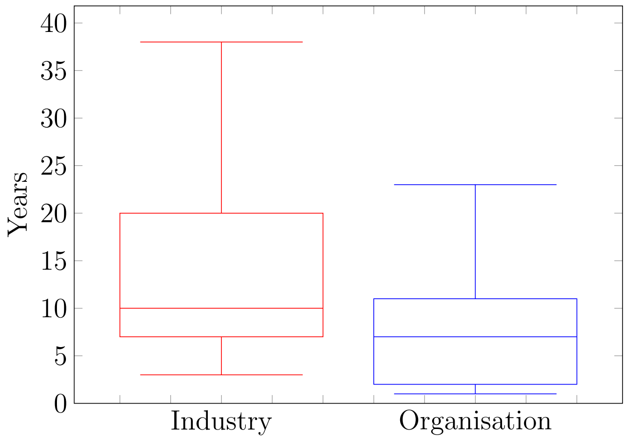


Figure 4.2: Distributions of years worked in industry and at current organisation.

The box on the left in Figure 4.2 shows that 50% of the employees have worked in their current industry for more than ten years. Furthermore, the top of the left hand box shows that 25% of the employees have worked in their current industry for between ten and 20 years while 25% have worked in their current industry for more than 20 years. Half of the remaining 50% of managers have worked in their current industry for between seven and ten years and the remainder of the managers (25%) have been in their current industry for between three and seven years.

The box on the right in Figure 4.2 shows the distribution of the number of years that the managers have been working at their current organisations. Notice that the line indicating the median in the right hand box is the same height (i.e. has the same y-value) as the line indicating the first quartile in the left hand box. This shows that 50% of the managers have worked at their current organisation for seven years or less. Half of this 50% have only worked at their current organisations for one to two years. Neither of the two boxes in Figure 4.2 has values indicated with individual markings (such as dots or crosses) which underlines that there are no values outside the respective fences.

All of the information pertaining to time spent by managers both in their current industries and at their current organisations is shown on the box and whisker graph. Furthermore, the statistics were reported using the inherent format of box and whisker graphs. For more information regarding box and whisker graphs, please refer to Section 2.7.2.

Ten of the managers have spent all of their working years at the same organisation whereas ten have changed employers. Once again the information collected in these two demographic questions is only used to characterise the sample and does not serve as input for the framework.

4.3.4 Highest Qualification

The highest level of education completed by the employees was the last demographic response collected purely to characterise the sample. Managers were asked to indicate their highest qualification on a list consisting of the following options: high school, diploma, bachelor's degree, master's degree, doctorate (PhD) and other. Figure 4.3 shows a summary of the responses.

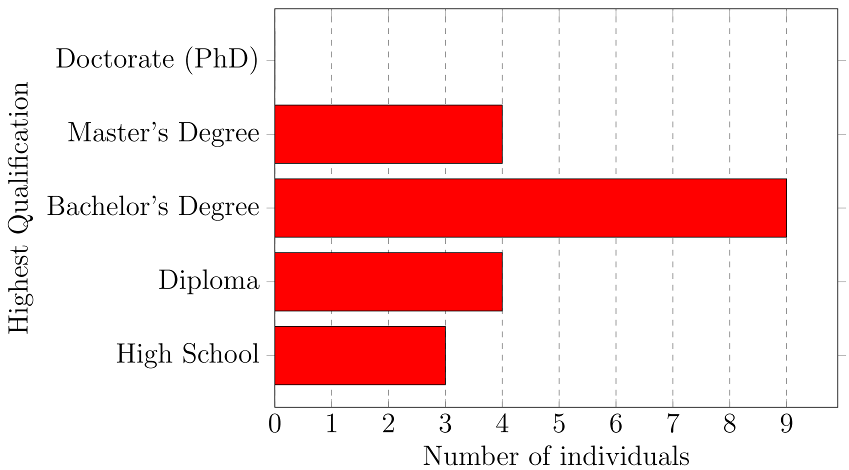


Figure 4.3: Distributions of years worked in industry and at current organisation.

Notice that the responses are ordinal - meaning that they have a natural ordering, but that the differences between qualifications cannot be reduced to a numerical value. High school is the highest qualification for three of the managers whereas four managers indicated that their highest qualification is a diploma. Of the remaining 13 managers, nine managers indicated their highest qualification to be a bachelor's degree and four indicated it to be a master's degree. None of the managers indicated a doctorate (PhD) or any other type of qualification not listed to be their highest completed level qualification.

4.3.5 Department

The reason for enquiring about the department in which each manager works is two fold. Firstly, it once again provides details regarding the characteristics of the sample and hence the generalisability of the findings. Secondly, the department in which employees work was identified by the framework as a possible attribute that could be used to form clusters of employees who would typically be exposed to the same information (see Sections 3.2.1.1 and 3.5.3). Managers were provided with a list of options of the typical departments in PAM organisations as discussed by Hastings (2009) in Section 2.1.2 of this thesis. All of the managers consulted were asked to indicate which department most closely resembled their own department in their organisation. Table 4.1 shows a summary of the results.

Table 4.1: Departments from which the 20 managers originate.

Department	Managers
Engineering	12
Asset Management	7
Finance	1

It is clear that the majority (60%) of the managers work in the engineering departments of their respective organisations. Of the remaining eight managers, seven work in the AM department and one works in the finance department. This means that the findings of this content validation are heavily biased towards engineers and asset managers and should be interpreted as such.

4.3.6 Management Level

The management level of each manager was the last descriptive attribute collected during the consultations. As with the question regarding the department of each manager (refer to Section 4.3.5), the responses were collected to detail the sample as well as to form clusters. The managers were once again provided with a list of possible management levels (and non-managerial levels) and asked to indicate the level which most closely resembles their position in their particular organisation. The list of possible options included: top management, middle management, low level management, entry level employee and trainee. Table 4.2 shows the number of managers in each management level.

Table 4.2: Managerial levels of sample.

Management Level	Individuals
Top	1
Middle	18
Low	1

Table 4.2 indicates that 18 managers (90%) occupy middle management positions. One of the remaining two managers occupies a top management position and the other is in a low level management position. No entry level employees or trainees were consulted and hence the sample is referred to as managers rather than employees. Once again the sample is biased towards employees in middle management positions and the responses should be interpreted accordingly.

4.4 Worked Example of Framework

In this section, the proposed information encoding framework discussed in Chapter 3 and summarised in Figure 3.1 will be applied to a hypothetical scenario. It is assumed that the reader is familiar with the contents of the proposed information encoding framework and Chapter 3 can be consulted should anything in the worked example be unclear. The fact that the researcher was not part of the organisations from which the managers originate meant that the generic framework discussed in Chapter 3 was used to collect preferences. This also means that some of the steps were not executed in exact accordance with the framework. However, these steps will still be discussed to provide a comprehensive overview of the use of the framework albeit at a high level of granularity.

Only a single contact session was possible with each manager and the information obtained was processed after each contact session. Consequently, none of the steps could be repeated if there were any erroneous responses even though it might be proposed in the framework.

4.4.1 Phase 1: Contextualisation

The contextualisation phase is aimed at tailoring the framework to the organisation where it is being implemented. Since a generic framework was applied to multiple organisations as opposed to a single organisation as per the framework's design, the contextualisation step was not particularly relevant and thus the steps will only be discussed as an overview.

4.4.1.1 Step 1: Map Organisation

The first step in the framework includes the identification of employee attributes that could possibly serve as filters when isolating specific groups of employees. As a generic framework was used, the only two attributes which were collected with the aim of being used to form clusters were the departments from which the managers originated as well as their level in the managerial hierarchy in the organisation. The departments and management levels of managers in the sample are discussed in Sections 4.3.5 and 4.3.6 respectively.

For the sake of protecting the anonymity of the managers in the sample, the departments and managerial levels of the employees are reported collectively. It should be noted that the combinations of options regarding departments and managerial levels provided to the managers are based on common examples in literature. The options provided to managers might therefore not be accurate representations of their organisational structure. Consequently, managers were asked to indicate the option which was most relevant to their actual department and managerial level. This underlines the importance of mapping the organisational structure and adapting the information collected by the framework to it.

Mapping the organisation, according to the framework, also includes determining the data presentation capabilities of the organisations. This could not be done for each organisation before the consultations. However, managers were asked to mention formats which did not form part of the list of options of the generic framework but to which they have been exposed before in their organisations. Figure 3.2 in Chapter 3 shows a check list which can be used to review all the formats that can be reproduced by the organisation. The check list was reproduced in Figure 4.4 to indicate how the missing formats (as indicated by the managers) could be added.

Only two formats were added to the generic list, namely: pie graphs and waterfall graphs. Pie graphs are discussed extensively in Section 2.7.6, but it was decided not to include it in the list of generic options as it does not convey enough information to be considered informationally equivalent to the other formats in the generic list. Nevertheless, pie graphs are still included as a format to be added in future. Waterfall graphs are very similar to histograms, but instead of starting every bar on the horizontal axis, it starts at the height (value) of the previous data point and extends to the height of the current data point. If the current data point is higher than the previous one, the bar is normally a particular colour and if it's lower than the previous data point it is a different colour. Waterfall graphs are also only intended to be used with interval data and thus they can be seen as an alternative to histograms.

Presentation format	Reproducible?
Bar Graph (Grouped)	Yes
Bar Graph (Simple)	Yes
Bar Graph (Stacked)	Yes
Box and Whisker	Yes
Column Graph (Grouped)	Yes
Column Graph (Simple)	Yes
Column Graph (Stacked)	Yes
Histogram	Yes
Histogram (Multiple)	Yes
Line (Area) Graph	Yes
Line Graph	Yes
Point Graph (Dot Plot)	Yes
Radar Graph	Yes
Table	Yes
Pie Graph	Yes
Waterfall Graph	Yes

Figure 4.4: Format check list populated with responses.

Notice that all of the presentation formats in Figure 4.4 were indicated as being reproducible. This is because the managers are not normally tasked with designing reports and it would be naive to assume that they would be aware of all the possible presentation formats in their organisations. Following this logic, there are possibly a variety of presentation formats which can be produced by the organisations from which the managers originate, but of which they are not aware. For this reason, the steps involving presentation formats in the framework should be done by or with the assistance of report designers.

4.4.1.2 Step 2: Consider Organisational Context of Tasks

The second step of the framework aims to find domain specific examples of the 16 data tasks. Since the same generic framework was applied to all six organisations, the most common attribute shared by the organisations was their interest in PAM. As a result, all of the example tasks were related to PAM in one way or another. Two background scenarios were used in order to diversify the examples used in the questionnaire and to avoid benefiting a single industry.

The background scenario used for data tasks related to nominal or ordinal data types was that of a textile industry. It was decided to present the information for tasks aimed at nominal or ordinal data types as the causes and hours of downtime at one or more textile factories. The causes and hours of downtime at a single factory and three factories were used to facilitate data tasks of single and multiple datasets respectively. The data task examples used to collect the preferences of the managers for the task codes related to nominal and ordinal data types are shown in Table 4.3.

Table 4.3: Data tasks used for nominal and ordinal data during consultations.

Task Code	Example
Single dataset	
Background	The graphs show the causes and hours of downtime at a textile factory in 2012.
TC3	Which one of Mechanical and Supplier caused more downtime in 2012?
TC7	What was the approximate combined downtime attributed to Staff and Unions in 2012?
TC11	Which cause was responsible for the most downtime in 2012?
Multiple datasets	
Background	The graphs show the causes and hours of downtime at three textile factories (Factory A, Factory B and Factory C) in 2012.
TC4	Which factory had the most downtime (in hours) attributed to Suppliers in 2012?
TC8	What was the approximate combined downtime caused by Electrical problems for Factory A and Factory C in 2012?
TC12	Which cause was responsible for the single smallest amount of hours downtime and at which factory?

In the case of interval or continuous data types, the background used for tasks was that of mining outputs similar to the data in Table 2.7 in Section 2.7. The mining outputs for twelve months at a single mine and at three mines were presented were used to facilitate data tasks of single and multiple datasets respectively. The data task examples used to collect the preferences of the managers for the task codes related to interval and continuous data types are shown in Table 4.4.

These backgrounds and examples of the data tasks were used when the presentation preferences were collected in Section 4.4.2.1.

4.4.1.3 Step 3: Match Presentation Formats to Tasks

The third step in the framework requires all the presentation formats which can be reproduced by the organisation to be matched to tasks which they can facilitate. In other words, if an information presentation format was indicated

Table 4.4: Data tasks used for interval and continuous data during consultations.

Task Code	Example
Single dataset	
Background	The graphs show the tons of ore processed by a mine in each month of 2014.
TC1	What is the approximate range of the outputs (difference between highest and lowest output)?
TC5	In which month was more ore processed in 2014, May or July?
TC9	Approximately how many tons of ore was processed in the first quarter of 2014 (Jan, Feb and Mar)?
TC13	In which month of 2014 was the least ore processed?
TC15	Was the trend in outputs for 2014 increasing, decreasing, reversing or noise?
Multiple datasets	
Background	The graphs show the tons of ore processed by three mines (Mine A, Mine B and Mine C) in each month of 2014.
TC2	Which mine had the smallest variation in output in 2014?
TC6	Which mine processed more ore in June 2014, Mine A or Mine B?
TC10	How many combined tons of ore was processed by the three mines in December 2014?
TC14	For each mine, indicate the month in which the most ore was processed in 2014.
TC16	Classify the trends of the outputs of 2014 as increasing, decreasing, reversing or noise for each mine.

as being reproducible by the organisation where the framework is being implemented, each one of the 16 data tasks which can be facilitated by that format should be identified.

A matrix can be set up to indicate all the task codes that are facilitated by a particular presentation format. Table 4.5 shows how the reproducible presentation formats have been matched to the task codes which they facilitate. Notice that the column headings in Table 4.5 correspond to the presentation formats indicated as reproducible in Figure 4.4. The only new formats added to the generic version of Table 4.5 are pie graphs and waterfall graphs. As waterfall graphs are a variation of histograms, the same tasks are facilitated by it.

Table 4.5: Presentation formats that facilitate each task code.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	Pie Graph	Waterfall Graph	TOTAL
TC1	✓	✓	✓	✓			✓			✓	✓		✓	✓	✓	✓	**11**
TC2	✓	✓	✓		✓	✓		✓	✓	✓		✓	✓	✓			**11**
TC3	✓	✓	✓	✓			✓								✓		**6**
TC4	✓	✓	✓		✓			✓									**5**
TC5	✓	✓	✓	✓			✓				✓		✓	✓	✓	✓	**10**
TC6	✓	✓	✓		✓			✓				✓	✓				**7**
TC7	✓	✓	✓	✓			✓										**5**
TC8	✓	✓	✓		✓			✓									**5**
TC9	✓	✓	✓	✓			✓				✓		✓	✓		✓	**9**
TC10	✓	✓	✓		✓			✓				✓	✓				**7**
TC11	✓	✓	✓	✓			✓								✓		**6**
TC12	✓	✓	✓		✓			✓									**5**
TC13	✓	✓	✓	✓			✓				✓		✓	✓	✓	✓	**10**
TC14	✓	✓	✓		✓			✓				✓	✓				**7**
TC15	✓	✓	✓	✓			✓				✓		✓	✓		✓	**9**
TC16	✓	✓	✓		✓	✓		✓	✓			✓	✓	✓			**10**

As was discussed in Section 3.2.3, a check mark (✓) at the intersection of a column and a row shows that information presentation format in that column can support the execution of the data task in that row. If not, the intersecting cell is filled with grey to indicate that it is not a viable option. The value in the *TOTAL* column indicates the number of possible presentation formats for each task code.

However, as has been stressed before, the framework used to collect preferences was not tailored to the six organisations from which the managers originated. Hence, although the pie graph and waterfall graph formats have been added to the check list in Figure 4.4, they were not included in the neural training phase discussed next.

4.4.2 Phase 2: Neural Training

This step aimed to collect and store the information presentation preferences of the managers. The data collected in this phase forms the foundation on which all recommendations made by the framework are based and can be used repeatedly when designing reports.

4.4.2.1 Step 4: Determine Employee Preferences

All of the managers were required to indicate their primary and secondary presentation format preferences for each task code as well as formats which they would like to be avoided in reports if possible. Preferences for the 16 data tasks were collected by means of a paper based questionnaire. Each data task had an introductory section followed by the various presentation formats which support the execution of that particular task (as indicated in Table 4.5, but without the newly added formats). The introductory section included:

- the task name;
- the data type and number of datasets to which that the task is being applied;
- a short description of the task;
- what is required from the user in that task;
- a practical example; and
- an example task.

After the introduction, fictional information related to the background was presented in all the presentation formats that were identified as appropriate to support the particular task code. The information was either based on a single dataset or multiple datasets, depending on the task code.

Managers were asked to follow the criteria outlined in Table 3.6 when indicating Primary (P), Secondary (S) and Avoid (A) responses. A summary of the P, S and A responses as well as the handling of erroneous responses follows next.

Primary Preferences Managers were asked to indicate their primary preference of format in which they would like information to be presented to them when executing each of the 16 different task codes. Of the 20 managers consulted, only 15 indicated exactly one primary preference for each task. Of the remaining five managers, three managers indicated two primary preferences for one of the 16 task codes and the remaining two managers did not provide a primary preference at all for one of the 16 task codes. If a task had no or more than one primary preference indicated by a manager, the response of that manager for that task was ignored. The combined primary preferences as indicated by the managers are shown in Table 4.6.

Notice that the primary preferences have merely been summed arithmetically to give an overview of the combined responses. In other words, the value *14* indicated at the intersection of the *TC7* row and the *Table* column indicates that out of the 20 managers, 14 indicated that they would prefer information to be presented in tabular format when having to perform task code TC7.

Table 4.6: Primary preferences of consulted managers.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	**TOTAL**
TC1	1	0	1	0			7			3	4		3	1	**20**
TC2	2	0	1		0	0		3	2	3		2	5	2	**20**
TC3	2	0	0	3			15								**20**
TC4	6	0	0		1			12							**19**
TC5	3	0	1	1			9				5		1	0	**20**
TC6	4	0	0		0			12				1	3		**20**
TC7	14	0	1	1			4								**20**
TC8	10	1	0		0			8							**19**
TC9	12	0	0	0			3				3		2	0	**20**
TC10	14	0	1		0			2				1	2		**20**
TC11	2	1	0	0			16								**19**
TC12	7	2	0		0			10							**19**
TC13	3	2	0	0			8				3		4	0	**20**
TC14	1	0	0		1			7				3	7		**19**
TC15	1	0	0	2			4				4		9	0	**20**
TC16	1	1	0		0	0		1	0			5	10	2	**20**

The column labelled *TOTAL* indicates that five erroneous responses were discarded. The errors occurred at task codes TC4, TC8, TC11, TC12 and TC14 as they did not have a full complement of 20 responses.

Secondary Preferences Managers were also asked to indicate an alternative format in which they would like information to be presented for each task code if their primary preference was not available. Only 13 managers indicated exactly one secondary preference for each task. If a manager did not indicate a secondary preference for a task or if more than two secondary preferences were indicated, the responses were discarded. No task received a full complement of correct secondary responses and the number of disregarded responses for each task code ranged from one to three. Nevertheless, the correct responses were summed and are show in Table 4.7.

Once again the value shown in the cell where the row of a task code and the column of a presentation format intersect indicates the number of managers who indicated that they preferred that particular presentation format for that task code if their primary preference was not available. For example, the cell at the intersection of the row of task code *TC1* with the column of *Table* indicates that two managers prefer to execute the task in task code TC1 using information presented in tabular format if their respective first choice

Table 4.7: Secondary preferences of consulted managers.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	**TOTAL**
TC1	2	2	0	2			8			0	3		2	0	**19**
TC2	0	0	2		0	0		1	2	2		6	3	2	**18**
TC3	6	2	3	4			4								**19**
TC4	4	0	3		5			5							**17**
TC5	5	0	0	3			3				5		3	0	**19**
TC6	5	0	2		1			4				0	6		**18**
TC7	1	1	0	4			13								**19**
TC8	4	1	0		5			9							**19**
TC9	2	1	0	3			9				3		1	0	**19**
TC10	2	0	0		1			12				0	3		**18**
TC11	6	3	2	5			2								**18**
TC12	4	4	0		4			6							**18**
TC13	2	2	1	3			2				2		6	0	**18**
TC14	1	1	1		2			2				3	7		**17**
TC15	0	4	1	0			4				2		5	3	**19**
TC16	0	0	0		1	0		2	1			6	6	3	**19**

presentation formats are not available.

Formats to be Avoided After the primary and secondary preferences were selected by the managers, the managers were asked to indicate the presentation formats which they would like to be avoided if possible. Unlike with the primary and secondary preferences, there were no restrictions on the number of presentation formats which the managers wanted to be avoided. The only condition that still had to be satisfied was that a single presentation format could not be indicated to be more than one of the three possible choices (P, S or A). Each cell in the matrix in Table 4.8 shows the total number of managers who indicated that a specific combination of task code and presentation format should be avoided.

Since there were no restrictions on the number of presentation formats that a manager could request to be avoided, the sum of avoid responses shown in the *TOTAL* column for each task code has no meaning and is only included as it is part of the template. It should be noted that even though the numbers in Table 4.8 are positive, they will ultimately be encoded and subtracted from the primary and secondary preferences. Furthermore, even though there were restrictions on the responses that managers could give for each task code, there was no guarantee that the managers would comply with the restrictions. The

Table 4.8: Avoid responses of consulted managers.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	**TOTAL**
TC1	4	4	5	1			0			7	0		1	3	**25**
TC2	3	2	4		0	2		0	0	4		1	0	1	**22**
TC3	4	2	6	0			0								**12**
TC4	5	4	5		1			1							**16**
TC5	2	6	4	1			0				0		0	5	**18**
TC6	3	6	3		2			0				8	1		**23**
TC7	2	6	8	1			0								**17**
TC8	2	5	3		1			0							**11**
TC9	2	4	7	0			0				1		3	3	**20**
TC10	0	5	6		0			0				4	1		**16**
TC11	3	4	4	0			0								**11**
TC12	1	5	6		1			0							**13**
TC13	4	1	7	0			0				0		0	6	**18**
TC14	6	3	7		1			0				5	0		**22**
TC15	7	5	8	2			1				1		0	3	**27**
TC16	7	4	6		3	2		2	2			1	0	2	**29**

validation step of the framework which aims to address this issue is discussed next.

4.4.2.2 Step 5: Validate Responses

This step aims to ensure that the information entered into the framework is usable when performing the subsequent steps. Section 3.3.2 proposes that the information should be collected in electronic format to reduce the possibilities for erroneous inputs. However, for practical reasons, each manager's information was collected using paper based questionnaires during a single contact session. This means that the information collected could only be processed after it has been collected and that erroneous data could only be identified, but not changed. This section will thus, in addition to discussing the efforts made to design validation elements into the questionnaires, also discuss the importance of the validation step in the framework as a whole and how it could have improved the quality of the data collected from the managers.

Validation was designed into the framework in two ways. The first was to provide the managers with a list of options of departments and management levels. Managers were then required to indicate a single department and a single management level which most closely resembles their own. This ensured that there were a finite number of different responses that could be provided by the managers. This is particularly important for isolating managers from the same department or on similar management levels when forming clusters as different organisations might have varying names for the same departments and management levels. However, providing the managers with a limited set of choices for a question does not prevent them from selecting more than one answer or no answer at all. Fortunately, as can be seen in Tables 4.1 and 4.2, the full complement of managers consulted indicated a single department and a single management level which most closely resembles their own.

Another way in which validation was designed into the framework was by having every presentation format in Step 4 be accompanied by three boxes which could be marked by managers to indicate whether that particular format was a P, S or A. Managers were also given the option not to mark any of the three options if the format was neither their first nor secondary preference while they also didn't require it to be avoided if possible. This prevented managers from mistakenly indicating that a format is more than one of P, S or A as they could see when a box was already marked. Furthermore, by requiring managers to indicate the preferences next to each format and not on a separate page, no mistakes could arise when transferring preferences from the page showing the examples to the page where they had to be indicated.

Despite attempts to minimise incorrect responses, there were still problems with the quality of the data collected. For example, consider the far right column of Table 4.6 marked *TOTAL*. Any cell that does equal 20 indicates that, for that task code, not all 20 managers indicated exactly one primary preference. In other words, the difference between 20 and the number in the cell indicates the number of managers who either indicated more than one primary preference for that task code or no primary preference at all.

This could have been prevented if the preferences were collected in an electronic format with built in conditions that the responses had to satisfy before they could be saved. One way of doing this is by including a warning message when an employee attempts to enter more than one primary or secondary preference or if the employee attempts to move to the next task code without having indicated a primary and/or a secondary preference. Finally, by including a condition that each format can only be at most one of P, S and A (with the option of not being any of those), it will ensure that the criteria in Table 3.6 are met.

Out of the 640 primary and secondary preferences indicated by the managers, there were 31 instances where the number of primary or secondary preferences indicated for a task code was not equal to one. All of these 31 incorrect responses could have been avoided if the responses were collected electronically.

4.4.2.3 Step 6: Create or Save in Repository

In this step, the information presentation preferences are saved in a repository which allows easy retrieval. Section 3.3.3 in Chapter 3 outlines the advantages of storing the information in a database which allows the retrieval of the preferences of clusters by subjecting the entries to multiple criteria. As the information in this worked example will only be used once, the advantages of designing an information system did not justify the disadvantages such as development cost and time. For this reason, the information was stored in a Microsoft Excel™ spreadsheet and manipulated accordingly.

All of the responses received from the managers were recorded individually (as to allow the isolation of clusters in Section 4.4.4.3), but only the collective results are reported. Table 4.9 shows the combined preferences (P, S and A) of the managers after they have been encoded and saved.

Table 4.9: Combined preferences of 20 managers after encoding.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC1	-4	-6	-8	0			22			-8	11		6	-4
TC2	-2	-4	-4		0	-4		7	6	0		8	13	4
TC3	2	-2	-9	10			34							
TC4	6	-8	-7		5			27						
TC5	7	-12	-6	3			21				15		5	-10
TC6	7	-12	-4		-3			28				-14	10	
TC7	25	-11	-14	4			21							
TC8	20	-7	-6		3			25						
TC9	22	-7	-14	3			15				7		-1	-6
TC10	30	-10	-10		1			16				-6	5	
TC11	4	-3	-6	5			34							
TC12	16	-2	-12		2			26						
TC13	0	4	-13	3			18				8		14	-12
TC14	-9	-5	-13		2			16				-1	21	
TC15	-12	-6	-15	0			10				8		23	-3
TC16	-12	-6	-12		-5	-4		0	-3			14	26	3

As discussed in Section 3.3.3, the following formulas were used to encode the preferences:

Primary preference (P) = 2 points
Secondary preference (S) = 1 point
Avoid (A) = -2 points

Table 4.9 was created by summing the results in Tables 4.6, 4.7 and 4.8 after they have been encoded. This can be seen when considering the value of *-4* in the cell at the intersection of the row of task code *TC1* with the *Table* column in Table 4.9 (shaded yellow). Table 4.6 reports that one manager indicated a table as his or her primary information presentation format when completing task code TC1. Table 4.7 indicates that two managers indicated that using a table for task code TC1 was their second choice (secondary preference). Finally, Table 4.8 shows that four managers indicated that they would like to avoid information presented in tabular format when completing task code TC1. The calculation of the value of *-4* at the intersection of the row of task code TC1 with the Table column in Table 4.9 is summarised in Table 4.10.

Table 4.10: An example showing how an encoded value is calculated.

Response	Weighting	Number of occurrences	Total points
Primary preference	2	1	2
Secondary preference	1	2	2
Avoid	-2	4	-8
Combined total (value in Table 4.9)			**-4**

The Microsoft Excel™ spreadsheet used in this worked example was set up in such a way that the preferences of certain managers could be removed in order to isolate the preferences of the target audience. Note that the statuses of the preferences entered into the worksheet were indicated as *unconfirmed*.

4.4.3 Phase 3: Confirmation

Phase 3 of the framework aims to ensure that the recommendations made by the framework are based on the right information. In other words, it aims to ensure that the preferences stored in the repository are the actual preferences of the individuals consulted in Step 4 of the framework. This step, however, is optional and will not prevent the framework from delivering an output.

4.4.3.1 Step 7: Confirm Preferences

Recall that in Section 3.4.1 it was explained that there are two possible reasons why the stored responses might not be a true reflection of the actual preferences of the managers. Firstly, an error could have occurred when the responses were transferred from the questionnaires collecting the preferences to the repository. Secondly, the managers could have misunderstood the requirements of the task codes when indicating their information presentation preferences for those task codes.

The first possibility for storing erroneous responses in this worked example was minimised by double-checking the responses after they were entered into the Microsoft Excel™ spreadsheet. However, this does not rule out the possibility for these kinds of errors completely.

Chapter 3 explains that the second possibility for storing erroneous data can be minimised by providing managers with a sample report containing all 16 task codes as well as the primary and secondary preferences they indicated for each task. Managers could then indicate whether they still agree with their initial choices or if they should be changed. Unfortunately, since the managers were only available for a single consultation (contact session), there was not an opportunity for the managers to confirm their preferences. This means that the statuses of the preferences stored in the Microsoft Excel™ worksheet remained *unconfirmed*. As mentioned before, the framework indicates that this step is optional and thus the fact that it was not completed in the worked example is not detrimental to its outcome.

The contextualisation, neural training and validation phases are all executed to collect and store organisation specific information. However, once they have been completed, they will not be repeated every time information is reported, although they might be reviewed every one to two years.

4.4.4 Phase 4: Application

The application phase outlines how reports should be created using the information encoding framework. Contrary to the first three phases, it is recommended that the steps in this phase should be repeated every time information is reported. To illustrate this phase in the worked example, a hypothetical scenario will be created wherein fictional datasets will have to be reported to a cluster of the managers. All of the information regarding the datasets are fictional and can be thought of as information related to three asset intensive factories that produce some kind of product. The 20 managers consulted will represent the employees of this fictional company along with their attributes (departments and management levels) and their information presentation preferences. Each step in the application phase of the framework will be followed closely to illustrate how the responses collected from the managers will aid in selecting formats in which to encode information in a fictional report.

4.4.4.1 Step 8: Design Report

According to the framework, the first step in the application phase is designing the report. This means that designers should decide on what they want to convey in the report and which datasets can be used to support their message. A summary of the datasets which will be used in the report is thus the output of this step. Table 4.11 shows information about five fictional datasets to which the steps in the application phase of the framework will be applied.

Table 4.11: Summary of datasets used in worked example.

No.	Dataset Name	Period	Source	Datasets	Data Type
1	Total output per factory	2014	Multiple	Single	Nominal
2	Factory A five-year output	2010-2014	Factory A	Single	Interval
3	Monthly output Factory A	2014	Factory A	Single	Interval
4	Sources of downtime per factory	2014	Factory A	Multiple	Nominal
5	Critical HS&E incidents per factory	2010-2014	Multiple	Multiple	Interval

4.4.4.2 Step 9: Identify Data Tasks in Report

Once the datasets that need to be included have been identified and summarised, the main data task that needs to be facilitated by each needs to be identified. This will normally be decided by the designers of the report as they can match the main reason for including the dataset in the report to one of the data tasks discussed in Section 3.2.2. Furthermore, the task code of that data task in the framework also has to be determined and Table 3.3 (also in Section 3.2.2) shows the task codes as a function of the task name, number of datasets and data types. Table 4.12 shows how each dataset has been assigned a task code based on its number of datasets, data type, and task name.

Table 4.12: Datasets with task names and task codes assigned to each.

No.	Dataset Name	Datasets	Data Type	Task Name	Task Code
1	Total output per factory	Single	Nominal	Compare Discrete Value	TC3
2	Factory A five-year output	Single	Interval	Identify Trend	TC15
3	Monthly output Factory A	Single	Interval	Characterise Distribution	TC1
4	Sources of downtime per factory	Multiple	Nominal	Find Extreme Values	TC12
5	Critical HS&E incidents per factory	Multiple	Interval	Identify Trend	TC16

Now that the task codes have been identified, the most preferred information presentation format for each can be retrieved.

4.4.4.3 Step 10: Identify Target Audience

This step is critical to the effectiveness of the recommendations made by the framework. Designers should pay special attention when identifying the targeted recipients of the reporting process as including or excluding the wrong employees in the clusters could greatly decrease the effectiveness of the framework. The target audience of this fictional report has been chosen so as to include the most prominent department and the management level. Table 4.1 shows that twelve of the 20 managers indicated that they worked in the engineering departments of their organisations. Of these twelve managers, ten were middle managers with one of the remaining two being a high level manager and the other a low level manager. The target audience for this fictional report was thus the ten middle managers in the engineering department.

4.4.4.4 Step 11: Match Audience Preferences to Tasks

This step combines all of the information collected in the previous steps to recommend an information presentation format for each dataset identified in Step 8 (Section 4.4.4.1). These recommendations will attempt to satisfy as many of the information presentation preferences of the target audience as possible. In order to do that, the preferences of the target audience need to be isolated from those of the rest of the managers. The preferences of the ten managers who did not form part of the target audience were filtered out in the Microsoft Excel™ spreadsheet in which they were stored. Table 4.13 shows the combined encoded P, S and A responses of the ten middle managers in the engineering department - the target audience of the fictional report.

Table 4.13: Combined P, S and A preferences of target audience after encoding.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph
TC1	-5	-1	-4	0			15			-8	3		4	0
TC3	-2	-3	-4	5			17							
TC12	11	1	-6		0			11						
TC15	-8	1	-7	0			7				5		11	-4
TC16	-8	0	-6		-4	-2		-1	-2			8	13	3

The cell with the highest value for each task code has been shaded green to indicate the presentation format which is recommended for that particular task code. Notice that row of task code *TC12* has two cells which contain the value *11* which is also the highest value in that row. This means that there are two formats which are considered to be equally appropriate for task code TC12 by the framework, namely tables and grouped column graphs. Section 3.5.4 mentions three steps which should be followed sequentially to determine which one of the tied formats should be used. These steps are:

1. The presentation format with the smallest number of *-2*'s (avoid) responses should be chosen.

2. If there is still a tie, the format with the most *1*'s (secondary preferences) should be used.

3. If there is still a tie, it means that the number of P, S and A responses for the tied formats are the same. The designer can then use his or her own discretion in deciding on one of the tied formats.

Table 4.14 shows the number of primary, secondary and avoid responses for task code TC12 (not encoded) as indicated by the ten managers in the target audience.

Table 4.14: Number of P, S and A preferences of target audience for TC12.

Task Code	Table	Point Graph (Dot Plot)	Radar Graph	Bar Graph (Simple)	Bar Graph (Grouped)	Bar Graph (Stacked)	Column Graph (Simple)	Column Graph (Grouped)	Column Graph (Stacked)	Box and Whisker	Histogram	Histogram (Multiple)	Line Graph	Line (Area) Graph	**TOTAL**
Primary															
TC12	4	2	0		0			4							**10**
Secondary															
TC12	3	1	0		2			3							**9**
Avoid															
TC12	0	2	3		1			0							**6**

Neither one of the first two steps recommended to break the tie are applicable as none of the managers requested that tables or grouped column graphs be avoided when performing task code TC12 and both formats were the secondary choices of three managers. Since the formats are tied and both have the same number of S and A responses, it means that both formats have the same number of P responses as well. Consequently, as indicated by the third step recommended when there is a tie, the designers of the report can use their own discretion when deciding whether to use a table or a grouped column graph when encoding the information for task code TC12.

The column on the far right shows the total number of primary, secondary and avoid preferences that have been indicated by the ten managers in the target audience. Notice that of the ten managers, only nine indicated a secondary preference for task code TC12. If the manager who did not indicate a secondary preference had indicated either a table or a grouped column graph as his or her secondary preference, that format would have been selected. This illustrates the importance of validating the responses to ensure that a single primary and a single secondary response is collected for each task code for each manager.

Table 4.15 shows the recommended format for each dataset identified in Step 8 of the framework. These recommended information presentation formats are based on the preferences indicated by the target audience in Step 4 of the framework. All of the datasets, except the *Sources of downtime per factory* dataset, have a single recommended presentation format. The designers of the report can decide whether they would like to use a grouped column graph or a table to depict the information in the *Sources of downtime per factory* dataset.

Table 4.15: Condensed example of the framework output showing the ideal presentation format for each task.

No.	Dataset Name	Task Code	Presentation Format
1	Total output per factory	TC3	Column Graph (Simple)
2	Factory A five-year output	TC15	Line Graph
3	Monthly output Factory A	TC1	Column Graph (Simple)
4	Sources of downtime per factory	TC12	Column Graph (Grouped) **or** Table
5	Critical HS&E incidents per factory	TC16	Line Graph

4.4.4.5 Step 12: Create Report

The last step in the framework is not only creating the report, but also creating the appropriate presentations for each dataset. Contextualising the framework (Phase 1) ensures that all of the recommended presentation formats are producible by the organisation. Appendix D shows information of each dataset to be reported as well as an example of the format recommended by the framework.

4.5 Validation of Framework

At the end of the consultation with each manager, they were asked to assess a variety of the aspects in the framework using a questionnaire (shown in Appendix C). It is important to realise that, although the 20 managers who were consulted were asked to assess the framework, the validation discussed in this section does not form part of the proposed framework. Unfortunately, one of the 20 managers indicated that he or she did not have time to complete the validation section and thus only 19 responses were received from the managers for all of the questions but one. One of the questions only had 18 responses. All of the elements which the managers were asked to assess and their responses are subsequently discussed.

4.5.1 Data Communication Problems in Organisations

The framework will serve no purpose if it is designed to solve a problem which does not actually exist in industry. As the framework was created to solve information communication problems in data-intensive industries, it first had to be established whether the organisations from which these managers originate are indeed data-intensive. All 19 of the managers confirmed that their organisations are indeed data-intensive with 89% of the managers indicating that they expected the amount of data in their organisation to increase in future.

Only 58% of the managers were of the opinion that factual data is available for the majority of the decisions made in their organisations. Furthermore, a mere 63% of the managers believed that these decisions are generally based on and supported by factual data. Even more concerning is the fact that 68% of the managers stated that some decisions are based on opinions rather than data even when there is data available that can be processed to support the decisions.

From the above responses it is clear that the majority of managers expect the amount of data in PAM organisations to increase, but that there are still decisions being made based on opinions rather than on the factual data even though it is available. However, the full complement of managers indicated that data will be used more frequently in their organisations if it was packaged in a more usable format. This shows that there is a need for an information encoding guideline and that its successful implementation can assist management in considering facts rather than opinions when making business decisions.

The proposed framework will only be accepted by the employees if its benefits are clear to the users. As a result, the value of the framework, as perceived by the managers, will be discussed next.

4.5.2 Benefits of the Framework

In order to determine the value that can be created by the framework, the managers were queried about the current data presentation practices in their organisations. Out of the 19 managers, only 10 (53%) indicated that they were satisfied with the way that data is currently being presented in their organisation. However, 74% of the managers indicated that the current presentation formats used by their organisations match their personal preferences.

Finally, the managers were questioned about the value of reports tailored to their presentation preferences would be. Even though 74 % of the managers indicated that the current presentation formats used in the organisations match their personal preferences, all 19 managers still indicated that reports tailored to their personal information presentation preferences would be of value to them.

Now that the benefits have been discussed, the viability of implementing such a framework in a PAM environment needs to be considered.

4.5.3 Viability of the Framework

Although it is important that the framework's benefits are clear to employees, the implementability of the framework in organisations also has to be investigated. Firstly, the managers were asked whether they believed that the proposed information encoding framework can address and improve the problem it was designed for. 95% of the managers believed that the framework could improve data communication in their respective organisations.

A further 89% of the managers indicated that they believed that their coworkers would be prepared to fill in the questionnaire if they knew that the way in which information was communicated to them would be tailored to their personal preferences if they did. In total, 18 of the 19 (95%) managers reported that, in their opinion, their respective organisations could benefit from the proposed framework.

These responses suggest that the framework is indeed implementable in PAM organisations. Furthermore, it is important to note that the framework needs to be adopted across the whole organisation if it is to be successful. The large proportion of managers who indicated that they believed their coworkers would be prepared to fill in the questionnaire is very promising as the implementability of the framework is, to a large degree, directly dependent on the support of all the employees in an organisation.

4.5.4 Components within the Framework

In order to make the framework as effective as possible, the managers were asked to assess assumptions of the framework. In some cases, these assumptions were the premise on which some of the steps in the framework were based. By confirming these assumptions, the need for certain steps in the framework can be validated. Moreover, having an assumption revoked means that future research can investigate the effect of removing or altering the step(s) based on that particular assumption. The opinions of the managers with regard to three assumptions were collected and are discussed next.

4.5.4.1 Creating Employee Clusters

This section investigates the supposition that different clusters of employees have unique information preferences and that they are exposed to different data. As the managers were asked to assess the generic framework, only the clustering components in that framework were evaluated. In the generic framework, clustering was only done using different levels of management and departments as input.

The first assumption that was assessed was that employees on different management levels in the organisational hierarchy are exposed to different data. All 19 managers indicated that, according to them, different levels of management are exposed to different data. Furthermore, all of the managers stated that they expected the information presentation preferences of the different levels of employees to vary.

In addition to the data exposure and presentation preferences of different levels of employees, those same aspects were investigated for different departments. In the case of departments, 18 out of the 19 (95%) of the managers thought that different departments are exposed to different data. Moreover, all of the managers expected the preferences of different departments to vary.

The vast majority of the managers thus agree that different levels of management and departments are subjected to different data and that the presentation preferences thereof might vary from department to department as well as from management level to management level. It is thus reasonable to assume that the formation of clusters is a necessary step in the framework and that there is no one-size-fits-all information presentation solution or recommendation for the whole organisation.

4.5.4.2 Determining Data Presentation Capabilities

One of the main reasons for executing the contextualisation phase in the framework is to ensure that the information presentation formats recommended by the framework are indeed producible by the organisation where it is being implemented. Additionally, it is important to include information presentation formats in the framework which are used by the organisation, but which do not form part of the generic list of formats tested. For these reasons, managers were consulted to determine whether their organisations, according to them, can reproduce the formats in the generic framework and whether there are formats which they would like to be included. Sixteen out of 19 (84%) of the managers indicated that they believed that their organisations would be capable of reproducing the presentation formats in the generic framework.

As expected, the list of presentation formats used in the generic framework is by no means complete. Nine out of the 19 (47%) of the managers indicated that their organisations used presentation formats which do not appear in the generic framework to present similar information as in the questionnaire. Managers were also asked to provide the names of these graphs which did not appear in the framework. Most of recommendations made by the managers were not applicable or only different names for the same formats in the generic list, although various managers indicated that they thought that pie graphs and waterfall graphs could have been included in the framework. As the managers are not generally tasked with creating reports it is also possible that they were not aware of the names of the formats which were missing and hence they did not mention them.

Figure 4.4 in Section 4.4.1.1 shows how the pie graph and waterfall graph formats have been added to the generic list. As mentioned before, the single contact session with each manager meant that these two additional graphs could not be presented as an option to the managers when determining their preferences. However, the fact that pie graphs display too little information to be considered informationally equivalent to the other formats resulted in its initial exclusion from the list of appropriate graphs. The waterfall graph format is thus the only format that would have been added as an option if the contextualisation phase could have been executed.

The assumption that the list of presentation formats is incomplete has thus been confirmed. The fact that three of the 19 managers interviewed indicated that there are certain presentation formats in the framework which can, according to them, not be reproduced by their respective organisations illustrates the importance of determining the data presentation capabilities of the organisation. If employees indicate a format which cannot be reproduced by their organisation as one of their preferences, the output of the framework will be meaningless.

4.5.4.3 Finding Domain Specific Examples of Tasks

The last assumption made by the framework which was investigated is the assumption that presenting employees with domain specific examples of the data tasks would improve their understanding. This is the only validation question where only 18 (instead of 19 as with all the other questions) of the managers provided a response. Exactly half (nine) of the managers indicated that they thought domain specific examples of tasks would improve the understanding of employees while the other half did not think that it would have an effect. The fact that some of the managers expected understanding to increase when using domain specific examples of the tasks suggests that it might be worthwhile to include the step. Including this step, however, does not require any additional effort (or time) by the employees indicating their preferences, only by the person(s) implementing the framework.

4.6 Discussion of Preference Responses

During the consultations, responses were collected in three parts. The discussion of the demographic information collected and the validation responses are deemed to be of adequate comprehensiveness. Although the preference responses of the managers were used in the worked example, they have not been discussed. Consequently, a summary of the preference responses is warranted. Recall that the primary preferences of all the managers in the worked example are shown in Table 4.6. Table 4.16 shows a more detailed analysis of the primary preferences indicated by the managers.

Table 4.16: Analysis of primary preferences in worked example.

Task Code	Total responses	Possible formats[A]	Formats selected[B]	Proportion selected[C]	Most responses for format[D]	Highest percentage responses per format[E]
TC1	20	9	7	78%	7	35%
TC2	20	11	7	64%	5	25%
TC3	20	5	3	60%	15	75%
TC4	19	5	3	60%	12	63%
TC5	20	8	6	75%	9	45%
TC6	20	7	4	57%	12	60%
TC7	20	5	4	80%	14	70%
TC8	19	5	3	60%	10	53%
TC9	20	8	4	50%	12	60%
TC10	20	7	5	71%	14	70%
TC11	19	5	3	60%	16	84%
TC12	19	5	3	60%	10	53%
TC13	20	8	5	63%	8	40%
TC14	19	7	5	71%	7	37%
TC15	20	8	5	63%	9	45%
TC16	20	10	6	60%	10	50%

A: Number of formats presented as alternatives for task code. **B**: Number of different formats marked as primary preference at least once for that task code. **C**: Proportion of formats marked as a primary preference at least once out of all the alternatives. **D**: The most number of primary preferences indicated for one format. **E**: The largest number of responses for a single format as a percentage of the total responses for the task code.

The first statistic which is of interest is the column indicating the *Formats selected*. It shows that for each of the 16 task codes, there were at least three different formats marked as primary preferences. Furthermore, the number of different formats indicated as primary preferences for the task codes TC1 and TC2 was seven. In other words, the ideal format in which information should be presented for task codes TC1 and TC2 could be one of seven different possibilities. This indicates the variability in the primary presentation format preferences within the sample of managers consulted.

The next statistic of interest is the column indicating the *Proportion selected*. This column indicates the number of formats which were indicated as a primary preference by at least one manager as a proportion of all the available options and aims to provide context for the values reported in the column indicating the *Formats selected*. Interestingly, for every task code, at least half of the formats provided as options to the managers were selected as primary preferences.

Finally, the column labelled *Highest percentage responses per format* shows the number of people who indicated the most popular format in the task code as their primary preference. It aims to show the degree to which the majority of the managers agree on the most appropriate format for a task code. For example, the value of 84% reported for task code TC11 indicates that 84% (16 out of 19) of the managers had the same primary preference for that task code. However, the value of 25% reported for task code TC2 indicates that the highest number of managers who indicated the same presentation format as their primary preference for task code TC2 was five.

The fact that there are seven different primary preferences indicated for the exact same task code implies that a designer of a report would have had to include seven different information presentations to satisfy everyone's primary preferences. Additionally, the value of 25% in the *Highest percentage responses per format* column shows that if the most popular format was selected for that task code, 75% of the managers would have received information encoded in a format which is not their primary preference. The preference responses collected confirm that the information presentation needs of individuals differ and that there is a requirement for a structured method which can be applied to address this matter. Furthermore, it suggests a possible reason why the current information communication practices in organisations might overwhelm employees.

4.7 Implications of Selected Content Validation Method

Throughout this chapter it has been mentioned that certain aspects related to the consultations with the managers have influenced the outcomes reported by the study. This section aims to summarise these influences and to comment on the generalisability of the results. Finally, this section will serve as background to some of the limitations and recommendations discussed in Sections 5.2 and 5.3 in Chapter 5.

The organisations selected to take part in the content validation was a sample of convenience based on their accessibility and relationship with the ACRG. In order to limit the intrusiveness of the study, only one contact session with the managers was requested and it was agreed that the results would be reported collectively. As a result of these characteristics of the consultations, there are certain limitations to the results reported.

Firstly, as the six organisations consulted have working relationships with the ACRG (a research group who aims to provide scientific solutions to problems in industry), these organisations might be more susceptible to new practices recommended by research than other organisations. In other words, the responses regarding the implementability of the framework might be biased towards organisations who are more likely to experiment with new methods as opposed to organisations relying on traditional methods.

Secondly, the sample size was relatively small and only consisted of individuals in management positions. The sample size was governed by the availability of individuals from the organisations, but as the validation was conducted in the form of management consulting and not a survey, the number of managers was not of high importance. As all of the individuals were in management positions, the responses are also heavily biased towards the individuals who would typically benefit from the proposed framework. However, the framework is intended to serve as a tool to improve information communication across the entire organisation and, as such, future studies should validate the framework by consulting employees in all levels of the organisations.

Another limiting factor of the research was the fact that a generic framework was used in the validation process. Consequently, the managers who were consulted did not experience the benefits associated with the contextualisation phase of the framework, but were merely informed what they would be. Furthermore, the single contact session with each manager did not allow the managers to evaluate the recommendations which serve as the output of the framework.

However, despite these limiting factors, the preferences provided by the managers and their assessment of the framework achieved the desired aims of the content validation. As such, the limitations above should be considered as an indication of opportunities that can be addressed in future research rather than shortcomings of the current study.

4.8 Chapter Summary

In this chapter, the information encoding framework is applied to and validated by 20 managers from six different data-intensive organisations. The chapter commences with a description of the six organisations that were consulted as well as how the responses of the 20 managers from these organisations are reported. A brief outline of the structure of the consultations is provided followed by an in depth discussion of each of the three parts.

First, the demographic information of the managers is provided to contextualise the findings and to demarcate the generalisability of the results. It is reported that the managers are predominantly males (75%) and that the sample is biased towards middle managers and that the majority of the managers work in the AM or engineering departments of their respective organisations.

Next, the preferences collected from the managers are discussed in the form of a worked example. Although a generic framework was used in the worked example, the steps were adhered to as closely as possible to illustrate how the four phases of the framework (contextualisation, neural training, confirmation and application) are intended to be executed.

This is followed by a discussion of the responses to the validation questions in Appendix C. These validation questions address the problem statement, the perceived benefits of the framework, the viability of the framework as a solution to the problems as well as certain components within the framework. All of the managers indicated that reports tailored to their individual preferences would be of value to them and 95% of the managers believed that the proposed framework could improve data communication in their organisations. Thereafter, the primary responses indicated by the managers are analysed and discussed.

A reflection on the implications that the selected validation process had on the outcome of the worked example and the validation results concludes the chapter. This chapter addressed and achieved the seventh and final objective identified in Section 1.3:

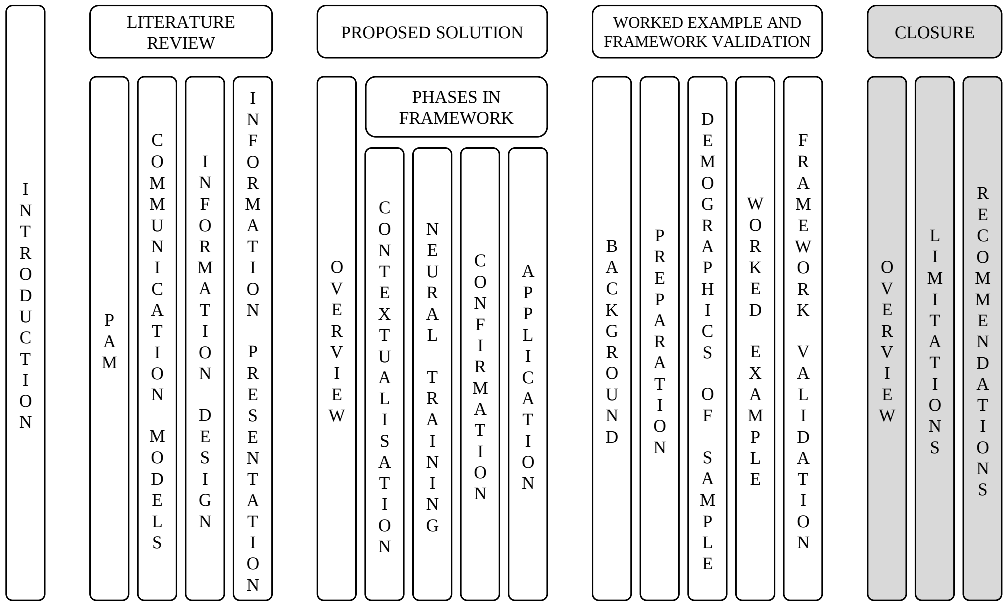
7. Validate the proposed solution with regards to the problem statement.

The next chapter serves as a conclusion to this thesis and mentions limitations of this research as well as recommendations for future research.

Chapter 5

Closure

This chapter aims to bring the study to a close and to summarise the contents and findings of the previous chapters. An overview of the study is provided to contextualise the findings, followed by the limitations encountered in the thesis and recommendations for subsequent studies in the same field.



5.1 Overview

Data capturing capabilities of organisations have improved significantly in recent times. The introduction of the internet has brought about an explosion of digital data which can be used by managers to gain new insights into their businesses. In engineering environments, the increase in data availability can be attributed to embedded sensors in large assets and the continuous monitoring of processes. This data can be translated into information to support factual decision making. In fact, research has found that the speed at which information is made available and the ease with which it can be interpreted is correlated to the quality of decisions made by managers. Moreover, the large amounts of data available have made factual decision making imperative to an organisation's ability to stay competitive. The interpretation of this data, however, is often a difficult task which results in managers disregarding it and basing decisions on their own opinions rather than the factual data.

Previous research has found that managers regularly feel overwhelmed by the amount of data provided to them and that the processing thereof is too difficult. Visualisations have been cited as a possible means of reducing the cognitive strain imposed on an individual when interpreting information. Advancements in information visualisations have resulted in a wide variety of information presentation formats being available to designers of reports. However, cognitive load theory suggests that the choice of format in which information is encoded affects the load placed on an individual's cognitive processes in two ways. Firstly, the cognitive load induced by information encoded in a format with which the individual is familiar is less than that induced by information encoded in a format with which the individual is not familiar. In other words, the cognitive load is reduced when formats with which individuals are familiar are used. Secondly, encoding information in a format which inherently supports the requirements of the task that needs to be executed using the information also reduces the cognitive load. Unfortunately, these are often two competing factors when deciding on an information presentation format as individuals are not always familiar with the most appropriate format for a task.

Another consideration when deciding on a format in which to encode information is the fact that information is frequently reported to more than one individual. Consequently, the collective preferences of the target audience of a report have to be considered as opposed to the preferences of a single individual. This can be a difficult task as literature has shown that the information preferences of individuals frequently differ.

This study proposes an information encoding framework which recommends a presentation format which considers both the requirements of the task that has to be completed and the preferences of the target audience. The development of the framework was preceded by an extensive literature review. In this literature review, PAM was shown to be a data-intensive domain and the importance of factual decision making by managers was highlighted. An investigation into three popular communication models showed that the success of the communication process is determined by the receivers of the information and that they should be the primary consideration when designing communications. Thereafter, the design of information was discussed at the hand of the cognitive load and cognitive fit theories. The literature review concludes with a discussion of information presentation formats and lists a number of popular graphics used to convey numerical information.

The proposed information encoding framework consists of twelve steps that have been divided into four phases. Each phase and its subsequent steps are discussed in a structured and detailed manner. The purpose and example output of each step is described and, as the framework was derived from the body of knowledge discussed in the literature review, the theoretical groundings of the majority of the steps are reported. In order to determine whether the framework is a viable solution to the problems faced by industry, it was validated by managers from data-intensive industries.

Twenty managers from six different data-intensive organisations were approached to both provide inputs for the framework and to validate the framework. A worked example illustrates how the preferences collected from the managers would have been used if the framework was applied in industry. The validation took place in the form of answers to a questionnaire attempting to confirm the validity of the problem statement, the expected benefits of the framework, the viability of the framework in industry as well as certain assumptions made by the framework. It was found that a mere 53% of the managers consulted are satisfied with the way in which information is presented to them in their organisations. As expected, a large proportion (68%) of the managers indicated that certain decisions are still based on the opinions of the decision makers rather than factual information, even when there is data available to support the decisions. The need for an intervention in the way information gets conveyed in organisations was confirmed as all the managers believed that information would be used more frequently if packaged in more usable formats. This need was further substantiated when considering the variety of first choice information presentation format preferences indicated by the managers for the same tasks. For two of the 16 task codes investigated, there were as many as seven different first choice presentation formats indicated by the managers. Moreover, for one of the task codes, the highest proportion of managers who indicated the same presentation format as their first choice was as low as 25%.

The viability of the information encoding framework as a solution to the problem statement was also investigated and 95% of the managers consulted were of the opinion that the proposed framework could improve data communication in their organisations. In addition, 89% of the managers believed that their co-workers would be prepared to use the framework if they were made aware of the expected benefits.

Therefore, the null hypothesis is rejected as information reported in a format which considers both the preferences of the individuals in the target audience and the characteristics of the task that needs to be completed using the information is of value to managers in data-intensive environments. Furthermore, upon completion of this thesis it can be confirmed that all of the research objectives in Section 1.3 have been achieved.

1. The role of data and leadership in AM environments have been studied and a comprehensive understanding was attained.

2. The fundamental principles of the communication process have been established.

3. The influences of various elements of visualisations on an individual's cognitive processes have been investigated.

4. A list of tasks which typically need to be facilitated by data has been created.

5. A variety of formats in which data can be encoded have been investigated and the appropriateness of each format to convey certain data types has been established.

6. A framework was developed which recommends a presentation format based on the preferences of the target audience as well as the characteristics of the task which needs to be supported by the information.

7. The proposed framework was confirmed to be a viable solution to the problem statement.

In conclusion, the proposed information encoding framework developed in this study was identified as a possible means to improve information communication in organisations. Recall that the sample of managers used in the content validation was a sample of convenience and not a randomised sample. As such, the results of the validation cannot be generalised to the entire data-intensive industry; instead, the validation results can be considered representative of the six organisations approached. Although the results of this study cannot be generalised to all data-intensive organisations, it is expected that there will only be marginal differences between organisations. The framework proposed by this study will enable designers of reports to select a format which considers both the preferences of the target audience and the requirements of the task that needs to be executed using the information.

5.2 Limitations

An essential part of any research is the identification of limiting factors within the design and the execution of the research. These factors possibly influenced the outcomes of the research. The limitations encountered during the conception, application and validation of the information encoding framework are listed below.

- Choice of presentation formats – The information presentation formats investigated in Section 2.7 and used in the proposed solution were limited to those which could be produced using Microsoft Excel™. There are a wide variety of presentation formats available to organisations, but in order to keep the number of preferences for each task manageable, those in Microsoft Excel™ were deemed adequate. Furthermore, formats which need to be created uniquely for a dataset, such as infographics and textual summaries, were not included in the framework. These formats need to be created by designers and SMEs and there is no guarantee that all organisations will have such personnel.

- Choice of data tasks – The choice of data tasks used in this thesis was determined subjectively. Only three lists of tasks could be found in literature and any task which did not appear in at least two of these lists was omitted. This was done to ensure that the tasks in the framework were prevalent in literature. However, it also means that there might be certain data tasks that have to be executed by individuals which are not in the list of primary tasks described in Table 3.2 in Section 3.2.2.

- Presentation formats considered in framework – The framework only focuses on the main task that needs to be executed using the information. As there was no consideration of any other tasks that had to be facilitated by the information, only formats that were informationally equivalent were considered.

- Design of the validation – The content validation discussed in Chapter 4 of this thesis was subjected to various constraints that influenced the generalisability of the results (see Section 4.7). Since the contact with managers was limited to a single session, none of the feedback loops in the information encoding framework were used in the worked example. Furthermore, the limited number of managers consulted (20) meant that the validation could not be confirmed statistically. This does not imply that the results are invalid, but merely that their statistical significance was not confirmed.

- Method used in collecting responses – The responses in Chapter 4 were collected using questionnaires in hard copy (paper) format. This did not allow the study to benefit from any of the advantages associated with electronic questionnaires. However, the resources required to develop an electronic questionnaire were not justified for the small sample of managers who were consulted.

- Framework used during consultations – Since there was only a single contact session with the managers during the consultations discussed in Chapter 4, a generic framework was used. This means that none of the steps in the contextualisation phase could be executed to tailor the framework to organisations from which the managers originated. However, the conceptual framework was validated by the managers and not the generic one used to determine their preferences.

- Validation of framework – As a result of the single contact session and the variety of managers consulted, the framework was only validated conceptually. This means that, although the framework's methodology and the theoretical output of the framework were validated, managers could not evaluate a report based on the preferences they indicated.

Although some of these limitations affected the outcomes of this thesis, steps were taken to mitigate the effects where possible. Some of these limiting factors can be avoided or overcome in future studies by following the recommendations in the next section.

5.3 Recommendations for Future Research

The design and execution of the study aimed to answer the research question and to meet all of the objectives outlined in Section 1.3. Although all of the objectives were met, opportunities for improvement and areas that could be researched further were identified during the research process.

1. A more comprehensive list of data tasks can be created using all tasks mentioned in literature. Although it would be impractical to include an infinite number of tasks in the framework, those tasks which are not prevalent in industry can be identified and removed from the initial comprehensive list.

2. The framework can be adapted to collect preferences for combinations of data tasks. By doing this, the number of presentation format alternatives considered by the framework can be increased. The new presentation formats will not have to be informationally equivalent as they will only have to support the combination of data tasks identified for each task code.

3. The framework can be used and validated in a single organisation. This will allow the user of the framework to execute the steps in the contextualisation phase of the framework to tailor the framework to the organisation where it is being applied. Secondly, by implementing the framework at a single organisation, the feedback loops in the framework can be used to ensure that the correct responses are collected. Notice that the feedback loops require more than one contact session with employees. Finally, responses can also be collected electronically with built in conditions that the responses have to satisfy. This would reduce the number of invalid responses supplied by employees.

4. In order to investigate the null hypothesis of this thesis, the value of the framework was assessed by the managers who would typically benefit from the recommendations made by it. However, the designers of reports are the most likely users of the framework. This means that the implementability (practicality) of the framework should thus also be confirmed by the designers.

5. Although the current study attempted to match the presentation formats used to encode numerical information in reports to the preferences of individuals, there is no guarantee that individuals would be able to complete the task at hand successfully using their preferred presentation format. For this reason, future studies can compare the response times and accuracies when using formats preferred by individuals to those which are recommended by literature for each one of the data tasks.

All of the above mentioned recommendations present opportunities to add to the body of knowledge contained in this thesis.

List of References

Abu Doush, I., Pontelli, E., Simon, D., Son, T.C. and Ma, O. (2009). Making Microsoft Excel^{TM} accessible: multimodal presentation of charts. In: *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pp. 147–154. ACM.

Adelheid, A. and Pexman, P. (2010). *Displaying your findings: A practical guide for creating figures, posters and presentations.* Washington, DC: American Psychological Association.

Aggarwal, C.C. (2013). *Managing and mining sensor data.* Springer Science & Business Media.

Agrawala, M., Li, W. and Berthouzoz, F. (2011). Design principles for visual communication. *Communications of the ACM*, vol. 54, no. 4, pp. 60–69.

Agresti, A. and Kateri, M. (2011). *Categorical data analysis.* Springer.

Al-Hakim, L. (2006). *Challenges of managing information quality in service organizations.* IGI Global.

Albers, M.J. and Mazur, M.B. (2014). *Content and complexity: information design in technical communication.* Routledge.

Allen, B. and Kim, K.-S. (2001). Person and context in information seeking: Interactions between cognitive and task variables. *The new review of information behaviour research*, vol. 2, no. November, pp. 1–16.

Amadi-Echendu, J. and Amadi-Echendu, A. (2015). Legal aspects of engineering asset management. In: *Engineering Asset Management-Systems, Professional Practices and Certification*, pp. 1797–1805. Springer.

Amar, R., Eagan, J. and Stasko, J. (2005). Low-level components of analytic activity in information visualization. In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 111–117. IEEE.

Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A. and Silva, C.T. (2011). A user study of visualization effectiveness using EEG and cognitive load. In: *Computer Graphics Forum*, vol. 30, pp. 791–800. Wiley Online Library.

Annesley, T.M. (2010). Bars and pies make better desserts than figures. *Clinical chemistry*, vol. 56, no. 9, pp. 1394–1400.

Bacim, F., Ragan, E., Scerbo, S., Polys, N.F., Setareh, M. and Jones, B.D. (2013). The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs. In: *Proceedings of Graphics Interface 2013*, pp. 25–32. Canadian Information Processing Society.

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of cognitive neuroscience*, vol. 4, no. 3, pp. 281–288.

Barrett, D.J. (2006). Strong communication skills a must for today's leaders. *Handbook of business strategy*, vol. 7, no. 1, pp. 385–390.

Barry, B. and Fulmer, I.S. (2004). The medium and the message: The adaptive use of communication media in dyadic influence. *Academy of Management Review*, vol. 29, no. 2, pp. 272–292.

Bauer, T. and Erdogan, B. (2012). *An Introduction to Organizational Behavior*. Flat World Knowledge, Inc.

Berlo, D.K. (1960). *The process of communication: An introduction to theory and practice*. New York: Holt, Rinehart and Winston.

Best, L.A., Steward, B. and McGuire, K.P. (2008). External representations and problem solving competence: Do graphs improve problem solving in students? *Readings in Education and Technology: Proceedings of ICICTE 2008*.

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A. and Pfister, H. (2013). What makes a visualization memorable? *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2306–2315.

Bowman, J.P. and Targowski, A.S. (1987). Modeling the communication process: The map is not the territory. *Journal of Business Communication*, vol. 24, no. 4, pp. 21–34.

Braithwaite, D.W. and Goldstone, R.L. (2013). Flexibility in data interpretation: effects of representational format. *Frontiers in psychology*, vol. 4.

Brous, P., Overtoom, I., Herder, P., Versluis, A. and Janssen, M. (2014). Data infrastructures for asset management viewed as complex adaptive systems. *Procedia Computer Science*, vol. 36, pp. 124–130.

BSI (2008a). *PAS 55-1:2008. Asset management – part 1: specification for the optimized management of physical assets*. British Standards Institute, London.

BSI (2014a). *BS ISO 55000:2014. Asset management – Overview, principles and terminology*. British Standards Institute, London.

BSI (2014b). *BS ISO 55001:2014. Asset management – Management systems – Requirements*. British Standards Institute, London.

BSI (2014c). *BS ISO 55002:2014. Asset management – Management systems – Guidelines for the application of ISO 55001.* British Standards Institute, London.

Card, S.K., Mackinlay, J.D. and Shneiderman, B. (1999). *Readings in information visualization: using vision to think.* Morgan Kaufmann.

Chan, S.Y. (2001). The use of graphs as decision aids in relation to information overload and managerial decision quality. *Journal of information science*, vol. 27, no. 6, pp. 417–425.

Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, vol. 8, no. 4, pp. 293–332.

Chen, C.-H., Härdle, W.K. and Unwin, A. (2007). *Handbook of data visualization.* Springer Science & Business Media.

Conati, C., Carenini, G., Hoque, E., Steichen, B. and Toker, D. (2014). Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In: *Computer Graphics Forum*, vol. 33, pp. 371–380. Wiley Online Library.

Ecles, R., Nohria, N. and Berkley, J. (1992). Beyond the hype: Rediscovering the essence of management. *Boston: Harvard Business School Press.[Links].*

Engle, R.W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, vol. 11, no. 1, pp. 19–23.

Falschlunger, L., Lehner, O., Eisl, C. and Losbichler, H. (2015). Development of a data visualization model based on information processing theory. In: *Forschungsforums der österreichischen Fachhochschulen*, pp. 1–7.

Fausset, C.B., Rogers, W.A. and Fisk, A.D. (2008). Visual graph display guidelines. Tech. Rep., Georgia Institute of Technology.

Feldman-Stewart, D., Kocovski, N., McConnell, B.A., Brundage, M.D. and Mackillop, W.J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, vol. 20, no. 2, pp. 228–238.

Few, S. (2007). Save the pies for dessert. *Visual Business Intelligence Newsletter*, pp. 1–14.

Few, S. (2013). Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*

Fischer, M.H., Dewulf, N. and Hill, R.L. (2005). Designing bar graphs: Orientation matters. *Applied Cognitive Psychology*, vol. 19, no. 7, pp. 953–962.

Fishwick, P.A. (2008). *Aesthetic computing.* Mit Press.

Fiske, J. (1990). *Introduction to communication studies.* Routledge.

Flensburg, P. (2009). An enhanced communication model. *The International Journal of Digital Accounting Research*, vol. 9, pp. 31–43.

Franzblau, L.E. and Chung, K.C. (2012). Graphs, tables, and figures in scientific publications: the good, the bad, and how not to be the latter. *The Journal of hand surgery*, vol. 37, no. 3, pp. 591–596.

Gao, J., Lin, S. and Koronios, A. (2006). Data quality in engineering asset management organizations-current picture in Australia. In: *ICIQ*, pp. 467–484.

Gelman, A. (2011). Why tables are really much better than graphs. *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, pp. 3–7.

Gillan, D.J., Wickens, C.D., Hollands, J.G. and Carswell, C.M. (1998). Guidelines for presenting quantitative data in hfes publications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 40, no. 1, pp. 28–41.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, vol. 16, no. 6, pp. 511–525.

Goldberg, J. and Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information visualization*, p. 1473871611406623.

Hager, P.J., Scheiber, H.J. and Corbin, N.C. (1997). *Designing & delivering: Scientific, technical, and managerial presentations.* John Wiley & Sons.

Hansen, C.D., Chen, M., Johnson, C.R., Kaufman, A.E. and Hagen, H. (2014). *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization.* Springer.

Harger, J.R. and Crossno, P.J. (2010). Comparison of open-source visual analytics toolkits. In: *IS&T/SPIE Electronic Imaging*, pp. 82940E–82940E. International Society for Optics and Photonics.

Harper, S., Michailidou, E. and Stevens, R. (2009). Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Transactions on Applied Perception (TAP)*, vol. 6, no. 2, p. 10.

Harris, R.L. (1999). *Information graphics: A comprehensive illustrated reference.* Oxford University Press.

Hart, C. (2001). *Doing a literature search: a comprehensive guide for the social sciences.* Sage.

Hartley, J., Cabanac, G., Kozak, M. and Hubert, G. (2014). Research on tables and graphs in academic articles: Pitfalls and promises. *Journal of the Association for Information Science and Technology.*

Hastings, N.A.J. (2009). *Physical asset management*. Springer Science & Business Media.

Hastings, N.A.J. (2015). *Physical Asset Management: With an Introduction to ISO55000*. Springer.

Hawley, S.T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T. and Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient education and counseling*, vol. 73, no. 3, pp. 448–455.

Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. *Information research*, vol. 9, no. 1, pp. 9–1.

Heylighen, F. (1999). The growth of structural and functional complexity during evolution. *The evolution of complexity*, pp. 17–44.

Hollands, J. and Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology*, vol. 12, no. 2, pp. 173–190.

Huang, S.Y., Chung, Y.-H. and Cheng, W.-W. (2008). The effects between numerical tabulations and graphs of financial information on the judgment of investors. *Review of Business Information Systems (RBIS)*, vol. 12, no. 4.

Huang, W., Eades, P. and Hong, S.-H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, vol. 8, no. 3, pp. 139–152.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, vol. 52, no. 12, pp. 5186–5201.

IAM (2014). Asset management - an anatomy. Tech. Rep., The Instutute of Asset Management.

International Accounting Standards Board (2014). *International Financial Reporting Standards: A Guide Through IFRS*. International Accounting Standards Board, London.

Johnson, C.N. (2002). The benefits of pdca. *Quality Progress*, vol. 35, no. 5, p. 120.

Jongen, R., Smit, J. and Janssen, A. (2008). Application of statistical analysis in the asset management decision process. In: *Condition Monitoring and Diagnosis, 2008. CMD 2008. International Conference on*, pp. 1284–1287. IEEE.

Jongsawat, N. and Premchaiswadi, W. (2011). A study of two different experimental settings for group awareness information in a web-based group decision support system. *International Journal of Information Technology & Decision Making*, vol. 10, no. 02, pp. 231–268.

Joshi, N.H., Venkatesh, C.S. and Joshi, H. (2014). Methods of data presentation in scientific communications. *Journal of Chronotherapy and Drug Delivery*, vol. 5, pp. S1–S11.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, vol. 23, no. 1, pp. 1–19.

Kashihara, A., Oppermann, R., Rashev, R., Simm, H. *et al.* (2000). A cognitive load reduction approach to exploratory learning and its application to an interactive simulation-based learning system. *Journal of Educational Multimedia and Hypermedia*, vol. 9, no. 3, pp. 253–276.

Kelton, A.S., Pennington, R.R. and Tuttle, B.M. (2010). The effects of information presentation format on judgment and decision making: A review of the information systems research. *Journal of Information Systems*, vol. 24, no. 2, pp. 79–105.

Kerick, S. and Allender, L. (2004). Effects of cognitive workload on decision accuracy, shooting performance, and cortical activity of soldiers. Tech. Rep., Defense Technical Information Center.

Koronios, A. and Lin, S. (2006). Information quality in engineering asset management. *Information Quality Management: Theory and Applications: Theory and Applications*, p. 221.

Koronios, A., Lin, S. and Gao, J. (2005). A data quality model for asset management in engineering organisations. In: *Proceedings of the 10th International Conference on Information Quality*. Massachusetts Institute of Technology, Cambridge, USA.

Koschat, M.A. (2005). A case for simple tables. *The American Statistician*, vol. 59, no. 1.

Kozak, M. and Hartley, J. (2013). Publication fees for open access journals: Different disciplines – different methods. *Journal of the American Society for Information Science and Technology*, vol. 64, no. 12, pp. 2591–2594.

Kozak, M., Hartley, J., Wnuk, A. and Tartanus, M. (2015). Multiple pie charts: Unreadable, inefficient, and over-used. *Journal of Scholarly Publishing*, vol. 46, no. 3, pp. 282–289.

Kress, G. (2004). Reading images: Multimodality, representation and new media. *Information Design Journal*, vol. 12, no. 2, pp. 110–119.

Kreye, M.E., Goh, Y.M., Newnes, L.B. and Goodwin, P. (2012). Approaches to displaying information to assist decisions under uncertainty. *Omega*, vol. 40, no. 6, pp. 682–692.

Kumar, S. and Phrommathed, P. (2005). *Research methodology*. Springer.

Law, A.S., Freer, Y., Hunter, J., Logie, R.H., Mcintosh, N. and Quinn, J. (2005). A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. *Journal of clinical monitoring and computing*, vol. 19, no. 3, pp. 183–194.

Lee, B., Plaisant, C., Parr, C.S., Fekete, J.-D. and Henry, N. (2006). Task taxonomy for graph visualization. In: *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–5. ACM.

Lee, D. (1993). Developing effective communications. *University of Missouri Extension, CM109 http://extension.missouri.edu/publications/DisplayPrinterFriendlyPub.aspx*.

Lehner, O., Falschlunger, L., Losbichler, H. and Eisl, C. (2015). Development of a data visualization model based on information processing theory.

LeRoy, H. and Snitkin, S. (2004). Asset information management – a calm prerequisite. Tech. Rep., ARC Adisory Group.

Lin, S., Gao, J. and Koronios, A. (2006). The need for a data quality framework in asset management. In: *Australian Workshop on Information Quality*, vol. 1. Citeseer.

Louit, D.M., Pascual, R. and Jardine, A.K. (2009). A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. *Reliability Engineering & System Safety*, vol. 94, no. 10, pp. 1618–1628.

Massart, D.L., Smeyers-Verbeke, A.J., Capron, X. and Schlesier, K. (2005). Visual presentation of data by means of box plots. *LC-GC Europe*, vol. 18, no. 4, pp. 215–218.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAfee, A. and Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, vol. 90, no. 90, pp. 60–6.

Minnaar, J.R., Basson, W. and Vlok, P.J. (2013). Quantitative methods required for implementing PAS 55 or the ISO 55000 series for asset management. *South African Journal of Industrial Engineering*, vol. 24, no. 3, pp. 98–111.

Mintzberg, H. (1973). The nature of managerial work.

Mostyn, G.R. (2011). Cognitive load theory: What it is, why it's important for accounting instruction and research. *Issues in Accounting Education*, vol. 27, no. 1, pp. 227–245.

Naisbitt, J. (1982). *Megatrends: ten new directions transforming our lives*. Warner Books.

Neely, M.P., Lin, S., Gao, J. and Koronios, A. (2006). The deficiencies of current data quality tools in the realm of engineering asset management. Proceedings of the Twelfth Americas Conference in Information Systems.

Nistal, A.A., Van Dooren, W., Clarebout, G., Elen, J. and Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: a critical review. *ZDM*, vol. 41, no. 5, pp. 627–636.

Nowell, L., Schulman, R. and Hix, D. (2002). Graphical encoding for information visualization: an empirical study. In: *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pp. 43–50. IEEE.

Oakland, J.S. (2008). *Statistical process control.* Routledge.

Oblinger, D. (2003). Boomers gen-xers millennials. *EDUCAUSE review*, vol. 500, no. 4, pp. 37–47.

Okan, Y., Garcia-Retamero, R., Galesic, M. and Cokely, E.T. (2012). When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition & Computation*, vol. 12, no. 2-3, pp. 195–218.

Ouertani, M.-Z., Parlikad, A.K. and McFarlane, D.C. (2008). Towards an approach to select an asset information management strategy. *IJCSA*, vol. 5, no. 3b, pp. 25–44.

Paas, F., Renkl, A. and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, vol. 38, no. 1, pp. 1–4.

Paas, F.G. and Van Merriënboer, J.J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review*, vol. 6, no. 4, pp. 351–371.

Pantelias, A., Flintsch, G.W., Bryant, J.W. and Chen, C. (2008). Asset management data practices for supporting project selection decisions. *Public Works Management & Policy*.

Peebles, D. and Ali, N. (2009). Differences in comprehensibility between three-variable bar and line graphs. In: *Proceedings of the thirty-first annual conference of the cognitive science society*, pp. 2938–2943.

Penrose, H.W. (2008). *Physical Asset Management for the Executive: Don't Read This If You Are on an Airplane.* Success by Design.

Petkosek, M. and Moroney, W. (2004). Guidelines for constructing graphs. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, pp. 1006–1010. SAGE Publications.

Pettersson, R. (1989). *Visuals for Information Research and Practice.* Educational Technology Publications, Englewood Cliffs, NJ.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, vol. 5, no. 2, pp. 27–45.

Phares, E.J. (1991). Introduction to psychology. *New York: Harper Collins Publishers*, vol. 4.

Platts, K. and Hua Tan, K. (2004). Strategy visualisation: knowing, understanding, and formulating. *Management Decision*, vol. 42, no. 5, pp. 667–676.

Porat, T., Oron-Gilad, T. and Meyer, J. (2009). Task-dependent processing of tables and graphs. *Behaviour & Information Technology*, vol. 28, no. 3, pp. 293–307.

Potter, K., Hagen, H., Kerren, A. and Dannenmann, P. (2006). Methods for presenting statistical information: The box plot. *Visualization of Large and Unstructured Data Sets*, vol. S-4, pp. 97–106.

Prasad, G.V. and Ojha, A. (2012). Text, table and graph–which is faster and more accurate to understand? In: *Technology for Education (T4E), 2012 IEEE Fourth International Conference on*, pp. 126–131. IEEE.

Qing-Ling, D., Shu-Min, C., Lian-Liang, B. and Jun-Mo, C. (2008). Application of pdca cycle in the performance management system. In: *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on*, pp. 1–4. IEEE.

Quirke, B. (2007). *Making the connections: using internal communication to turn strategy into action*. Gower Publishing, Ltd.

Ramakrishnan, R. and Gehrke, J. (2000). *Database management systems*. McGraw Hill.

Ratnayake, R.C. (2013). Sustainable performance of industrial assets: the role of pas 55-1&2 and human factors. *International Journal of Sustainable Engineering*, vol. 6, no. 3, pp. 198–211.

Robertson, G., Fernandez, R., Fisher, D., Lee, B. and Stasko, J. (2008). Effectiveness of animation in trend visualization. *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1325–1332.

Saary, M.J. (2008). Radar plots: a useful way for presenting multivariate health care data. *Journal of clinical epidemiology*, vol. 61, no. 4, pp. 311–317.

Sant'Anna, A.M.O. and Douglas, A. (2015). Framework of decision in data modeling for quality improvement. *The TQM Journal*, vol. 27, no. 1.

Schonlau, M. and Peters, E. (2008). Graph comprehension: An experiment in displaying data as bar charts, pie charts and tables with and without the gratuitous 3rd dimension. *Social Science Research Network Working Paper Series*.

Schonlau, M. and Peters, E. (2012). Comprehension of graphs and tables depend on the task: empirical evidence from two web-based studies. *Statistics, Politics, and Policy*, vol. 3, no. 2.

Schramm, W. (1954). Procedures and effects of mass communication. *Mass Media and Education*, pp. 113–38.

Scott, D.W. (2009). Sturges' rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 3, pp. 303–306.

Seufert, T., Jänen, I. and Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Computers in Human Behavior*, vol. 23, no. 3, pp. 1055–1071.

Shaft, T.M. and Vessey, I. (2006). The role of cognitive fit in the relationship between software comprehension and modification. *MIS Quarterly*, pp. 29–55.

Shannon, C. and Weaver, W. (1963). The mathematical theory of communication-urbana. simpson eh 1949. *Measurement of diversity–Nature*, vol. 163, p. 688.

Shannon, C.E. and Weaver, W. (1949). The mathematical theory of information.

Shmueli, G., Patel, N.R. and Bruce, P.C. (2011). *Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with xlMiner*. John Wiley and Sons.

Siirtola, H. (2014). Bars, pies, doughnuts & tables: Visualization of proportions. In: *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014-Sand, Sea and Sky-Holiday HCI*, pp. 240–245. BCS.

Simon, H.A. (1978). On the forms of mental representation. *Perception and cognition: Issues in the foundations of psychology*, vol. 9, pp. 3–18.

Solomon, P. (2002). Discovering information in context. *Annual review of information science and technology*, vol. 36, pp. 229–264.

Spatz, C. (2007). *Basic statistics: Tales of distributions*. Cengage Learning.

Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies*, vol. 64, no. 11, pp. 1115–1131.

Speier, C. and Morris, M.G. (2003). The influence of query interface design on decision-making performance. *MIS quarterly*, pp. 397–423.

Spence, I. and Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, vol. 5, no. 1, pp. 61–77.

Steele, J. and Iliinsky, N. (2011). *Designing Data Visualizations: Representing Informational Relationships*. O'Reilly Media Inc.

Steinberg, W.J. (2010). *Statistics alive!* Sage Publications.

Stewart, B.M., Cipolla, J.M. and Best, L.A. (2009). Extraneous information and graph comprehension: Implications for effective design choices. *Campus-Wide Information Systems*, vol. 26, no. 3, pp. 191–200.

Swanson, M. and Curry, S. (1989). Results of an asset engineering program. *Information & Management*, vol. 16, no. 4, pp. 207–216.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, vol. 12, no. 2, pp. 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, vol. 22, no. 2, pp. 123–138.

Sweller, J., Van Merriënboer, J.J. and Paas, F.G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, vol. 10, no. 3, pp. 251–296.

Tang, V. (2006). *Corporate decision analysis: an engineering approach*. Ph.D. thesis, Massachusetts Institute of Technology.

Toker, D., Conati, C., Carenini, G. and Haraty, M. (2012). Towards adaptive information visualization: on the influence of user characteristics. In: *User Modeling, Adaptation, and Personalization*, pp. 274–285. Springer.

Tory, M. and Moller, T. (2004). Human factors in visualization research. *Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 1, pp. 72–84.

Tuovinen, J.E. and Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of educational psychology*, vol. 91, no. 2, p. 334.

Van den Honert, A., Schoeman, J. and Vlok, P. (2013). Correlating the content and context of PAS 55 with the ISO 55000 series. *South African Journal of Industrial Engineering*, vol. 24, no. 2, pp. 24–32.

Van Der Meulen, M., Logie, R.H., Freer, Y., Sykes, C., McIntosh, N. and Hunter, J. (2010). When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. *Applied Cognitive Psychology*, vol. 24, no. 1, pp. 77–89.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, vol. 22, no. 2, pp. 219–240.

Wang, Y.-H. and Kuo, T.-H. (2010). A financial assets and liabilities management support system. *Contemporary Management Research*, vol. 6, no. 4.

Wang Baldonado, M.Q., Woodruff, A. and Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. In: *Proceedings of the working conference on Advanced visual interfaces*, pp. 110–119. ACM.

Welman, C., Kruger, F. and Mitchell, B. (2005). *Research methodology*. Oxford University Press.

Wilson, T.D. (2000). Human information behavior. *Informing science*, vol. 3, no. 2, pp. 49–56.

Woodhouse, J. (2014). Briefing: Standards in asset management: PAS 55 to ISO 55000. *Infrastructure Asset Management*, vol. 1, no. 3, pp. 57–59.

Yang, H., Li, Y. and Zhou, M.X. (2014). Understand users' comprehension and preferences for composing information visualizations. *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 21, no. 1, p. 6.

Zaccaro, S.J., Rittman, A.L. and Marks, M.A. (2002). Team leadership. *The Leadership Quarterly*, vol. 12, no. 4, pp. 451–483.

Zhu, B. and Chen, H. (2008). Information visualization for decision support. In: *Handbook on Decision Support Systems 2*, pp. 699–722. Springer.

Appendices

Appendix A

Variations on Bar and Column Graphs

The graphs depicted in this appendix were also created using the information in Table 2.7. They were created to serve as summaries of the four graphs which appear in the generic framework, but which are not discussed explicitly in Section 2.7.

A.1 Bar Graph Variations

There are two variations of simple bar graphs which can be used to display more than one dataset or data series on the same graph, namely: grouped bar graphs and stacked bar graphs. Each variation is displayed and briefly described next.

A.1.1 Grouped Bar Graphs

Also referred to as: *clustered, side-by-side or multiple bar graphs.*

Figure A.1 shows an example of a grouped bar graph. Data series are plotted side-by-side on the same bar graph. Each data element of each data series is plotted as an individual bar (Harris, 1999). Bars of the same data series are filled with the same colour, shade, or pattern which is generally described in a legend. The bars that correspond to the first independent variable on the vertical axis from each data series are placed side-by-side to form a cluster. Similarly, the bars that correspond to the second independent variable on the vertical axis are also placed in a cluster. This patter continues until all of the data elements appear on the grouped bar graph.

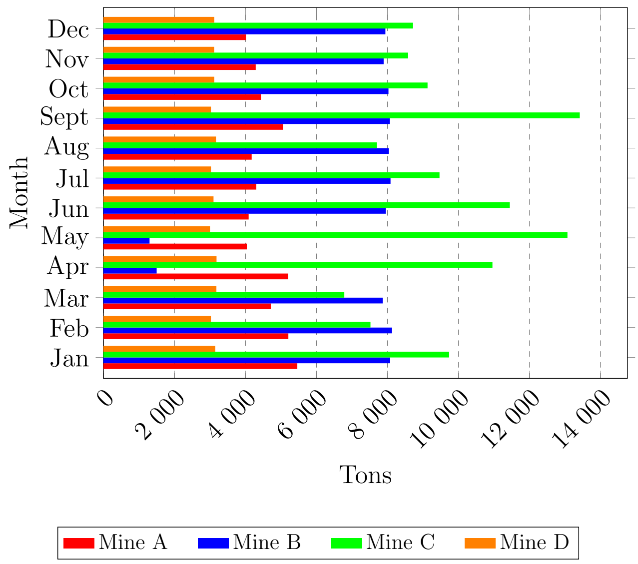


Figure A.1: An example of a grouped bar graph.

A.1.2 Stacked Bar Graphs

Also referred to as: *segmented, extended, divided, composite or subdivided bar graphs.*

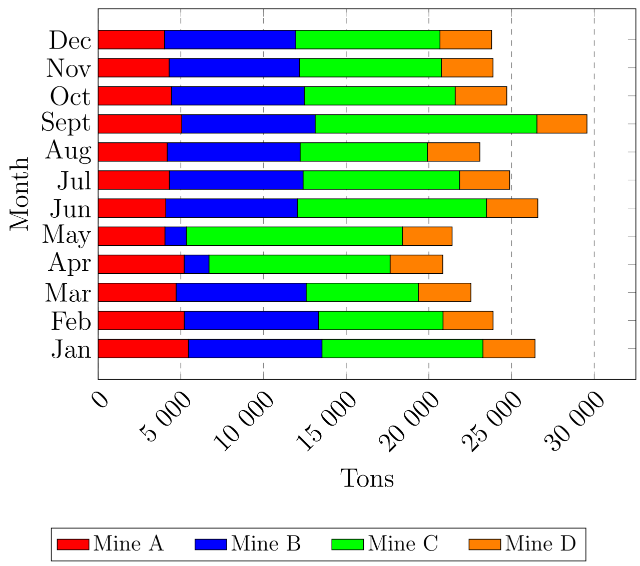


Figure A.2: An example of a stacked bar graph.

Figure A.2 shows an example of a stacked bar graph. Harris (1999) explains that stacked bar graphs are also used to present multiple data series, but that the series are stacked en-to-end rather than side-by-side as with grouped bar graphs. The ends on the far right of a group of stacked bars represent the sums of all the bars for a specific independent variable. Furthermore, the horizontal scale is always linear, quantitative and originates at zero. As with grouped bar graphs, elements of the same data series are filled with the same colour, shade, or pattern which is once again generally described in a legend.

A.2 Column Graph Variations

Since bar graphs and column graphs are closely related, the same variations of simple bar graphs are applicable to simple column graphs. Consequently, the two variations of simple column graphs which can be used to display more than one dataset or data series on the same graph are: grouped column graphs and stacked column graphs. Each variation is displayed and briefly described next.

A.2.1 Grouped Column Graphs

Also referred to as: *clustered, side-by-side or multiple column graphs.*

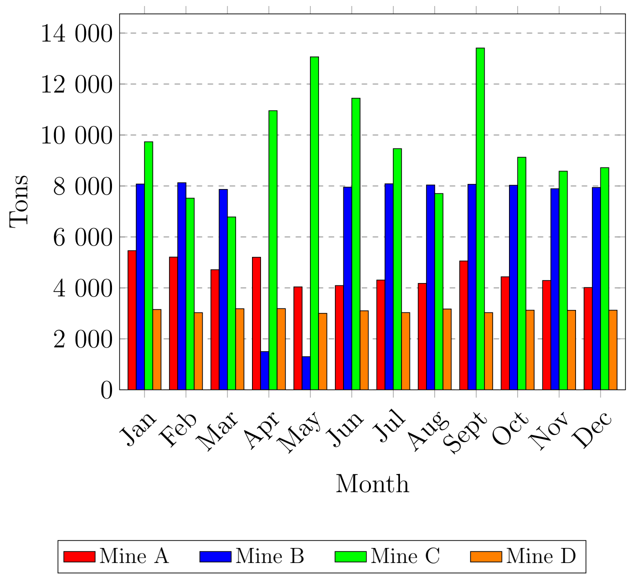


Figure A.3: An example of a grouped column graph.

Figure A.3 shows an example of a grouped column graph. The elements in the grouped column graph are identical to those in the grouped bar graph, with the only difference being the orientation of the axes and bars.

A.2.2 Stacked Column Graphs

Also referred to as: *segmented, extended, divided, composite or subdivided column graphs.*

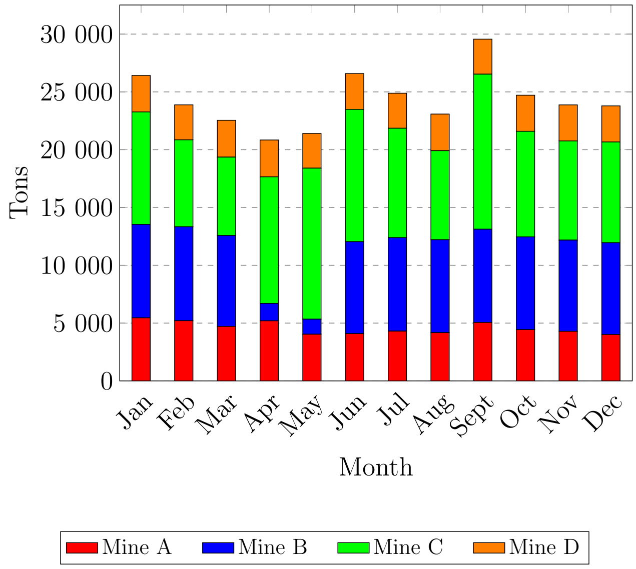


Figure A.4: An example of a stacked column graph.

Figure A.4 shows an example of a stacked column graph. The elements in the stacked column graph are once again identical to those in the stacked bar graph, with the only difference being the orientation of the axes and bars.

Appendix B

Questionnaire to Collect Individual Information Presentation Preferences

The next two pages show extracts of the questionnaire used in the validation. It illustrates how preferences for task code TC2 were recorded during the validation.

Task 2: Compare distribution

Data Type: Interval/Continuous *Number of Datasets: Multiple*

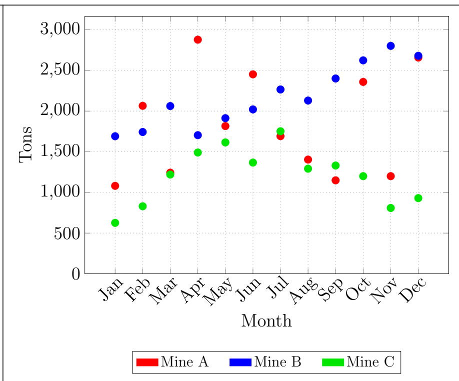
Description: Characterising a distribution typically involves determining the variation and the range within a dataset. With multiple datasets, the ranges and variations of the datasets need to be compared.

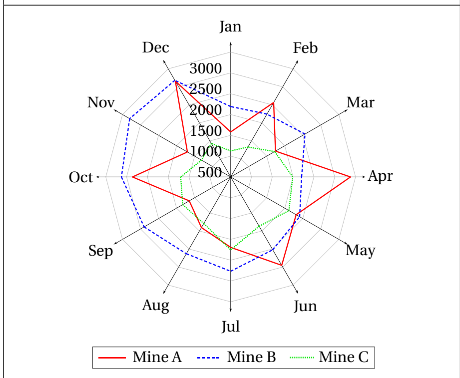
User Requirements: Determine and compare the ranges of the distributions.

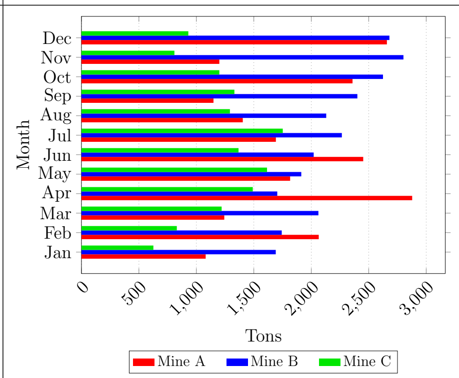
Practical Example: The graphs show the tons of ore processed by three mines *(Mine A, Mine B and Mine C)* in each month of 2014.

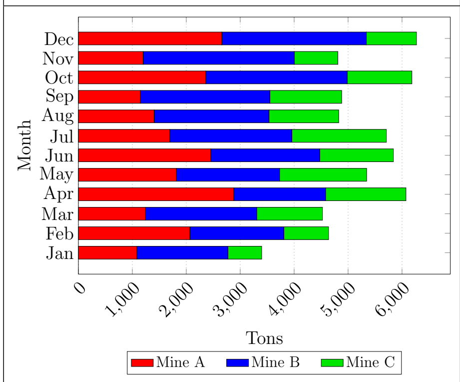
Example Task: Which mine had the smallest variation in output in 2014?

Month	Mine A	Mine B	Mine C
Jan	1 081	1 691	626
Feb	2 065	1 743	830
Mar	1 243	2 062	1 220
Apr	2 878	1 704	1 491
May	1 816	1 913	1 615
Jun	2 452	2 021	1 367
Jul	1 692	2 266	1 752
Aug	1 404	2 130	1 292
Sep	1 149	2 401	1 331
Oct	2 359	2 624	1 200
Nov	1 200	2 802	809
Dec	2 658	2 680	930









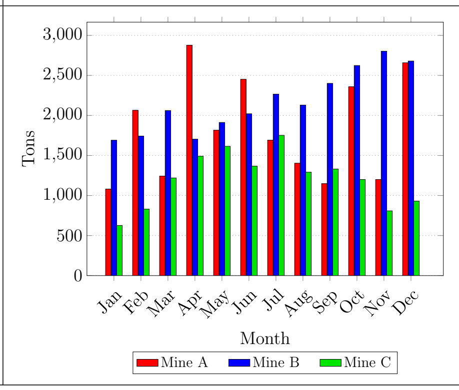


Figure B.1: First page used in determining the user preferences of task TC2.

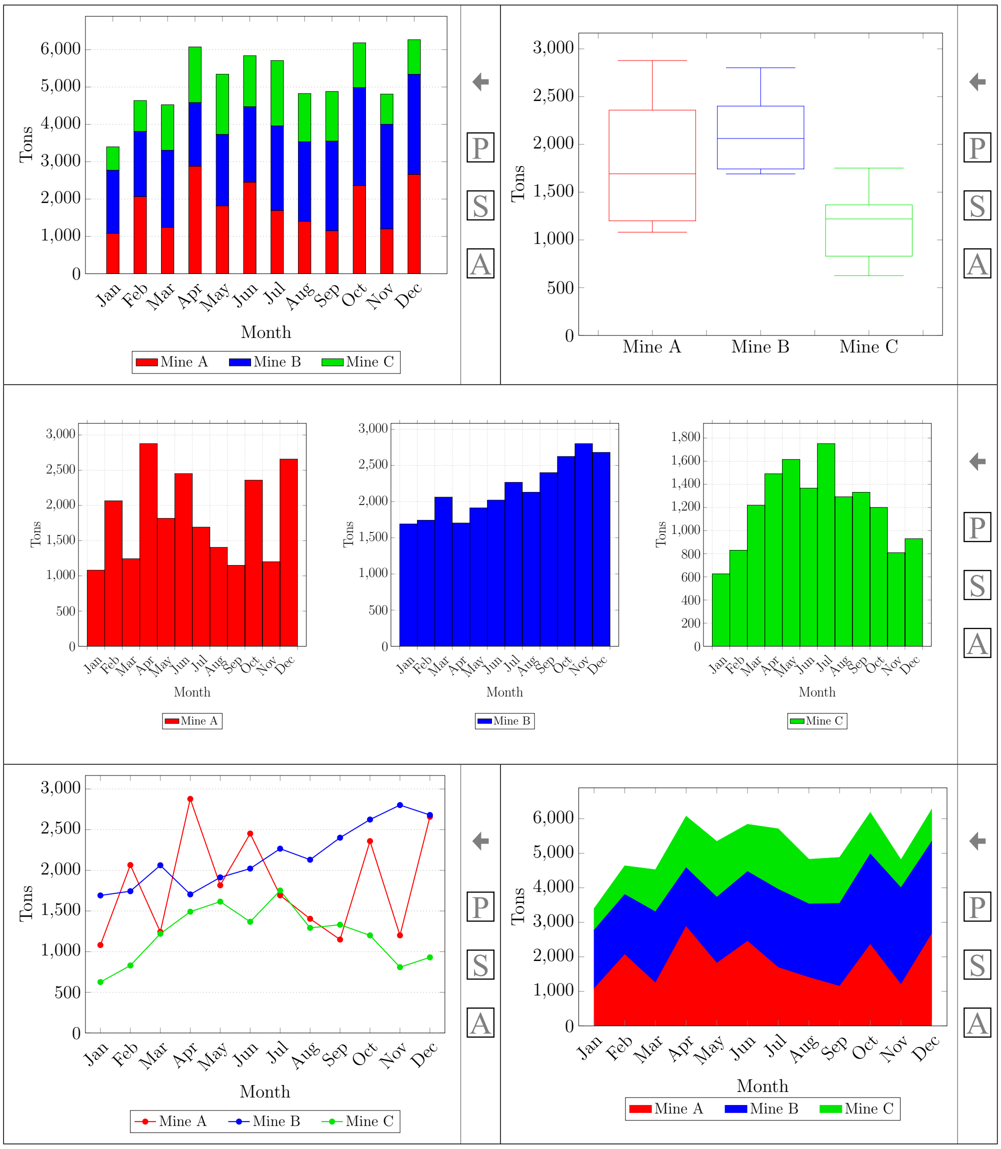


Figure B.2: Second page used in determining the user preferences of task code TC2.

Appendix C

Feedback on Proposed Framework

The following two pages will be used to validate the framework. Please indicate your answer to each question by placing a tick in the appropriate box (e.g. ☑). If a question is not applicable to your department directly, please consider your organisation in general.

Data communication problems in organisations		
Is your organisation data-intensive (makes use of a lot of data)?	☐ YES	☐ NO
Do you expect the amount of data in your organisation to increase in the future?	☐ YES	☐ NO
In your opinion, is factual data available for most decisions that need to be made in your organisation?	☐ YES	☐ NO
Are decisions in your organisation generally based on and supported by factual data?	☐ YES	☐ NO
Are some decisions made in your organisation based on opinions rather than factual data, even though there is data that can be processed and used to support the decisions?	☐ YES	☐ NO
Do you think data will be used more frequently if it is packaged in a more usable format?	☐ YES	☐ NO

Clear benefit for the users		
Are you satisfied with the way in which data is current being presented in your organisation?	☐ YES	☐ NO
Do you think the current presentation formats used in your organisation match your personal preferences?	☐ YES	☐ NO
Would reports that are tailored to your personal information presentation preferences be of value to you?	☐ YES	☐ NO

The viability of the framework		
Do you think the proposed framework can improve data communication in your organisation?	☐ YES	☐ NO
Do you think your co-workers would be prepared to fill in the questionnaire if they knew that the way information is communicated will be tailored to their personal preferences if they do?	☐ YES	☐ NO
Do you think that your organisation could benefit from such a framework?	☐ YES	☐ NO

Creating employee clusters		
Do you think that different levels of employees are exposed to different data?	☐ YES	☐ NO
Do you think that different departments (of employees) are exposed different data?	☐ YES	☐ NO
Do you think that different levels of employees have different information presentation preferences?	☐ YES	☐ NO
Do you think that different departments (of employees) have different information presentation preferences?	☐ YES	☐ NO

Determining data presentation capabilities		
Do you think your organisation is capable of producing all of the presentation formats in the questionnaire?	☐ YES	☐ NO
Are there any information presentation formats used by your organisation to present similar information/data which do not appear in the questionnaire?	☐ YES	☐ NO

If yes, please name: ______________________________

____

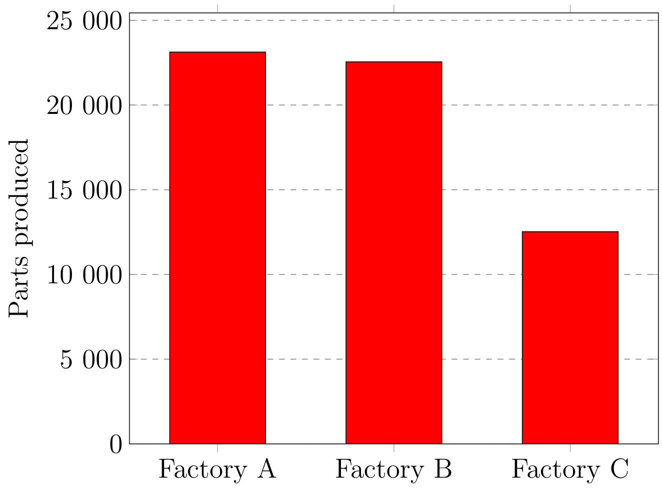
____

Finding domain specific examples of tasks		
Do you think that a person's understanding of the tasks in the questionnaire will improve if the examples provided are from his or her domain (e.g. mining examples, textile examples, pharmaceutical examples, etc.)?	☐ YES	☐ NO

Appendix D

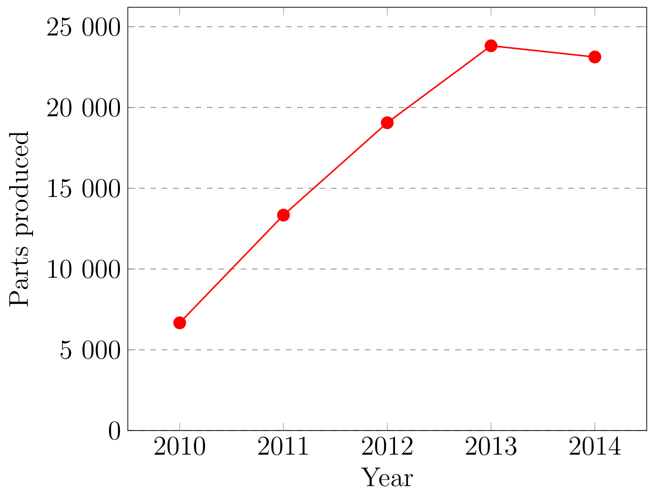
Worked Example Output

TOTAL OUTPUT PER FACTORY			
No.:	1	**Data Type:**	Nominal
Period:	2014	**Task Name:**	Compare Discrete Value
Source:	Multiple	**Task Code:**	TC3
Datasets:	Single	**Presentation Format:**	Column Graph (Simple)



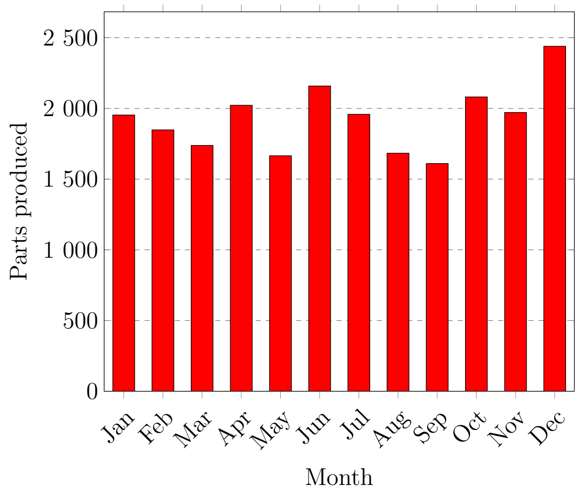
Factory A five-year output

No.:	2	**Data Type:**	Interval
Period:	2010-2014	**Task Name:**	Identify Trend
Source:	Factory A	**Task Code:**	TC15
Datasets:	Single	**Presentation Format:**	Line Graph



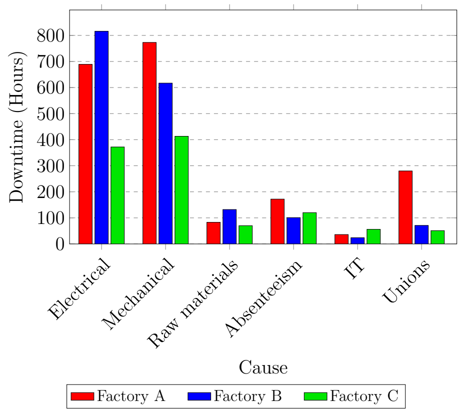
Monthly output Factory A

No.:	3	**Data Type:**	Interval
Period:	2014	**Task Name:**	Characterise Distribution
Source:	Factory A	**Task Code:**	TC1
Datasets:	Single	**Presentation Format:**	Column Graph (Simple)



Sources of downtime per factory

No.:	4	**Data Type:**	Nominal
Period:	2014	**Task Name:**	Find Extreme Values
Source:	Factory A	**Task Code:**	TC12
Datasets:	Multiple	**Presentation Format:**	Column Graph (Grouped) <u>or</u> Table



<u>or</u>

Cause	Factory A	Factory B	Factory C
Electrical	689	816	372
Mechanical	773	617	413
Raw materials	83	132	70
Absenteeism	172	101	120
IT	36	24	56
Unions	280	71	51

Critical HS&E incidents per factory

No.:	5	**Data Type:**	Interval
Period:	2010-2014	**Task Name:**	Identify Trend
Source:	Factory A	**Task Code:**	TC16
Datasets:	Multiple	**Presentation Format:**	Line Graph