Improving data quality and management in one South African SME for data analytics

PP Mabotja

orcid.org/0000-0001-7923-2282

Dissertation accepted in fulfilment of the requirements for the degree *Master of Science in Computer Science* at the North-West University

Supervisor: Dr S van der Linde

Co-supervisor: Dr S Gilliland

Graduation: June 2022

Student number: 30414954

DECLARATION

I, Phakisa Pertunia Mabotja, hereby declare that the dissertation entitled

Improving data quality and management in one South African SME for data analytics

is my own work and was prepared by me at North-West University under the guidance of Dr Suné van der Linde and Dr Sonja Gilliland for the award of MSc degree in Computer Science. This dissertation has not been previously submitted to this or any other institution for a degree or other qualifications. I declare that this dissertation represents my ideas in my own words and where others’ work has been included or quoted, I have adequately referenced and acknowledged the original sources. I have adhered to the university research ethics guidelines, obtained the appropriate ethical clearance, identified all risks that may arise in conducting the study, and acknowledged my obligations and rights to participants.

DEDICATION

I would like to dedicate this paper to my little baby boy, my angel Botshelo Mabotja. I wasn’t always there when you needed me, to watch your favourite cartoons with you and to play with you as much as I wanted to. Thank you for those cute hugs and kisses, they meant a lot to me especially when I felt overwhelmed. I love you so much.

ACKNOWLEDGEMENTS

- Above all, I would like to thank God Almighty for being my source of strength and guidance throughout the study. I will forever be thankful for your enduring love, the wisdom and power you have given me. I have accomplished this milestone through your love, your blessings and faithfulness.
- I would like to express my deepest gratitude to my supervisors who have supported and assisted me throughout this journey. Dr Sonja Gilliland and Dr Suné van der Linde, your immense knowledge and experience have helped me to learn a lot about research work and helped me to grow as an individual. The completion of this study could not have been possible without your knowledge and expertise. Thank you for your unwavering support and guidance.
- Additionally, I would like to thank Dr J.T. Janse van Rensburg for your contributions in this study and to everyone at NWU who was involved in making this research study a success.
- To the participants and everyone at one SA SME who have been part of this study, thank you for your time, your support and for sharing your knowledge and experience with me. Your contributions made this research study a success.
- To my brother, Collen Mawela, thank you for always being there for me when I needed you the most. I'm deeply grateful for your guidance, contributions, constructive criticism, practical suggestions and insightful comments throughout this journey. "Ngiyabonga bhuti".
- To my sister and my friend, Zanele Sithole, your words of encouragement, valuable advises and support have made it possible for me to be where I am today and celebrate this achievement with you. Thank you for believing in my abilities.
- I am privileged to have my loving family who were there for me throughout this study. To my loving parents, Noko Mabotja and Joel Mabotja, your love and support have carried me throughout this journey. To my siblings Kgabo, Makolobe and Mmanoko, thank you for listening when I needed someone to talk to, for motivating and supporting me till the end. *"ke a leboga Bakone!".*
- This has been an incredible and challenging journey that I'm proud that I had taken and I have accomplished. It wasn't easy, I will forever be very grateful that my hard work, persistence, dedication and perseverance have held me up to this stage to achieve this milestone.

ABSTRACT

In today's competitive and complex environments, organisations are collecting enormous amounts of data that require intelligent systems and technologies for storage, management, integration, processing, analysing and usage to assist in the making of informed business decisions. In this era of technological advancement, data are regarded as a critical organisational asset and investment to assist organisations with innovation, support of organisational activities, measurement of business performance, analysis of customer behaviour and most importantly, decision-making. However, if the organisational data is of poor quality, is unreliable, and from untrusted sources, it becomes a challenge to extract valuable information from such data. Although large enterprises have resources to manage large volumes of data, most small and medium-sized enterprises (SMEs) are struggling to use their data for performance measurement and decision-making processes. Poor data quality and not having proper tools and technologies to manage their enormous volumes of data prevent them from gaining competitive advantage over rivalries. This study followed a design research methodology where an artefact was designed to assist one South African SME with capturing, storing and management of data. The study further proposes guidelines and technologies that one SA SME and other SMEs can implement to better manage data and improve data quality.

Keywords:

Data quality, Data management, Small and Medium Sized Enterprises (SMEs), Data warehouse, Data analytics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1.1 Introduction

The information technology (IT) landscape is constantly changing. With every advancement in technology comes a shift in the way organisations do business and small and medium-sized enterprises (SMEs) have been urged to develop data-driven business models that can assist them to innovate, keep pace with changing technologies and trends, and be competitive with their larger counterparts (Bianchini & Michalkova, 2019:7; Kalan & Unalir, 2016:2; Kfouri & Skyrius, 2017:97; Parapi & Masykuri, 2020:6; Popovič *et al.*, 2019:10; SAS, 2020; Van der Krogt *et al.*, 2020:2). Therefore, SMEs are constantly faced with the challenge to discover new and innovative ways to enhance and adapt to the rapid business and IT transformations (Bianchini & Michalkova, 2019:7; Curraj, 2018:3; Van der Krogt *et al.*, 2020:2).

Today, local and global organisations are collecting extensive amounts of data that require intelligent systems and technologies for storage, management, integration, processing, analysing and usage to assist in the making of informed business decisions. In fact, "*businesses are becoming more data intensive*" (Liu *et al.*, 2018:840). Modern technologies such as artificial intelligence (AI), cloud computing, big data and social networking contribute to excessive accumulation of data (Baharuden *et al.*, 2019:1; Bianchini & Michalkova, 2019:7; Williams & Tang, 2020:29). In the year 2020, 2,5 quintillion bytes of data are created by humans every day (Bulao, 2020; IBM, 2020) and it is estimated that 463 exabytes of data will be generated daily by humans as of year 2025 (Bulao, 2020; Vuleta, 2020).

For years, large organisations have acknowledged the importance of their existing data and started investing in systems that can assist them in managing and analysing their data (Kfouri & Skyrius, 2017:97). Due to digital transformations and collection of large volumes of data, organisations are implementing technologies such as databases, data warehousing and business intelligence systems (DW/BI), data mining and database management systems such as Microsoft SQL Server, SAP, IBM DB2, MySQL, and Oracle to manage data. SMEs are not only regarded as the core pillars of any national economy but also play a role in the social development of countries (Assarlind, 2011:1; Olszak & Ziemba, 2012:138; Scholz *et al.*, 2010:2; Van der Krogt *et al.*, 2020:3). Therefore, data analytics have become important for SMEs as it enables extraction of useful insight from raw data (Bianchini & Michalkova, 2019:9; Dominguez *et al.*, 2019:1).

Runkler (2012:2) defines data analytics as the application of computer systems to the analysis of vast data sets for the support of decision-making purposes. Data analytics can be referred to as a "*process that involves the use of statistical techniques (measures of central tendency, graphs* *and so on), information system software (data mining, sorting routine) and operations research methodologies (linear programming) to explore, visualise, discover and communicate patterns and trends in data*" (Schniederjans *et al.*, 2015:4). The main purpose of data analysis is to optimise processes and enhance the competitive position of the company (Runkler, 2012:1).

The present study sought to first explore the literature to determine how SMEs are storing and managing large volumes of data in this information age. The study uses the literature to report on various aspects that contribute to poor data quality in SMEs and discuss the impacts of poor data quality in SMEs. It further analyses the literature to determine the challenges of poor data quality and benefits of high data quality. Different BI models, data warehousing and ETL concepts suitable for SMEs are also explored.

Second, the proposed empirical research uses a case study at a steel manufacturing company in South Africa where an artefact was designed to assist with capturing of data into a database in one South African (SA) SME. The study further proposes guidelines and technologies that can be implemented to assist with data analytics and decision-making processes. It should be noted that, even though this study is primarily focused on the steel roll manufacturing company, the findings and conclusions might have relevance and add value to other SMEs.

The next section will cover the background and motivation (1.2) followed by the problem statement in section 1.3. The research aim and objectives are presented in sections 1.4.1 and 1.4.2, followed by the purpose of the research (section 1.4.3) and research questions in section 1.4.4. Assumptions and limitations of the study are presented in section 1.5.7. The significance of the study (1.6) is then presented, along with the dissertation layout (1.7).

1.2 Background to study

Data management theory has been in existence for decades and it is regarded as one of the most critical and innovative areas of computer science (Grillenberger & Romeike, 2017:2). Over the years, researchers have explored the concepts of data management and data quality and its impact on decision-making processes (Bianchini & Michalkova, 2019:7; Gesmann-Nuissl & Kirchner, 2018:2; Gray, 1996:1; Grillenberger & Romeike, 2017:2; Inmon, 2002:179; Jesiļevska, 2017:89; Kademeteme *et al.*, 2017:31; Liu *et al.*, 2020:2; Tayi & Ballou, 1998:54; Timmerman & Bronselaer, 2019:1; Wang & Strong, 1996:6; Williams & Tang, 2020:26). During the 1950s,` organisations were handling their data management processes manually and were using punched cards to record data (Gray, 1996:2). During this era, organisations did not have computer applications to store and manage data efficiently.

High-level programming language such as COBOL surfaced into the information technology space in the mid-1950s to 1970s and made it possible to write computer programs that read, analysed and transformed individual records (Gray, 1996:3). In the early 1960s, manual file systems were used to manage small amounts of data, and master files housed on magnetic tapes were used to store large volumes of data at a low cost; however, one major disadvantage was that they had to be accessed sequentially (Gray, 1996:2; Inmon, 2002:2). Batch transaction processing systems were responsible for processing data that were captured and stored on cards and magnetic tapes (Gray, 1996:3). These data management methods were daunting and extremely time-consuming to use.

Master files and magnetic tape grew extensively in the mid-1960s, resulting in huge amounts of redundant data, complexities in developing and maintaining programs, the need for large amounts of hardware to support all master files, issues when updating records and the need to synchronise data upon update (Gray, 1996:2; Inmon, 2002:2). The information management system (IMS) DBMS was developed by IBM in the late 1960s (Ramakrishnan & Gehrke, 2000:6). DBMS was introduced to solve issues that were created by the master files and to enable programmers to easily store and manage their data (Inmon, 2002:4). DBMS enables efficient access to data, data independence, data administration, ensure data integrity and security, concurrent access and crash recovery, and reduced application development time (Ramakrishnan & Gehrke, 2003:9).

Furthermore, in 1970 Edgar Codd proposed the relational data model as the new data representation framework (Ramakrishnan & Gehrke, 2000:6). This data representation model represents entities and relationship consistently (Gray, 1996:5). Database schema was introduced in database design to guide the development of physical database systems. Client-server computing and relational databases emerged between the 1980s and 1995 (Gray, 1996:5). Relational databases are designed using a unified language called SQL query language that is used for data definition, data manipulation and data navigation (Ramakrishnan & Gehrke, 2000:52). The SQL query language for relational databases has developed, and it is currently the standard query language (Gray, 1996:6; Ramakrishnan & Gehrke, 2000:52).

The popularity of relational DBMSs transformed the commercial landscape which saw the relational model claiming its position as the dominant DBMS paradigm (Ramakrishnan & Gehrke, 2000:6). Relational database was and is currently designed based on the entities and relationships defined in the relational model (Gray, 1996:5). There was great advancement in many areas of database systems and transformed data management.

During the 1980s, fourth generation languages (4GLs) began to surface and the notion that there is more that could be done with data than simply processing online transactions resulted with management information systems (MIS) being introduced (Ramakrishnan & Gehrke, 2003). Inmon (2002:5) states that organisations did not have a single database that could serve both analytical and operational transaction processing at the same time. MIS, currently known as decision support systems (DSS), assist organisations to analyse their data for decision-making purposes (Inmon, 2002:4).

Further development emerged between 1990 and 2000, extending DBMS to support new types of data such as spatial, multimedia and data analysis techniques such as OLAP, data warehousing and data mining (Vargas-Solar *et al.*, 2017:329). The concept of database management continues to gain importance as enormous data is brought online and made accessible via the computer network where it is driven by exciting visions such as multimedia databases, streaming data and interactive video (Ramakrishnan & Gehrke, 2003:7). This resulted in the birth of distributed databases, online analytical processing, and online complex processing (Vargas-Solar *et al.*, 2017:329). DBMSs have entered the internet age where queries are generated through Web-accessible form, answers are formatted using HTML, and DBMSs are used to store data that are accessed through the web browser (Ramakrishnan & Gehrke, 2000:7). Currently, the data management market is dominated by key object-relational database management systems like MS SQL server, DB2 and Oracle (Vargas-Solar *et al.*, 2017:329).

Today, cloud computing and other technologies such as smartphones, artificial intelligence (AI), big data, Internet of things (IoT) and social networking contribute to the excessive accumulation of data that traditional storage systems are unable to store and manage (Bianchini & Michalkova, 2019:7; Razbonyalı & Güvenoğlu, 2016:2558; Williams & Tang, 2020:29). The quantity of information available to organisations is literally exploding, and the inability to manage and discover information that is applicable to a given question in this vast amount of data is becoming a distraction and liability rather than an asset (Ramakrishnan & Gehrke, 2000:3). Large enterprises have resources to manage such volumes of data; however, SMEs are struggling to store, manage and analyse their data (Bianchini & Michalkova, 2019:7; Papachristodoulou *et al.*, 2017:70).

Small and medium-sized enterprises are the largest industrial contributor to the economy employment generation, output growth and wealth creation, and are the core pillars of any national economy (Assarlind, 2011:1; Kfouri & Skyrius, 2017:96; Nkwe, 2012:29; Olszak & Ziemba, 2012:138; Ramukumba, 2016:19; Scholz *et al.*, 2010:2). Small and medium-sized enterprises have shown massive growth in South Africa, they account for 55% of all jobs, compromise over 90% of African business operations, and contribute to more than 50% of African employment and gross domestic product (GDP) (Nkwe, 2012:29; Ramukumba, 2016:19; Van Scheers, 2011:5048). Small and medium-sized enterprises are constantly proving to be the engine of growth and are regarded as the major source of technological innovations and new products (Nkwe, 2012:29).

However, there seems to be a number of challenges that contribute to the failure of SMEs in South Africa and worldwide (Ramukumba, 2016:24). SMEs are constantly faced with issues related to unforeseen changes such as globalisation problems, technological innovation, climate change, market competition and business dynamisms (Ali *et al.*, 2017:152; Bustos & Vicuña, 2016:218; Curraj, 2018:3; Karanasios, 2011:4; Zainun Tuanmat & Smith, 2011:209). Ramukumba (2016:24) also identified issues such as a lack of management skills, finance, developing relationship with customers, access to markets, appropriate technology etc. Additionally, SMEs are facing critical challenges in accessing and analysing pertinent data (Bianchini & Michalkova, 2019:6).

Prior research has indicated that poor data quality is a critical issue in many SMEs and it impedes them from obtaining the best value from their data (Ballou & Tayi, 1999:56; Bianchini & Michalkova, 2019:25; Redman, 1998:80; Tayi & Ballou, 1998:73; Wang *et al.*, 1992:3; Wang & Strong, 1996:6). SMEs experience challenges in protecting the quality of their data. Poor data quality compromises decision-making processes, leads to customer dissatisfaction and employee job dissatisfaction, and intensify operational costs since resources such as time are spent diagnosing and fixing the errors (Aljumaili, 2016:3; Haug *et al.*, 2011:173; Jesiļevska, 2017:90; Mahanti, 2018:28; Redman, 1998:80).

The way organisations collect, manage, capture, integrate, process and load their data (Kademeteme *et al.*, 2017:31; Singh & Singh, 2010:42) can compromise data quality if data are not handled with care. Moreover, large volumes of data, consolidating data from various sources, disparate data stores, lack of legacy data standards, incomplete or missing data, heterogeneous data structures for the same customer, and actual data values being different from meta-label formats could lead to poor data quality (Gudivada *et al.*, 2017:14; Razbonyalı & Güvenoğlu, 2016:2559; Singh & Singh, 2010:182; Vosburg & Kumar, 2001:21).

According to Tayi and Ballou (1998:56), no one can anticipate all the conditions that could compromise the integrity of an organisation’s data; however, the first step in understanding how data are corrupted is to acknowledge the fact that data have many attributes and dimensions. Data quality requires understanding which dimensions of data quality are critical to the user (Wang *et al.*, 1992:4). The concept of data quality is divided in four dimensions; accuracy, timeliness, completeness and consistency (Haug *et al.*, 2011:171; Singh & Vashishtha, 2015:182). Data must satisfy quality dimensions in order to be considered of high quality (Batini

& Pernici, 2006:52; Haug *et al.*, 2011:171; Laranjeiro *et al.*, 2015:5; Singh & Singh, 2010:41; Wang & Strong, 1996:19).

To protect the quality of data and obtain important information out of large and complex datasets, SMEs should implement technologies that can simplify data management procedures and extract useful information in a timely fashion (Ramakrishnan & Gehrke, 2000:3). SMEs require better storage systems, improved data management capabilities, efficient data retrieval, and agile analysis and reporting, which can assist management in decision-making processes (Guarda *et al.*, 2013:187). According to Curraj (2018:51), SMEs need to adapt to the rapid transformations in order to innovate as well as improve performance and competitiveness.

Yeoh and Popovič (2016:1) postulate that DW/BI was developed as innovation mechanisms that can offer data integration and analytical capabilities that can enable stakeholders at various organisational levels with key information needed to make informed decisions. Large and small medium enterprises can use this technology to better manage their data, perform agile analysis and reporting, and make informed decisions (Curraj, 2018:3). Digital technologies are offering SMEs new opportunities to take part in the global economy, to innovate, increase performance and grow (Bianchini & Michalkova, 2019:6; Curraj, 2018:3). Business intelligence is regarded as an innovative technique that can be used to digitize SMEs to enable them to discover insight from their data, improve performance, improve data support, reduce costs, increase competitiveness, improve decision support, analyse customer behaviour, and increase revenue (Curraj, 2018:3; Scholz *et al.*, 2010:5).

Today, data are regarded as important assets of any organisation (Kalan & Unalir, 2016:4). SMEs are operating in highly competitive, unstable and uncertain markets. These changes have compelled SMEs to not only compete among themselves but also with larger enterprises (Kfouri & Skyrius, 2017; Zainun Tuanmat & Smith, 2011:208). However, Nenzhelele and Pellissier (2014:97) discovered that SMEs find it hard to compete with large enterprises. Adopting DW/BI solutions has become critical in today’s hyper-competitive markets where businesses are seeking to become more agile, proactive and efficient in decision-making processes (Ali *et al.*, 2017:155; Kfouri & Skyrius, 2017:97).

1.3 Problem statement

Large organisations are no longer the only ones interested in technologies and systems that can enable them to gain a competitive advantage over their rivalries, SMEs are now collecting enormous amounts of information and require systems with such capabilities (Bianchini & Michalkova, 2019:7; Guarda *et al.*, 2013:187; Gudfinnsson, 2019:2; Kfouri & Skyrius, 2017:98;

Popovič *et al.*, 2019:4). However, SMEs are struggling to use their business data to gain insight into information that will enable them to make informed decisions, to measure performance, innovate, reduce costs, analyse customer trends and gain competitive advantage over their rivalries (Curraj, 2018:17; Guarda *et al.*, 2013:187; Papachristodoulou *et al.*, 2017:70; Ponis & Christou, 2013:4; Razbonyalı & Güvenoğlu, 2016:2556).

Complexities such as handling large volumes of data, poor data quality, inadequate storage capabilities, data stored in disparate sources, system migrations, dissimilar data structures and data entry errors make data analysis and the extraction of information for decision making purposes a complicated and daunting process (Eckerson, 2002:10; Mahanti, 2019:5; Papachristodoulou *et al.*, 2017:70; Razbonyalı & Güvenoğlu, 2016:2556; Vosburg & Kumar, 2001:21). Such issues may result in management and users lacking trust and confidence in the quality of their data.

For organisations to evolve with technology, to keep up with the dynamic environment in which they operate and at the same time remain competitive, they require business systems, technologies and tools that can aid them with information to base decisions on and give them a competitive edge (Gudfinnsson, 2019:1; Venter, 2005:149). However, SMEs lag behind in the use of intelligence systems that enable them to efficiently capture, store, manage and extract knowledge for decision making purposes because they find it challenging to adopt new technologies and take advantage of enormous amounts of data (Bianchini & Michalkova, 2019:7; Gudfinnsson, 2019:2; Raj *et al.*, 2016:6; Scholz *et al.*, 2010:1).

The research study is based on one SA SME which specialises in engineering casting and manufacturing of steel mill rolls and rings. This SA SME as referred to in this research study is struggling to use their data to make informed decisions that will enable them to gain a competitive advantage over their rivalries. The following problems experienced by the SME are currently denying them to innovate, measure performance and reduce cost:

- Data capturing issues (human error issues);
- Data are stored in disparate locations;
- Dissimilar data structures;
- Poor data quality;
- Poor data management;
- Inefficient access and retrieval of data;
- Multiple users unable to access the system simultaneously and capture data; and
- Management do not have confidence in the data.

1.4 Research objectives

The primary objective of this study is:

- **To improve data quality and management in one South African SME for data analytics.**

In order to reach the primary objective of this study, the following theoretical and practical objectives have been formulated:

1.4.1 Theoretical objectives

- To explore technologies that SMEs can implement to better manage and store their data.
- To investigate the impacts of poor data quality on SMEs.
- To identify the challenges of poor data quality and benefits of high data quality in SMEs.
- To explore different BI models, data warehousing and ETL concepts suitable for SMEs.

1.4.2 Practical objectives

- Propose guidelines and technologies for improvement of data quality in one South African SME.
- Design an artefact for one South African SME to store and manage data for data analytics.

1.4.3 The research purpose

The purpose of this research was to improve data quality and management in one South African (SA) SME for data analytics. In the process, an artefact was designed with consequent guidelines that one SA SME can implement to better manage their data and improve data quality for analytics and decision-making purposes.

1.4.4 Research questions

Main research question

> How can SMEs exploit technology to better manage data and improve the quality of their data for data analytics?

The study aims to answer the following supporting research sub-questions indicated by SQ1 to SQ3:

- SQ1: Which applications are SMEs using to store and manage their data?
- SQ2: What are the impacts of poor data quality in SMEs?
- SQ3: What are the challenges of poor data quality and benefits of high data quality in SMEs?

1.5 Research methodology

The main aim of this study was to develop an artefact for one SA SME to assist with capturing, storing and managing data for data analytics purposes and to propose guidelines and technologies that one SA SME can use to improve data quality. The study followed the design science research methodology in order to address and answer the research questions, because one SA SME was trying to solve a real-world problem that required the design of an artefact as a solution. A case study approach was selected as an appropriate strategy.

1.5.1 Design science research

Design science research (DSR) in information systems (IS) is a problem-solving paradigm that produces new knowledge and insight by the construction and evaluation of innovative artefacts (Hevner *et al.*, 2004; Kuechler & Vaishnavi, 2012:396). According to Vaishnavi *et al.* (2004/2019:1), state that design science research is a "*lens or set of synthetic and analytical techniques and perspectives for performing research in IS*". Hevner *et al.* (2004:84) state that in information systems research, the aim is to understand and solve critical business problems through the construction of a technology-based artefact. This methodology involves meticulous processes of constructing artefacts to solve identified problems, to make research contributions and to communicate the results to the relevant audience (Hevner *et al.*, 2004:77).

The research must produce an artefact designed to address a problem, be relevant, and its utility, quality and efficacy must be rigorously evaluated (Hevner *et al.*, 2004:85). This study uses a design research process model developed by Vaishnavi *et al.* (2004/2019:12) as illustrated in

Figure 1-1, which follows five key steps: (1) awareness of the research problem, (2) suggestion of the problem solution, (3) development of the artefact, (4) evaluation of the artefact and (5) conclusion on the research followed by a discussion of each step.

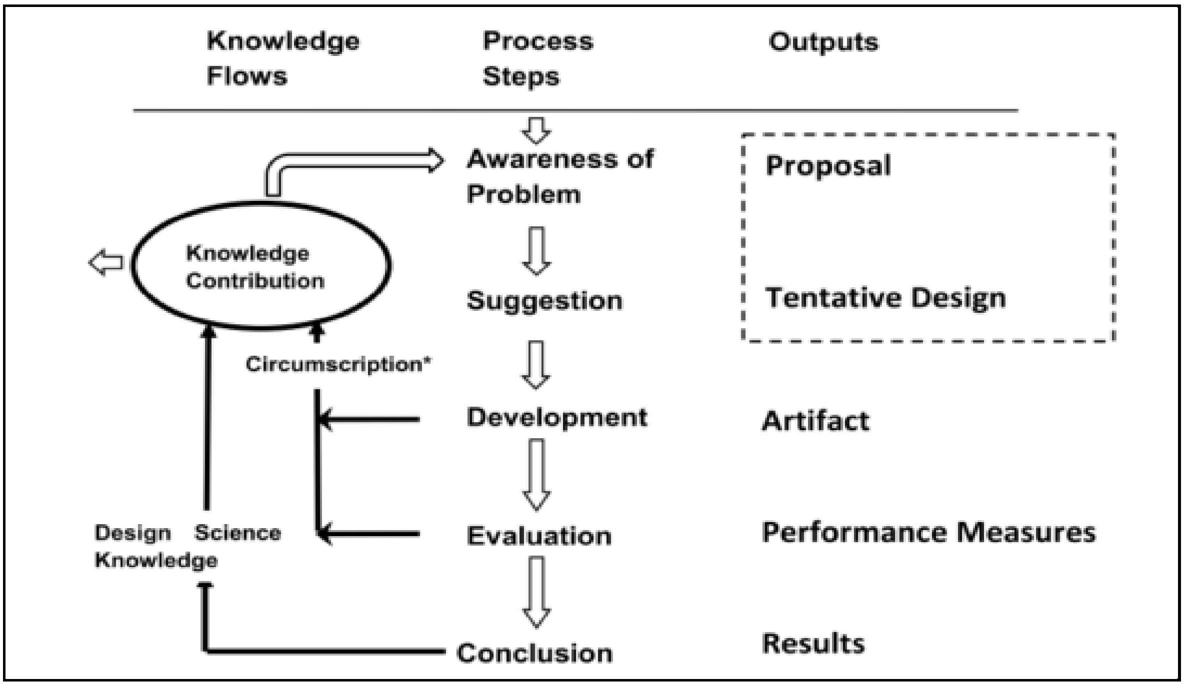


Figure 1-1: Design Science Research Process Model (Vaishnavi *et al.*, 2004/2019:14).

Awareness of problem

In this phase, the awareness and identification of the problem together with the probable solution are presented. The identification of the problem may stem from numerous sources, including new development in industry or identification of issues in a reference discipline. "*The types of problems that are relevant for a design science research tend to be problem-solving focused in their approach as opposed to questions or problems that are answered through explanation*" (Vaishnavi *et al.*, 2004/2019:12). The research problem is defined and it will be used to develop artefacts such as architectures, design principles, constructs, frameworks, models, methods and instantiations. The output of this phase is usually a proposal to solve a problem.

Suggestion

During the suggestion phase, the proposal produced in the awareness phase will be used to make suggestions. Feasible methods that can be applied or used to achieve the goal are researched and preliminarily evaluated. The researcher will research and evaluate feasible ways of achieving the goal. The suggestion phase "*is a creative step wherein new functionality is envisioned based on a novel configuration of either existing or new and existing elements*" (Vaishnavi *et al.*, 2004/2019:12). These suggestions may however be insufficient to be solutions for the problem or suffer from important knowledge gaps. A tentative design or a final design is produced at the end of this stage.

Development

The development phase deals with the development of the preliminary design. The design produced in the suggestion phase will be used to develop the artefact. The output of this phase is an artefact.

Evaluation

Once the development phase is completed, the artefact must undergo a thorough testing process. This phase focuses on assessing the functionality of the artefact. "*The utility, quality and efficacy of the design artefact must be rigorously demonstrated via well-executed evaluation methods*" (Hevner *et al.*, 2004:83). The artefact is evaluated based on the criteria outlined in the first phase, i.e., awareness of the problem phase. Accuracy, reliability, completeness, performance, usability, functionality and fit with the organisation are some of the attributes that are used to evaluate the IT artefacts (Hevner *et al.*, 2004:85). Any form of deviations either qualitative or quantitative are meticulously noted.

Conclusion

Once the evaluation phase is completed, the conclusion phase follows, which indicates the end of the design science research project or end of the research cycle. In this phase, it is ideal to reflect on the challenges encountered, lessons learned, techniques that were used that worked or did not work in finding a solution to the problem. According to Vaishnavi *et al.* (2004/2019:16) abstraction allows comprehensive and appropriate conclusions to be reached through applying the knowledge acquired from the research effort to be achieved in the process of communicating and sharing the results to the larger knowledge base. Design science knowledge is the output of design science research (Vaishnavi *et al.*, 2004/2019:16). Design science knowledge and circumscription indicates the knowledge contribution emerging from new knowledge production. The circumscription process "*is especially important to understanding design science research process because it generates understanding that could only be gained from the specific act of construction*" (Vaishnavi *et al.*, 2004/2019:15).

Development and evaluation of an artefact usually occur iteratively until the original problem is addressed or solved.

1.5.2 Design science research applied to this study

Figure 1-2 illustrates the design science research model applied in this study. The research model outlines the research problem using the main cycle and three sub-cycles followed in the suggestion phase. Additionally, more details are provided about each cycle.

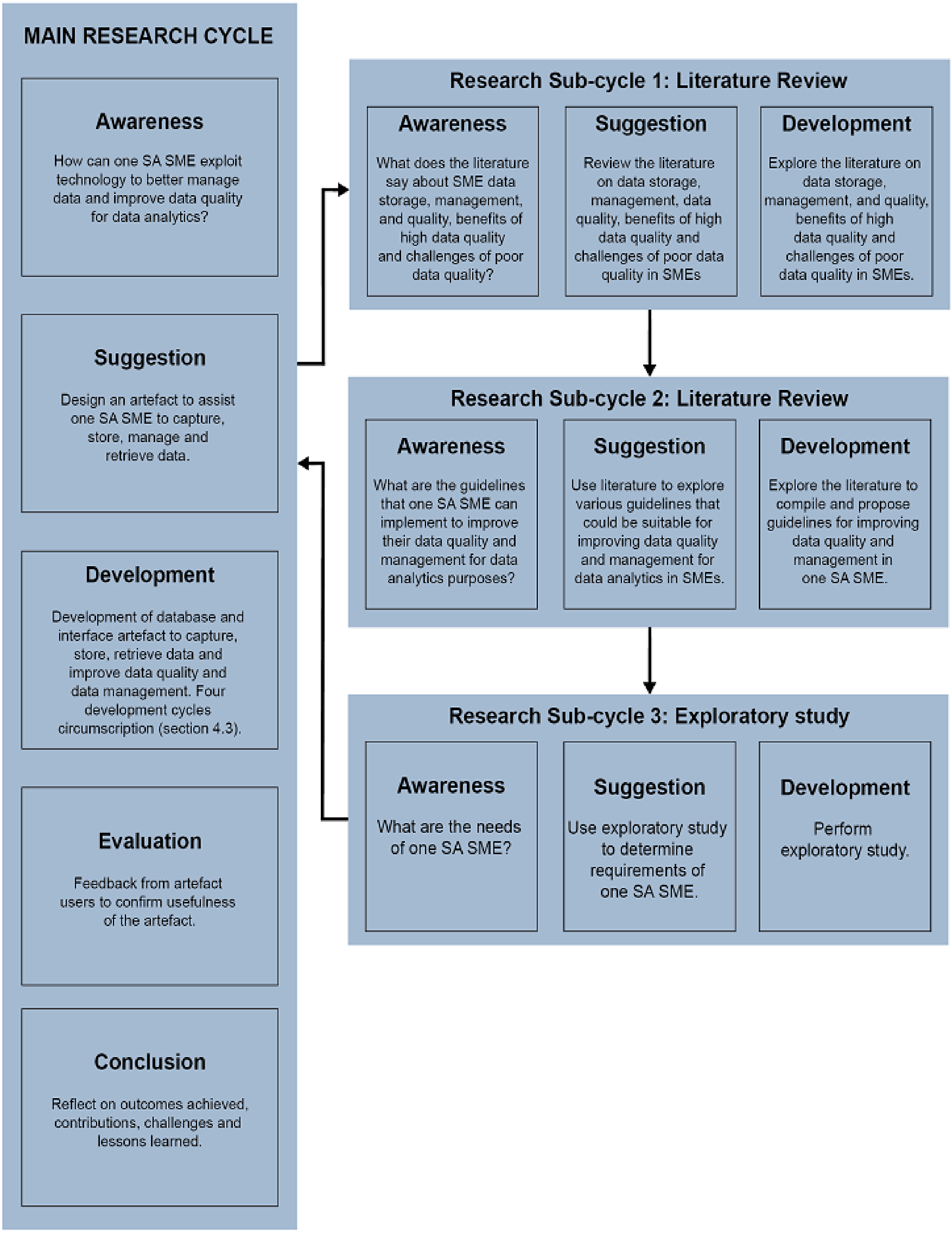


Figure 1-2: Design science research applied in this study

Awareness

Awareness occurred when one SA SME expressed a need for better data quality and management. The research problem was defined as how one SA SME can exploit technology to better manage data and improve data quality for data analytics purposes.

Suggestion

The suggestion was first to assist the organisation with capturing, storing, and managing data. In preparation for the design and development of an artefact, the literature was consulted to gain insight into the problem and suggested solution. Three design science research sub-cycles needed to be completed.

Sub-cycle 1

The awareness phase of sub-cycle 1 is concerned with what literature says about data storage, management and data quality, and challenges and benefits of data quality. The suggestion was to review literature related to the problem in hand and analyse how SMEs are storing and managing their data. In the development stage, the researcher explored literature related to the study and determined how SMEs are storing and managing data, dealing with poor data quality, challenges of poor data quality and benefits of high quality data.

Sub-cycle 2

The awareness of the problem in this phase is which guidelines and technologies are available that one SA SME can implement to improve their data quality and management for data analytics purposes. The suggestion was to review the literature and explore various technologies and guidelines that one SA SME can implement to assist with improving data quality and management. The researcher explored the literature and proposed guidelines and technologies that one SA SME can implement.

Sub-cycle 3

The awareness of the problem in this phase focused on the data management and quality needs of one SA SME. The suggestion was to perform an exploratory study and identify various needs of one SA SME. In the development phase, an exploratory study was completed to identify various needs and data management requirements of one SA SME.

Development

In the development phase, four circumscriptive phases were completed as a result of new awareness arising from evaluation while constructing the different phases of the artefact. An existing inefficient data storage facility and user interface had to be improved. A data warehouse had to be implemented, as well as a new reporting solution.

Evaluation

The functionality of the artefact was rigorously evaluated. In the evaluation phase of the main research cycle, a questionnaire was given to IT administrators and data capturers at the SME to give feedback on the functionality of the artefact and other deliverables of the project as a whole. Interviews were also conducted with the IT administrator and data capturers at the SME.

Conclusion

An artefact to improve the data quality and management in one SA SME was implemented to assist with data analytics. Furthermore, guidelines were proposed to improve data quality in SMEs. Also, in this phase it was necessary to reflect on outcomes achieved, contributions made, challenges encountered and lessons learned.

1.5.3 Participants

The participants in this research study included IT department employees of the SME chosen for the study. Data capturers, data management team, data analysts and IT infrastructure team assisted the researcher with understanding the problem, gathering requirements and developing the artefact. These participants were selected because they have a better understanding of data capturing and management processes in the company. They were familiar with the current state of the SME and had a better understanding of the organisational data needs. Most of these participants were involved in the daily activities of capturing and managing production data and were familiar with the data-related issues that the SME is currently experiencing.

1.5.4 Data collection and analysis

Qualitative research focuses on "*why*", "*how*" and "*what"* questions associated with motives, human behaviour, views and barriers", whereas quantitative methods are appropriate for problem quantification and testing of theories, intervention and new treatments and are suited for questions such as "*when*", "*how much*" and "*how many*" (Neergaard *et al.*, 2009:2). Qualitative research enable researchers to examine the feelings, perceptions and thoughts of research participants, to develop an understanding of the meaning that people allocate to their experience, whereas quantitative research is used to determine the number of people who partake in particular behaviour (Sutton & Austin, 2015:226). Qualitative methods allow researchers to acquire a better understanding of how and why such behaviours take place (Sutton & Austin, 2015:226). Quantitative methods presume that variables can be measured objectively, whereas qualitative methods are presumed as only partially objective, and more subjective accounts of the world can be produced and consequently be understood in various ways (Mehrad & Tahriri, 2019:4). Mixed method research integrates the two methods, qualitative and quantitative, to gather data required for answering the research questions. Research studies that use more than one method to improve accuracy of the findings are said to be utilising the mixed or multiple methods approach (Kumar, 2019:56).

In the main DSR suggestion phase, sub-cycle 3, an exploratory study was conducted to understand the data quality and management needs of one SA SME. The researcher used interviews as a primary method of collecting qualitative data. The researcher conducted interviews with IT personnel of the SME to better understand and gain insight into the problem. The feedback received from the questions and audio-recorded interviews enabled the researcher to thoroughly understand the needs of one SA SME. The researcher then designed and developed a solution that addressed their needs and assisted with solving their current issues.

In the evaluation phase of the main DSR cycle, interviews were conducted with two of the project leaders at the SME. Additionally, IT personnel were required to complete a questionnaire regarding the evaluation process of the artefact to obtain their feedback. A mixed method approach was followed. The feedback was used to evaluate the functionality of the artefact and to determine if the solution met their expectations.

The quantitative results of the questionnaires were analysed using descriptive statistics to evaluate the satisfaction of the users. The researcher used statistical software for the social sciences (SPSS). According to Ali and Bhaskar (2016:663), descriptive statistics paints a picture of the relationship between variables in a sample or population whereby it produces a summary of data in the form of mean, median and mode.

The qualitative data collected were analysed using content analysis, in particular using the guidelines for content analysis as described by Zhang and Wildemuth (2009:3), as detailed below.

Step 1: Prepare the data

The researcher needs to prepare the feedback received for analysis. In this process, audio-recorded interview feedback will be transcribed and checked. Transcribing is a burdensome process of converting the spoken words to the written words to ease analysis (Sutton & Austin, 2015:228). Qualitative content analysis is utilised to examine transcribed interviews to disclose and model people's information-related behaviour and thoughts (Zhang & Wildemuth, 2009:3).

Once the audio and video-recorded interviews are transcribed, they need to be coded. Sutton and Austin (2015:228) state that coding involves discovering topics, issues, similar and different items that are disclosed within the participant's narratives and explained by the researcher (Sutton & Austin, 2015:228). ATLAS.ti. will be used to analyse transcribed interview feedback.

Step 2: Define the unit of analysis

Qualitative content analysis mainly uses individual themes as the unit for analysis in preference to physical linguistic units such as sentences, words, or paragraphs that are frequently used in quantitative content analysis (Zhang & Wildemuth, 2009:3). The unit of analysis refers to "*the basic unit of text to be classified during content analysis*" (Zhang & Wildemuth, 2009:3). In this step, the researcher will choose a theme to be applied as the unit for analysis.

Step 3: Develop categories and coding scheme

The researcher needs to develop a coding scheme inductively from data. According to Zhang and Wildemuth (2009:3), categories and coding schemes can be obtained from three sources, such as previously related studies, data and theories. The researcher will establish a coding manual to document the definitions for assigning codes, category names and examples.

Step 4: Test your coding scheme on a sample of text

The researcher will use sample data to test the category definitions. The coding consistency needs to be verified once the sample is coded. Furthermore, Zhang and Wildemuth (2009:3) state that the researcher will continue to code sample text, verify coding consistency and revise coding rules up to a point where adequate coding consistency is attained.

Step 5: Code all the text

Once the researcher is satisfied with the coding rules and adequate consistency is attained, the coding rules will be applied to the entire collection of the text. As data is collected throughout the process, the coding manual will be updated.

Step 6: Assess your coding consistency

Once the researcher has coded the rest of the data set, coding consistency needs to be rechecked.

Step 7: Draw conclusions from the coded data

During this step the researcher will make sense of the categories identified and their properties, and make assumptions and conclusions based on the meanings derived from the data. It is regarded as the most crucial stage in the analysis process and the researcher's reasoning abilities plays a critical part in its success.

Step 8: Report your methods and findings

Qualitative content analysis is not concerned with producing counts and statistical significance but focuses on unearthing patterns, themes and categories critical to a social reality (Zhang & Wildemuth, 2009:5). In this stage, practices and decisions related to the methods and coding process that were utilised to initiate the trustworthiness of the study are explained and reported by the researcher. As it is very challenging to present research findings from qualitative content analysis, the researcher will make an effort to balance between description and interpretation.

1.5.5 Ethical considerations

Applicable ethical considerations:

- Ethical clearance was obtained from the North-West University.
- A signed agreement between NWU and one SA SME was obtained to ensure the legitimacy of the research.
- There was a written code of conduct compiled between the researcher and the SME used in the research.
- The researcher ensured that confidentiality, anonymity and privacy was maintained throughout the research.
- An agreement was made that all resources such as business data received from the organisation would be protected and handled with care.
- The researcher signed a code of conduct as per North-West University regulations.

1.5.6 Rigour of the study

Research methods need to be evaluated before the researcher makes any conclusions. Cypress (2017:254) advocates that qualitative studies must be conducted with rigour due to the potential subjectivity of the nature of this type of research. In qualitative research, authenticity, credibility, criticality and integrity are strategies that can be used to enhance rigour (Neergaard *et al.*, 2009:4).

- ❖ **Authenticity**: The IT personnel of one SA SME must be able to speak and express their views openly; their voices must be heard and their perceptions must be accurately represented.
- ❖ **Credibility**: The research method must capture and portray a truly insider perspective.
- ❖ **Criticality**: The researcher must ensure that every decision should have a reflection of critical appraisal applied to it. It must be relevant and trustworthy.
- ❖ **Integrity**: Researcher bias must be reflected on, validations and member checks can be carried out by the IT personnel of one SA SME, and collected and analysed data must be peer reviewed.

Rigour refers to the state of being thorough and accurate (Cypress, 2017:254). Authenticity rather than reliability is often the problem in qualitative research (Seale & Sillverman, 1997:379). Cypress (2017:254) emphasises that reliability and validity are two primary aspects of all research. Trustworthiness is described by Cypress (2017:254) as authenticity, quality and truthfulness of discoveries of qualitative research. Once the researcher had transcribed the interview voice recordings using ATLAS.ti, the supervisors checked the transcribed interviews.

A checklist developed by Hevner and Chatterjee (2010:20) is available to researchers to assess a design research project and to ensure that a project addresses the primary aspects of DSR.

Design science research checklist

Table 1-1: Eight design science research checklist questions - adopted from (Hevner & Chatterjee, 2010:20).

Questions
1. What is the research question (design requirements)? 2. What is the artefact? How is the artefact represented? 3. What design processes (search heuristic) will be used to build the artefact? 4. How are the artefact and the design processes grounded by the knowledge base? What, if any, theories support the artefact and the design process? 5. What evaluations are performed during the internal design cycles? What design improvements are identified during each design cycle? 6. How is the artefact introduced into the application environment and how is it field tested? What metrics are used to demonstrate artefact utility and improvement over previous artefacts? 7. What new knowledge is added to the knowledge base and in what form (e.g., peer reviewed literature, meta-artefacts, new theory and new methods)? 8. Has the research question been satisfactory addressed?

1.5.7 Assumptions and limitations of the study

The main limitation of this study is that it was based on one SA SME, and the guidelines to improve data quality and data management were developed based on the needs of this SME. The SME has a limited number of staff responsible for data management. Therefore, another limitation would be the limited number of participants interviewed and small number of responses obtained from questionnaires. Covid-19 restrictions resulted in virtual meetings and workshops that was originally planned as being face-to-face.

1.6 Significance of the study

The significance of this study is:

- ❖ This study proposes guidelines and technologies that one SA SME can implement to improve their data quality and management processes.
- ❖ The practical outcomes produced an artefact that was implemented and assists the one SA SME with efficient capturing, storing and management of data.
- ❖ Through the study the SME is provided with insight into poor data quality, challenges associated with poor data quality, benefits of high data quality, and efficient managing and storing of data when processing big data.
- ❖ The study not only provided the one SA SME with a solution to their data quality and management for data analytics, but the results and the guidelines may prove to be useful in other contexts too.

1.7 Dissertation layout

Chapter 1: Introduction

This chapter introduced the research study by discussing the background of the research environment to give the reader an introduction on what the research paper is about and various aspects that the researcher will cover. This chapter described the problem statement, purpose and objectives, together with the research questions and aims of the study. Furthermore, the research methodology, data collection, ethical considerations and rigour of the study are discussed.

Chapter 2: Literature review

This chapter describes the key concepts of the study where in-depth explanation of data quality, data management, business intelligence, data warehouse, data analytics and ETL will be covered.

Chapter 3: Research methodology

This chapter will explore research paradigms, research approaches, research strategies and design science research models applicable to this study. Subsequently, data collection techniques, data analysis, ethics in design research, and the study plan are described in detail.

Chapter 4: Design and development

This chapter will discuss the design and development processes followed in constructing the artefact. Additionally, the case environment is also explained to understand the SME involved in the study.

Chapter 5: Demonstration and evaluation

This chapter discusses the design science research followed in the study, data collection techniques, sampling, the results from the interview and questionnaire, as well as the guidelines for improving data quality and management.

Chapter 6: Research conclusion and recommendations

This chapter describes the objectives accomplished in the study, theoretical and practical objectives of the study, and limitations of the study, rigour of the research, recommendations and conclusions. Based on the findings, the researcher will present a conclusion to the study and make recommendations for further research.

1.8 Conclusion

In this section the researcher provided an introduction of the study by discussing the key concepts associated with this research study. A brief overview of concepts such as data quality, database, data warehouse, data management and data analytics was provided in this chapter. The researcher explained the advancement and impact of IT in SMEs and how SMEs are using these concepts to enhance business performance, increase revenue and gain competitive advantage.

The research study was based on one SA SME which specialises in engineering casting and manufacturing of steel mill rolls and rings. In this section, the researcher indicated that one SA

SME was struggling to use their data to make informed decision that could enable them to innovate and gain competitive advantage over their rivalries. Some of the issues that were discussed were dissimilar data structures, data capturing and data management issues, poor data quality and data management within the SME. The research problem guided the conceptualization of the research objectives.

The primary objective of the study was to improve data quality and management in one SA SME for data analytics. To achieve the primary objective, the theoretical objectives and practical objectives were discussed. The theoretical objectives involved exploring literature to investigate technologies that SMEs can implement to improve data management and storage, to investigate the impacts of poor data quality, identify the challenges poor data quality and benefits of high data quality and to explore extract, transform and load (ETL), business intelligence (BI) and data warehouse (DW) concepts. Furthermore, the practical objectives involved constructing an artefact for one SA SME to capture, store and manage data and to propose guidelines and technologies for improvement of data quality in one SA SME.

The chapter also discussed the research purpose and questions. The development of the artefact was guided by the design research process model developed by Vaishnavi *et al.* (2004/2019:12) which followed five key steps: (1) awareness of the research problem, (2) suggestion of the problem solution, (3) development of the artefact, (4) evaluation of the artefact and (5) conclusion phase. The main research questions and sub questions were addressed using the main research cycle, sub-cycle one, two and three adopted from design research process model developed by Vaishnavi *et al.* (2004/2019:12).The participants, data collection and analysis procedures, ethical considerations, rigour of the study and limitations and assumptions of the study were discussed. The dissertation layout is also presented where the next chapter will explore the literature on concepts associated with this research study.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

A literature review is critical in research because it enables the researcher to acquire thorough understanding of the research topic, identify key issues related to the topic, and investigate how issues have been addressed through research studies (Hart, 1998:1). Randolph (2009:2) emphasises that conducting a literature review is a technique of denoting an author's knowledge about a distinct field of study, vocabulary, theories, phenomena and key variables. Moreover, Mauch and Park (2003:123) state that the literature review should make it indisputably clear to the reader that there are incomplete pieces to the body of research, what those pieces are, and that the proposed study is directly aimed at completing one or more of those missing pieces. According to Randolph (2009:2) the literature review assists with gaining methodological insights, eluding fruitless approaches, exploring new lines of inquiry, delimiting the research problem, searching support for grounded theory and determining recommendations for further research. The conclusion of the literature review should indicate how the proposed study would add to the subject's knowledge base (Mauch & Park, 2003:123).

According to Raj *et al.* (2016:1), the ability of an SME to access, store, manage and analyse large though seemingly unrelated volumes of data, is critical to business decision-making. However, such data need to be processed by going through conversion stages until the data are deemed useful (Duan *et al.*, 2017:327).The progressive relationship between data, information, knowledge and wisdom is demonstrated using a pyramid in Figure 2-1 (Duan *et al.*, 2017:331). The figure illustrates the transitions from data to information, to knowledge, which leads to wisdom. Data are the foundation of the transition process according to Bellinger *et al.* (2004:3); you cannot have information without data (Davenport & Prusak, 1998).

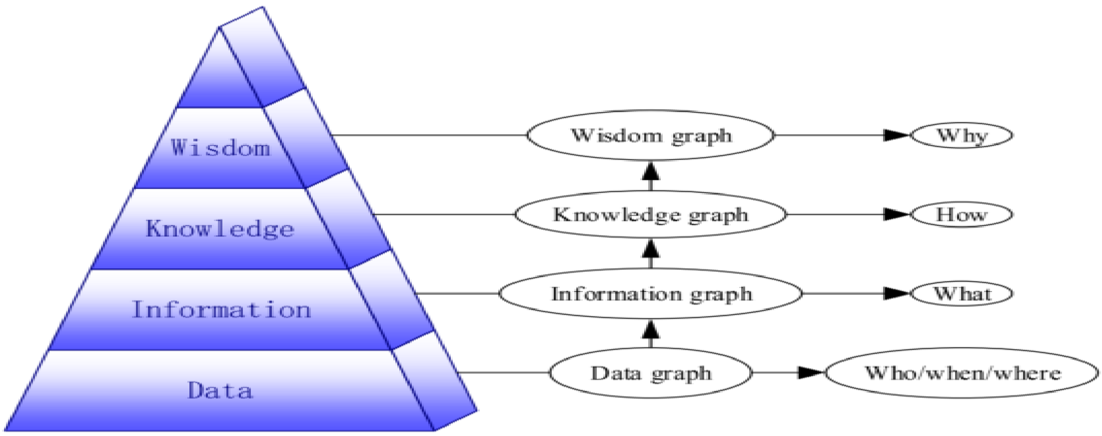


Figure 2-1: Relationship among data, information, knowledge and wisdom (Duan *et al.*, 2017:331).

In order for organisations to obtain value from their data, data must be transformed into valuable information, and this information into knowledge that can enable organisations to gain competitive advantage, to enhance productivity, increase revenue, present new business opportunities and assist management in making informed decisions effectively and efficiently (Bianchini & Michalkova, 2019:9). Large, medium and small enterprises need technologies and tools to analyse data to extract knowledge that can aid businesses in decision-making processes.

The purpose of the literature review in this study was to position the study within the relevant research field, to explore the significance of data quality, data storage and management in SMEs, to gain an understanding of the data quality dimensions and methodologies, and identify concepts and technologies that were used to address similar issues through completed research studies.

Section 2.1 gives an introduction of this chapter. Section 2.2 discusses data, information in section 2.3, knowledge in section 2.4 and wisdom in section 2.5.

Section 2.6 gives an overview of data quality and definitions, as well as the benefits of high-quality data. It further discusses the data quality dimensions in 2.6.1; explores data quality methodologies in 2.6.2; discusses big data and its benefits in 2.6.3; and poor data quality is discussed in 2.6.4 where the researcher explores the causes and impacts of poor data quality.

Section 2.7 discusses data management where various application and processes that SMEs can adopt to manage data quality are explored. Section 2.8 explores SMEs by discussing the role of SMEs in society, gives a broad definition of SME in South Africa, and their significance in economic development.

Furthermore, section 2.9 discusses data analytics, data warehousing and business intelligence (DW/BI) concepts (2.9.1), section 2.9.2 explores the ETL process followed by the conclusion in section 2.10.

2.2 Data

Today, data is quickly becoming a fundamental piece of doing business and a key economic driver around the globe (Experian, 2018:2). Data are no longer regarded as a by-product of an organisation's IT system and applications, but are currently considered the most valuable organisational asset and resource that have a real, measurable value (Mahanti, 2018:2). It is worth noting that data have become an enterprise-wide corporate asset that are used virtually in all organisational activities and constitutes the basis for operational, tactical and strategic level decisions (Haug *et al.*, 2011:169; Moore, 2018; Razbonyalı & Güvenoğlu, 2016:2556; Sidi *et al.*, 2012:300). Furthermore, Limpeeticharoenchot *et al.* (2020:88) state that data are considered the new ‘oil’, and are regarded as valuable assets for large, small, and medium-sized enterprises. Sherman (2014:3) mentioned, “*[d]ata is [sic] the key, the ticket and the Holy Grail all rolled into one*”.

Davenport and Prusak (1998:2) defines data as *“a set of discrete, objective facts about events”.* Duan *et al.* (2017:327) maintain that data existing as distinct elements have no semantics. According to Ahsan and Shah (2006:4), data is raw and it can originate any form that can either be usable or unusable and has no significance beyond its existence. Additionally, Duan *et al.* (2017:328) emphasise that data without context and on their own have no meaning; they are simply represented through symbols of noticeable properties of the world. Data serves as the foundation from where information and knowledge are elicited. Furthermore, Duan *et al.* (2017:332) emphasise that data can be used to answer questions directed by “*Who/When/Where*”, whereas information answers questions directed by “*What*”; subsequently knowledge is capable to answer questions directed by “*How*”, then wisdom addresses “*Why*” questions (Duan *et al.*, 2017:332).

Recently, the amount of data within organisations have grown tremendously as organisations are creating, gathering, storing and managing more data than ever before (Liu *et al.*, 2018:840; Mullins, 2017:54). As a result, such data are presented in various formats as they are coming from various sources. A broad spectrum of possible data representation and a basic classification was proposed by various authors to implicitly or explicitly distinguish three types of data (Batini & Pernici, 2006:53; Parapi & Masykuri, 2020:1; Sidi *et al.*, 2012:301; Williams & Tang, 2020:30; Zhu & Cai, 2015:3):

- ❖ **Structured data** refers to when the data are categorised in elementary items and represented in standardised format that can be interpreted by a grammar. Such data are stored in regular, predictable formats that simplify retrieval and problem solving e.g., relational tables and statistical data (numerical or text inputs).
- ❖ **Unstructured data** are communicated in natural language and there is no particular structure or domain types outlined e.g., documents, video, audio, etc. The data are stored in non-relational databases, they may be textual or non-textual, human or machine generated e.g., questionnaires with free text answering and body of email.
- ❖ **Semi-structured data** refer to data that do not adhere to the standard structure of data models related with relational databases, and are not prepared for computations through metadata or tags e.g., XML, spreadsheets and financial reports.

Furthermore, three categories can be used to classify data when organisations consider data as a product (Sidi *et al.,* 2012:301):

- ❖ **Raw data items** are smaller data units which are utilised in creating information and components.
- ❖ **Component data** items are formulated from raw data elements and temporarily stored until a final product can be manufactured.
- ❖ **Information product** are the outcome of carrying out manufacturing activity on the data.

Once the data are generated and captured, the next step is to analyse the data to discover useful and high quality information that may lead to discovery of knowledge (Liu *et al.*, 2018:840). Kamensky (2018) states that data are not useful until they are refined into information and insight to support a business process or organisational decision-making processes (DAMA-UK, 2013:4). According to Ehrlinger *et al.* (2019:6), information is often used interchangeably with data even though they are not synonymous because data "*refers to plain facts*" and information describes "*the extension of those facts with facts and semantics*". Information contains, data but data is not necessarily information (Ahsan & Shah, 2006:4).

2.3 Information

The main objective of collecting data, information and knowledge is to be able to make informed decisions (Ahsan & Shah, 2006:3). High quality information enables organisations to formulate better decision strategies and unveil business patterns for decision making (Jaya *et al.*, 2017:2647; Mahanti, 2019:8). However, informed decisions can only be achieved if the data that is being used in the decision-making process is of high quality. Raw data need to be extracted from external and internal sources, transformed to meet acceptable standards, and stored into a single source system to be used for querying and analytical purposes. Bianchini and Michalkova (2019:9) state that raw data need to be cleaned, standardised, integrated and organised before it is regarded as information that could be used for analytics.

One of the most critical assets of any organisation is its information, hence it is used for operational record keeping and for analytical decision making (Kimball & Ross, 2013:2). Data becomes information once it has been assigned meaning through defining relational connection (Duan *et al.*, 2017:328). Information simply describes data that has been processed where "meaning" can be useful but does not have to be (Ahsan & Shah, 2006:4; Duan *et al.*, 2017:328). Davenport and Prusak (1998:3) state that information should change the way the receiver discerns something, to have an impact on his behaviour and judgement; it is crucial material required for eliciting and establishing knowledge (Duan *et al.*, 2017:329). Once information is interpreted, put into context or meaning has been added to it, then it becomes knowledge (Ahsan & Shah, 2006:3).

2.4 Knowledge

Knowledge refers to information that is organised and structured based on cognitive processing and validation (Ahsan & Shah, 2006:3). According to Bellinger *et al.* (2004:2), knowledge is the applicable collection of information such that its purpose is to be useful. Information serves as source from which knowledge is obtained. According to Liew (2013:2), knowledge "*is a product of human intelligence, intellectual activities and cognitive conscience*". We can intuitively say that knowledge originates from the interpretation of information which has been captured from disparate places (Liew, 2013; Nyaboga & Mwaura, 2009:20).

Enormous volumes of data increase the potential of enterprises to discover knowledge (Jaya *et al.*, 2017:2652). Bellinger *et al.* (2004:3) posit that knowledge constitutes a pattern that describes a high level of predictability of what might happen next. According to Duan *et al.* (2017:329) the objective of knowledge is to better our lives and to enable stakeholders to make more informed and accurate decisions. In fact, knowledge is critical in decision-making processes.

2.5 Wisdom

Wisdom is generated based on the knowledge one has acquired and it embodies mainly an understanding of fundamental principles integrated within the knowledge that are essentially the foundation for the knowledge being what it is (Bellinger *et al.*, 2004:3). Wisdom refers to the process of being able to judge or discern between good or bad, right or wrong (Bellinger *et al.*, 2004:3). Wisdom is an "*extrapolative and non-deterministic, non-probabilistic process*" that utilise past level of consciousness, especially special kinds of human programming such as codes, morals and ethics. Additionally, Bellinger *et al.* (2004:3) advocate that wisdom goes far beyond understanding by generating understanding that did not exist previously.

Based on knowledge generated during data analytics, organisational management needs to apply wisdom to make decisions and determine what will happen next. The key objective in collecting data is to elicit information and knowledge that can assist organisations to gain a competitive advantage, present new business opportunities, and assist management in making informed decisions effectively and efficiently.

2.6 Data quality

Over the past years, the quality of data has been important for organisations, and numerous studies have been intensively conducted focusing on the meaning of data quality, how it can be measured, how it can be achieved, and its significance in business success and decision-making processes (Batini *et al.*, 2009:16:12; Eckerson, 2002:3; Geiger, 2004:1; Ghobadian & Gallear, 1996:83; Li *et al.*, 2011:140; Loshin, 2001:10; Redman, 2004:22; Russom, 2006:6; Vaziri & Mohsenzadeh, 2012:60; Wang & Strong, 1996:6). Today, due to digital transformations and technological complexities, data quality persists to be a critical issue and a topic of interest to many researchers as large enterprises and SMEs are collecting extensive amounts of data and are more concerned about the quality of data (Bianchini & Michalkova, 2019:7; Jesiļevska, 2017:90; Mahanti, 2019:6; Timmerman & Bronselaer, 2019:1). More recently, over the last five years, data quality (DQ) in big data has gained greater prominence and has evolved as a well-researched topic in the literature (Coleman *et al.*, 2016:1; Del Vecchio *et al.*, 2018:6; Heinrich *et al.*, 2018a:1).

Before we analyse data quality and its role in business success, we must first understand what data quality is. However, there is no consensus in the literature on a single definition of data quality, rather the definitions are based on what data quality means to those who use the data. In the 1990s, the Total Data Quality Management (TDQM) group of MIT University led by Professor Richard Y. Wang conducted in-depth research in the area of data quality (Zhu & Cai, 2015:2). From the study, Wang and Strong (1996:6) define data quality as "*data that are fit for use by data consumers*". According to Loshin (2001:27)**,** data consumers can be grouped into three categories:

- ❖ **Operational data consumers –** they are responsible for internal processes or any transaction processing activity, message passing routing and workflow activities that needs input to operate.
- ❖ **Tactical and strategic data consumers –** refers to those who utilise processed information to make strategic and tactical decisions. This includes marketing, sales management, acquisitions, enterprise resource planning, mergers, etc.
- ❖ **External consumers –** refers to those who obtain information processed by the information factory. Information such as customer billing, invoices, geographic data, sales teams and data supplied by government agencies are included in this category.

The data warehousing institute report (TDWI) elucidates that data usability is important for ensuring that business leaders have trust and confidence in their data and their decision making is based on reliable and accurate information (TDWI, 2020). Data quality is a crucial concepts in businesses even though it is regarded as a complicated approach that requires numerous business quality practices and data management techniques. For effective data quality management, organisations need to develop a formal data governance program and adopt a data stewardship approach (Mahanti, 2018:431). However, TDWI (2020) emphasises that data trust goes beyond simple data quality metrics.

Heinrich *et al.* (2018b:2) describe data quality as "*agreement between the data views presented by an information system and that same data in the real world*". According to Jesiļevska (2017:89) and Mahanti (2019:5), the concept '*fitness for use'* depends on the application for data, the attribute of quality that is essential for that particular purpose and on users' expectations of what they define to be valuable and functional information. Fitness for use implies that organisations need to look beyond traditional concerns associated with data accuracy (Tayi & Ballou, 1998:54). In other words, data is fit for use when it satisfies acceptable levels across all dimensions, for all applicable business processes through the entire organisation (Pitney Bowes, 2015:4). According to Loshin (2001:48) data quality is defined based on how data consumers desire to utilise the data.

According to ISO/IEC (2008), a usual prerequisite associated with all IT projects is the quality of the data which are interchanged, processed and utilised between the computer systems and users and among computer systems themselves. As described in Figure 2-2, data are of high quality if they are comprehensive, complete, relevant, easy to read, current, accessible, accurate, easy to interpret, and consistent with other sources (Singh & Singh, 2010:41). Fitness for use simply implies that data are free from defects and possess desired features (Juran & Godfrey, 1999:34:39; Redman, 2004:22).

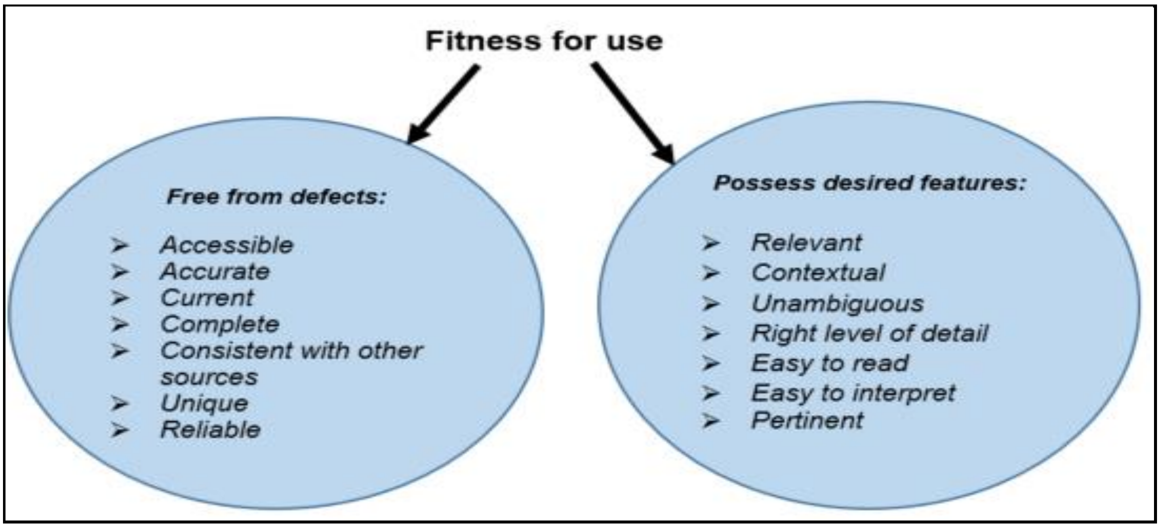


Figure 2-2: Data are of high quality if they are "fit for use" (Juran & Godfrey, 1999:34:39; Mahanti, 2018:10; Redman, 2004:22).

Data quality is referred to as "data integrity" and it advocates maintaining and assuring the accuracy and consistency of data throughout its entire lifecycle (Hafeez *et al.*, 2019:306). Moreover, the concept of data quality can be understood by classifying it into subcategories and dimensions (Haug *et al.*, 2011:171). Data quality concepts can be described and evaluated using dimensions or attributes such as accuracy, consistency, timeliness and reliability (Ballou & Pazer, 1999:854).

Consequently, Mullins (2013:489) states that data quality is an overarching industry issue. According to Gudivada *et al.* (2017:14), common data quality issues include missing and incomplete data, outdated data, duplicate data and inaccurate data. Due to technological advancement, social and mobile channels and rapid growth in data generation aggravates the data quality issues (Gudivada *et al.*, 2017:14). Data quality continues to impede SMEs from obtaining the best value from their data (Heinrich *et al.*, 2018b:2). Furthermore, Heinrich *et al.* (2018a) emphasise that it is essential for organisations to assess and assure the quality of the underlying data in order to achieve informed and effective decisions. In fact, Vosburg and Kumar (2001:25) emphasise that it is every employee's responsibility to protect the integrity of the data. Additionally, organisations must develop a data-driven culture and mind-set required to maintain and sustain data quality (Mahanti, 2018:408; Mullins, 2017:55; Vosburg & Kumar, 2001:25).

According to Mahanti (2019:6) "*high-quality data is not a nice-to-have requirement but a must-have requirement*". High quality data are essential to enable organisations to construct effective business strategies, provide excellent customer service, comply with regulations, operational efficiency, assist in generating returns, discover business patterns, and provide effective decision-making (Jaya *et al.*, 2017:2647; Mahanti, 2018:2). Pitney Bowes (2019) state that high quality data produce better analytics, increased insight and greater opportunity. High data quality allows users to have more confidence in the results they generate, reducing the risk in the outcomes and increasing efficiency in reporting, data access and decision making (Forbes Insight, 2017:10). According to Parapi and Masykuri (2020:3), organisations can use high quality data to predict future trends, detect anomalies, classification of dissimilar characteristics, and association of complementary behaviours.

Data quality is a complex concept because 'quality' depends on the conditions in which it is applied along with its fitness for use (Jesiļevska, 2017:89; Scannapieco *et al.*, 2005:13). For this study, data quality refers to data that are "fit for use" by data consumers and meet acceptable levels across dimensions such as validity, accuracy, consistency, uniqueness, completeness and timeliness.

2.6.1 Data quality dimensions

Usually, when people think about data quality, they associate it with accuracy; however, data quality goes beyond data accuracy (Mahanti, 2019:5; Scannapieco *et al.*, 2005:6; Sidi *et al.*, 2012:303; Wang & Strong, 1996:6). Organisations need to consider and analyse other significant dimensions in order to achieve high data quality (Mahanti, 2019:5; Sidi *et al.*, 2012:303). Wang and Strong (1996:21) argue that in order for organisations to be able to improve data quality, they need to understand what quality means to those who use the data. In fact, to measure data quality, organisations need to measure one or more data quality dimension depending on the situation, context and task for which data will be utilised (Mahanti, 2019:5). Numerous attributes can be used to measure or evaluate the quality of data and usually the criteria for measuring the quality of data differ based on the type of data, nonstandard or incomplete data, what is possible technologically, data use and business requirements, how tolerant the technology and the business are to flawed (Russom, 2006:5).

Data quality is typically referred to as a multi-dimensional and hierarchical concept (Mahanti (2019:5), where a single feature of it is described by data quality dimensions (Ehrlinger *et al.*, 2019:2). Data Management United Kingdom Group (DAMA-UK) state that data quality professionals widely use data quality dimension to represent an attribute of data that can be measured against predefined standards to discover the quality of data (DAMA-UK, 2013:3). A data quality dimension is a "*characteristic or part of information for classifying information and data requirements*" (Sidi *et al.*, 2012:302).

Vaziri and Mohsenzadeh (2012:55) argue that a dimension represents a single feature of data quality such as "accuracy", "consistency", "timeliness" or "completeness". In fact, a dimension captures a particular facet of quality (Scannapieco *et al.*, 2005:6). Wang and Strong (1996) define data quality dimension "*as a set of data quality attributes that represent a single aspect of construct of data quality*". The only question that should be answered when measuring the quality of data should be "*Are data of high quality or not*?" (Timmerman & Bronselaer, 2019:1). Data quality measurement is a necessity for thorough and strategic data quality improvement (Ehrlinger *et al.*, 2019:1).

According to Mahanti (2018:129) measuring data quality dimension for a data set requires understanding the data set and its constituent data elements altogether, understanding the context of data use and the attributes of the data elements such as data type, size and default values. Furthermore, Mahanti (2018:136) assert that metadata is the initial input required for measuring the data quality. However, three data granularity levels should be taken into consideration (Mahanti, 2018:130):

- **Data element –** a characteristic of a real-world entity that can be assessed and measured.
- **Data record –** a group of characteristics (defined by data elements) that illustrate a real-world entity occurrence.

- **Data set –** refers to a group of data represented in rows and columns extracted from database tables or data files to achieve a specific task.

Numerous attempts have been made to identify data quality dimensions. Researchers Loshin (2001:113) and Nyaboga and Mwaura (2009:16) identified completeness, accuracy, timeliness and consistency as the four dimensions of data quality concerning data values. The DAMA-UK working group in their report on defining data quality dimensions recommended six core dimensions for assessing and measuring data quality. Figure 2-3 below outlines the six primary dimensions as accuracy, validity, timeliness, consistency, uniqueness and completeness (DAMA-UK, 2013:7).

Figure 2-3: Six core data quality dimensions (DAMA-UK, 2013:7).

- **Data accuracy –** describes the degree to which data correctly describe the "*real world*" event or object being described. The data values need to correspond with an identified source that represent the correct information. In a relational model, value comes from a data domain, so accuracy can also be defined as the proximity the value of a characteristic to a particular value in the attribute domain, which is regarded a true value (Nyaboga & Mwaura, 2009:16). According to (Hafeez *et al.*, 2019) the recorded value should be correct, free from errors, valid, complete, truthful, reliable and reflective of the observation.

  Primary question – Do the data give accurate representation of the data set?

- **Completeness –** describes the expectation that specific attributes are expected to have assigned values in a dataset. The percentage of stored data against the potential of "100% complete". In a relational model, a mandatory attribute needs a value; optional value is defined by an attribute that may or may not have a value. It will include the assessment of the absence of blank, null or empty values or the existence of non-blank values (Nyaboga & Mwaura, 2009:16).

  Primary question – Are all data sets and data items captured?
- **Consistency –** refers to one data set having consistent data values with another data set. Comparison between two or more representations of a thing against the definition should not produce any dissimilarities.

  Primary question – Can the datasets across all data stores be matched?
- **Currency/Timeliness –** describes the degree to which data provides a true reflection of reality from a required point in time. Nyaboga and Mwaura (2009:16) emphasise that it can be measured as the time between when information is anticipated to be acquired and when it is readily accessible for utilisation.

  Primary question – Does the data represent reality from that specific point in time?
- **Uniqueness –** is the contrary of an assessment of the level of duplication. Data item will not be stored more than once based on how that item is identified.

  Primary question – is there a single representation of the data?
- **Validity –** describes the degree to which data are sustainable and they adhere to the syntax (data type, format, range) of its definition.

  Primary question – Does the data correspond with the rules?

Singh and Singh (2010:41) study on descriptive classification of causes of data quality problems in data warehousing, found completeness, accuracy, integrity, consistency, conformity and validity as dimensions for measuring data quality in data warehouses. However, Scannapieco *et al.* (2005:6) and Sidi *et al.* (2012:304) argue that dimensions such as accuracy, completeness, currency and timeliness are more commonly referenced dimensions than others. Furthermore, Scannapieco *et al.* (2005:7) argue that there are many other subjective dimensions that have been proposed to characterise data quality such as interpretability, reputation, believability and reputation. In relation to information systems, researchers identified dimensions such as relevancy, usability, independency, precision and reliability (Sidi *et al.*, 2012:302).

Additionally, DAMA-UK (2013:15) states that even though dimension accuracy, validity, timeliness, consistency, uniqueness and completeness are deemed satisfactory in measuring data quality, the data can still fail to accomplish the objective. As a result, organisations need to check other data quality considerations and ask additional questions about the data such as (DAMA-UK, 2013:15):

- **Usability of the data –** Are the data at the correct level of precision? Are the data simple, pertinent, understandable, maintainable and accessible?
- **Timing issues with the data (beyond timeliness itself) –** Is it the right time to legitimate the change requests?
- **Confidence in the data –** are data security, data governance and data protection in place? How is the trustworthiness of the data? Is it verifiable or verified?
- **Flexibility of the data –** are the data compatible and comparable with other data? Do the data have useful classification? Can the data be repurposed and are the data easy to manipulate?
- **Value of the data –** is there a cost or benefit case associated with the data? Are the data impeccably used? Based on corporate image or message, do the data contradict or support it?

Even though data quality dimensions do not depend on each other, there is a relationship that exists among them; classifying one dimension to be more important than the others for a particular application can result in negative consequences on the other dimensions (Scannapieco *et al.*, 2005:10). DAMA-UK (2013:5) are of the view that each dimension is likely to have a unique weighting, and in order for organisations to accurately measure the data quality they need to discover how each data dimension contributes to the data quality as a whole.

2.6.2 Data quality methodologies

For decades, there has been lots of research on data quality (DQ) techniques, frameworks and various methodologies for data quality assessment and improvement. Researchers have developed methodologies and models to assist organisations in managing and enhancing the quality of the data (Günthera *et al.*, 2019; ISO/IEC, 2008:584; Vaziri & Mohsenzadeh, 2012:65; Wang, 1998:60). To ensure high quality data, it is imperative that large and small-medium enterprises have processes, methodologies and resources to aid with monitoring, analysing and maintaining the quality of data (Li *et al.*, 2011:140).

According to Vaziri and Mohsenzadeh (2012:55), data quality methodology is defined as a set of techniques and guidelines designed for the assessment and improvement of data quality in a particular application or organisation. ISO/IEC (2008) defines a data quality model as "*a set of characteristics which provide a framework for specifying data quality requirements and evaluating* *data quality*". There are two types of techniques that organisations can adopt to improve their data quality, namely (Jaya *et al.*, 2017:300):

- Process driven – refers to a strategy that redesigns the processes responsible for producing and modifying data in order to improve its quality. A process driven technique has two main techniques: process redesign and process control. Process redesign ensures that new processes will be added in order to produce high quality when all contributors of poor data quality are eradicated. In the process control, data will be checked and managed among the manufacturing processes.
- Data driven – refers to a strategy applied to enhance the quality of data through directly altering the data value. It involves data improvement techniques such as data and schema integration, error localization and correction, source trustworthiness, standardization or normalisation, acquisition of new data, record linkage as well as cost optimisation (Jaya *et al.*, 2017:300).

Despite the significance of data quality and various methodologies for data quality assessment, not all organisations take data into consideration in the decision-making processes (Günthera *et al.*, 2019:583). It is imperative for organisations to evaluate and measure data quality by means of well-founded metrics and methodologies (Heinrich *et al.*, 2018a:1). According to ISO/IEC (2008), it is crucial for organisations to manage and improve the quality of data because of:

- the need for processing data which are not instantly re-usable because of systematic ambiguity;
- accession of data from organisations of which quality of data production process is undisclosed or weak;
- data scattered among various users and owners;
- the existence of flawed data contributing to unsatisfactory information, worthless results and unsatisfied customers;
- coexistence of legacy architectures and computer systems with distributed systems, designed and obtained at different times with different standards; and
- the existence of information systems where data changes regularly (ISO/IEC, 2008).

The earliest work in data quality by Wang and Strong (1996:6) proposed a preliminary framework for data quality which addressed the followings aspects: the data consumer must be able to access the data, interpret data, find accurate data from reputable sources, and find the data relevant and timely to be used in decision-making processes.

Total data quality methodology (TDQM) was introduced in the 1990s to present the concepts of TDQM cycle and information products (Wang, 1998:60). According to Wang (1998:60), TDQM methodology was developed to simplify the implementation of an organisation's overall data quality policy formally as communicated by top management and deliver a high-quality information product (IP) to information consumers. Vaziri and Mohsenzadeh (2012:65) agree that TDQM methodology views data as a product entity. Figure 2-4 represents the TDQM cycle that involves four process steps: defining, measuring, analysing and improving information quality Wang (1998:60), similarly to the Deming cycle that follows plan, do, check and act stages (Vaziri & Mohsenzadeh, 2012:65).

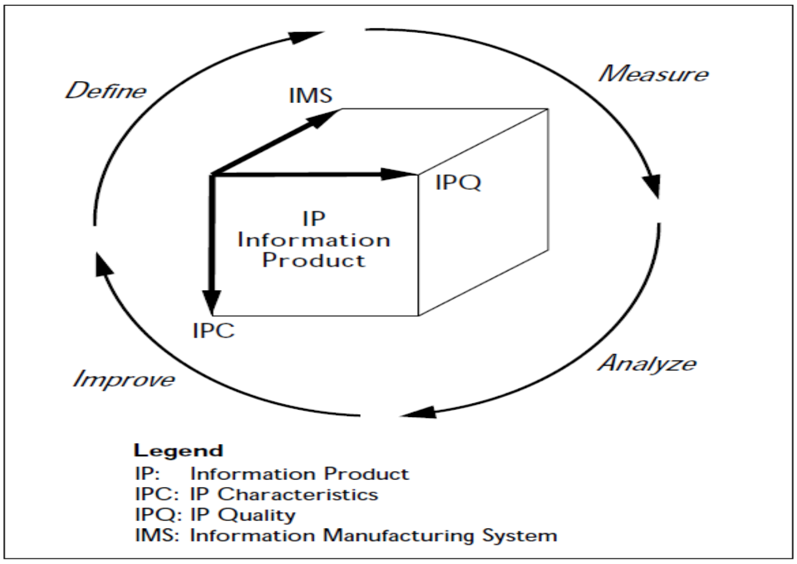


Figure 2-4: A schematic of the TDQM methodology (Wang *et al.,* 1998:60).

To apply the TDQM methodology, organisations must firstly define the characteristics of the information product (IP), then measure information quality (IQ) using IQ measures such as accuracy, timeliness, completeness and consistency (Wang *et al.*, 1998:64). Wang *et al.* (1998:64) emphasise that the key to measuring IP is to develop information quality metrics. Once the measurement results have been obtained, the next step is to analyse IP which requires the IP team to investigate the root cause of current information quality issues. Once the analysis phase is completed, the IP improvement phase ensues, which require the IP team to determine key areas for improvements. In agreement, Heinrich *et al.* (2018a:2) developed five requirements for data quality metrics:

- **Requirement 1 –** the minimum and maximum metric values must exist;
- **Requirement 2 –** refers to the required interval-scaled metric values;
- **Requirement 3 –** the determination of the metric values and quality of the configuration parameters are needed;

- ❖ **Requirement 4 –** it is crucial to ensure that sound aggregation of metric values are achieved; and
- ❖ **Requirement 5 –** It is essential to ensure that the economic efficiency of the metric is attained.

Decision-making under economically-oriented management of data quality and uncertainty can be supported by applying these requirements. In fact, applying data quality measures may enhance the level of data quality and thus bring benefits to the organisation (Heinrich *et al.*, 2018a:7). Wang *et al.* (1998:95) further emphasised that organisations should treat information as a product by following four basic principles; first, organisations need to understand thoroughly the consumers' needs, second, maintain information as the product of a well-defined production process. Third, organisations must maintain information as a product through a life cycle and last, organisations must appoint an information product manager (IPM) to manage the information processes and resulting product. Wang *et al.* (1998:101) advocate that the primary goal of treating information-as-product approach is to provide quality information to the consumer.

Heinrich *et al.* (2018a:1) agree that in order to measure quality of information product is to develop information quality metrics. Furthermore, a study by Williams and Tang (2020:29) present a data quality methodology that can be implemented by large enterprises and SMEs to manage the quality of data. The methodology follows four core activities:

- ❖ State reconstruction and data profiling
- ❖ Data measurement and assessment
- ❖ Data cleansing
- ❖ Data quality monitoring

According to Günthera *et al.* (2019:584), data acquisition is hard and master data management is very challenging in sheet metal manufacturing (SMM), as operations are influenced by various aspects such as sheet thickness, material type, the estimation of processing times is burdensome, and cutting edge requirements. Günthera *et al.* (2019:583) denote that experience alone is not sufficient in making production planning and control (PPC) decisions, but needs to be supported by high data quality. Günthera *et al.* (2019:583) propose a general-purpose methodology for DQ assessment applicable to SMEs, which is expressed in the context of production planning and control in sheet manufacturing (SMM). The methodology chooses context-related data and evaluate DQ with a set of generic metrics that can be modified to different tasks and domains (Günthera *et al.*, 2019:583). The methodology was successfully implemented to assess DQ of manufacturing execution system (MES) data of various sheet metal manufacturers and of enterprise resource planning (ERP) (Günthera *et al.*, 2019:583). Research work by Vaziri and

Mohsenzadeh (2012:60) propose a questionnaire-based data quality methodology using relevant dimensions. The methodology is of the assumption that appropriate dimensions have been identified for the particular organisation and it uses three groups of participants: information professionals (IPs), independent experts (IEs) and information consumers (ICs) to answer a questionnaire associated with the significance of the popular dimensions in the current organisation.

Questions cover all the appropriate dimensions and ask the participant to rate each dimension from scale 0 to 10, where 0 is rated as “Not True” and 10 as “Completely True” (Vaziri & Mohsenzadeh, 2012:60). Vaziri and Mohsenzadeh (2012:60) propose four questions for each single dimension. For instance, when testing for data “**accessibility**”:

Table 2-1: Four questions for each dimension (Vaziri & Mohsenzadeh, 2012:60).

Type of question	Question when using dimension. e.g., **“accessibility”**
1. Direct question	
Are the data in your organisation [dimension name]?	E.g., are the data in your organisation accessible?
2. Reverse question	
Are the data in your organisation [opposite of dimension]?	E.g., are the data in your organisation inaccessible?
3. Synonym question	
Are the data in your organisation [dimension synonym]?	E.g., are the data in your organisation reachable?
4. Definition question	
Are the data in your organisation [definition of dimension]?	E.g., are the data in your organisation accessible?

Once the appropriate dimensions have been measured, the next task is to identify a suitable method to improve data quality (Vaziri & Mohsenzadeh, 2012:61). Vaziri and Mohsenzadeh (2012:61) emphasise that the TDQM cycle can be easily applied to their questionnaire-based methodology despite that the TDQM cycle was not followed when devising the questionnaire-based methodology. According to Sousa *et al.* (2012:798), the quality management methodology adopted by large enterprises cannot be implemented in SMEs with the expectation of obtaining great outcomes due to their distinctive features. Thus, organisations need to continually enhance their processes by implementing adequate tools and methodologies (Sousa *et al.*, 2012:794).

Ali *et al.* (2020:2) developed a framework that can be adopted to assist with improving data quality by following the four phases of data cleansing: data quality assessment, improving data quality, data cleansing using the ETL (extract, transform and load) process, and the performance evaluation process. This framework can be adopted by organisations to assist with improving data quality in one SA SME. In the data quality assessment phase, factors related to the safety and quality of data such as data management, data quality check, data security and integrity, as well as data efficiency and effectiveness are reviewed. For the data quality check, dimensions such as accuracy, completeness, timeliness, consistent, uniqueness and validity will be used to measure the quality data by following the four questions outlined by Vaziri and Mohsenzadeh (2012:60).

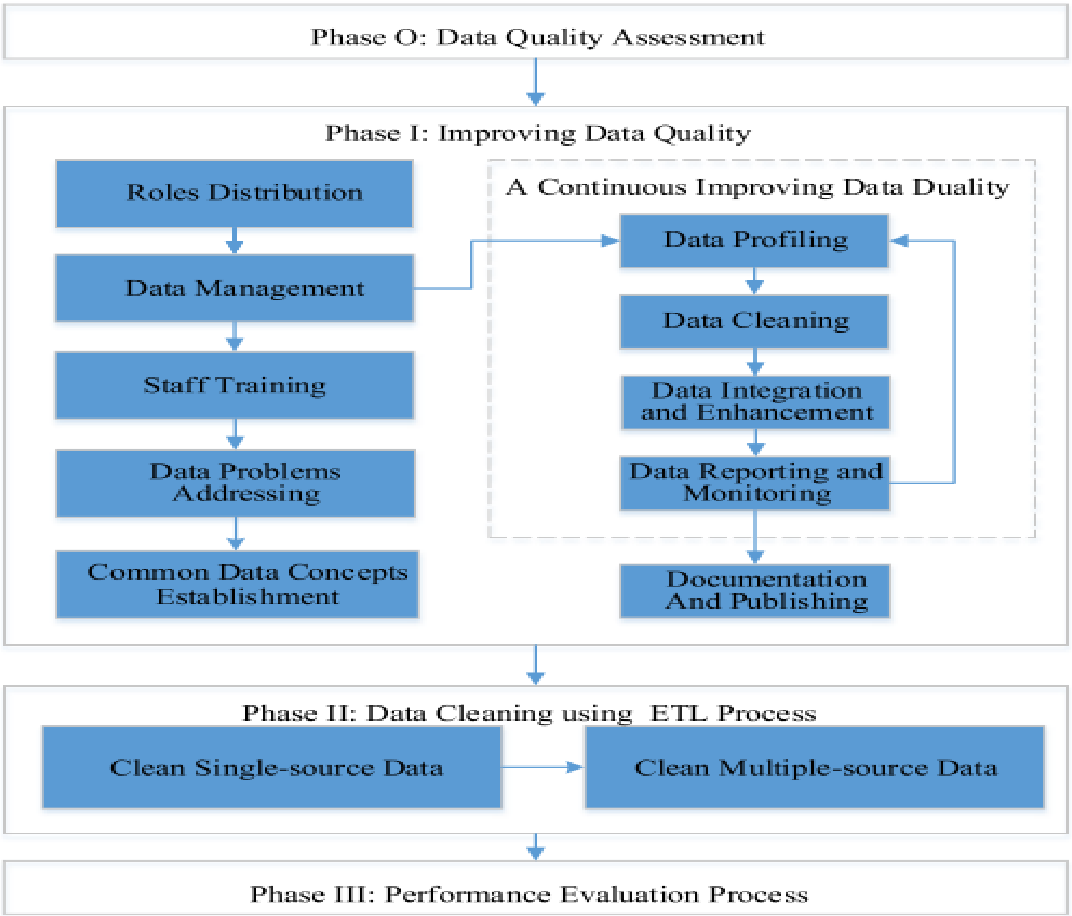


Figure 2-5: The phases of the data cleansing framework (Ali *et al.*, 2020:2).

The researcher will follow a process driven strategy ((Jaya *et al.*, 2017:300) because the research study requires the researcher to redesign the processes responsible for producing and modifying data in order to improve its quality, by developing an artefact (user interface) that will assist with capturing data. In phase II, data will be cleansed using the ETL process and loaded into the data warehouse which will be used as a cleaned single source of data for analytics.

2.6.3 Big data

The concept of big data can be acknowledged as a revolutionary process in the field of IT and has gained recognition among scholars, academics and business communities in an endeavour to identify how it may be leveraged to develop business opportunities, innovation solutions, and gain a competitive advantage (Del Vecchio *et al.*, 2018:6). The emergence of big data offers both challenges and opportunities for data quality research (Gudivada *et al.*, 2017:17).

In fact, big data is regarded as a fundamental economic asset to achieve big value (Kalan & Unalir, 2016:5). Organisations such as Amazon, Netflix, Facebook, Google, Apple, etc. have developed their own business models and big data strategies to assist with leveraging big data. Big data are defined based on four characteristics, often called the four Vs, of big data: volume (scale of data), velocity (high speed at which data are accumulated), variety (the certainty and credibility of the data) and veracity (diversity of data, e.g., structured and unstructured) (Heinrich *et al.*, 2018a:6; IBM, 2020).

With the appropriate strategy, technology, resources, business talent and management, SMEs can utilise big data to understand today and plan tomorrow (Kalan & Unalir, 2016:6). According to Kalan and Unalir (2016:6), large and start-up enterprises are seeking ways to monetize their own big data with the belief to obtain new revenue by focusing on the four Vs:

- ❖ Volume-based value – leveraging cloud computing will aid SMEs with accessing enormous data in a cost-effective manner, which can additionally provide the capability of in-depth analysis of latest and several years of historical context behind the data.
- ❖ Velocity-based value – SMEs need rapid analytics to handle the dynamic nature of big data and make the informed decisions.
- ❖ Variety-based value – SMEs need to be mindful about customer behaviour and expectations to achieve value by analysing diverse data.
- ❖ Veracity-based value – inaccurate data leads to incorrect results, therefore, data accuracy has a huge impact in decision-making.

Bianchini and Michalkova (2019:7) advocate that big data can transition how organisations operate by enabling them to gather accurate information about customers, competitors and suppliers, and additionally utilise this information to make informed strategic decisions. Large enterprises have resources and capabilities to analyse big data; however, one challenge that many SMEs are facing is that they may not have the same capacity to handle big data. Additionally, Kalan and Unalir (2016:6) emphasise that due to the volume, veracity, velocity and variety of data, traditional data processing tools are not usable for SME data storage, management and visualisation. Managing large volumes of data in large enterprises can be cumbersome and complex, but it can be very overwhelming for SMEs lacking appropriate IT resources (Smithers, 2013:44). Data quality is also critical in the big data era. Since data quality has been explored focusing on data as represented in the relational model and then traditionally stored in the database, it is evident that the traditional tools were not designed for massive volumes and the variety of data that we face today (Kalan & Unalir, 2016:5).

An online poll led by Harris Interactive on behalf of SAP discovered that 76% of SMEs view big data as an opportunity for growth (Smithers, 2013:44). However, due to the shortage of IT resources and staff in place, storing big data within traditional databases can be challenging, costly and it can destroy the organisation infrastructure and revenue (Smithers, 2013:44).

SMEs must leverage big data to achieve business growth for improved business decision making, to increase business efficiency and capitalise on new opportunities (Smithers, 2013:45). SMEs with the correct big data strategy in place can analyse both the unstructured data and structured data to realise benefits that are crucial to business development (Smithers, 2013:44). In today's business, SMEs that are capable of accessing the right information efficiently and accuracy are enabled to respond with the correct answers at the right time (Smithers, 2013:44). It is crucial for today's SMEs to develop a data strategy to assist with managing data growth, identify new opportunities and increase business efficiency (Smithers, 2013:44). Kalan and Unalir (2016:1) in their study about leveraging big data technology for SMEs, discovered that since SMEs do not have the resources, skills and technology to manage big data, they can use public cloud and open-source tools to manage large amounts of data (Kalan & Unalir, 2016:6).

However, Kalan and Unalir (2016:6) states that SMEs need relevant skills to take advantage of cloud-based technologies and apply three levels of business models, such as:

- Data as a service (Daas) – which enables customers to mine their own data;
- Information as a Service (Iaas) – which provides insight based on the analysis of processed data; and
- Answers-as-a-Service (AaaS) – which assists with providing higher-level answers to particular questions

There is a growing realisation that within big data resides a goldmine of information that can enhance organisational income, assist to identify future customers, and provide better service to current customers (Smithers, 2013:44). With the right technology that can provide real-time analytics, teams can process both historical data and in addition create predictive models to assist with making strategic, immediate decisions (Smithers, 2013:44). Such technology can enable SMEs to discover new sources of revenue to drive business growth further (Smithers, 2013:45). Many SMEs are acknowledging that big data is a field they must keep close watch on; however, they need to clearly understand the impact of data to their business on a day-to-day basis before exploring big data solutions (Smithers, 2013:45).

The SME involved in this study should not be left behind in exploiting big data and gaining benefits such as improved business decision making, to increase business efficiency and capitalise on new opportunities and business growth. Cloud-based and open-source big data tools for SMEs are available and can enable organisations to store, manage, improve data quality, and perform data analytics.

2.6.4 Poor data quality

In the information age, organisations need to manage the quality of data with the same attention to detail and diligence they allocate to managing finances, as well as have a clear understanding of how data and analytics affect their strategic objective (Bianchini & Michalkova, 2019:11; Eckerson, 2002:8; Forbes Insight, 2017:17). Despite the existence of wide-ranging studies on data quality, data quality remains a critical issue in many companies as they are still struggling to improve and manage the quality of their data (Jesiļevska, 2017:170; Mullins, 2017:54). Haug *et al.* (2011:172) allude that as the data volumes are increasing, so will the complexity in managing the data. In order for enterprises to understand the significance of data quality, they need to understand the impacts and implications of poor quality data (Mahanti, 2018:28).

However, organisations continue to overestimate their data quality and underestimate the impact that inconsistencies and errors can have on their income, and constitute a critical cost factor for many organisations (Haug *et al.*, 2011:170; Jesiļevska, 2017:170; Moore, 2018; Pitney Bowes, 2015:3). In this era, organisations cannot overlook or excuse negligent handling of data (Kimball *et al.*, 2008:323). "*Garbage in, garbage out*" (GIGO) is a popular belief that is commonly used in information systems (IS) and data quality management (DQM) to emphasise the impact and cost of poor data quality in organisations. The notion simply indicates that capturing inaccurate data into the system leads to poor outputs.

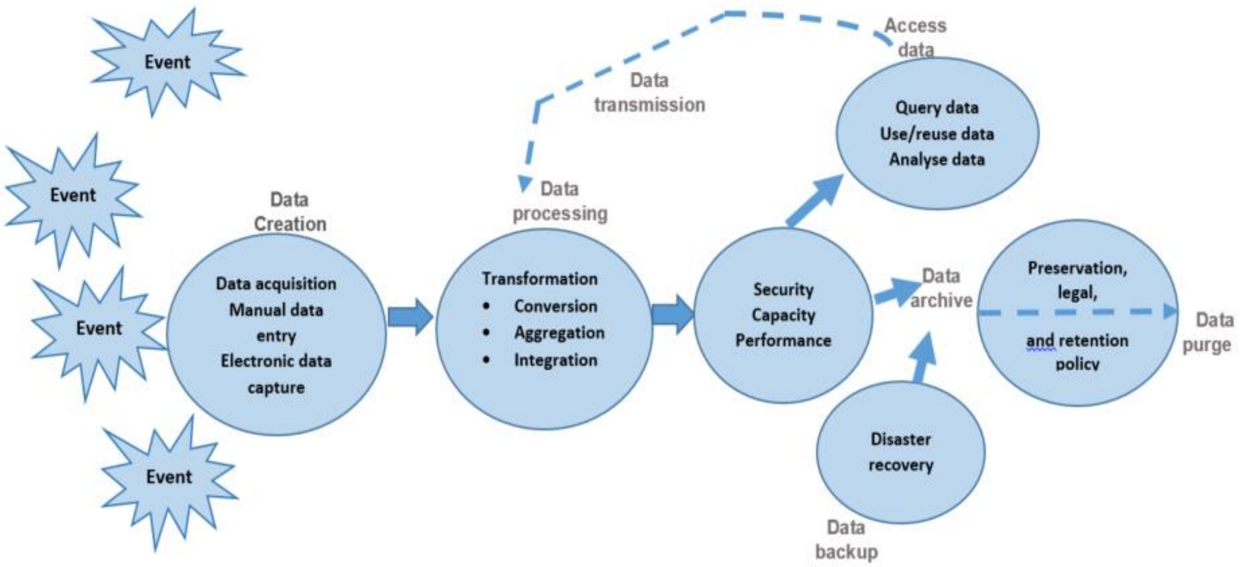


Figure 2-6: The data life cycle (Mahanti, 2018:12).

Data quality issues can creep into every phase of the data life cycle, from data creation, processing, storage, through archive to purging, as outlined in Figure 2-6 (Mahanti, 2018:12). Therefore, senior management must acknowledge existing data quality issues and address them as early as possible before they spread to other systems within the organisation (Mahanti, 2018:403; Mullins, 2017:55). However, once data are captured incorrectly in the data creation stage, they need to undergo cleansing, conversion, transformation and integration before they could be stored in databases or data warehouses. Russom (2006), Ali *et al*. (2020:5) and Mahanti (2018:15) agree that data entry by employees is the major cause of poor quality data. In agreement, Singh and Singh (2010:42) state that data quality can be compromised depending on how the data are received, captured, integrated, maintained, processed and loaded. If organisations do not have validations and checks in the data capturing system that prohibits human error, then organisational data will not be free from defects (Vosburg & Kumar, 2001:22). Furthermore, organisations must adopt formal data governance procedures and adopt a data stewardship approach where a certain individual within the organisation is assigned the responsibility to care of organisational data (Mahanti, 2018:431). However, Mahanti (2018:388) argues that "*aiming for 100% clean data is a mistake*" because trying to achieve data that are completely free of defects is a costly endeavour and is hardly attainable.

Mahanti (2018:15) identified the following as causes of poor data quality: manual data entry, data integration and migration, lack of common standards, inadequate validation in the data capturing process, aging of data, multiple uses of data and lack of shared understanding, business data ownership and governance issues, inefficient business process management and design, organisational changes, system upgrades, data purging and data corruption by hackers.

Disparate departmental, individual and organisational data stores that were utilised by organisational users over the years lead to poor data quality (Vosburg & Kumar, 2001:21). Additionally, Informatica (2006) emphasises that fragmented data systems and distributed IT systems can cause data duplication and a lack of conformity across systems and other discrepancies.

Organisations end up having duplicated data and various versions of data stored in various locations. On the other hand, data from different sources (Excel files, flat files, ODBC connection to source databases) may require manual consolidation of multiple files, which can compromise the quality of data (Singh & Singh, 2010). Various data silos in organisations contribute to data discrepancies, producing multiple versions of the data. Furthermore, dissimilar data structures for the same customer (multiple account numbers, spelling discrepancies, and address variations), lack of legacy data standards, incomplete or missing data, duplicate data, use of free form fields and actual data values not matching meta-labels lead to poor data quality (Vosburg & Kumar, 2001:21).

Informatica (2006) emphasises that poor data quality in large and small-medium enterprises is caused by:

- Inadequate data handling and procedures;
- External and third-party data that may not be suitable for your business data standards or may be of indefinite quality;
- It is imperative for organisations to follow data entry and maintenance procedures. Not observing such procedures may result in compromising the quality of data;
- Utilising data gathered for one specific application in other business processes and systems and;
- Errors originating when migrating data from one system to another.

Large volumes of data that are stored and utilised by organisations can contribute to poor data quality. In some cases, you find that organisations "*are data rich but information poor*" (Bernus & Noran, 2017:1). Today, organisations seem to be complaining that they have spent years compiling excessive amounts of data to support their operations but they feel that they are not getting any return on their investment (Obeidat *et al.*, 2015:29). They are rich in data but poor in information. This simply emphasises that organisations have large volumes of historical and current data, but lack processes and technologies to transform the data into meaningful information to produce insight that can enable them to make informed decisions, gain competitive advantages, innovate, understand customer behaviour, support organisational activities, minimise operational costs and increase revenue, and measure business performance.

Data quality issues occur when there are inconsistencies between the expected meaning of the data value according to its producer and interpreted meaning of a data value according to its consumer (Bertossi *et al.*, 2011). Information generated from poor data will produce poor analytics, which in turn will lead to poor business decisions. The insight and conclusions that can be formulated from data will only be as good as the data used to attain them (Forbes Insight, 2017:3). However, when data are found to be of poor quality, they need to be transferred to a data processing phase where they will be cleansed and transformed to meet acceptable standards. This can be a costly and time-consuming process. Poor data quality increases operational costs as resources and time are spent detecting and fixing errors (Haug *et al.*, 2011:173). It is therefore critical to fix errors at the early stages of data acquisition.

Loshin (2011:5) and Mahanti (2018:31) describe four categories that can be used to assess the degree to which poor data quality impacts businesses as: financial, productivity, risk and compliance, and confidence and satisfaction. Twenty years ago the data quality expert Redman (1998:80) have already discovered that poor data quality damages employee morale, leads to less effective decision-making, breeds organisational mistrust, makes it more burdensome to align the enterprise, and reduces the ability to make and execute strategy. Eckerson (2002:7) agrees that poor data quality can put an organisation at a competitive disadvantage. Pitney Bowes (2015:4) and Haug *et al.* (2011:173) argue that poor data quality can damage business reputation, hinder important initiatives, derail performance, delay critical reports, cause mistrust in company data, damage valuable customer relationships, and put the entire organisation at risk. Poor data quality exterminates business value. Moore (2018) and Kalan and Unalir (2016:5) denotes that poor data quality undermines business strategies.

Poor data quality can have a significantly negative impact on the productivity of a business (Jesiļevska, 2017:168). It also leads to poor analytics, which in return yields poor business decisions. Data quality impedes large and small-medium enterprises from obtaining the best value from their data (Heinrich *et al.*, 2018b:2). According to KPMG's 2016 Global CEO Outlook report, 84% of CEOs are concerned about the quality of the data they are basing their decisions on (KPMG, 2016:42). At the Gartner data and analytics summit of 2018 in Frankfurt, Ted Friedman emphasised that as organisations are accelerating their digital business efforts, poor data quality still persist to be a key contributor to crisis in business value and information trust which leads to poor financial performance (Moore, 2018).

Mullins (2017:54) confirms that data quality proceeds to be a pervasive issue. Poor data quality can cost organisations lots of money. A report on the return of investment of data quality by Pitney Bowes (2015:4) indicated that inaccurate data costs the US businesses $700 billion every year. Moreover, IBM's big data and analytics hub 2016 report indicated that poor data quality costs the

US economy $3,1 trillion per year (IBM, 2016). Gartner's 2017 data quality market survey discovered that poor data quality costs organisations $15 million (Moore, 2018). A study by Jesiļevska (2017:170) indicates that poor data quality can cause organisations to incur costs when cleaning and ensuring high master data quality, and also incur costs resulting from faulty managerial decisions that were based on data that are not cleaned. Irrespective of the size of the business, poor data quality persists to be a hurdle in the digital age (Forbes Insight, 2017:2; KPMG, 2016:24).

2.7 Data management

Due to enormous data volumes and the tedious and vulnerable nature of the data quality tasks, it is essential for organisations to have tools for cleaning, transforming, integrating and aggregating data, as well as assessing and monitoring the data quality. Organisations across various sectors are seeking ways to manage and improve data quality (Gudivada *et al.*, 2017:15). Organisations of any size can implement systems or technologies that can assist with enhancing and managing data. SMEs can choose to develop customised data management tools in-house if they have the appropriate skills and resources, or they can use open-source technologies or purchase commercial data quality and management tools.

According to Mahanti (2018:317) and Williams and Tang (2020:29), data quality management (DQM) is the management of people, technology, policies, processes, standards and data within an organisation, with the objective of improving the dimensions of data quality that are most crucial to the organisation. According to Mahanti (2018:317), organisations can use the data quality improvement program driven with the Six Sigma approach to manage and improve data quality.

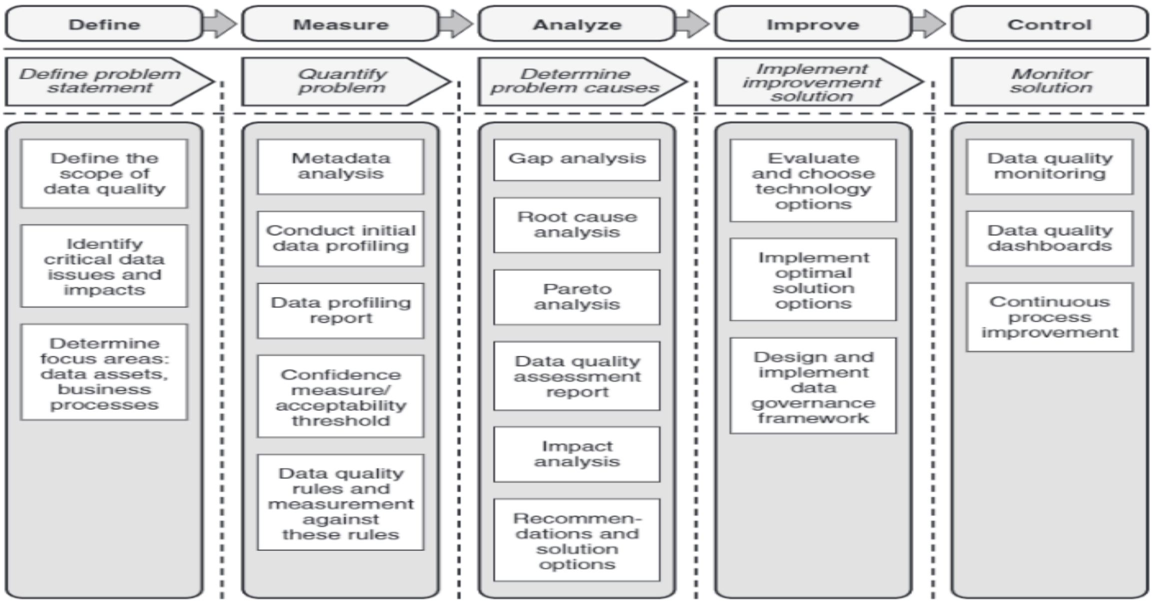


Figure 2-7: Data quality management using DMAIC (Mahanti, 2018:324).

As outlined in Figure 2-7, the Six Sigma DMAIC follows five phases (define, measure, analyse, improve and control) to improve data quality (Mahanti, 2018:324). SMEs can adopt this approach to assist with defining and identifying critical data quality issues, quantifying data quality issues, determining root causes of data quality issues, evaluating and choosing technology or tools for improving the issues, and provide continuous monitoring of process improvement.

Grabova *et al.* (2010:1) mention that SMEs need cheap, flexible, simple, lightweight and efficient solutions to manage data quality, and as a result it is not advisable for SMEs to adopt and use tools destined for large-scaled enterprises because the two sectors are considerably different. SMEs need cheap, flexible, simple, lightweight and efficient solutions. However, Grabova *et al.* (2010:4) and Moyo and Loock (2016:251) state that there are existing cloud-based BI tools that are appropriate for SMEs in terms of price and flexibility, and dominant software vendors such IBM, SAS, Informatica, Oracle, Trillium Software, Talend and Information Builders which offer data quality and management tools (Gudivada *et al.*, 2017:13).

In big data concepts, traditional database management systems are incapable of handling large data volumes, and as a result organisations can use Apache Hadoop for storing and analysing unstructured data and support low-cost, reliable, distributed programming because it is an open source code used for calculating big data analytics in an easily scalable medium (Razbonyalı & Güvenoğlu, 2016:2560).

Proactive and reactive data quality management approaches can be implemented to manage data quality (Mahanti, 2018:319). Proactive data quality management approaches are used for eliminating data quality issues before they surface, while reactive approaches involve responding to data quality issues after they have surfaced (Mahanti, 2018:319).

Ehrlinger *et al.* (2019:4) conducted a survey to evaluate existing data quality tools with respect to their DQ measurement and monitoring functionalities. Ehrlinger *et al.* (2019:4) identified 667 software tools dedicated to data quality, from which 13 tools were evaluated in their study focusing on three functionality areas: data quality measurement in terms of metrics, data profiling and continuous data quality monitoring. Tools such as IBM InfoSphere information Server for Data Quality, Microsoft BI suite, Informatica Data Quality and Talend Open Studio for data quality can be implemented by SMEs to improve data quality, management and analytics.

Based on the existing IT infrastructure in one SA SME, Microsoft BI suite will be used to improve data quality. Microsoft BI suite is the appropriate solution because the SME is currently using a Microsoft product and one of their databases is using Microsoft SQL server 2017; Microsoft BI products cover the entire stack of BI components which include SQL Server Management Studio (SSMS) and SQL Server Data Tools (SSDT) (Raj *et al.*, 2016:2); studies by Raj *et al.* (2016:2) have shown that SMEs can use Microsoft BI suite to improve data quality and enable SMEs to maintain a competitive advantage.

Furthermore, SMEs can use interfaces, embedded with validations and constraints to capture data, and control the quality of data that is accepted into the database. However, Ali *et al.* (2020:5) emphasise that user interfaces must satisfy a set of conditions:

- **Easy entry –** the ability of the user to use the data entry forms to capture data efficiently without wasting time trying to figure out how it works and mobilising it.
- **Effectiveness –** Ensuring that the electronic and paper input are able to accomplish the task for which they were effectively prepared.
- **Data accuracy –** data entry forms can be filled out correctly.
- **Regularity –** The interface elements design must be kept in order to ensure tracing of user attention.
- **Gravity –** the user must be given a feeling of pleasure and comfort when using the model.

As the recent generation user interfaces are developed to receive and reject data values, the interfaces should also allow specifying a degree of confidence in inclusion to the value itself (Gudivada *et al.*, 2017:14). According to Mahanti (2018:344), data validation checks embedded in user interfaces and databases assist with revealing certain quality problems, and one of the following four options can be chosen depending on the type of problem:

- ❖ **Accept the data –** Sometimes, the best option can be to accept the data with the errors only if the errors are within acceptable limits.
- ❖ **Reject the data –** The best approach is to reject the data if the data issues or errors are deemed severe. The rejected data can be captured in a reject file which would later be analysed to determine a course of action.
- ❖ **Correct the data –** When various versions of a customer name are presented, one of them could be selected to be the master in order for the data to be integrated.
- ❖ **Insert a default value –** Default values such as "*unknown*" can be captured in a field when we are unsure of the correct value.

One SA SME needs to take note that data quality management is not a once-off process, it needs a defined strategy that is carried out continuously, as well as suitable tools to implement and automate the strategy (Ehrlinger *et al.*, 2019:1). Organisations must create enterprise-wide data quality awareness programs and data quality training (Mahanti, 2018:407). Furthermore, Mahanti (2018:380) describes the key principles that organisations can apply to manage data quality:

- ❖ Organisations need to focus on important data elements when cleaning, assessing, or monitoring data quality;
- ❖ Apply data validation rules and strong data quality controls on data capturing systems/applications;
- ❖ If possible, enterprise-wide data standards must be established;
- ❖ A centralised data quality issue log must be used to capture, prioritise and maintain data quality issues;
- ❖ Data governance frameworks must be established;
- ❖ Rather than attempting to solve all data quality problems, try to solve critical data issues that have maximum impact on the business;
- ❖ Data cleansing processes should not solely be used to solve data quality problems. The primary cause behind data quality issues must be investigated and implement improvement initiatives and preventive measures to stop data quality issues from happening in the first place; and
- ❖ Automate data capturing as much as possible.

Databases

Organisations are implementing databases as data quality management mechanisms. According to Nyaboga and Mwaura (2009:20), most organisations use databases to perform data management. Microsoft SQL Server applications have a database system and other components providing services such as data storage, management, data profiling, administration and security, which SMEs can implement. Microsoft SQL Server has the functionalities and capabilities required for database and data management processes. According to Ramakrishnan and Gehrke (2003:4), a database management system (DBMS) is software designed to assist organisations to maintain and utilise large collections of data. A database "*is a collection of data, typically describing the activities of one or more related organisations*" (Ramakrishnan & Gehrke, 2003:4). Structured Query Language (SQL) is a multifunctional language developed for the management of a relational database, as well as to enable developers to define data structure and to create, modify, and delete data (Denchev, 2017:9). Additionally it assists with restricting access to database elements, data management operations such as block copying and updating, and creates database backups and transaction processing (Denchev, 2017:9).

SMEs can use databases to improve data quality. According to Mullins (2017:55), database administrators (DBA) can implement constraints in databases as a mechanism to improve quality. Databases make use of constraints to self-enforce data quality to a particular extent (Mullins, 2017:55). Triggers and check constraints can enforce business rules on data components. Furthermore, referential integrity constraints can enforce primary foreign key relationships to ensure that a foreign key cannot be entered while the parent primary key does not exist. The DBA can also implement a not null constraint to ensure that critical information is captured into the database and that the data is complete. Unique constraints can be implemented to ensure that duplicate records are not entered into the database. However, Gudivada *et al.* (2017:14) argue that integrity constraints (ICs) are not a solution to prevent bad data, they are only one step of a multi-step process for ensuring data quality.

The best rule of data quality enforcement is to impede any error prior occurring (Nyaboga & Mwaura, 2009:20). Nyaboga and Mwaura (2009:20) propose other methods such as identifying and checking to make sure values are unique, use special values for unknowns to reduce and eliminate confusion, and ensuring validity across systems by using version numbers. Another method that can be used is data quality audits, which requires detecting data errors by listing all types of errors and their frequency of occurrence. Additionally, random sampling is the cheapest method of tracking errors and it can eliminate additional costs. Gudivada *et al.* (2017:14) mention several steps that can be implemented to enforce data quality in databases, which include:

- Check the compatibility of data schema;
- Ensure maintenance of the schema quality;
- Verifying data entry in fields;
- Check data dependencies; and
- Enforce data constraint if known ahead of time.

2.8 Small and Medium-sized Enterprises

All around the world, SMEs have been recognised as the key engine to job creation, poverty alleviation and economic development (Gherghina *et al.*, 2020:1; Kilimis *et al.*, 2019:2140; Llave, 2019:19; Ngek, 2014:253). According to Ghobadian and Gallear (1996:83), SMEs "*are the life blood of modern economies*" and they contribute to entrepreneurship and innovation (Bernardino, 2015:1964). SMEs are essential in achieving industrial and economic development objectives in such a way that they contribute to employment creation, poverty mitigation and the generation of potential entrepreneurs (Nkwe, 2012:33). According to Ramukumba (2016:20), the role of SMEs become more critical in developing countries as they have the potential to reduce poverty, enhance income distribution, facilitate export growth and create new employment. Moreover, SMEs take a substantial portion of all enterprises in any economy and they are regarded as key drivers for economic growth and development (Curraj, 2018:16; Kfouri & Skyrius, 2017:96).

Nowadays, characteristics and definitions of SMEs are based on country or region, annual turnover, employment level, and the annual balance sheet (Nkwe, 2012:30). Researchers, Kfouri and Skyrius (2017:98) in their study on factors influencing the implementation of BI among SMEs in Lebanon, define SMEs as enterprises that employ less than 250 people, whose annual turnover doesn't exceed 50 million EUR, and whose annual balance sheet doesn't total above 43 million EUR. An SA small enterprise is defined as a distinct and separate business entity together with its branches or subsidiaries, if any, including cooperative enterprises classified as micro, a small or a medium enterprise, and managed by one or more owners (Department of small business development, 2019:110). Additionally, the South African small enterprise is predominantly carried on in any sector or subsector of economy mentioned in standard industrial classification and satisfying the total number equivalent to full-time paid employees and the total annual turnover outlined in the schedule (Department of small business development, 2019:110).

Small and medium-sized manufacturing firms are under appalling competitive pressure due to customer requirements and increased global competition (Ghobadian & Gallear, 1996:83; Sahoo & Yadav, 2018:541). Digital technologies are presenting SMEs and entrepreneurs with new opportunities to engage in the global economy, to innovate and grow (Bianchini & Michalkova, 2019:6). In this regard, entrepreneurs must sustain high levels of innovativeness and transform their business models to meet the dynamics of technology (Curraj, 2018; Kfouri & Skyrius, 2017:97). In fact, to survive and prosper, all enterprises need to establish mechanisms allowing them to exercise conscious and sustained effort to consistently improve all facets of their operations (Ghobadian & Gallear, 1996:83).

When SMEs survive and grow, it is an indication that they are responding positively and adapting successfully to the needs of the markets (Ngek, 2014:254). However, prior studies elucidate that not all SMEs can respond positively to globalization, technological complexities, customer requirements and changing markets; SMEs are struggling to adopt new technologies and increase business performance as compared to large enterprises (Bianchini & Michalkova, 2019:7; Curraj, 2018:17; Llave, 2019:19; Raj *et al.*, 2016:41).

Digitalisation is a significant trend reshaping societies and economies (Bianchini & Michalkova, 2019:6). A survey regarding SMEs' digitalisation concepts in the federal states of Brandenburg in Germany focused on fifty SMEs, of which 51% were steel and metal processing companies. Ten of these SMEs made a decision to invest in digitalisation and started implementing the process (Kilimis *et al*., 2019:2142). From the ten SMEs, three were selected for a case study to evaluate trade-offs, potential benefits and barriers hindering the implementation of digitalisation technologies (Kilimis *et al.,* 2019:2142).

According to Kilimis *et al.* (2019:2142), out of the three companies selected for the case study, two SMEs, one was providing construction service and the other metal processing, indicated that they used exchange files, e.g., Excel files, to store business information (logistics, rental lists of devices and equipment, fitter schedules) and technical data (machine specification, tool designs and tool manufacturing plans). These files were maintained manually and made it impossible to update information from sites in real time. Therefore, it was impossible to ensure constant flow of updated data, to provide accurate short-term daily plans or long-term plans, to track availability of equipment and devices, up-to-date data, as well as security. Tracking availability of devices and equipment was time-consuming because there was no database to manage the process. Furthermore, the computer-aided design/computer-aided manufacturing files were updated manually because there was no interface for CAD/CAM to access the ERP.

Kilimis *et al.* (2019:2142) state that four industry 4.0 Internet of Things (IoT), Internet of Service (IoS), smart factory and cyber physical systems (CPS) were presented to assist the SMEs with the digitalisation. The SMEs can implement a database of the machine tools' features to reduce the amount of time spent on selecting tools. Additionally, an integrated ERP/PPS/MES (Enterprise Resource Planning/Physical production System/Manufacturing Execution System) system was implemented to assist with planning and managing the workflow (Kilimis *et al.*, 2019:2142). Today, SMEs are operating in environments with ever-growing complexities that impose various complications spanning environmental, social and technological aspects that significantly impede their success (Kfouri & Skyrius, 2017:97). Furthermore, Curraj (2018:17) emphasises that changes in the environment create more uncertainty in SMEs than in large enterprises.

As compared to large enterprises, SMEs have different production processes, management styles, inventory systems, purchasing practises, capital availability and negotiating power (Gadenne & Sharma, 2005:2). SMEs are vulnerable and not vigorous enough to withstand the aggression of global and economic competition, as they have limited financial resources, internal Information Technology (IT) resources and competencies (Llave, 2019:19). Additionally, Llave (2019:19) states that SMEs differ from large enterprises in terms of culture, structure, management, decision-making, ownership, processes and procedures. Curraj (2018:48) found that a lack of resources, time and know-how are obstacles hindering SMEs from adopting new ways of action and new techniques to develop their operations. Additionally, SMEs have distinctive characteristics, which also depend on specific cultural, political and economic context (Curraj, 2018:16). Table 2-2 below depicts that SMEs, compared to large enterprises, have a small market share, limited resources, lower net income, and limited deployment of ICTs, amongst other things.

Table 2-2: Comparison of features between large enterprises and SME - adopted from (Curraj, 2018:44)

Large Enterprises	**SMEs**
Large firm size	Small scale firms
Ample resources (capital, equity, credit)	Limited resources (internal funding)
Experience, portfolio and specialization	Limited experience/no specialization
Complex processes to manage customer relation	Proximity to customer relations
Higher market share	Small market share
Brand recognition	No brand recognition or very limited
Higher net income growth, lower costs/high efficiency	Lower net income and no economy of scale
Complex production lines/processes	Flexible production and absorption of demands
Very complex in structure and management	Flexibility and agility due to no complex structures
Legitimacy, social ties and networks	No or limited legitimacy and acceptance, no social ties and limited networks
Liability of being too big	Liability of smallness and newness
Slower decision-making and bureaucratic	Faster decision-making
Difficulty to respond to changing external environment	Responsiveness to external environment
Deployment of advanced ICTs to manage business processes	Limited deployment of ICTs, but flexibility of adaptation

In South Africa, the unemployment rate crisis is deepening, and job creation is critical to economic growth and political stability. SMEs are the core pillars of any national economy as they frequently employ more people than large enterprises (Assarlind, 2011:1). SMEs contribute to over 50% of African employment and gross domestic product (GDP), and constitute over 90% of African business operations (Ramukumba, 2016:19). SMEs represent 99% of businesses in the

European Union (EU) and created 85% of new jobs in the last five years (Gherghina *et al.*, 2020:1). In the year 2014, Ramukumba (2016:19) indicated that SMEs constitute 55% of all jobs in South Africa. Gherghina *et al.* (2020:1) state that there were 23 million SMEs in 2015 that produced 90 million jobs, generating a higher benefit of 3,9 billion Euros (EUR). According to the Organisation of Economic Co-operation and Development (OECD), the South African National Development Plan (NDP) estimates that 11 million jobs need to be created by the year 2023, of which 90% of these jobs are expected to come from new and expanding SMEs (OECD, 2020).

Even though SMEs contribute to growth and development, the Department of small business development discovered that South Africa has a low creation rates of successful SMEs; 70% to 80% of small businesses fail in the first year of operating and half of the survivors endure only for the next five years (OECD, 2020). SMEs are operating in unpredictable and dynamic business environments, and not many are able to adapt to changing environments and survive. Additionally, SMEs in South Africa and all over the world are failing due to a number of challenges (Ramukumba, 2016:19). SMEs are facing challenges such as access to finance, lack of knowledge, skills and networks required in a particular industry, access to markets, developing relationships with customers, and appropriate technology (Ramukumba, 2016:25). Ramukumba (2016:33) recommends that managers and owners of SMEs need to maintain technological advantage, and take the time to build social and business networks and find resources. They need to understand the importance of financial and marketing management and skills development.

2.9 Data analytics

Despite the size of the organisation, the ability of organisations to aggregate, elaborate and analyse data is becoming a fundamental competitive advantage and resource that could bring innovation, productivity, quality, efficiency and customer satisfaction (Del Vecchio *et al.*, 2018:9). The KPMG 2016 Global CEO Outlook report indicates that data and analytics is the highest investment for CEOs (KPMG, 2016:19). Furthermore, the 2018 global data management benchmark report indicates that data and analytics ranked highest as the key source to unlocking business opportunities in the coming years (Experian, 2018).

Bianchini and Michalkova (2019:6) advocate that digital technologies offer SMEs and entrepreneurs new opportunities to participate, innovate and grow in the global economy. Data analytics has become a fundamental driver of enterprise competitiveness (Bianchini & Michalkova, 2019:7). Data analytics signify the totality of data-based inference methodology utilised for the sole purpose of analysing, predicting and controlling processes in industry and business (Coleman *et al.*, 2016:1). In fact, high-quality data "*is the key to interpretable and* *trustworthy data analytics and the basis for meaningful data-driven decisions*" (Ehrlinger *et al.*, 2019:1).

A study by Bianchini and Michalkova (2019:11) indicates that data analytics provide SMEs with a wide range of opportunities such as a better understanding of production processes, the needs of partners and clients, and of overall characteristics of national and local markets. Furthermore, in manufacturing industries, data analytics are used by operations managers to analyse historical production data, establish relationships and patterns among distinct processes and inputs, and then enhance the factors that prove to have greatest impact on yield (Bianchini & Michalkova, 2019:11).

2.9.1 Data warehousing and business intelligence

Data Warehousing and Business Intelligence (DW/BI) has evolved to become an essential centrepiece of any enterprise, including SMEs' decision support (Grabova *et al.*, 2010:1; Scholz *et al.*, 2010:1). All over the world, a data warehouse (DW) is regarded as the core of decision support systems used by large and small-medium sized enterprises (Grabova *et al.*, 2010:1). Bernardino (2015:1695) advocates that BI and Decision Support Systems (DSS) can assist SMEs to be competitive in the global world. BI was primarily developed as a system to solve analytical tasks, reduce costs, improve the quality of processes and performance, and additionally it was considered to be a solution for better decision-making processes (Olexová, 2014:96).

The father of data warehouse concepts Inmon (2002:31) defines a data warehouse as a "*subject-oriented, integrated, non-volatile and time-variant collection of data in support of management's decisions*". De Silva (2005:1.1) refers to DW/BI as a management tool that assists executives to gain access to critical information required to make informed business decisions and to establish business strategy for the future. According to Kalan and Unalir (2016:3), BI is an umbrella term for tools, methodologies and applications that enable organisations to gather data from various resources, augment it for analysis, execute the queries against the data, and visualise the results to make informed decisions. Zafary (2020:61) refers to BI as a collection of technologies, tools, abilities and approaches that aid managers with understanding business conditions. For the purpose of this study, DW/BI is defined as tools, applications, and technologies to collect structured and unstructured data from disparate sources, clean and transform the data, store it in a DW for querying, analysing, reporting and decision-making purposes.

Over the past years, large enterprises have been using BI as core strategy for innovation, growth and improving business processes. However, according to Raj *et al.* (2016:43), SMEs misperceive BI as technology that is only applicable to large enterprises, and therefore they are lagging behind in adopting Business Intelligence (BI) to assist in decision-making processes (Raj *et al.*, 2016:6; Scholz *et al.*, 2010:1). Large enterprises are exploiting BI technology to support decision-making processes, whereas SMEs are slow on adopting the technology.

Organisations can follow the high level DW/BI system architecture model to develop in-house DW/BI solutions. As per Figure 2-8, the process involves collecting data from various sources such as flat files, CSV files, ERP systems, relational databases, etc., and transform the data using an ETL tool (Extract, Transform, Load) before loading the data into the data warehousing. Once the data are cleaned and loaded into the DW, the DW will be used as the primary and trusted source of quality information for querying, data analytics and producing standard and dashboard reports.

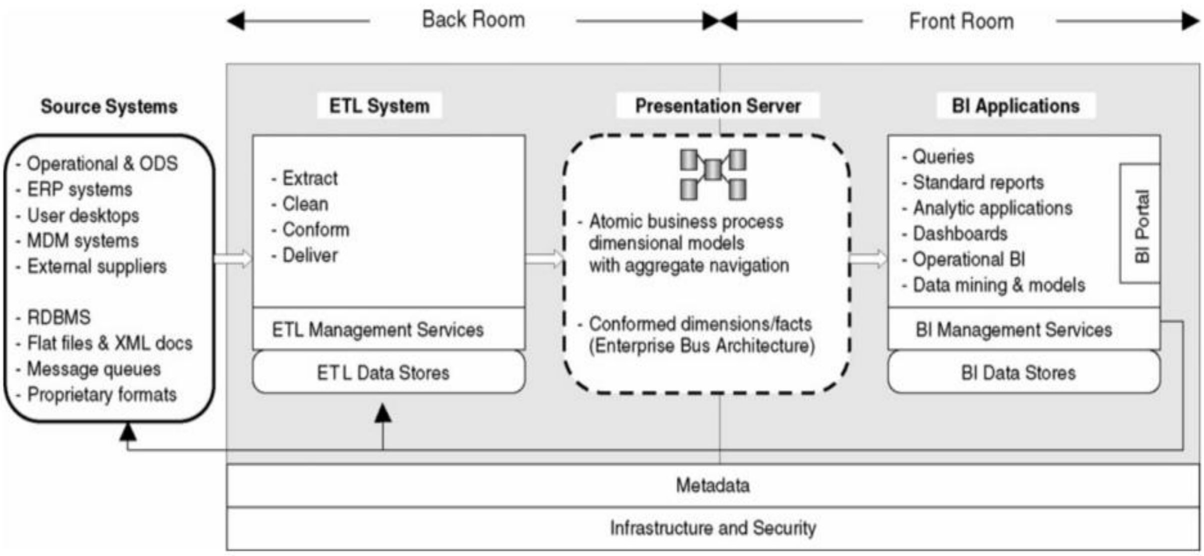


Figure 2-8: High level DW/BI system architecture model (Kimball *et al.*, 2008:101).

SMEs can implement BI technologies to provide improvement in data support through easy access to high quality data, to improve data cleansing and validation processes which leads to reports produced using high-quality data, and provide improvements in decision support by generating rich visuals to support effective and efficient decision-making processes (Raj *et al.*, 2016:44). Furthermore, Raj *et al.* (2016:44) emphasise that a BI solution can assist with identifying risks and rectifying them efficiently, assist with cost and time-saving through using scorecards and dashboards as they enable visuals of data analytics to be interpreted easily, and retrieve data.

Data Warehouse/Business Intelligence (DW/BI) can provide organisations with tangible or intangible benefits (Kimball *et al.,* 2008:28). Tangible benefits such as increased revenue, reduced time to collect data and produce analysis, reduced customer churn, eliminated cost of producing legacy reports, reduced defects rate and reduced time to market new products can be realised (Kimball *et al.*, 2008:29). Furthermore, Kimball *et al.* (2008:29) state that DW/BI can provide intangible benefits such as greater confidence in the data with less debates about accuracy, and eliminate inefficiencies by providing one version of the truth about the data, improved customer satisfaction, improved employee satisfaction and morale, greater information consistency and standardisation, faster response to changing conditions, enhanced ability to analyse alternatives, provide more timely access to information, and enable organisations to remain competitive.

In agreement, BI provides availability of high quality and correct information required for analytics and to support business decisions and operations (Zafary, 2020:60). BI initiatives enable organisations to gain insight from large volumes of data generated by applications such as supply chain management, web analysis and customer relationship management (Bernardino, 2015:1695). DW/BI requires a strong foundation of high quality data; without quality data the project will fail (Kimball *et al.*, 2008:20).

However, in spite of the broad benefits presented by DW/BI technologies, SMEs are struggling to access and adopt these technologies (Bernardino, 2015:1695). Issues such as complexity, high price, redundant functionality, high requirements for a hardware infrastructure, low flexibility to handle fast changing dynamic business environments, and irrelevant functionality make DW/BI frequently inaccessible or insufficient to SMEs (Grabova *et al.*, 2010:9). Most SMEs lack the cost associated with adopting BI, technical know-how and expertise needed to choose and adopt a suitable BI solution, lack of hardware infrastructure, lack of knowledge and understanding of data warehousing and dimensional modelling (Raj *et al.*, 2016:43).

Additionally, SMEs perceive BI technologies as complex solutions to use. Four general factors that can influence the adoption of a new technology by SMEs have been identified as: (1) the characteristics of the organisation, (2) the competitiveness and management strategies of the organisation, (3) the influences of the internal and external parties on the adoption decision process, and (4) the characteristics of new technologies adopted (Scholz *et al.*, 2010:2).

In this research study, Kimball methodology will be followed in designing the DW/BI solution. Dimensional modelling is a widely accepted approach for DW/BI presentation and is referred to as a logical design technique for structuring data to provide fast query performance and is intuitive to business users (Kimball *et al.*, 2008:204).

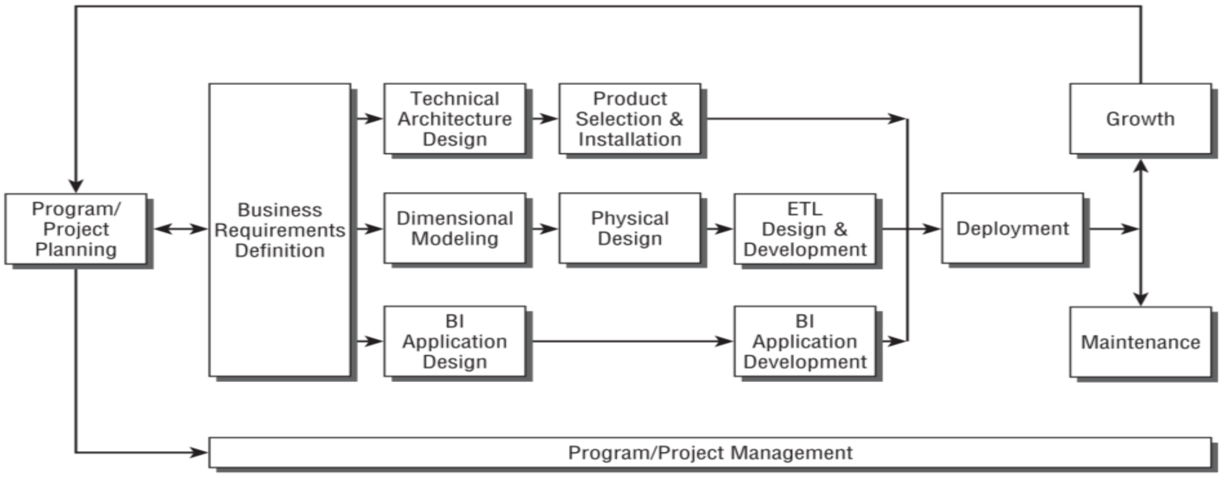


Figure 2-9: Kimball lifecycle approach (Kimball & Ross, 2013:404).

Scholz *et al.* (2010:6) conducted an exploratory study on 214 German SMEs in the state of Saxony who had deployed BI solutions, used the factor and cluster analysis to examine the challenges and benefits of BI adoption in SMEs. Scholz *et al.* (2010:6) identified improvements in decision support, data support and savings in personnel and costs as the general BI benefit factors. Furthermore, challenges associated with BI adoption in SMEs include challenges depending on usage (e.g. BI solution is too complicated), challenges with interfaces, challenges depending data quality and the solution (Scholz *et al.*, 2010:8).

A case study by Olexová (2014:105) on BI adoption in the retail chain identified issues in the organisation, such as poor information sharing and analysis, the use of Excel files by management to acquire information, management information were sometimes changed or delayed due to a lack of relevant data, the need to reduce the costs of stock management and pricing and planning were not efficient due to defects in the market change forecast. After adopting the BI solution, the company was able to acquire up-to-date and better quality information, make faster, better-informed decisions based on high quality information, have an improved ability to anticipate earlier changes on the market, as well as better pricing and stock management optimisation (Olexová, 2014:104). Additionally, it enhanced the value of management in the company and assisted with improving business processes. The key major benefit of adopting BI in the retail chain is improved decision-making (Olexová, 2014:105).

2.9.2 ETL

The ETL process is a three-stage process that involves cleansing and integrating data to improve quality and decision-making processes. The ETL (extract, transform, load) process involves gathering data from disparate data stores, send the source data through a series processing stems to improve the quality and integrity of data, then load it into a data warehouse (Kimball *et* *al.*, 2008). The objective of data integration is to provide a 360-degree view of the enterprise by ensuring that all systems work together effortlessly (Kimball *et al.*, 2008:324). Once the data are consolidated, standardised and summarised, they provide a single version of the truth and hold tremendous potential for reporting (Ram, 2013:28; Vosburg & Kumar, 2001:7).

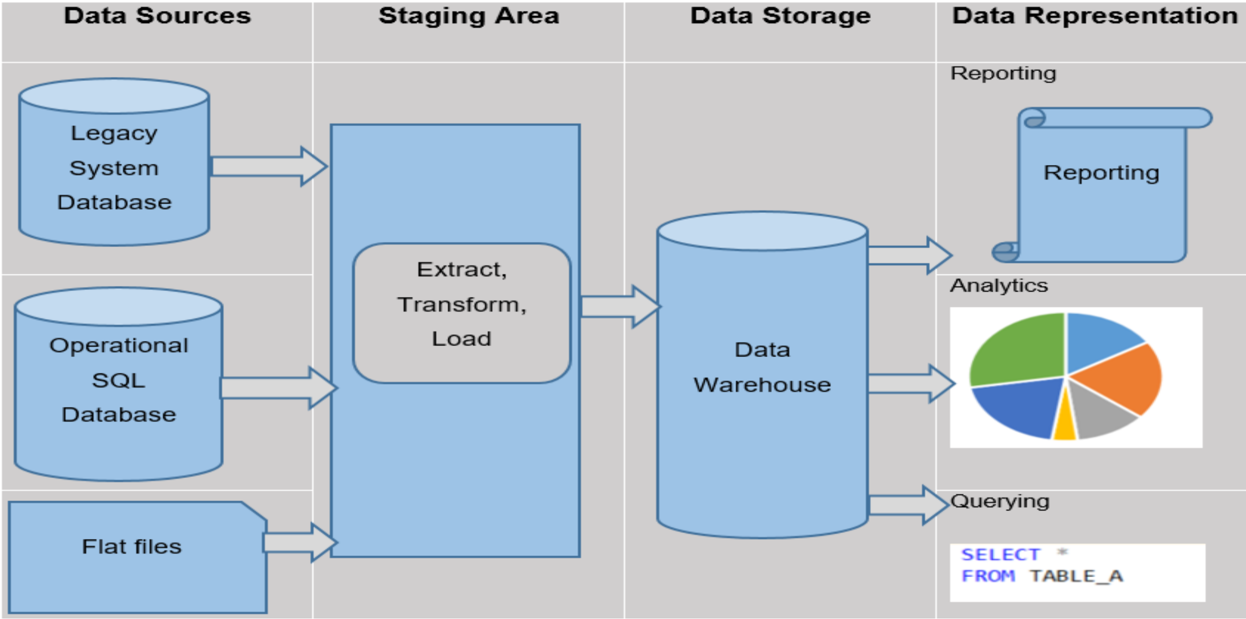


Figure 2-10: ETL Process described by author.

Figure 2-10 above illustrates the ETL process; data are extracted from disparate sources, transformed into standard format before they are loaded into a data warehouse in preparation for analysis, reporting and data mining. The ETL process is relevant to this study because one SA SME has data quality issues and data stored in disparate locations, which present tremendous challenges in data management and analytics.

Data inconsistencies need to be fixed and conversions need to be made before data can be loaded successfully into the data warehouse. Kimball *et al.* (2008:321) state that the ETL process is regarded as the biggest task of building the data warehouse because it is time-consuming and complex. The ETL developer is responsible for developing an end-to-end production process to extract, transform and load data into the data warehouse (Kimball *et al.*, 2008:36). Kimball *et al.* (2008:321) state that the ETL system is the base of the DW/BI project and is often estimated to take up to 70% of the time and effort of developing a DW/BI environment. One SA SME can use the ETL process to clean and transform the data to eliminate any inconsistencies.

2.10 Related Study

A case study by Raj et al. (2016:45) in the UK-based company known as “AGGORA”, specialising in catering equipment, required a solution to improve their ability to manage, analyse and utilise data. They also needed a data management system to manage the quality of their data and to produce reports for the management team to analyse and make business decisions. The company also had legacy data issues with data residing in various locations affecting the efficiency of report generation. To solve all these issues, the leadership team of the company decided to implement a BI solution even though the AGGORA group did not have sufficient expertise in BI technologies.

According to Raj et al. (2016:45), the objectives of the project were to analyse client KPI and satisfaction, and to analyse engineers’ performance and productivity (manage engineers’ time and quality of their work). Additionally, the aim was to analyse financial figures to establish and measure the economic state of the SME and their asset management (analyse historical information about clients and use this information to enhance the service provided to clients).

Data were extracted from legacy systems and from their in-house service management system relational database using an SQL server. The database stored equipment management information together with cost and sales and engineers’ timesheet information. The source data were transformed into a form that would provide intelligence to the users before it was loaded onto the data warehouse (Raj et al., 2016:45).

Furthermore, Raj et al. (2016:45) emphasise that the SME followed a five-stage implementation process, namely planning, data collection, data analysis, data distribution and feedback. To develop the solution, SQL server management studio and SQL server tools were used to develop the BI solution. For data visualisation, Power BI was used to self-serving reports while SQL Server Reporting service was used to generate standard reports. The Company was already using Microsoft SQL Server 2012 for their databases, so Microsoft BI suite was the convenient option to use.

Once the requirement gathering phase was complete and the objectives were in order, the data source was transformed before loaded into the data warehouse. A dimensional modelling technique, using facts and dimensions, was implemented in developing the data warehouse. The Kimball methodology steps for designing a data warehouse were followed (Raj et al. (2016:46): a business process was selected, the grain was declared, relevant dimensions were identified, and facts were identified (Kimball & Ross, 2013:70). The DW followed a hybrid method which is a combination of a star schema and snowflake in order to maintain data consistency.

To improve the quality of data, the ETL (extract, transform, and load) process was used to extract data from the source systems, transforming it into a desired format before it was loaded into a DW. The ETL tool from Microsoft SQL Server Data Tools was used for data transformations. Once the data were loaded into the DW, Power BI was used in this case study to connect to the DW and perform data analysis and data visualisation.

Business intelligence was exhibited to the end users as the end product of data analysis. Data analysis results were distributed to company managers through a cloud-based PowerBI solution. Furthermore, users could generate and deploy reports and dashboards and share them with other users for efficiently. Power BI is easy to use, so with some basic training and without any technical assistance, users were capable of producing the required business intelligence. Power BI produces rich data visualisations and support an intuitive process for on-demand reports (Raj et al., 2016:48). In order to generate standard reports, Microsoft SQL Server Reporting Services (SSRS) was used. Standard KPI reports such as Engineering performance and productivity were created using SSRS.

Feedback was collected from users at the end of each implementation iteration and it was used to inform the next iteration of implementation. The steering committee consisted of managers (business operations managers, commercial and technical directors) from several levels in the organisation, who are also users of the BI solution, provided feedback. The steering committee confirmed the quality of the data, the accuracy of the intelligence, and gave feedback on areas that would need improvement in the future.

The implementation of the solution was a success and presented the SME with a vast number of benefits. Adopting the BI solution has been advantageous to the SME. Now, management has obtained freedom to access and explore business information without requesting IT support. The IT team has been liberated to improve data quality and the granularity information was available. Organisations are enabled to make informed decisions efficiently because they have access to reliable information, and the BI solutions present them with a transparent picture of the state of their SMEs.

Post implementing the solution, tangible benefits such as cost savings, increased efficiency, productivity and revenue totalled £29 262 within 6 months and the projected value within 12 months totalled £65 404 (Raj et al., 2016:47). Based on the results, Raj et al. (2016:47) emphasise that it is imperative for organisations to understand that the process of implementing a BI solution is not trivial but an iterative process. The case study proved that BI solutions empower SMEs to better understand their revenue and current business performance. Lack of sufficient funds prevent SMEs from investing and adopting BI solutions; however, the study proved that SMEs can use Microsoft BI suite which is affordable and easy to use.

Furthermore, Raj et al. (2016:47) advocate that high-quality data is critical in order for organisations to generate more accurate business intelligence. In this study, high quality data was achieved using the ETL processes’ transformation and cleansing steps. For efficient retrieval of data and having a trusted data source that can be used for analytics, a data warehouse was properly designed following guidelines on dimensional modelling. According to Raj et al. (2016:47), Microsoft suite of BI tools could be an appropriate solution for SMEs interested in implementing the BI solution, especially when they are already using a Microsoft business product. Additionally, if the SMEs are interested in leveraging big data, there is a wide range of tools and algorithms available to produce advanced business intelligence.

For this project to have been successful, Raj et al. (2016:47) used the top-down approach to deploy the first set of BI solutions to top managers. This approach enabled top-level management to better understand the benefits associated with BI and be more interested and supportive of a broader exploitation of BI within the organisation. Additionally, before any implementation phase could commence, the objectives and KPIs of the business were well-defined. SMEs need to have a better understanding of their existing IT infrastructure prior to making any new investments. Thus, to manage the cost of implementing the solution and enable the solution to be accessed by a wide range of users, a limited set of BI tools were deliberately used. Raj et al. (2016:48) state that it is critical for an IT solution to be user-friendly to ensure that more buy-in from end-users is obtained.

The AGGORA case study presented the same data quality and management issues as one SA SME. The Microsoft tools implemented at this SME can be beneficial to one SA SME as they are also using Microsoft SQL Server and Microsoft Access to store and manage data. However, it is critical for the data managers, owners and capturers of one SA SME to understand the progressive relationship between data, information, knowledge and wisdom in order to treat their business data as valuable organisational assets and protect its integrity. As data integrity is crucial in decision-making processes, SMEs need to have an understanding that raw data must go through transformation processes to yield information, and also understand how knowledge and insight is acquired. Consequently, one SA SME must take note that data analytics requires high quality to efficiently produce knowledge and insight that can enable efficient decision-making.

2.11 Conclusion

The researcher analysed research work in relation to this study to gain more knowledge and understanding on how SMEs are using technology to improve data quality and management. The literature review explored concepts such as data quality, data management and data analytics to answer the theoretical research questions of the study. Subsequently, the relationship between data, information, knowledge and wisdom is discussed and will give the SME in-depth knowledge on how knowledge and wisdom is acquired.

Numerous researchers indicated that organisations need to treat data as an organisational asset. However, in order to treat data as a valuable asset, SMEs need to understand what data quality means to its consumers. Consequently, if data are not protected and managed accordingly, this may lead to poor data quality as outlined in section 2.6 and section 2.6.4. The various causes of poor data quality, impacts of poor data quality and benefits of high data quality from other SMEs are discussed in this section to provide the researcher with critical knowledge to assist with determining an appropriate solution in one SA SME.

As a result, understanding the value data add to an organisation is very critical in ensuring that data are handled accordingly, and are protected and managed respectively. "*Fitness for use*" is the most widely used definition for data quality. In this study the researcher will refer to data quality as data that are "*fit for use*" by data consumers and meet acceptable levels across dimensions such as accuracy, consistency, timeliness, uniqueness, validity and completeness. The six-core data quality dimension outlined in Figure 2-3 will be used in this study to measure data quality.

Furthermore, it is evident that SMEs are interested in utilising business data to guide and influence decision making processes, improve business strategies, attract new customers and compete with larger counterparts. Section 2.9 provides the researcher with a broader understanding of tools and techniques that one SA SME can implement to improve data quality, data management, improve customer satisfaction, increase revenue and make informed decisions.

The researcher will develop an interface embedded with validations to enable effective capturing of data, develop a database with various constraint to manage and improve data integrity, and a data warehouse to store cleansed and transformed data. The "AGGORA" case study outlined section 2.10 proved that data analytics are crucial in SMEs. The Kimball methodology will be followed in developing the data warehouse. Additionally, the researcher will use Microsoft BI suite to develop the database, data warehouse, perform data transformations and data cleansing using (SQL server integrated services), SQL server reporting services to design reports, and use Power BI for data analytics.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

The main objective of this study was to develop an artefact for one South African (SA) SME to assist with capturing, storing and managing data in the quest of improving data quality for data analytics purposes. The researcher had to understand different research methodologies to make an informed choice for the research. This will create an understanding of the appropriate methodology that was chosen as suitable for the study and enabled the author to formulate the research structure and identify techniques and procedures to be used in collecting, interpreting and analysing the data.

Thus, research is a process for collecting, analysing and interpreting information to discover answers to particular questions by using procedures, methods and techniques that have been tested for their reliability and validity, and the process is undertaken within a framework of a set of philosophies and is designed to be objective and unbiased (Kumar, 2019:38). During this process researchers seek to systematically and with the support of data find the answer to a question, and achieve the resolution of an issue or a broad understanding of a phenomenon and develop knowledge in a particular field of study (Hevner & Chatterjee, 2010:3; Saunders *et al.*, 2019:130)

According to Kumar (2019:83), there are two important decisions in the research process that the researcher needs to make, "*what you need to find out*" (research problem and research questions) and "*how to go about in finding the answers*" (study plan).The process to discovering answers to research questions such as "*why, what, from where, when and how*" constitutes the research methodology (Kumar, 2019:86; Scotland, 2012:9). Additionally, Kumar (2019:76) emphasises that a research study is conducted to achieve four objectives:

- ❖ to discuss a problem or issue, situation, phenomenon (descriptive research);
- ❖ to explore or demonstrate a relationship between two or more variables (correlational research);
- ❖ to describe why particular things occur the way they do (explanatory research); and
- ❖ to assess the feasibility of conducting a study or investigating a subject area where nothing or little is known (exploratory).

There are various methodologies in information systems research that authors can select from to assist with conducting research studies. Thus, it is imperative for a researcher to use an appropriate method and methodology when exploring a particular phenomenon in order to reach factual conclusions and achieve research objectives.

The research onion as illustrated in Figure 3-1 provides a broad layout of this chapter that the researcher has followed in addressing the research problem and developing knowledge.

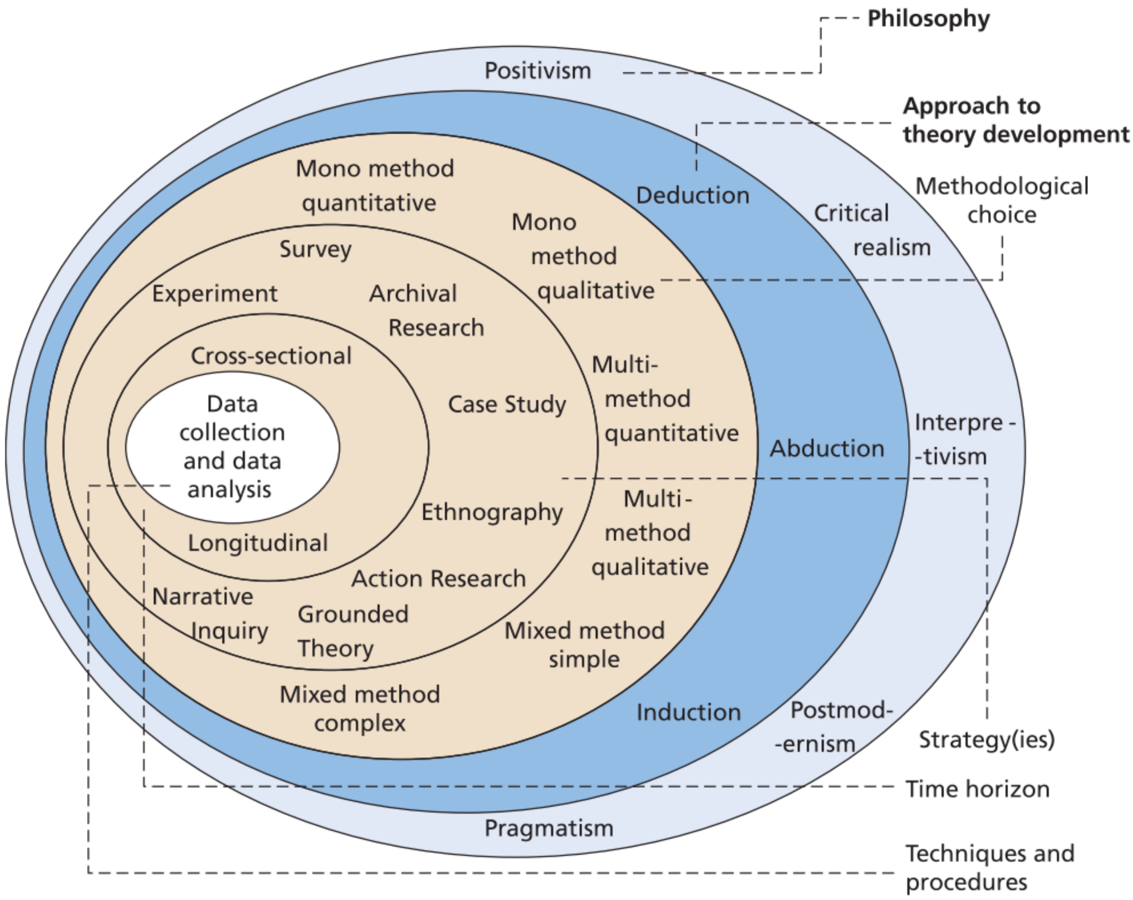


Figure 3-1: The research onion (Saunders et al., 2019:130).

Section 3.2 of this chapter will focus on the research paradigms and philosophical assumptions used in information systems and considered for the research. Furthermore, philosophical assumptions such as ontology, epistemology, axiology and methodology will be discussed to gain knowledge on each assumption, their characteristics and questions they seek to answer.

Section 3.3 of the study will focus on research approaches such as qualitative (3.3.1), quantitative (3.3.2) and mixed methods (3.3.3). Research strategies that the researcher can use as a plan of action to achieve research goals are discussed in section 3.4.

Section 3.5 explores various data collection techniques that researchers can use to collect data for the research study.

Section 3.6 explores ways in which researchers can analyse the collected data in a research study. Section 3.7 discusses design science research in information systems, its objectives and its role in solving real-world problems and the guidelines for conducting DSR in information systems

Section 3.8 discusses three design science research models considered for the research. The evaluation methods in design science research artefacts are discussed in section 3.9. Research ethics applicable to the study are discussed in section 3.10 followed by the study plan in section 3.11.

Rigour of the research study is explained in section 3.12, and lastly, the conclusion (3.13).

3.2 Research paradigms

In order for researchers to conduct well-founded research, it is significant to use paradigms to direct and guide inquiries and to ensure that the researcher does not dwell much in his or her own philosophical know-how (Rahi, 2017:1). Paradigms enable researchers to understand the reality of the world and the beliefs about the nature of reality, what is known about it and how this knowledge can be acquired (Rehman & Alharthi, 2016:51). Denscombe (2010:326) refers to a paradigm as a collection of beliefs and practices associated with a specific research style, and believes that research is conducted in accordance with a particular world-view and general philosophy. Guba and Lincoln (1994:107) define a paradigm as a "*set of basic beliefs that represent a worldview that defines, for its holder, the nature of the world, the individual's place in it, and the range of possible relationships to that world and its parts*". Mertens (2010:7) describes it "*as a way of looking at the world"* and is composed of particular philosophical assumptions that direct and guide thoughts and actions.

Theoretical paradigms such as positivism (post-positivist), critical theory, interpretivism, deconstructivism, emancipatory, pragmatism, transformative and constructivism have been discussed in various literatures (Creswell, 2013:23; Mackenzie & Knipe, 2006:3; Rahi, 2017:1). Guba and Lincoln (1994:112) proposed four paradigms that researchers can use in qualitative research; positivist, post-positivist, critical theory and constructivist. Peffers *et al.* (2007:5) indicate that the interpretive research paradigm has also been accepted in the IS discipline.

However, Hevner *et al.* (2004:79) and Cleven *et al.* (2009:2) feel that information systems research is conducted using two complementary but distinct paradigms:

- **behavioural science –** research is addressed through the development and justification of theories that describe and predict phenomena in relation to identified business needs.

Behavioural science draws its origins from natural science paradigms, and its objective is to discover the truth which informs design (Hevner, 2018:5). Furthermore, Hevner and Chatterjee (2010:5) indicate that behavioural science begins with an hypothesis, then researchers collect data and decide too either prove or disprove the hypothesis.

- **design science –** research is addressed through developing and evaluating artefacts designed to satisfy identified business needs. It is a problem-solving paradigm and its objective is to produce an artefact, "*the goal is utility*" (Hevner, 2018:5). Once the artefact is constructed, it should be evaluated. Hevner and Chatterjee (2010:5) emphasise that the knowledge produced by design science research provide guidance on how the artefact can be enhanced.

Matthews and Ross (2010:34) state that a paradigm tends to address the interests and focus of scientists or research communities from a specific discipline or allow them to share a set of theory-based beliefs about the world. Within each paradigm, the researcher can share a general view about the social reality, also called "*ontology*", and the best tools for research or "*epistemology*" (Denscombe, 2010:326).

Philosophical assumptions

When researchers are conducting a study, they make several assumptions throughout stages of the research that shape researchers' understanding of their research questions, the research methods to apply as well as how the findings should be interpreted (Saunders *et al.*, 2019:130). Furthermore, Saunders *et al.* (2019:130) mention that these assumptions are about human knowledge, about the realities researchers encounter in their own research, and about the extent to which researchers' own values influence their research processes.

Creswell (2013:18) advocates that it is important for researchers to understand the philosophical assumptions that inform a qualitative research study and be able to articulate them in a research study, because it outlines how we formulate our problem and research questions, and how we try to find information to answer the questions. Reviewers construct philosophical assumptions about a study when they assess it, and these assumptions are strongly rooted in our training and reinforced by the scholarly community in which we operate.

According to (Rahi, 2017:1), seven various philosophical assumptions associated with paradigms exist: ontology, epistemology, axiology, rhetoric, methodology, strategies of inquiry, and methods. Scotland (2012:9) is of the opinion that a paradigm is made up of four components: ontology, epistemology, methodology and methods, and each paradigm is based on its own ontological and epistemological assumptions. However, Creswell (2013:20) defines the four philosophical assumptions made by researchers undertaking qualitative research as outlined in Table 3-1 below. The four philosophical assumptions made by researchers include ontological, epistemological, axiological and methodological assumptions, which are discussed next.

Table 3-1: Philosophical assumptions with implications for practice – adopted from (Creswell, 2013:21).

Assumption	**Questions**	**Characteristics**	**Implications for practice (Examples)**
Ontological	What is the nature of reality?	Reality is multiple as seen through numerous views.	Researcher reports dissimilar perspectives as themes develop in the findings.
Epistemological	What is the relationship between what is being researched and the researcher?	Provides subjective evidence from participants. The researcher seek to minimise distance between himself or herself and that being researched.	Researcher depends on quotes as evidence from participants. The researcher will collaborate, spend time with participants in the field and become and "insider".
Axiological	What is the role of values?	Researcher acknowledges that research is value-laden and biases exists.	The researcher openly discusses values that shape the narrative and add his or own interpretation in conjunction with the interpretations of participants.
Methodological	What is the process of research? What is the language of research?	Researchers make use of inductive logic, studies the topic within its context and utilises an emerging design.	The researcher works with details prior to generalisation, describe in detail the context of the study and regularly revises questions from experiences in the field.

Ontology

In research, ontology refers to "*what can be assumed about the nature and reality of the social phenomena that make up the social world*" (Matthews & Ross, 2010:23). Ontology is a study that discusses the nature of reality or social phenomena and its characteristics (Creswell, 2013:20; Rahi, 2017:1; Vaishnavi *et al.*, 2004/2019:8). As outlined in Table 3-1, ontology focuses on natures of reality, and researchers conducting qualitative research must embrace and report on dissimilar realities. However, people see things differently and their knowledge about the world is influenced by various sources such as their own experiences, common sense of how things are, their beliefs, their values, and those they regard as authorities from the diverse aspect of their social world (Matthews & Ross, 2010:17). Thus, the ontological question asks "*What is the nature of reality?"* (Creswell, 2013:21; Guba & Lincoln, 1994:108) *and, therefore, what is there that can be known about it*?" (Guba & Lincoln, 1994:108). Matthews and Ross (2010:17) ask *"What is there to study and why do people see things differently?"*

Epistemology

As indicated in Table 3-1, epistemology refers to the theory of knowledge and how we come to know this reality (Dawson, 2009:18; Krauss, 2015:759; Matthews & Ross, 2010:23). It provides a rationale and justification for what can be known and what criteria knowledge must satisfy in order to be regarded as knowledge rather than beliefs (Matthews & Ross, 2010:26). Vaishnavi *et al.* (2004/2019:8) agree with this description, and advocate that epistemology seeks to explore the nature of knowledge.

In a qualitative study, it is essential for the researcher to conduct the study in the field where participants reside and work (Creswell, 2013:20). This will enable the researcher to stay close to the individuals being studied in order to understand what they say and believe. The epistemological question asks "*What is the nature of the relationship between the knower or would-be knower and what can be known*?" (Guba & Lincoln, 1994:108), and "*How do we know what we know*?" (Krauss, 2015:758). Creswell (2013:20) asks "*what is the relationship between what is being researched and the researcher?"*

Axiology

Saunders *et al.* (2019:134) and Vaishnavi *et al.* (2004/2019:8) define axiology as the study of values and ethics. Axiology seeks to understand the role of values in research and that such values need to be articulated openly, enabling researchers to include their own interpretation and bias (Creswell, 2013:20). The researcher publicly articulates the values that structure the narrative and adds his own interpretation in conjunction with the interpretations of the participants (Creswell, 2013:20). When conducting a qualitative study, the researcher positions themself in the study by acknowledging the value-laden nature of the study, and actively reports their biases and values together with the value-laden nature of the information that was gathered from the field (Creswell, 2013:20). However, it is up to the researcher to make a decision on how to handle their own values together with the participants' values.

Furthermore, Saunders *et al.* (2019:134) state that one of the primary axiological choices that researchers will encounter is the extent to which they wish to view the influence of their own values and beliefs on their research as a positive thing. For example researchers have to articulate their values as a foundation for making judgements about their choice of research method (Saunders *et al.*, 2019:134). Stating their own value position can be beneficial when deciding what is ethically significant and when there are queries related to the decisions they have made (Saunders *et al.*, 2019:134). Thus, the axiology asks "*what is the role of values*" (Creswell, 2013:20).

Methodology

A methodology is a strategy that guides the researcher in selecting the research method, provides a plan of how the study should be conducted, the type of data needed for answering the research questions, and which tools will be used in collecting and analysing the data**.** According to Silverman (2013:225), methodology refers to a plan the researcher follows in studying a particular phenomenon. Methodology in qualitative research is characterised as inductive, studies the topic within its context, utilises an emerging design, and is shaped by the researcher's experience in gathering and analysing data (Creswell, 2013:22).

The methodological question asks "*How can the inquirer (would-be knower) go about finding out whatever he or she believes can be known"* (Guba & Lincoln, 1994:108) or *"What is the process of research?"* Creswell (2013:21). Saunders *et al.* (2019:130) advocate that well-thought-out and consistent assumptions initiate a trustworthy research philosophy which will underpin the researcher's methodological choice, research strategy, data collection and data analysis processes. Research paradigms such as positivism, critical social research, interpretivism and design science research (DSR) are discussed in the next section.

3.2.1 Positivism

A positivist paradigm is frequently used by researchers in the IS environment to study the world. Positivism is also referred to as post-positivism, empirical science, scientific method and quantitative research (Rahi, 2017:1). Positivism and its successor post-positivism is based on the belief that the world can be studied in the same manner as the natural world using a method that is value free (Mertens, 2010:10). According to Rahi (2017:1), the adherents to positivism believe that true knowledge can be acquired through observation and experiment. Additionally, positivists use observations and measurements to test a theory in order to predict and control forces that surround us (Mackenzie & Knipe, 2006:3).

Post-positivism possesses features of being logical, reductionistic, empirical, cause-and-effect oriented, and deterministic based on *a priori* theories (Creswell, 2013:24). Post-positivist researchers perceive inquiry as a series of logically related steps, believe in numerous perspectives from participants rather than a single reality, and adopt rigorous methods of qualitative data collection and analysis (Creswell, 2013:24). However, Mertens (2010:11) states that post-positivists still hold beliefs about the significance of objectivity and generalisability, but recommends that researchers alter their claims to understandings of the truth based on probability rather than certainty. Furthermore, Mackenzie and Knipe (2006:3) indicate that positivists and post-positivists are mainly aligned with quantitative data collection methods (scales, experiments, tests, quasi-experiments) and analysis.

3.2.2 Critical social research

In information systems, critical social research is concerned with social issues including power, social control, freedom and values with respect to the development, utilisation and impact of information technology on social issues (Myers & Klein, 2011:17). Despite tremendous diversity of critical research, Myers and Klein (2011:17) discovered that critical research is marginalised in IS discipline, and some researchers do not recognise it as a legitimate approach in the IS discipline where it is considered to lack theoretical basis, as its objectives are deemed vague. Critical research has three elements (Myers & Klein, 2011:20):

- the insight element focuses on interpretation and acquiring insight;
- the critique element is mostly concerned with the geology of knowledge, critique and the social practices of control reproduction; and
- the transformation element is concerned with proposing improvements to the conditions of human existence, social theories and existing social arrangements.

3.2.3 Interpretivism

The interpretive paradigm is often referred to as constructivism, social constructivism or qualitative research paradigm (Creswell, 2013:24; Rahi, 2017:1). Creswell (2013:24) and Rahi (2017:1) emphasise that interpretive researchers centre their attention on the particular contexts in which individuals live and work in order to understand the cultural and historical settings of the participants. The interpretive paradigm believes that true knowledge can only be acquired through extensive understanding and interpretation of cultural settings of participants (Rahi, 2017:1). However, Mackenzie and Knipe (2006:3) advocate that interpretivism seeks to understand the world through human experience, recommending that reality is socially constructed. Researchers use interpretivism to view the world through the eyes of the respondents (Rashid *et al.*, 2019:4). According to Rashid *et al.* (2019:4), interpretivists answer questions related to dependability, conformability, credibility and transferability.

3.2.4 Design science research

Design science research (DSR) is a crucial and legitimate information systems (IS) research paradigm that involves the construction of numerous socio-technical artefacts such as modelling tools, decision support systems, methods of IS evaluation, IS change interventions and governance strategies (Gregor & Hevner, 2013:337). According to Hevner (2018:5), information systems are complex, artificial and purposefully designed systems that are made up of structures, people, technologies and work systems that work together to collect, process and store information. Design science refers to an approach that develops design knowledge, and it was initially introduced by Buckminster Fuller in the 1960s to inform design activities in applied science disciplines such as computer science, information systems, economics, medicine, engineering and applied mathematics (Gregor *et al.*, 2020:6; Vom Brocke *et al.*, 2020:2). According to Vom Brocke *et al.* (2020:1), DSR is a problem-solving paradigm that seeks to enhance technology and science knowledge bases through developing innovative artefacts that solve real-world problems and enhance the environment in which they are instantiated. This paradigm involves a rigorous process of designing artefacts to solve critical organisational problems, to make contributions to the research field of study, to evaluate the design, and to communicate the outcome to a relevant audience (Hevner *et al.*, 2004:85).

Furthermore, Hevner and Chatterjee (2010:5) emphasise that the design science research paradigm is highly appropriate to IS research because it communicates two primary issues of the IS discipline:

- The primary, although controversial role of artefacts in IS research; and
- The perceived lack of professional applicability of IS research.

Subsequently, Hevner *et al.* (2004:87) advocate that effective DSR must deliver clear contributions in the areas of design artefact, design construction knowledge such as foundation, and design evaluation knowledge such as methodologies. The outcome of DSR includes both the constructed artefact and design knowledge that provide a broad understanding through design theories of why the artefacts improve or disrupts the applicable application contexts (Vom Brocke *et al.*, 2020:1).

Furthermore, Hevner and Chatterjee (2010:5) state that design science research in IS addresses what are regarded to be wicked problems that are distinguished by:

- a crucial dependence upon human social abilities such as teamwork to produce effective solutions;
- a crucial dependence upon human cognitive abilities such as creativity to produce effective solutions;
- intrinsic flexibility to change design processes as well as design artefacts; and
- inconsistent requirements and constraints based on ill-defined environmental contexts.

Gregor and Hevner (2013:337) state that a design artefact is deemed effective and complete when it satisfies the constraints and demands it was meant to solve. In DSR, the design artefacts are required to fulfil certain goals through their material properties and through their functional affordance (Gregor *et al.*, 2020:7).

3.2.5 Positioning the study within DSR

Table 3.2 outlines the philosophical underpinnings of the frequently used paradigms in IS based on the four components of a research paradigm (Vaishnavi *et al.*, 2004/2019:9).

Table 3-2: The philosophical assumptions of the frequently used paradigms in Information Systems (Creswell, 2013:36; Guba & Lincoln, 2005:193; Saunders ***et al.***, 2019:145; Vaishnavi ***et al.***, 2004/2019:9).

Basic belief	**Post positivism**	**Interpretivism**	**Critical social theory**	**Design science research**
Ontological (the nature of reality)	Reality is real but it is comprehendible. A single reality that is knowable, real, independent, external and probabilistic. Due to lack of absolutes, researchers may not have the ability to understand it or get to it.	There are numerous realities which are socially constructed through our lived experiences and interactions with others. Complex, rich Socially constructed through language and culture.	Virtual reality is shaped by cultural, political, social, economic, ethnic, gender values.	There are numerous contextually situated world states which are socio-technologically enabled.
Epistemological (how reality is known)	Findings are true; It is dualist. Reality can only be approximated however it is constructed through research and statistics. Knowledge is made up of validated hypothesis that can be considered as laws or facts. It is objective and detached observer of truth.	It is subjective as knowledge and value are acquired through interactions. Reality is shaped by individual experiences and it is co-constructed between the researcher and the researched. Focuses on narratives, interpretations, perceptions and stories.	Advocates value-mediated finding and its subjective. Knowledge is established by the lived experience and the social relations that shape these experiences.	Based on knowing through making and construction occurs iteratively in a restricted context. The researcher knows a piece of information is factual and what it means through iterative circumscription.
Axiological (role of values)	The researcher's beliefs and biases need to be managed and not expressed or articulated in the study. It must be truthful; universal and prediction. The researcher must be neutral, independent, detached and must maintain objective stance.	Individual values are acknowledged and are negotiated among individuals. It must be understandable; descriptive Value-bound Researcher is reflexive.	Facts can never be separated from values and values of the researcher influence the research study.	Incorporates the values of the researchers and participants throughout the development process. Refers to the shared valuing of what researchers are anticipating to find in the pursuit of their efforts in a particular phenomenon.

Basic belief	Post positivism	Interpretivism	Critical social theory	Design science research
Methodological (approach to inquiry)	Involves the use of manipulative methodology or experiments to test hypotheses mainly in qualitative methods to find out whatever the researcher believes can be known. Deductive methods are essential, the objective is to construct new knowledge.	Involves participation and dialect, qualitative methods to discover solutions. Inductive methods of emergent ideas acquired through methods such as observing, interviewing and analysis of texts.	Concerned about the logical discussions and ideas.	Involves development and measuring the artefact's impacts on other complex systems.

The researcher chose DSR as a suitable paradigm for the study because in this study the researcher's objective was to solve a real-world problem by developing an artefact as a solution to a socio-technical research problem. The study fits with DSR because the researcher aimed to solve data quality and data management issues in one SA SME by means of developing a technology-based solution.

The ontological stance of my study was that the SME in which the study was conducted is perceived as a socio-technological environment. The reality was that the participants and the researcher would be involved throughout the study. The researcher and the participants would interact frequently with each other throughout the development process of the artefact, as the participants' input and feedback were critical in ensuring the success of the solution.

The researcher was aware of the data quality and management issues and will embark on an iterative process of constructing an innovative artefact to solve these problems. The epistemological stance of my study is addressed through "*knowing by doing*" – knowing the nature of knowledge through the iterative process of developing an artefact which can improve data quality and management for data analytics.

The axiological stance of my study is that as the researcher, I formed part of the study and interacted with participants when conducting interviews, so I was aware that my values as the researcher and those of the participants might come into play and influence the outcome of the study. Furthermore, the researcher conducted this case study with integrity and developed a trustworthy artefact that can assist the SME, reach applicable findings and conclusions that can contribute to the knowledge base, and assist other organisations. The participants were respected throughout the study and the researcher promptly communicated processes and any changes to them to maintain a trustworthy relationship. Meetings were held frequently with the participants to evaluate progress of the artefact development and to obtain their thought and feedback.

3.3 Research approaches

Depending on the nature of the problem the researcher is trying to solve, there are numerous research methods that the researcher can use in a research study to discover answers to a particular research problem. In research, a framework for classification of research has been designed to outline the types of research from different perspectives (Kumar, 2019:43). Figure 3-2 outlines the three perspectives that structure the basis of this classification as: applications of the findings of the research study, objectives of the study, and mode of enquiry used on conducting the study (Kumar, 2019:43).

Additionally, Figure 3-2 presents various ways in which research methodologies can be classified. However, the most popular and widely used methodologies are quantitative and qualitative (Kilani & Kobziev, 2016:1; Matthews & Ross, 2010:141; Mehrad & Tahriri, 2019:2). Another approach is the mixed method approach which incorporates both qualitative and quantitative methods in a way that is best for a particular research project (Matthews & Ross, 2010:141). The researcher followed a mixed method research approach in this study because the researcher interacted with the participants to gain in-depth understanding of the phenomenon from a participant's perspective, by collecting semi-structured data using interviews and observation to better understand their situation, experiences, feelings and perceptions. During the evaluation phase of the artefact, interviews and a questionnaire were utilised as data collection methods.

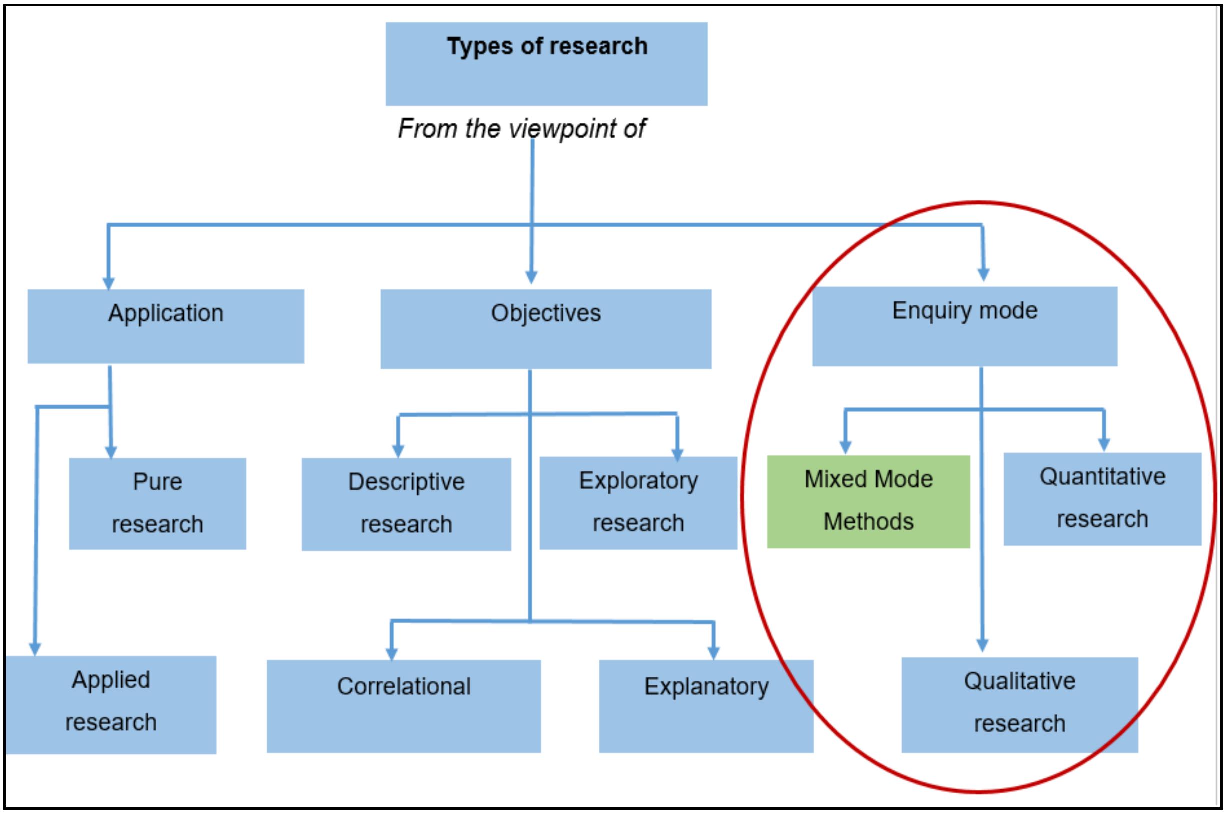


Figure 3-2: Types of research (Kumar, 2019:45).

The mode of enquiry perspective outlines three approaches that are used in research to discover answers to research questions (Kumar, 2019:48):

- The qualitative or unstructured approach;
- The quantitative or structured approach; and
- The mixed methods approach.

Some research questions may require the use of both qualitative and quantitative methods to be answered. Consequently, Silverman (2013:273) advocates that before researchers select a particular method, they need to consider the broader societal context in which this method is placed. However, the research problem itself should guide and determine whether the research study should be conducted using qualitative, quantitative or mixed method methodologies (Kumar, 2019:52). Additionally, researchers need to think carefully before selecting adopting mixed methods, as many models recommend that researchers cannot simply aggregate data in order to reach an overall truth (Silverman, 2013:273).

In the following section, the researcher discusses qualitative, quantitative and mixed method methods to determine a suitable study plan to be followed in answering the research questions and determining a solution to the research problem in this study. Based on the three research approaches, mixed method research study is a suitable research design for this study, because the researcher will closely interact with the participants to explore and understand their feelings, needs, perceptions, situation, beliefs and values, using interviews and observation. Additionally, the researcher will collect data during the evaluation of the artefact using a questionnaire. The aim of the researcher is not to answer the research questions by specifically quantifying the collected data, but rather to interpret the perceptions, feelings and experiences of the participants and test the functionality of the artefact.

3.3.1 Qualitative research

Qualitative research begins with the use of interpretive or theoretical frameworks and assumptions that inform the study of research problems, addressing the meaning individuals or groups ascribe to a human or social issue (Creswell, 2013:44). According to Kumar (2019:48), qualitative research is "*embedded in the philosophy of empiricism*" and conform to an flexible, open and unstructured approach of enquiry that aims to analyse diversity rather than to quantify. In qualitative research, the data are unstructured or semi-structured and may be in the form of phrases, pictures, words, artefacts, narrations and sentences (Kilani & Kobziev, 2016:3; Kumar, 2019:48; Mertens, 2010:3). Furthermore, the primary focus of a qualitative study is to understand, explain, discover, explore and clarify feelings, beliefs, perceptions, values, situations and experiences of a group of people, where the measurements of the findings are communicated in a narrative and descriptive manner rather than analytical manner, ensuring that no or less emphasis is placed on generalisation (Kilani & Kobziev, 2016:2; Kumar, 2019:50; Matthews & Ross, 2010:141).

Thus, a study is regarded as qualitative if its sole purpose is to describe a phenomenon, situation or event, and the information and data analysis is carried out to establish the variations in the phenomenon, situation or problem without quantifying variations (Kumar, 2019:52). Researchers conduct a qualitative study using data collection tools such as interviews, questionnaires and focus groups to gain access to the thoughts and feelings of research participants in order to develop an understanding of a particular phenomenon from another person's perspective and understand the meaning that people ascribe to in their expectations (Creswell, 2013:44; Dawson, 2009:14; Kilani & Kobziev, 2016:2; Sutton & Austin, 2015:226).

However, qualitative research requires the researcher to reflect upon situations before and during the research process in order to provide readers with context and understanding (Sutton & Austin, 2015:226). Additionally, when being reflexive, researchers should not avoid their own biases, but should rather reflect upon and clearly articulate their positions, beliefs, world view, biases and perspectives in order to allow readers to better understand the criteria through which questions were asked, data were collected and analysed, and findings were presented (Sutton & Austin, 2015:226).

3.3.2 Quantitative research

Quantitative research is "*embedded in the philosophy of rationalism*" and conforms to a structured, rigid and predetermined set of procedures to analyse with the objective of quantifying the magnitude of variation in a phenomenon and emphasise the measurement of variables and the objectivity of the process (Kumar, 2019:48). Additionally, the approach believes in substantiation on the basis of vast sample sizes and yield significance to the validity and reliability of findings. Findings are communicated in an analytical and aggregated manner by formulating conclusions and inferences that can be generalised (Kumar, 2019:48).

The quantitative approach enables researchers to gather and work with data that are structured (Matthews & Ross, 2010:141), and it is more concerned with numerical data such as interval, ratio or percentages, and utilising items such as diagrams and graphs to achieve appropriate results (Kilani & Kobziev, 2016:3; Mertens, 2010:3). Furthermore, Kilani and Kobziev (2016:3) and

Dawson (2009:20) state that a quantitative method enables researchers to answer questions such as *“how many”, “test”, “verify”, “to what extend” or “how often*”.

A study is regarded as a quantitative study if the researcher wants to quantify the variations in a situation, phenomenon, problem or issue, and the information is accumulated using predominantly quantitative variables where the analysis is geared to ascertain the volume of the variation (Kumar, 2019:52). Kumar (2019:83) states that qualitative and quantitative research methodologies both differ in the procedures, methods and models used for processing and analysing data, especially in their underpinning philosophy. The various features that are used to differentiate between quantitative and qualitative methods or using both in a mixed method mode are outlined in Table 3-3.

Table 3-3: Features of qualitative and quantitative approaches (Bryman & Bell, 2011:410; Kumar, 2019:55; Matthews & Ross, 2010:142).

Theme	Quantitative approach	Qualitative approach
Paradigms	The assumption is that the social world is real. Ontological and epistemological approaches are positivist.	Assumes that reality is a social construct. Ontological and epistemological approaches are interpretive.
Approach to inquiry	The approach to inquiry is structured/rigid/predetermined methodology.	The approach to enquiry is unstructured/flexible/open methodology.
Research question	Research questions may be presented as testable hypotheses.	Research questions may be developed utilising subsidiary questions.
Measurement of variables	Research questions can be answered by using statistical analysis or by counting events. The variation in a situation, phenomenon, and issue can be quantified to some extent.	Research questions can be answered by analysing participants’ experiences, beliefs and understandings, and by describing and explaining events. Variations in a phenomenon, situation, and issue are described.
Inquiry	Researchers usually know what they are looking for.	Researcher may only have a general idea of what they are looking for.
Researchers’ involvement	Objective such that the researcher is not part of the research.	Subjective such that the researcher is involved as a social being,
Data collection	Surveys or questionnaires used as data collection tools.	The researcher is the primary instrument for collecting data.
Data representation	Data is represented in the form of numerical or named codes, structured.	Data can exist in any form; unstructured.
Data type	Hard, reliable data.	Rich, deep data.
Data generalisation	Researcher can generalise from the data	Not usually possible for the researcher to generalise from the data. Searching to understand the values, beliefs, behaviour of the context in which the research is conducted.
Point of view	Point of view of researcher.	Point of view of participants.

Theme	Quantitative approach	Qualitative approach
Research format	Static.	Process.
Interactions with participants	Researcher are uninvolved with their subjects.	Researcher is closely involved with the individuals being investigated to genuinely understand the world from their point of view.

Despite the difference between qualitative and quantitative research, similarities also exist. Both quantitative and qualitative research are concerned with finding answers to research questions. Both strategies seek to ensure that deliberate distortion doesn't take place, argue for the significance of transparency, are concerned with reduction of data, and uses associating data analysis in the research literature (Matthews & Ross, 2010:148).

Matthews and Ross (2010:148) state that selecting a qualitative, quantitative or mixed method research approach is one of the decisions that a researcher needs to make to determine what data collection methods to use. However, both qualitative and quantitative approaches have their own strengths and weaknesses as outlined in Table 3-4.

Table 3-4: Strengths and weaknesses of the qualitative and quantitative methods (Kilani & Kobziev, 2016:3).

Method	Strength	Weakness
Qualitative	❖ Qualitative analysis provides a rich, complete and detailed description. ❖ Enables a good sight into a person's experience and behaviour and does not lessen complex human experience in numerical form. ❖ As compared to quantitative methods, qualitative approach can be faster. ❖ Ambiguities which are inherent in human language can be identified in the analysis. ❖ Qualitative methods can be cheaper as compared to quantitative research.	❖ Qualitative data is hard to analyse and requires a high level of interpretative skills. ❖ Excellent chance of bias. ❖ Difficult to draw conclusions from qualitative data. ❖ Provides low level of accuracy in terms of statistics. ❖ Qualitative data faces hardships in terms of comparisons.
Quantitative	❖ Enables classification of features, construction of more complex statistical models in an attempt to explain what is observed and counting them. ❖ Quantitative data represent numerical format which enables researchers to easily analyse it. ❖ Provides high level of accuracy. ❖ Compare measures of dispersion. ❖ Analysis can be presented graphically.	❖ Can be slow to implement and requires time. ❖ Not easy to implement. ❖ Can be expensive. ❖ Requires computer analysis. ❖ Low response rate.

It is worth noting that there is a growing recognition by many disciplines that both types of research are significant for a research project (Kumar, 2019:51). Furthermore, Dawson (2009:16) emphasises that all methodologies have their distinct advantages and disadvantages, neither is more superior than the other as both rely on the training, skills and experiences of the researcher. However, in qualitative research, a case study is one of the most widely used methodologies by researchers to conduct a detailed investigation about a particular phenomenon, with empirical material collected over a period of time from a well-defined case to produce analysis of the contexts and processes associated with the phenomenon (Rashid *et al.*, 2019:5).

3.3.3 Mixed method research

Researchers use mixed method research in a case study to develop in-depth evidence, practical understanding and conclusions associated with the complexity of a case (Creswell & Plano Clark, 2018:189). Mixed method research integrates two methods, qualitative and quantitative, to gather data required for answering the research questions. Research studies that use more than one method to improve accuracy of the findings are said to be utilising the mixed or multiple methods approach (Kumar, 2019:56). According to Rahi (2017:1), researchers have the freedom to use both qualitative and quantitative approaches, taking into consideration that the key is to find the significant techniques and procedures of research that can assist in solving the problem statement. Mixed method data collection in case study research is important in assisting researchers in understanding the complexity of a case (Creswell & Plano Clark, 2018:189).

Researchers in a case study must ensure that their case research study is based on several sources of evidence in order to accomplish the purpose and worth of the study (Yin, 2018:174). Mixed method research is based upon the notion that disparate paradigms and methods have disparate strength for particular situations; their integrated strength will result in enhancing the depth and accuracy of the findings (Kumar, 2019:57). Depending on the research problem that the researcher is addressing, in some cases the qualitative approach is better and in others quantitative is a better option. Thus, in some cases the researcher will need to combine the strength of different paradigms and methods to achieve great results.

Kumar (2019:57) states that mixed or multiple methods can be used when:

- the researcher wants to explore both perspectives;
- when accurate and complete information from a single source is hard to acquire;
- there is a necessity for good quality research;
- when the researcher needs to construct generalisation;
- when researchers need to discover an explanation for their findings;
- when the researcher wants to establish a good data collection instrument and ascertain the validity of the questions; and
- when the researcher tackles studies with numerous objectives.

According to Kumar (2019:64), the rationale underpinning the mixed or multiple method is fundamentally based upon two beliefs:

- the ability of the methods of the paradigm to produce accurate answers to all research questions in all situations; and
- the use of more than one method in most situations will produce a better and more accomplished picture of a situation or phenomenon than a single method alone.

3.4 Research strategies

In design science research, an artefact is regarded as the end-product produced to enhance an existing solution to a problem or provide a first solution to a critical problem (Gregor & Hevner, 2013:338; Venable *et al.*, 2017:77; Vom Brocke *et al.*, 2020:1). Design science research seeks to improve human knowledge with the construction of innovative artefacts and generate design knowledge through innovative solutions to real-world problems (Myers & Venable, 2014:1; Vom Brocke *et al.*, 2020:1). To achieve this, researchers can select an appropriate research strategy to be used in solving a real-world problem through developing an innovative artefact. Based on the research onion by Saunders *et al.* (2019:130) presented in Chapter 3, research strategies such as grounded theory, experiment, survey, archival research, case study, ethnography, action research and narrative inquiry can be used by researchers .

When researchers choose a strategy, they need to take into consideration certain factors that are critical for conducting the research study (Denscombe, 2010:9):

- **Feasibility –** will the chosen strategy allow me to complete the research study? "*Can it be done*"?
- **Suitability –** will the chosen strategy produce significant data required to answer the research questions?
- **Ethics –** will the chosen strategy allow me to be ethical when interacting with the participants?

Thus, it is critical for researchers to choose research strategies that can enable them to operate within an appropriate code of research ethics, ensure informed consent from participants, and avoid any harm to participants. A strategy such as a case study is suitable for this study because the researcher is exploring a phenomenon in-depth in a real-life context. In the next sections, I explain a few research strategies that researchers commonly use when conducting a research study and were considered when the most suitable research strategy for this study was selected.

3.4.1 Ethnography

Ethnography is an approach that researchers use to explore and interpret cultural behaviour by immersing themselves in the lives and cultures of the group involved in the study, often living with that group for a particular period of time (Dawson, 2009:18). According to Creswell (2013:102), the main focus with this approach is on positioning the individuals' stories within the context of their culture and culture sharing group. Therefore, this type of research study usually follows a case study design (Matthews & Ross, 2010:135). Ethnography requires researchers to perform "*participation observation*", which involves studying the participants in their real life environment and participating in group activities while conducting formal interviews, taking notes, observing behaviour, analysing, reflecting and writing reports, sometimes stretching over extended periods (Dawson, 2009; Sutton & Austin, 2015:226). Visual recording, such as taking photos and videos, is also used to collect data and the data are analysed as they are collected (Matthews & Ross, 2010:135). During this process the researcher uses a reflective diary and his reflections are incorporated with the collected data (Matthews & Ross, 2010:135).

3.4.2 Grounded theory

Grounded theory is an approach in which researchers use their collected data to develop inductive theoretical analyses and subsequently collect more data to verify these analyses (Silverman, 2013:215). According to Creswell (2013:105), a data collection method such as interviews with approximately twenty to sixty participants are primarily used to gather data. Furthermore, the collected data will be analysed through open coding, axial coding and selective coding (Creswell, 2013:105). By contrast, grounded theory is entrenched in an assumption that your hypothesis must be induced from close data analysis rather than starting with a prior hypothesis (Silverman, 2013:215). However, Silverman (2013:470) mentions that grounded theory has been criticised for its failures to accept implicit theories which can provide guidance to the work at an early stage.

3.4.3 Action research

Action research is a research approach that was initially introduced by Kurt Lewin in 1946 to address social system change by using action as a way to effect change and to generate knowledge about the changes (Hevner & Chatterjee, 2010:182). With this approach, organisations, institutions, people or communities collaborate with the action researcher to diagnose an issue and develop a solution based on the diagnosis (Bryman & Bell, 2011:413). This approach starts with a process of communication and agreement amongst a small group of people who desire to enhance a situation in a particular environment, then proceed to four stages of planning, acting, observing and reflecting, where the researcher will be acting as the facilitator

(Dawson, 2009:17). According to Coghlan and Brannick (2019:5), action research must include four characteristics:

- **Practical nature –** its main focus is on real-world problems and issues typically in an organisational setting and at work;
- **Change –** is regarded as a fundamental element of research (as a means of discovering more about phenomena and as a way of dealing with practical problems);
- **Cyclical process –** research includes a feedback loop in which the initial findings yield possibilities for change which are then implemented and assessed as a prelude for further investigation; and
- **Participation –** participants are the most important people in the research study and their participation is active, not passive.

Coghlan and Brannick (2019:9) and Bryman and Bell (2011:413) state that action research involves four steps that can be used in designing a solution:

- **Step 1: Constructing –** the researcher constructs the possible issues in the study environment. This step needs to be carried out carefully and thoroughly and any changes in constructing needs to be recorded and articulated accordingly.
- **Step 2: Planning action –** the researcher plans what actions can be implemented.
- **Step 3: Taking action –** the researcher implements the action that was planned in step 2.
- **Step 4: Evaluating action –** the results of the action implemented in step 3 are evaluated to check if the action was taken in an appropriate manner and matched the constructing. If the problem is not solved, then the cycle starts all over again from step 1.

As a result, the researcher conducting action research must have excellent group management skills as well as an understanding of group dynamics (Dawson, 2009:17). Denscombe (2010:5) emphasises that the main purpose of action research is to solve a practical problem and produce guidelines. Additionally, action research leads to re-education and contributes to academic theory and practical action (Bryman & Bell, 2011:413). Even though action research output is relevant, readable and interesting to practitioners and academic audiences, some researchers criticise it for its consequent lack of rigour and lack of repeatability, as well as for focusing too much on organisational action at the expense of research findings (Bryman & Bell, 2011:415).

Bryman and Bell (2011:415) emphasise that action research can include the collection of both qualitative and quantitative data. Advantages of action research include: participation of practitioners, professional self-development of practitioners, addresses practical issues in a positive way, transfers the research results directly into practice, and it provides a continuous cycle of development and change on-site in the workplace (Denscombe, 2010:134).

3.4.4 Case study

The case study approach has been widely used, mainly when a researcher wants to explore a phenomenon in-depth and present an explanation that can cope with the intricacy and delicacy of a real life situation (Rashid *et al.*, 2019:2; Zainal, 2007:1). Yin (2018:15) defines a case study as "*an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly defined"*. According to Bryman and Bell (2011:60), a case study is associated with a geographical location such as organisation or workplace, where an intensive, detailed examination of a situation is conducted. The researcher can choose to conduct a study of a social phenomenon such as an organisation, country, individual, industry, etc. In a case study, the researcher's main focus is to provide an in-depth investigation of a situation.

The purpose of case study research is to understand the complex relationship between factors as they operate within a specific social setting (Denscombe, 2010:5; Yin, 2018:15). Rashid *et al.* (2019:5) argue that the objective of a case study is to do intensive research on a particular case such as a group, individual, institution or community. Furthermore, Yin (2018:9) states that a case study seeks to answer 'how' and 'why' questions about a contemporary set of events over which the researcher has little or no control. Bryman and Bell (2011:60) explain that what differentiates a case study is that the researcher is usually concerned to elucidate the distinctive attributes of the case.

Zainal (2007:2), however, states that case study as a research tool receives a lot of criticism in terms of its lack for robustness. When choosing cases, the researcher should first and foremost choose a case where great opportunity of learning is anticipated (Bryman & Bell, 2011:60). A case study must appropriately link to the research question and should be a viable strategy that can be used to obtain implicit and explicit data from participants (Zainal, 2007:2). Additionally, it must prove that it adheres to the set of procedures with proper application, the case is connected to a theoretical framework, and that the "*chain of evidence*" either qualitative or quantitative are systematically recorded and archived, particularly when interviews and direct observations by researchers are used as primary sources of data (Zainal, 2007:2). Furthermore, there are several types of case studies that researchers can select from (Bryman & Bell, 2011:60; Cook & Kamalodeen, 2019:12; Creswell, 2013:99; Yin, 2018:324):

- **Intrinsic –** researchers undertake this type of case study to obtain insight into the particularities of a situation instead of obtaining insight into other cases or generic issues.
- **Instrumental –** these types of case studies focus on using the case as a means of gaining understanding on a broader problem or letting generalisation be challenged.
- **Collective cases –** these cases are undertaken in conjunction to investigate a general phenomenon.
- **Explanatory –** researchers use this type of case study to explain a phenomenon or a problem in a specific context. It will explain "*how*" and "*why*" certain situations came to be.
- **Exploratory –** data are gathered to determine where a topic is worthy of further investigation, and if so, research questions and data collection procedures to be used in subsequent studies are identified, which might or might not be a case study.
- **Descriptive –** A phenomenon is described in its real-world context.

However, Denscombe (2010:5) argues that cases are not randomly picked but are picked on the basis of known attributes. Additionally, the selection criteria used for selecting the cases need to be justified and made explicit as an important section of the methodology (Denscombe, 2010:5). Denscombe (2010:5) and Matthews and Ross (2010:139) state that a case study approach enables researchers to:

- focus on a small number of cases in-depth allowing the researcher to deal with delicacies and intricacies of complex social situations;
- view their research topics within a particular context;
- use a holistic approach in conducting their research and exploring their research topic in context; and
- to utilise multiple methods to capture the complex reality under scrutiny.

Rashid *et al.* (2019:2) indicate that a case study is a widely used methodology in qualitative research, and propose in Figure 3-3 four phases that researchers can follow when conducting a case study:

- **Foundation phase –** is the initial phase in conducting the case study and during this phase the researcher briefly discusses literature to set the tone of the research study. Philosophical consideration involves selecting a paradigm, the inquiry technique consideration involves selecting a research approach, and research logic involves selecting suitable research logic.

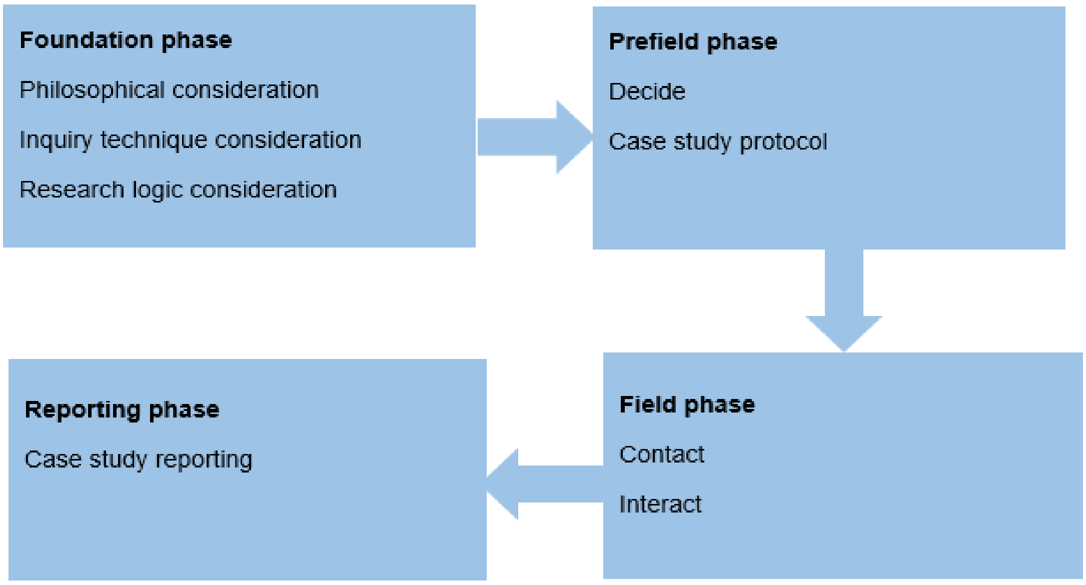


Figure 3-3: Four phases to conduct a case study (Rashid *et al.*, 2019:2).

- **Prefield phase –** is the second phase when conducting a case study and it involves discussing the operational details which are discussed using two sections to decide and case study protocol. In the decide step the researcher must select a suitable research method and ascertain whether the case study is the most appropriate method. The next step, case study protocol, enables the researcher to formulate a research protocol, which includes the research question, research method, and scope of research, focus, permission seeking, ethical considerations, interpretation process, and criteria for assessment.
- **Field phase –** is the third phase and includes two steps, contact and interact. Contact requires the researcher to go and collect potential material that can assist in creating strong findings. The researcher needs to have good knowledge of the cases and the participants who will be involved in the research. Interact involves collecting empirical material and interpreting it to assist with answering the research questions. Data collection methods such as semi-structured interviews, observations and document collection can be used.
- **Reporting phase –** is the last phase in conducting a case study and requires a researcher to write a case study report. Case study reporting is very important, so the researcher needs to ensure that the report structure and the storyline is presented accordingly. The quality of a case study doesn't rely only on data collection and analysis, but also on reporting (Rashid *et al.*, 2019:6). Furthermore, when reporting the case study, the researcher should take into consideration the case descriptions, participants' descriptions,

empirical material interpretation, relationship descriptions, details of field protocols and analysis and conclusion.

In this study the researcher followed a case study approach because the researcher was seeking to answer "*how*" and "*why*" questions by doing an in-depth investigation on one phenomenon. The researcher performed a thorough investigation on the data management in one SA SME, and an exploratory study to gather requirements for development of an artefact in one SA SME. It was based on a real-life situation where the researcher interacted with participants to collect data by using the mixed method. The researcher followed the four phases of conducting a case study to answer research questions as defined by Rashid *et al.* (2019:2) in Figure 3-3.

3.5 Data collection techniques

Once the research problem has been formulated, the study design developed, and the research strategy defined, the researcher needs to start gathering data required for determining a solution to a research problem. Depending on the research method chosen for the study, there are various data collection methods that researchers can implement to collect data. Research methods refer to tools that researchers use for empirical data collection (Dawson, 2009:27; Denscombe, 2010:326). However, Mackenzie and Knipe (2006:7) enunciate that the paradigm and research questions should determine which data collection and analysis methods will be relevant for a study.

Data collection in a case study is deemed substantial because it is getting data from various sources of information (Creswell, 2013:100). In case study research, interviews, participants' observation, direct observation, documentation, archival records and physical artefacts are six sources of evidence that are commonly used by researchers (Yin, 2018:156). In ensuring that the researcher produces a good case study, it is important that various sources of evidence are utilised (Yin, 2018:156). When choosing between interviews and questionnaires, researchers can apply criteria to assist them with making the decision: check the type of study population, verify the nature of the investigation and check how the geographical distribution of the study population is (Kumar, 2019:273).

The researcher needs to understand that some information that participants will share will be of sensitive nature or confidential. Thus, the researcher's responsibility is to protect the participants and their data, including all the data being collected (Sutton & Austin, 2015:227). Furthermore, Sutton and Austin (2015:227) emphasise that the method that the researcher will use to secure participants' data must be clearly articulated to participants, and needs to be approved by a relevant research ethics review board before commencing with the research study. Researchers must take into account the pros and cons of handling participants' data, and if they are not familiar with data handling they need to request advice from experienced qualitative researchers (Sutton & Austin, 2015:227). Ethical considerations need to be in place to protect participants from any form of harm and were discussed in section 3.10. The researcher will discuss data collection techniques applicable to qualitative studies next. It was necessary to determine suitable techniques to be used for collecting data in this study.

3.5.1 Interviews

Interviews are a vital source of case study evidence that researchers can utilise to obtain crucial insight about human affairs and actions from well-informed participants (Yin, 2018:164). Interviews are a popular data collection method where the interviewer tries to acquire information from interviewees by asking questions telephonically or face-to-face. According to Kumar (2019:280), an interview is defined as any person-to-person interaction between two or more individuals, either face-to-face or telephonically. Interviews are one of the dominant data collection methods that provide researchers an opportunity for direct interaction between the researcher and the research participants (Matthews & Ross, 2010:220). The researcher using interviews as data collection method has the freedom to choose the content and format of the question to ask the participants. However, the research questions can either be flexible (the interviewer has the freedom to formulate the questions based on the issue being investigated) or inflexible (the researcher must stick strictly to the questions beforehand including the sequence, wording, and the way in which they are asked). Furthermore, based on the degree of flexibility in terms of structure and standardisation, interviews can be classified as structured, semi-structured and unstructured (Dawson, 2009:27; Kumar, 2019:281; Matthews & Ross, 2010:220). Structured interviews are regarded as standardised, whereas unstructured and semi-structured interviews are regarded as non-standardised (Matthews & Ross, 2010:221).

Using semi-structured interviews is an appropriate data collection technique for this study because the researcher will not follow a rigid method in conducting the interviews; it should have some form of flexibility. The researcher will have predefined questions that will be followed in the process, however, there will be flexibility in the way questions will be asked, providing clarity where it is required and allowing discussions between participants and the researcher. Furthermore, more questions may arise during the session and the researcher will oversee the session to ensure that it is professional, manageable and participants are free to speak.

Yin (2018:161) states that audio recordings during a case study interview session are very important because they enable the researcher to obtain an accurate rendition of the interview session. However, the researcher can use an audio recording device in the interview only if the participant is comfortable with it and has given permission for the interview to be recorded (Yin, 2018:161). The researcher will use interviews to collect data from participants and sessions will be recorded to avoid losing any crucial information.

Structured interviews

In a structured interview, the researcher asks the interviewee a set of prearranged questions. applying the same wording and order of questions as outlined in the interview schedule (Kumar, 2019:283). Furthermore, structured interviews provide consistent information which assures the comparability of data. Matthews and Ross (2010:221) outlines characteristics of a structured interview:

- adheres to a regular set of questions for each interview;
- questions are asked in the same way, using the exact same words, probes for each interview; and
- participants are presented with a group of answers to select from.

Semi-structured interviews

Basically, semi-structured interviews are used to gather qualitative data when the researcher is interested in individuals' behaviour, experiences and understandings, and how and why they experience and come to understand the phenomenon the way they do (Matthews & Ross, 2010:221). According to Matthews and Ross (2010:221), semi-structured interviews are used by researchers to discover what people think about a particular social phenomenon they possess knowledge of (evaluation), to collect data which will assist the researcher to explain why people experience and understand a particular phenomenon the way they do (explanatory), and to discover what participants think is significant about the research topic (exploratory). Characteristics of semi-structured interviews include (Matthews and Ross (2010:221):

- regular set of topics or questions are followed for each interview;
- enables the participants to discuss the topic using their own words or respond to the questions; and
- distinctive methods or orders relevant for each interview may be used to introduce the questions.

Unstructured interviews

With unstructured interviews, the researcher has the flexibility to structure the questions in any format and there is no predetermined method of asking the questions. Dawson (2009:27) is of the view that researchers use unstructured interviews in an attempt to achieve a holistic understanding of the respondent's point of view. According to Kumar (2019:281), the researcher has the freedom to formulate questions, to use any wording and sequence when asking questions, to raise issues on the spur of the moment, and give explanations to questions when respondents require clarity. Kumar (2019:281) states that unstructured interviews are mainly used in qualitative research. Furthermore, Matthews and Ross (2010:221) explain that unstructured interviews focus on a broad scope for discussion and give participants freedom to discuss the research topic in their own way without any limitations.

Even though interviews are regarded as the common method of collecting data that provide the researcher with direct interaction with participants, this method has its own disadvantages. Kumar (2019:291) identified various advantages and disadvantages of interviews that researchers must take into consideration if they want to use interviews to collect data:

Table 3-5: Advantages and disadvantages of interviews (Kumar, 2019:291).

Advantages of interviews	Disadvantages of interviews
❖ Interview is critical for gathering in-depth information. ❖ Interview is more significant for complex situations. ❖ Interviewing has a broader application. ❖ Information can be supplemented.	❖ Interviews are expensive and time consuming. ❖ The interaction between the interviewer and the participants will determine the quality of the data. ❖ The quality in which the interviewer conducts the interview will determine the quality of the data. ❖ When many interviewers are utilised the quality of the data may differ. ❖ The researcher may present his/her bias.

Consequently, for the interviewer to be able to elicit desired information from the participants in an interview, the interviewer must plan in advance for the interview session, must have a thorough understanding of the issue in question, and must ensure that appropriate rapport is accomplished with the interviewee (Kothari, 2004:119). Furthermore, the interviewer must have an informal and friendly approach, have the ability to listen with understanding, act accordingly during the session, and at the same time show respect and curiosity (Kothari, 2004:119).

3.5.2 Questionnaires

A questionnaire refers to a compiled list of questions that the researcher distributes to respondents to provide answers. Kumar (2019:284) states that a questionnaire is a written list of questions which participants should answer. Questionnaires are less expensive and offer great anonymity. Despite that, researchers planning to use questionnaires as data collection method must be aware of the several disadvantages associated with questionnaires: low response rate, application is limited, include self-selecting bias, response to a question may be influenced by the issues, lack opportunity to clarify issues, spontaneous responses are not permitted, response cannot be supplemented with other information, and it is possible to consult others (Kumar, 2019:290).

However, Dawson (2009:91) advocates that researchers need to construct questionnaires in such a way that they do not frustrate, embarrass, annoy participants, or make participants feel uncomfortable to respond. When constructing the questionnaires, researchers should take the following into consideration (Dawson, 2009:91):

- Researchers should ensure that questions are kept brief and simple. Any use of jargon words, technical terms or words that have multiple meanings or could be misinterpreted should be avoided when constructing questionnaires. Make sure that participants will be able to answer the questions.
- The researcher must avoid using words with emotional connotation, vague words such as *often* and *sometimes* should be avoided.
- The researcher must start with easy-to-answer questions and followed by complex questions at the end.
- Researchers must avoid asking leading questions and negative questions.
- Researchers should not include any questions that may have any form of prestige bias. Questions that could influence the participants to be dishonest and give false answers or cause frustration, offence, sadness, anger, or embarrassment should be avoided.
- Different question formats should be used.
- The researcher must ensure that all possible answers are covered when constructing a closed-ended question.
- Where sensitive issues are present, the researcher may promise participants that anonymity and confidentiality will be maintained, however, it is best for the researcher to ask indirect questions rather than direct questions.
- Personal information should be asked at the end.

3.5.3 Focus groups

Focus groups have grown in popularity and been recognised as a legitimate data collection method in science (Matthews & Ross, 2010:236). Researchers use focus groups when they are interested in collecting deep, rich qualitative data about people’s ideas, understandings, and experiences, and are interested in why people experience the phenomenon the way they do (Matthews & Ross, 2010:235). They occur in several contexts or settings such as the workplace, a community centre, school, public building, non-spatial communities, and online, geographical communities, or with groups of people with common interest (Matthews & Ross, 2010:241).

Furthermore, focus groups are also referred to as discussion interviews led by a facilitator, whereby several people are requested to meet in a group to discuss a particular issue and the researcher will listen to, observe and record their discussions. During these sessions, the facilitator or moderator is in charge of the session and ensures that no one dominates the discussion whilst ensuring that all participants are making a contribution ((Dawson, 2009:30). Dawson (2009:30) states that the facilitator will introduce a topic, ask particular questions, control digressions and stop break-away conversions.

According to Dawson (2009:30), using a focus group can enable the researcher to obtain a broad range of responses during one meeting, and the group result is an effective resource in data analysis. Participants are able to ask questions among each other, participants' interaction is effective to analyse, and researcher bias is reduced (Dawson, 2009:30). However, Dawson (2009:30) warns that some researchers may find focus groups intimidating and strenuous to moderate. Additionally, some disadvantages associated with focus groups include that not all individuals in the session may contribute; some people may be nervous about speaking in front of others or are uncomfortable in group settings, certain people may contaminate an individual's views, it may be hard to extract individual views during the analysis, and venues and equipment required can be expensive (Dawson, 2009:30).

3.5.4 Observations

Observation data collection method is a common method to use in DSR studies to observe how people perceive and use the artefact. It requires the researcher to collect information by means of directly observing participants without asking them any questions (Kothari, 2004:96). In qualitative research, no framework is used when collecting data using observation, and the recording is carried out in a descriptive and narrative form (Kumar, 2019:311). Furthermore, Kothari (2004:96) states that this method is relevant in studies that deal with participants who are unable to provide verbal reports of their feelings. Silverman (2013:246) argues that this data collection method is crucial in understanding other cultures.

However, this data collection method is expensive, it provides limited information and sometimes unanticipated factors interfere with the observational task (Kothari, 2004:96). According to Kothari (2004:96), the researcher planning to use observation as a data collection method should keep the following in mind:

- ❖ What must be observed?
- ❖ How must the observations be recorded?
- ❖ How can the accuracy of the observations be ensured?

By contract, observation methods have a number of advantages: if used accurately, subjective bias is eradicated; it is independent of participants' willingness to respond; and the information acquired under this method relates to what is currently happening (Kothari, 2004:96). However, qualitative researchers argue that observation is an unreliable data collection method because different researchers can obtain dissimilar observations (Silverman, 2013:247).

3.6 Data analysis

Data collection methods enable researchers to acquire large sets of data that need to be analysed to determine the findings and reach conclusions. Depending on the nature of the research study, either qualitative or quantitative, there are various tools that researchers can utilise to analyse data. However, Kumar (2019:92) is of the view that to analyse the collected information depends upon the two things:

- ❖ The type of information (quantitative, qualitative, or attitudinal and descriptive)
- ❖ The way the researcher wants to communicate the findings to the readers.

Furthermore, Dawson (2009:124) explains that there are different processes involved in qualitative data analysis that could be followed:

- ❖ Researchers need to think about data from the moment they start gathering the information;
- ❖ The value of the collected data should be judged, especially data coming from unconfirmed sources;
- ❖ As the research study progresses, the researcher must interpret the data to develop a better understanding of what is going on; and
- ❖ Finally, the researcher needs to commence with the data analysis process.

Qualitative data analysis requires the researcher to respect the privacy of the participants, and avoid reporting only positive results participants (Creswell, 2013:59). According to Sutton and Austin (2015:227), qualitative research requires putting oneself in another person's position to better understand and see the world from that person's perspective, and as a result it is critical for the researcher to be truthful to participants when analysing, managing and presenting the data. With qualitative data, data analysis is an on-going process as the researcher doesn't have to wait for the data collection process to be completed first before commencing with data analysis. Instead, the researcher can analyse data throughout the data collection process. However, the researcher must first transform the data in the format that can be analysed effortlessly (Dawson, 2009:116).

If the researcher is using an interview or a focus group as data collection method, it is useful to complete a summary form pertaining all the details about the interview at the end of the session and attach it to the transcripts to assist with data analysis (Dawson, 2009:132). All audio recorded interviews need to be transcribed verbatim to convert spoken words into written words to facilitate data analysis (Sutton & Austin, 2015:228). Furthermore, Sutton and Austin (2015:228) indicate that once transcription has been completed, the researcher needs to read it while listening to the recording to:

- ❖ correct any errors and spelling mistakes;
- ❖ insert notation for any pauses, laughter and looks of discomfort;
- ❖ ensure anonymity by ensuring that names, places, significant events associated with participants cannot be identified or linked to them;
- ❖ add punctuations such as full stops, periods, quotations, commas; and
- ❖ include any other contextual information such as temperature or discomfort of the room that might affected the participants.

According to Yin (2018:175) when analysing and reporting on the interviews, the researcher's text would frequently have to point out the self-reported nature of the data by using phrases such as "as stated in the interviews", "as reported by the interviewees" or "she/he reported that …". Furthermore, Creswell (2013:59) advocates that when researchers are reporting the analysed data they must communicate clearly in a straightforward manner and in a significant language. Additionally, researchers should avoid any form of plagiarism, avoid disclosing information that would harm the participants and falsifying evidence, data, authorship, findings and conclusions (Creswell, 2013:59).

Qualitative data analysis methods such as thematic analysis, discourse analysis, content analysis and comparative analysis could be used by researchers to analyse data (Dawson, 2009:132). Researchers can use transcription editing software such as ELAN (EUDICO Linguistic Annotator) to assist with keeping data organised by linking media and data files. especially when videotaping interviews is complemented by transcriptions (Sutton & Austin, 2015:228). Furthermore, Saldaña (2013:29) and (Silverman, 2013:498) state that Computer-Aided Qualitative Data Analysis Software (CAQDAS) such as ATLAS.ti, Nvivo, AnSWR, Transana, Qualrus, etc. have statistical capabilities such as code frequency counts, word frequency counts, and the matrix display of qualitative data in Excel spreadsheet to assist with analysing the data. Qualitative data analysis software assist researchers in implementing the core feature of qualitative data analysis called coding, which involves grouping evidence and labelling ideas so that they provide an increasingly broader perspective (Creswell & Plano Clark, 2018:321). CAQDAS provides various advantages such as: speed in handling large volumes of data, demonstrating that research conclusions are based on rigorous analysis and facilitation of team research, as well as the development of consistent coding schemes (Silverman, 2013:506).

Furthermore, Saldaña (2013:9) states that researchers use coding to organize and group similarly coded data that share specific characteristics into categories or themes. According to Creswell and Plano Clark (2018:321), during coding the researcher breaks the text down into small units such as sentences, phrases or paragraphs, assigns a code to each unit, and then sorts the codes into themes. In qualitative inquiry, a code is a short phrase or word that "*symbolically assigns a summative, salient or evocative*" attribute for a portion of data and captures the essence of the data (Saldaña, 2013:3).

In this case study, the researcher transcribed all audio recorded interviews and used Computer-Aided Qualitative Data Analysis Software (CAQDAS), ATLAS.ti to analyse the data by grouping similarly coded data that share some characteristics into categories. Additionally, SPSS was used to analyse responses from the questionnaire.

3.7 Design science research

Design science is an information system paradigm associated with constructing innovative and high-quality artefacts to solve real-world problems. The researcher will follow DSR as an appropriate methodology in this study, because the objective of this study is to construct an IT artefact to solve a real-world organisational issue. According to Hevner *et al.* (2004:85) and Vom Brocke *et al.* (2020:1), the objective of design science research in information systems is to construct artefacts to solve critical organisational problems, make contributions to the research field study, evaluate the design and communicate the outcome to a relevant audience. The importance of design science research is widely addressed in the information systems discipline by researchers to assist with constructing new IT artefacts to solve real-world organisational problems and to increase knowledge (Hevner & Chatterjee, 2010:5; Vaishnavi *et al.*, 2004/2019:1; Vom Brocke *et al.*, 2020:1).

Vaishnavi *et al.* (2004/2019:1) define design science research as a "*set of synthetic or analytical techniques and perspectives for performing research in information systems*" to produce new knowledge and insight through constructing and evaluating artefacts. Furthermore, Hevner and Chatterjee (2010:5) is of the view that design science research is a research paradigm used by researchers to construct innovative artefacts to answer questions appropriate to human problems thereby contributing new knowledge to the body of scientific evidence. The design science paradigm "*seeks to extend the boundaries of human and organisational capabilities by creating new and innovative artefacts*" (Hevner *et al.*, 2004).

Artefacts are synthesized, can be distinguished based on goals, functions and adaptations, may emulate the appearances of natural things, and are frequently discussed in terms of both imperatives and descriptives (Hevner & Chatterjee, 2010:24). Furthermore, an artefact is an object that is artificial and constructed by humans to solve an important problem (Hevner & Chatterjee, 2010:5). Design science research is grounded in existing ideas acquired from knowledge described by experts and specialists in a particular field (Hevner & Chatterjee, 2010:18). In information systems research, IT artefacts are widely referred to as constructs ("*vocabulary and symbols")*, models ("*abstractions and representations*"), methods ("*algorithms and practices*"), instantiations ("*implemented and prototype systems*") and better design theories (Hevner & Chatterjee, 2010:5).

Researchers use design science research to address crucial unsolved problems in a unique or innovative manner by developing IT artefacts. Thus, the core principle of design science research is that knowledge, understanding and solutions to the design problem are obtained in the development and application of an artefact (Hevner & Chatterjee, 2010:5).

3.7.1 The design science research framework

The conceptual design science research framework presented in Figure 3-4 enables researchers to understand, execute and evaluate design science research (Hevner *et al.*, 2004). The DSR framework involves the environment, design, and knowledge base stages. According to Vom Brocke *et al.* (2020:3), the environment represents the problem space in which the phenomena of interest is found, and is made up of people, organisations, as well as existing or planned technologies. The design activities include multiple iterations of developing and evaluating the artefacts. Furthermore, Vom Brocke *et al.* (2020:3) state that the relevance of the research can be assured by positioning each research activity to address the business needs.

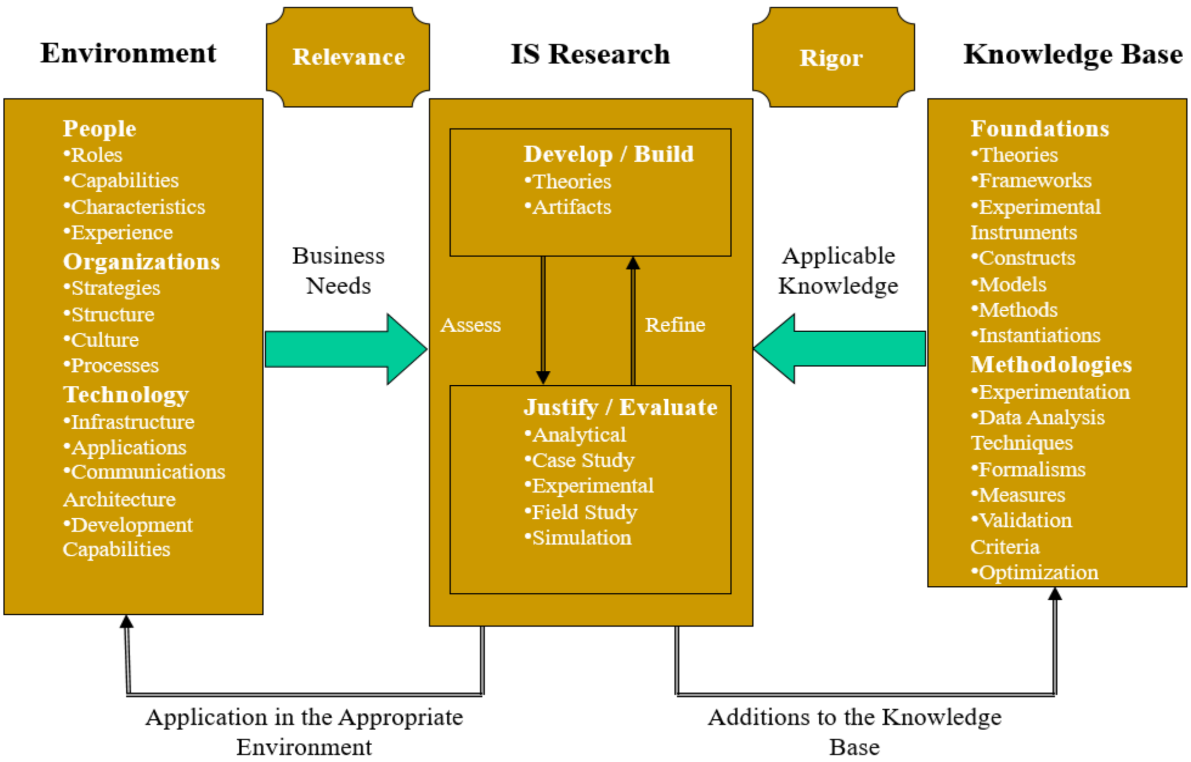


Figure 3-4: Design Science Research Framework (Hevner, 2018:8).

The knowledge base represents the methodologies involved in the evaluate phase and then foundations used in the build phase of a research study. Consequently, Vom Brocke *et al.* (2020:4) explain that existing methodologies and foundations can be applied by researchers to achieve rigour.

3.7.2 Guidelines for design science research in information systems

Hevner (2018:9) outlines eight guidelines for design science in information system research to assist researchers in determining a solution for the research project:

- **Guideline 1: Design an artefact –** design science research must produce a feasible artefact in the form of a model, method, construct or an instantiation to address a critical organisational problem.
- **Guideline 2: Problem relevance –** the goal of design science research is to solve critical and relevant business problems by developing technology-based solutions.
- **Guideline 3: Design evaluation –** appropriate and well-executed evaluation methods will be used to rigorously demonstrate the efficacy, utility and quality of a design artefact.
- **Guideline 4: Research contributions –** design science research must produce a verifiable and understandable contribution in the areas of the design artefacts, design foundations and design methodologies.

- **Guideline 5: Research rigor –** application of rigorous methods are critical in the development and evaluation of the design artefact.
- **Guideline 6: Design as a search process –** the utilization of available means to reach desired ends while satisfying laws in the problem environment are crucial when searching for an effective artefact.
- **Guideline 7: Communication of research –** design science research must be presented to both management-oriented and as well as technology-oriented audiences.

However, the most important guideline in this list is to produce an artefact to address a critical organisational issue (Hevner *et al.*, 2004:82). Furthermore, relevant quality attributes such as functionality, completeness, consistency, accuracy, reliability, performance, usability and fit with organisation can be used to evaluate the IT artefacts (Hevner *et al.*, 2004:85).

3.8 Three design science research models

This section will explore design science research models that researchers can follow when designing an artefact as a solution in solving a real-world problem. Three design science research (DSR) models were considered for this research study: design science research model (DSRM) (Peffers *et al.*, 2007:48), design science research cycles (Hevner & Chatterjee, 2010:16), and design science research model (Vaishnavi *et al.*, 2004/2019:14).

3.8.1 Design science research cycles (Peffers *et al.*, 2007:48)

Peffers *et al.* (2007:35) developed the DSRM (Fig. 3-5) to be consistent with concepts in prior literature about design science in IS, to provide guidance to researchers, to provide a nominal process for conducting design science research, and to provide a mental model for the presentation of its results. This design science model comprises six activities:

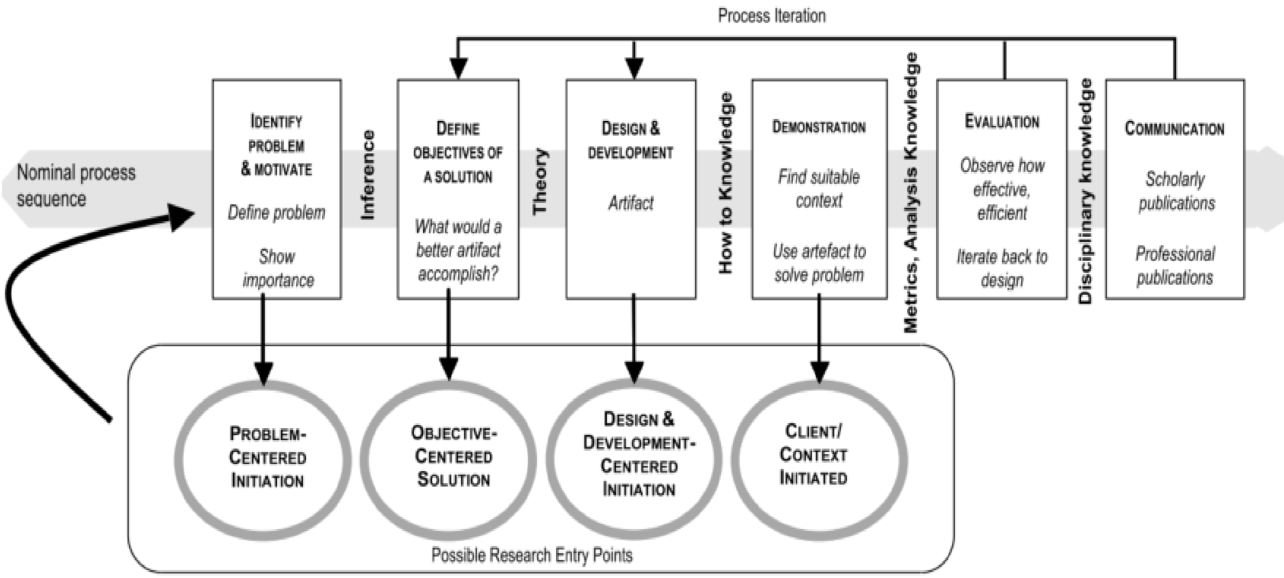


Figure 3-5: Design Science Research Methodology Process Model (DSRM) (Peffers *et al.*, 2007:48).

Activity 1: Problem identification and motivation

In this activity, a specific research problem will be identified and clearly defined to assist in defining the requirements that will be needed for the development of the artefact in order to provide a solution. Additionally, the value of the solution must be justified to assist with understanding the reasoning linked to the researcher's understanding of the problem, and additionally, to motivate the audience and the researcher to pursue the solution and accept the results (Peffers *et al.*, 2007:12). However, the identified problems do not automatically translate directly into objectives for the artefact because the design process is necessarily one of partial and incremental solution. Resources required are knowledge of the state of the problem and the significance of its solution.

Activity 2: Define the objectives for a solution

In this activity, the researcher must conclude the objectives for possible solutions from the problem definition and knowledge based on what is attainable and feasible. Furthermore, Peffers *et al.* (2007:12) state that the objectives should be deduced rationally from the problem specification either qualitatively or quantitatively to increase knowledge. Resources needed to achieve this include knowledge of the state of the problem and any existing solutions.

Activity 3: Design and development

In this activity, the researcher must create the artefact by determining the desired functionality of the artefact and its architecture and then developing the actual artefact. Resources needed include knowledge of theory for constructing the artefact.

Activity 4: Demonstration

In this activity, the functionality of the artefact must be clearly demonstrated, which could include its use in case study, simulation, proof, experimentation and other significant activities. Resources needed include effective knowledge on how to use the artefact or theory to resolve the problem.

Activity 5: Evaluation

The functionality of the artefact must be evaluated by observing, assessing and measuring how efficient the artefact supports a solution to the problem. This activity requires comparison of the defined objectives of the solution to the actual observed results from the functionality of the artefact as in the demonstration. Evaluation of the artefact can take any form such as quantifiable measures of system performance for example, response time or availability, the results of a satisfaction survey, client feedback or simulation, and comparison of the artefact's functionality with the solution objectives.

Once this activity is completed, the researchers can decide to try to improve the effectiveness of the artefact or leave further improvements to subsequent projects.

Activity 6: Communication

In this activity, the problem and its significance, the artefact, the rigor of the design, its utility and novelty, its effectiveness to researchers and other audiences must be communicated. The layout of this process can be adopted by the researcher in, for example, scholarly research publications. Resources needed will include knowledge of disciplinary culture.

3.8.2 Design science research cycles (Hevner & Chatterjee, 2010:16)

In any design science project, the researcher can gain insight by identifying and understanding the three important design science research cycles: relevance, rigor and design (Hevner & Chatterjee, 2010:16). It is worth noting that in every design science study the three research cycles must exist and be clearly identifiable. However, before an artefact could be released into field testing in the relevance cycle, it needs to be first tested in laboratory and experimental situations (Hevner & Chatterjee, 2010:19).

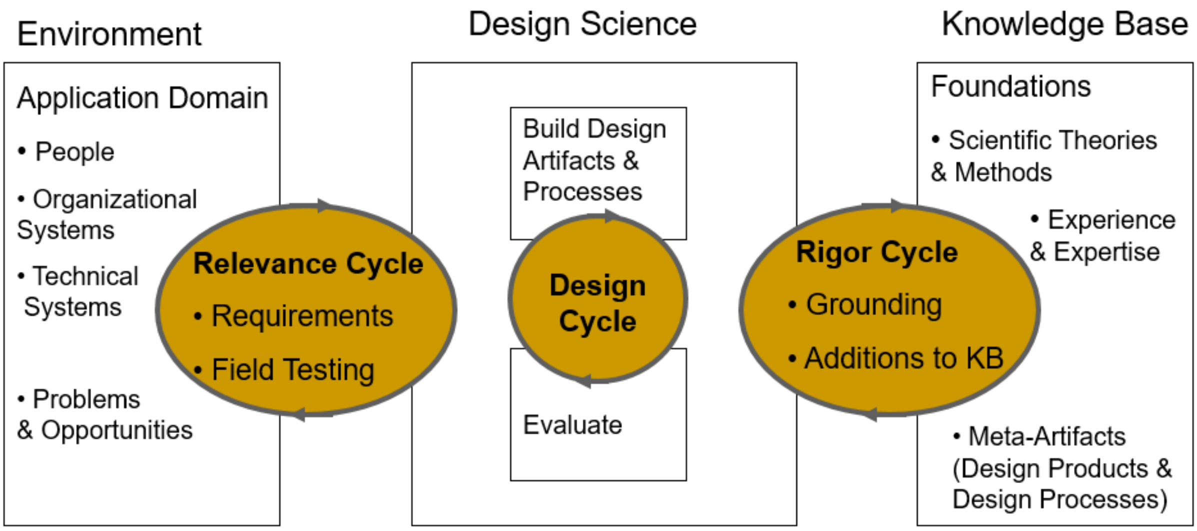


Figure 3-6: Design science research cycles (Hevner, 2018:10).

Figure 3-6 outlines the three design science research cycles that can assist the researcher to address a problem in design research in order to develop a solution.

Relevance cycle

The relevance cycle connects the design science activities with the contextual environment of the research project (Hevner & Chatterjee, 2010:17). Design science research commences in the relevance cycle with an application domain that provides the requirements for the research as input and describes justifiable criteria that will be applied in the evaluation of the artefact (Hevner & Chatterjee, 2010:17; Hevner, 2018:11). The problems and opportunities will be outlined in this cycle. The application environment includes organisational systems, people and technical systems that work together to solve a real-world problem.

One question that can be asked in this phase is "*Does the design artefact improve the environment and how can this improvement be measured*" (Hevner & Chatterjee, 2010:17). The outcome from design science research must be returned to the study environment application context for evaluation. After field testing, the results will determine if additional iterations of the relevance cycle are required. The relevance cycle will iterate to determine the deficiencies in the artefact qualities or behaviour, to restate the research requirements and provide feedback from field testing evaluation into research (Hevner, 2018:11).

Design cycle

The central design cycle is the "*heart of any design science research project*" and it is where the tedious work of design science research is conducted (Hevner & Chatterjee, 2010:17). The researcher needs to understand the dependencies of the design cycle on the relevance cycle and the rigor cycle, while admiring its relative independence during the actual execution of the research (Hevner & Chatterjee, 2010:17). The design cycle iterates between the fundamental activities of constructing and evaluating the design artefact (Hevner & Chatterjee, 2010:17; Hevner, 2018:13). Additionally, the design cycle activities iterate quickly between the construction of the artefact, its evaluation and successive feedback to enhance the design further. In this cycle, the researcher creates and refines the artefact as a product and process which will undergo a rigorous evaluation in a laboratory or in a controlled environment (Hevner, 2018:13). Good design science research must produce results of interest to both technology-focused audiences and management-focused audience (Hevner & Chatterjee, 2010:19). Consequently, a balance must be sustained between the effort spent in developing and evaluating the evolving artefact during then performance of the design cycle. Hevner (2018:13) emphasises that the design cycle will continue until the artefact is ready for field testing in the application environment and new knowledge relevant which must be included in the knowledge base is achieved.

Rigor cycle

The rigor cycle links the design science activities with the knowledge base's scientific experiences, foundations and expertise that informs the research project (Hevner & Chatterjee, 2010:17; Hevner, 2018:12). Design science draws from extensive knowledge bases of scientific theories and engineering methods that provide the grounding for rigorous design science research. Furthermore, this cycle links previous knowledge to the research project to provide for innovative design.

Additionally, Hevner (2018:12) explains that the knowledge base includes two types of additional knowledge:

- The experiences and expertise that describe the state of the art in the application domain of the research; and
- The existing artefacts and processes found in the application environment.

However, Hevner and Chatterjee (2010:17) urge that the design decision and design processes based on grounded behaviour or mathematical theories may not be feasible or relevant for a cutting-edge artefact, because such theories may be undiscovered or incomplete.

The ability to contribute to the knowledge base is important and will include any additions or extensions made to the methods and theories during the research, the creation of the new artefact, all experiences acquired from performing the iterative design cycle, and field testing the artefact in the application context (Hevner, 2018:12).

Design science research checklist

The checklist developed by Hevner and Chatterjee (2010:20) is available to researchers to assess the design research project and to ensure that their projects address the primary aspects of DSR.

Table 3-6: Eight design science research checklist questions – adopted from (Hevner & Chatterjee, 2010:20).

Questions
1. What is the research question (design requirements)? 2. What is the artefact? How is the artefact represented? 3. What design processes (search heuristic) will be used to build the artefact? 4. How are the artefact and the design processes grounded by the knowledge base? What, if any, theories support the artefact and the design process? 5. What evaluations are performed during the internal design cycles? What design improvements are identified during each design cycle? 6. How is the artefact introduced into the application environment and how is it field tested? What metrics are used to demonstrate artefact utility and improvement over previous artefacts? 7. What new knowledge is added to the knowledge base and in what form (e.g., peer reviewed literature), meta-artefacts, new theory and new methods)? 8. Has the research question been satisfactory addressed?

The eight checklist questions are presented in the three design research cycles (Figure 3-7) and listed (Hevner & Chatterjee, 2010:20).

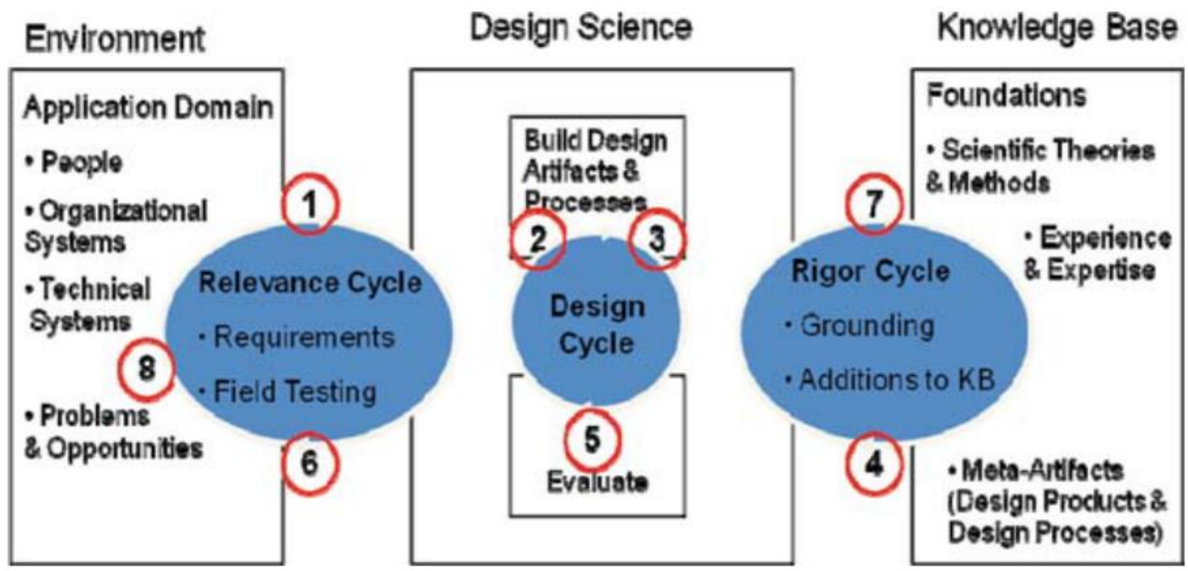


Figure 3-7: Checklist questions mapped to the three design research cycles (Hevner & Chatterjee, 2010:20).

3.8.3 Design Science Research Process Model (Vaishnavi *et al.*, 2004/2019:14)

The DSR model developed by Vaishnavi *et al.* (2004/2019:12) as outlined in Figure 3-8 uses five primary phases for developing an artefact: (1) awareness of the research problem, (2) suggestion of the problem solution, (3) development of the artefact, (4) evaluation of the artefact, and (5) conclusion on the research.

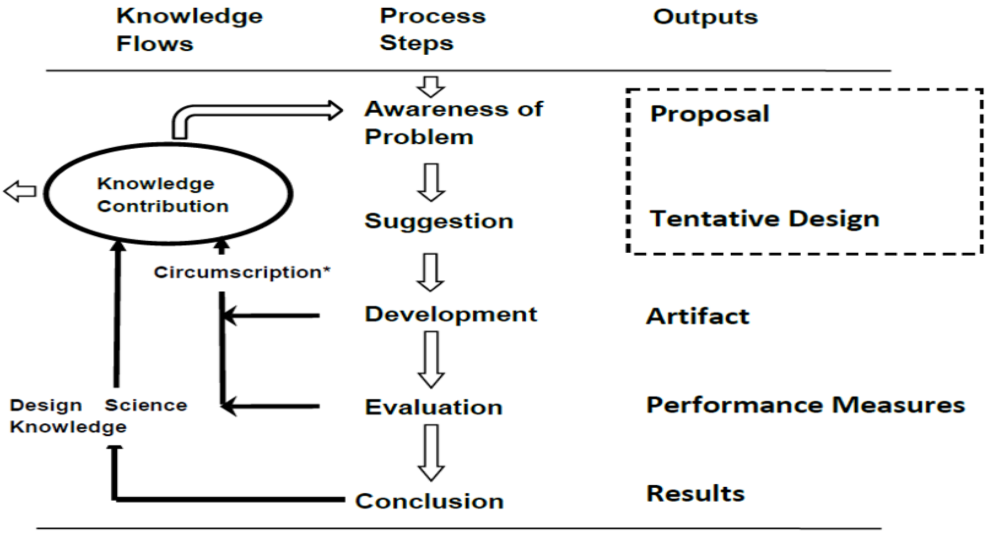


Figure 3-8: Design Science Research Process Model (Vaishnavi *et al.*, 2004/2019:14).

Awareness of the problem

The main objective of this phase is to identify a compelling research problem that requires a solution. A formal or informal proposal for a new research effort is produced.

Suggestion

Once the proposal has been developed, the researcher will explore feasible methods that can be used to achieve the goal. The proposal will guide the researcher in this phase and a tentative design will be the output of this phase.

Development

The objective of this phase is to develop the preliminary design of the solution. The tentative design produced in the suggestion phase will be used to construct the artefact. This may require software development probably utilising a high-level package or tool. The implementation technique will depend on the type of artefact being developed. An artefact will be produced at the end of this stage.

Evaluation

In this phase the artefact is evaluated based on the functional specifications documented in the problem awareness and identification phase. Development, evaluation, and suggestions are performed iteratively as indicated by the circumscription arrow in the research model. According to Vaishnavi *et al.* (2004/2019:15), this enables the deductive cognitive process as a supplementary premise regarding the artefact and its environment studied and/or known.

Conclusion

The conclusion phase informs the end of the project and the researcher reflects on what worked, what was learned, and what did not work to solve the problem. Communication is crucial in this stage as the researcher must communicate the outcome or the results of the project. Vaishnavi *et al.* (2004/2019:15) emphasise that abstraction, communicating the results, and contributing to the larger knowledge base allow the researcher to reach broad and appropriate conclusions based on knowledge acquired from the research project. The output of design science research includes models, methods, frameworks, instantiations, architectures, constructs, and design theories.

3.8.4 Preferred research model for the study

The design science research process model by Vaishnavi *et al.* (2004/2019:14) will be used in this research study to assist with solving a real-world problem in one SA SME. The model will guide the researcher in accomplishing the technical objectives of this study. This model as outlined in section 3.8.3 uses five phases, which include problem awareness and identification, suggestions, design, evaluation, and conclusion.

The core practical objective of this study is to develop an artefact to solve a data management and interface real-world problem in one SA SME. The process will involve these five phases until the end goal is reached. However, the development, evaluation and suggestions are performed iteratively until the artefact for capturing data into the database is complete. The outcome of the project will be a functional artefact and guidelines to assist the organisation with improving data quality and management. The researcher will communicate the output of the project not only to the organisation, but also through the research report to the academic field indicated in the research, and thus contribute to the larger knowledge base.

3.9 Evaluation of design science research artefacts

In design science research, the evaluation of design artefacts and theories is regarded as a fundamental activity that enables the researcher to assess and validate the artefact to obtain feedback and a better understanding of the problem for further development, to enhance the quality of the product and the design process, and to assure the rigour of the research (Hevner *et al.*, 2004:78; Venable *et al.*, 2017:77). The evaluation of design science research artefacts is a crucial process of assessing, validating and measuring the worth and significance of programs and products developed to serve as solutions for practical problems (Cleven *et al.*, 2009:1; Sonnenberg & Vom Brocke, 2012:1; Vom Brocke *et al.*, 2020:8). According to Hevner *et al.* (2004:85), researchers must apply well-executed evaluation methods to rigorously demonstrate the quality, efficacy and utility of the artefact.

Prior work has reported numerous evaluation methods and criteria as well as frameworks that can assist researchers conducting DSR studies to evaluate artefacts that were developed to address particular situations (Cleven *et al.*, 2009:2; Hevner *et al.*, 2004:85; Sonnenberg & Vom Brocke, 2012:5; Venable *et al.*, 2017:77; Vom Brocke *et al.*, 2020:8). Table 3-7 outlines various design evaluation methods that can be used.

Table 3-7: Design Evaluation Methods – adopted from (Hevner ***et al.***, 2004:86).

Design Evaluation Methods	
1. Observational	**Case study:** Involves an in-depth study of the artefact in an organisation or business environment.
	Field study: Observe the use of the artefact in various projects.
2. Analytical	**Static analysis:** Assess the artefact structure for static qualities such as complexity.
	Architecture analysis: Investigate the relevancy of the artefact into the technical IS architecture.
	Optimisation: Illustrate the inherent optimal properties of the artefact or supply optimality bounds on artefact behaviour.
	Dynamic analysis: Investigate the artefacts in use to determine dynamic qualities such as performance.
3. Experimental	**Controlled experiment:** The researcher studies the artefact in a controlled environment for qualities such as usability.
	Simulation: Execute the artefact using artificial data.
4. Testing	**Functional (black box) testing:** The researcher executes the interfaces of the artefacts to determine failures and identify defects.
	Structural (White box testing): The researcher performs a coverage testing of some metrics such as execution paths in the artefact implementation.
5. Descriptive	**Informed argument:** The researcher can apply information from the knowledge base such as relevant research to build a compelling argument regarding the artefact's utility.
	Scenarios: Create detailed scenarios to demonstrate the utility of the artefact.

Venable *et al.* (2017:77) developed a framework for evaluation design science (FEDS) to assist DSR researchers (especially novice researchers) to decide on an appropriate strategy or strategies to evaluate the artefacts developed within their DSR projects to address a particular phenomenon. The FEDS addresses two important aspects of evaluation strategies: (1) the functional purpose of the evaluation which includes formative and summative evaluation; and (2) the paradigm of the evaluation which includes artificial and naturalistic evaluation (Venable *et al.*, 2017:80). The framework is concerned about the functionality of the artefact in solving a problem and the quality of the knowledge contribution (Venable *et al.*, 2017:80). Furthermore, Venable *et al.* (2017:82) propose a four step process that researchers can adopt in choosing an evaluation strategy for a particular DSR project: (1) explicate the goals of evaluation, (2) choose the evaluation strategy, (3) determine the properties to evaluate, and (4) design the individual evaluation episode.

Consequently, Cleven *et al.* (2009:2) emphasise that DSR artefacts can be evaluated using variables and their respective values integrated and depicted in a morphological field as, outlined in Figure 3-9. Researchers can use variables together with their values to evaluate the artefacts.

Variable	Value				
Approach	Qualitative	Quantitative			
Artifact Focus	Technical	Organizational	Strategic		
Artifact Type	Construct	Model	Method	Instantiation	Theory
Epistemology	Positivism	Interpretivism			
Function	Knowledge function	Control function	Development function	Legitimization function	
Method	Action research	Case study	Field experiment	Formal proofs	
	Controlled experiment	Prototype	Survey		
Object	Artifact	Artifact construction			
Ontology	Realism	Nominalism			
Perspective	Economic	Deployment	Engineering	Epistemological	
Position	Externally	Internally			
Reference Point	Artifact against research gap	Artifact against real world	Research gap against real world		
Time	Ex ante	Ex post			

Figure 3-9: Variables and values for the evaluation of DSR artefacts (Cleven *et al.*, 2009:4).

Hevner *et al.* (2004:85) argue that reliability, completeness, usability, functionality, accuracy, performance and fit with the organisation are some of the attributes that are used to evaluate the IT artefacts. Gregor and Hevner (2013:350) state that artefacts are evaluated to prove their usefulness using criteria such as efficacy, validity, utility, and quality. According to Sonnenberg and Vom Brocke (2012:5), DSR artefacts can be evaluated using the criteria outlined in Table 3-8. The evaluation criteria can be used by researchers when they need to determine the progress accomplished by designing, constructing, and utilising the artefact in relation to the identified problem that they are trying to solve as well as the design objective to be reached.

Table 3-8: Evaluation criteria for DSR artefacts – adopted from (Sonnenberg & Vom Brocke, 2012:5).

Criteria	Construct	Model	Method	Instantiation
Completeness	X	X		
Ease of use	X		X	
Effectiveness				X
Efficiency			X	X
Elegance	X			
Fidelity with real world phenomena		X		
Generality			X	
Impact on the environment and on the artefacts' users				X
Internal consistency		X		
Level of detail		X		
Operationality			X	
Robustness		X		
Simplicity	X			
Understandability	X			

Subsequently, Sonnenberg and Vom Brocke (2012:5) developed a framework for evaluation in design science (FEDS) which researchers can use to evaluate their DSR artefacts.

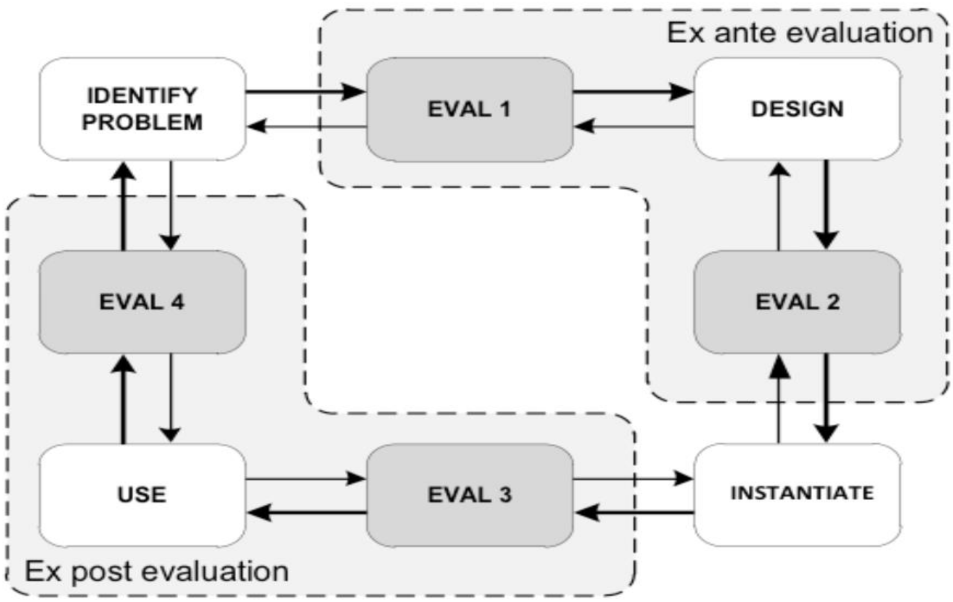


Figure 3-10: Cyclic high level evaluation activities within DSR process (Sonnenberg & Vom Brocke, 2012:8; Vom Brocke *et al.*, 2020:8).

Sonnenberg and Vom Brocke (2012:8) state that the cyclic high-level DSR process includes activities such as problem identification, design, construction and use, which are followed by an evaluation activity together with feedback loops from the evaluation activity to the preceding design activity as outlined in Figure 3-10.

- ❖ Evaluation 1 – The criteria used to evaluate the problem identification includes importance, feasibility, and novelty.
- ❖ Evaluation 2 – The criteria used to evaluate the solution design includes clarity, simplicity and consistency.
- ❖ Evaluation 3 – The criteria used to evaluate the solution instantiation includes robustness, ease of use and fidelity with the real-world phenomena and.
- ❖ Evaluation 4 – The criteria used to evaluate the solution in use includes efficiency, effectiveness, and external consistency.

According to Vom Brocke *et al.* (2020:9), researchers can decide to perform an evaluation before the instantiation of the artefact (*ex ante*) or perform the evaluation after the instantiation of the artefact (*ex post*).

In this research study, a combination of the attributes such as reliability, completeness, usability, functionality, accuracy, performance and fit with the organisation presented by Hevner *et al.* (2004:85) and the evaluation criteria by Sonnenberg and Vom Brocke (2012:5) as illustrated in Table 3-8 were used to evaluate the artefact. Additionally, the framework for evaluation in design science (FEDS) in Figure 3-10 was applied in this study.

3.10 Ethics in design research

Nowadays, research institutions and universities have made it a prerequisite for researchers undertaking a research study that involves real people to obtain permission from their own institutional review board or human objects' ethics committee (Myers & Venable, 2014:3). Ethics are principles used for regulating any form of misconduct in the research study. According to Denscombe (2010:339), ethics are there to protect any individual who will be participating in a research study from any form of harm and from researchers who might exploit participants.

In this study, a written permission and agreement between NWU and one SA SME to participate in this study was obtained to ensure the legitimacy of the study. A written permission was obtained to grant the researcher access to their premises, and access their systems and data. System access credentials such as usernames and passwords given to the researcher were handled with care throughout the study to avoid any unauthorised access to their systems. The researcher ensured that all sensitive and personal data were protected and adhered to according to the protection of personal information act (POPI Act). Personal information of participants was anonymised throughout the research study and a trusted employee in one SA SME acted as a gatekeeper for the study. Additionally, the researcher ensured that the organisation and the participants could not be identified by referring to the organisation as “one SA SME”, and to participants as “P1” and “P2”.

Furthermore, to ensure that confidentiality, anonymity, and privacy was maintained throughout the research study, a written code of conduct was compiled and signed by the researcher and the gatekeeper of the SA SME. Ethical clearance from NWU was obtained to perform the study and to ensure that participants were protected from any harm and exploitation. Confidentiality was maintained throughout the research study. Participation was voluntary and informed consent forms were signed by participants to provide written consent to participate in the study. In a case where data of sensitive nature were collected, the researcher signed an informed consent form that all original recordings will be destroyed after transcription. Transcriptions and data collected during the process will be kept in a secure location for five years and be destroyed after the time period has passed.

The researcher applied the principles of Denscombe (2010:339) by ensuring that the research study is conducted in an ethical manner, that an open, transparent and honest relationship with participants is maintained, that no participants are coerced to participate in the research study, and that participants are protected from any form of harm. The evaluation of the artefact was conducted ethically. The researcher allowed the instantiated artefact to be evaluated by participants in reality, not only focusing on the utility aspect, but also taking into consideration the quality of knowledge that the artefact will contribute (Venable *et al.*, 2017:87).

The researcher was open and explicit about the nature of the research study, and was truthful to participants about their roles in the study. The researcher communicated her intentions with the collected data for investigation. The research study complied with the laws of the land. The researcher obtained written permission of the study from one SA SME, obtained ethical clearance from NWU, and was careful with matters related to ownership of the data, intellectual property rights and copyrights. Furthermore, the researcher ensured that sensitive or personal data are secured and kept private and do not violate the POPI Act.

Myers and Venable (2014:11) recommend six ethical principles for design science research in information systems as outlined in Table 3-9, and suggest that the application of the six principles and ethical issues should be considered very early in the research study prior to conducting or implementing any activities that may impact individuals.

Table 3-9: A proposed set of ethical principles for design science research in information systems (Myers & Venable, 2014:11).

	Ethical Principle	Explanation
1.	Public interest	Once the artefacts are implemented and operational, the design science researchers should explicitly identify all stakeholders who might be affected by them. Any form of benefits and harm that they may encounter should be critically taken into consideration.
2.	Informed consent	Design science researchers need to obtain informed consent from all individuals participating in the research study.
3.	Privacy	Appropriate safeguards should be put in place to protect privacy of participants in the research study including those who might utilise or be affected by any software developed.
4.	Honesty and accuracy	Design science researchers should acknowledge inspiration from other sources rather than plagiarising ideas
5.	Property	An agreement regarding the ownership of the intellectual property right must be discussed at the beginning of the project. The ownership of any information collected should be determined including the rights that the researcher has in publishing the findings.
6.	Quality of the artefact	Design science researchers must ensure that the artefact is of quality and any risks must be addressed. Evaluation and testing must be adequately rigorous.

Based on Table 3-9, the researcher followed the six ethical principles of DSR in ensuring that a high-quality artefact is developed, rigorously tested and meets the expectations of the participants. In this study, written permission in the form of an agreement between NWU and one SA SME to participate in this study was obtained to ensure the legitimacy of the study. Written permission was also obtained to grant the researcher access to their premises, and access their systems and data. System access credentials such as username and password given to the researcher were handled with care throughout the study to avoid any unauthorised access to their systems. The researcher ensured that all sensitive and personal data were protected and adhered to, according to the protection of personal information act (POPI Act). Personal information of participants was anonymised throughout the research study and a trusted employee in one SA SME acted as a gatekeeper for the study. Additionally, the researcher ensured that the organisation and the participants could not be identified by referring to the organisation as “one SA SME” and to participants as “P1” and “P2”.

Furthermore, to ensure that confidentiality, anonymity, and privacy was maintained throughout the research study, a written code of conduct was compiled and signed by the researcher and the gatekeeper of the SA SME. Ethical clearance from NWU was obtained to perform the study (NWU-01780-20-A9) and to ensure that participants were protected from any harm and exploitation. Confidentiality was maintained throughout the research study. Participation was voluntary and informed consent forms were signed by participants to provide written consent to participate in the study. In a case where data of sensitive nature were collected, the researcher signed an agreement form that all original recordings would be destroyed after transcriptions. Transcriptions and data collected during the process will be kept in a secure location for five years and will be destroyed afterwards.

The researcher applied the principles of Denscombe (2010:339) by ensuring that the research study is conducted in an ethical manner, that an open, transparent and honest relationship with participants is maintained, that no participants are coerced to participate in the research study, and that participants are protected from any form of harm. The evaluation of the artefact was conducted in a rigorous but ethical way. The researcher allowed the instantiated artefact to be evaluated by participants in reality, not only focusing on the utility aspect but also taking into consideration the quality of knowledge that the artefact will contribute (Venable *et al.*, 2017:87).

Once the artefact is fully implemented and operational, the researcher will critically take into consideration any form of benefits and harm that may be encountered in using the artefact and address them. The participants and stakeholders who will be using the artefact will give the researcher feedback on the functionality, challenges and risks associated with the artefact. All issues will be documented and addressed.

The researcher must have a clear understanding of ethical principles and ensure that they are observed. However, researchers have the freedom to decide which set of ethical principles are more significant in a particular situation than others.

3.11 Study plan

This section of the study outlines the roadmap the researcher followed in addressing data quality and management issues in one SA SME. The main objectives of the study were to construct an artefact to store and manage data for data analytics, and subsequently propose guidelines and technologies for improvement of data quality in one South African SME.

The theoretical objectives of the study included:

- ❖ To explore technologies that SMEs can implement to better manage and store their data;
- ❖ To investigate the impacts of poor data quality on SMEs;
- ❖ To identify the challenges of poor data quality and benefits of high data quality in SMEs; and
- ❖ To explore different BI models, data warehousing and ETL concepts suitable for SMEs.

The practical objectives of the study were to:

- ❖ Propose guidelines and technologies for improvement of data quality in one South African SME; and
- ❖ Design an artefact for one South African SME to store and manage data for data analytics.

Theoretical objectives

The researcher explored the literature to investigate the impacts of poor data quality, to identify challenges of poor data quality, benefits of high data quality in SMEs, explored technologies that SMEs can implement to better manage and store data, and to explore different BI models, data warehousing and ETL concepts suitable for SMEs. Thus, to construct the foundation and motivate the significance of the research study, the researcher explored the following concepts:

- ❖ Data, information, knowledge;
- ❖ Data quality;
- ❖ Data management;
- ❖ Small and medium sized enterprises; and
- ❖ Data analytics.

Existing literature was reviewed to gain a thorough understanding and insight into the problem of the research through exploring concepts associated with this study, and to set the tone of the research study (Chapter 2).

Furthermore, the researcher used the literature as well as applied lessons learned to develop the artefact to formulate guidelines that one SA SME can use to improve data management and data quality. These guidelines will not be beneficial only to one SA SME only, but will be a great contribution to the knowledge base such that other organisations can adopt and use them to improve data quality and management issues.

Study approach

The design science research methodology and a case study research strategy were followed in this study to answer the primary research question:

- ❖ How can SMEs exploit technology to better manage data and improve the quality of their data for data analytics?

The researcher chose the DSR approach because the researcher wanted to find a solution to a real-world problem through the design and development of an artefact. Additionally, this research approach enabled the researcher to use mixed method data collection methods such as observation, interviews and questionnaires to understand the beliefs, perceptions, feelings and experiences of research participants in order to develop a better understanding of the phenomenon. Furthermore, guidelines were compiled to improve data management in one SA SMEs for data analytics.

The research strategy for this study was a case study research because the researcher was investigating a single phenomenon and conducting an in-depth investigation of a real-life situation. According to Yin (2018:9), a case study seeks to answer ‘how’ and ‘why’ questions, and in this study the researcher sought to understand how one SA SME can exploit technology to better manage and improve data quality for data analytics and reporting purposes. The researcher followed the four phases of conducting a case study by Rashid *et al.* (2019:2) as outlined in Chapter 3 section 3.4.4.

Semi-structured interviews and a questionnaire were used to collect data. All data collected from the interviews were transcribed and analysed using ATLAS.ti. Statistical software for the social sciences (SPSS) was used to analyse data collected using questionnaires. The qualitative data collected were analysed using content analysis, in particular Zhang and Wildemuth (2009:3) guidelines for content analysis of data.

Design science research for this study

The design science research structure for my study (Figure 3-11) was guided by the design science research process model of Vaishnavi *et al.* (2004/2019:14). The model contains the main research cycle and three sub-cycles. The purpose of this model was to guide the researcher to answer the research questions and develop a solution for the identified research problem in developing an artefact to solve a real-world problem. In the awareness phase of the main DSR cycle, the problem was identified and described, after which research objectives were compiled. Suggestions of how to answer the main research question were addressed in three research sub-cycles. Each research sub-cycle represents a research phase of problem awareness and identification phase, suggestion and development phase.

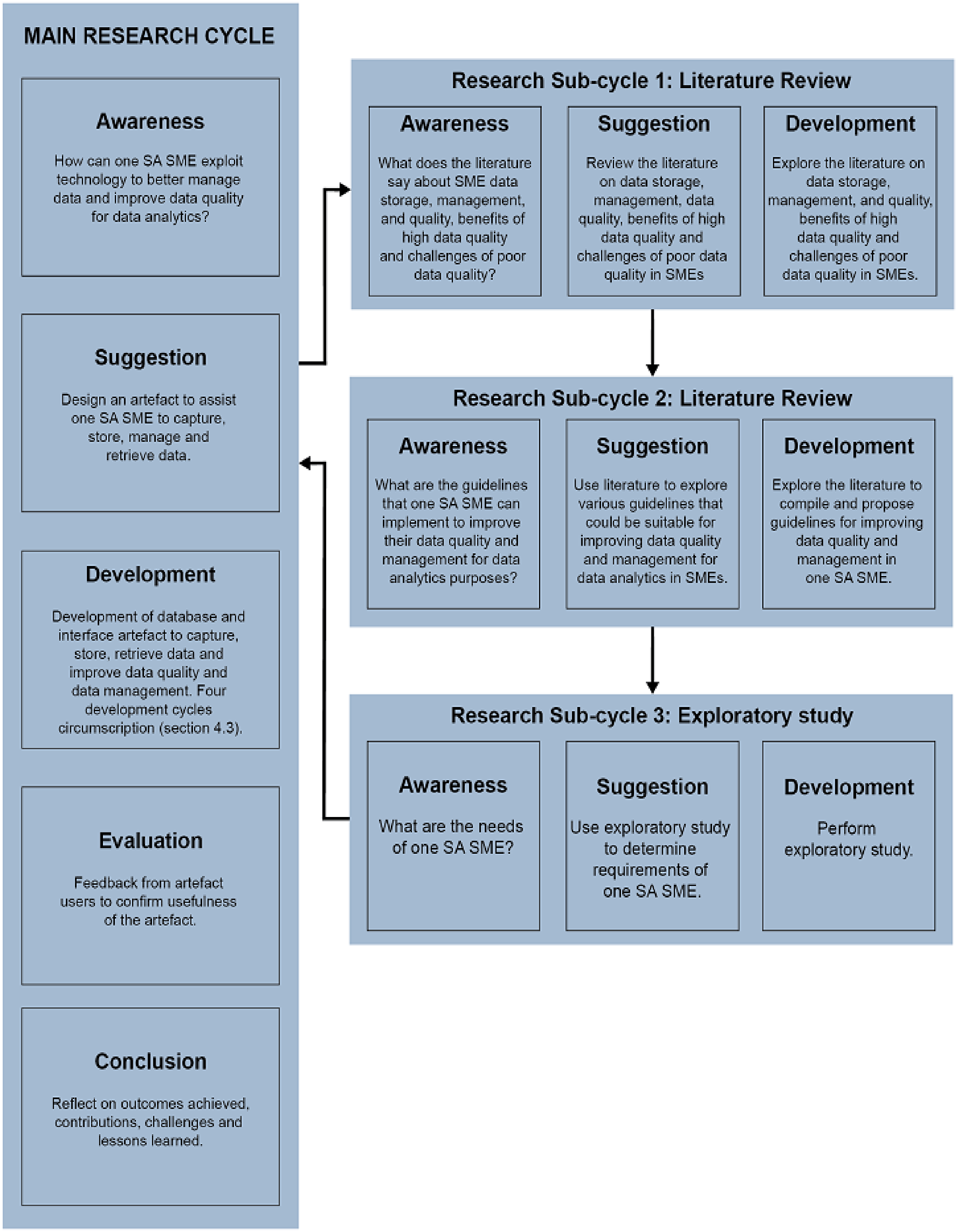


Figure 3-11: Design science research applied in the study.

Awareness

Awareness occurred when one SA SME expressed a need for better data quality and management. The research problem was defined as how one SA SMEs can exploit technology to better manage data and improve data quality for data analytics purposes.

Suggestion

The suggestion was first to assist the organisation with capturing, storing, and managing data. In preparation for the design and development of an artefact, the literature was consulted to gain insight into the problem and suggested solution. Three design science research sub-cycles needed to be completed.

Sub-cycle 1

The awareness phase of sub-cycle 1 is concerned with what literature says about data storage, management and data quality, and challenges and benefits of data quality. The suggestion was to review literature related to the problem in hand and analyse how SMEs are storing and managing their data. In the development stage, the researcher explored literature related to the study and determined how SMEs are storing and managing data, dealing with poor data quality, challenges of poor data quality and benefits of high-quality data. The theoretical objectives of the study were to explore technologies that SMEs can implement to better manage and store data, investigate the impacts of poor data quality in SMEs, identify the challenges of poor data quality and benefits of high data quality, and explore various BI tools suitable for SMEs.

Sub-cycle 2

The awareness of the problem in this phase is which guidelines and technologies are available that one SA SME can implement to improve their data quality and management for data analytics purposes. The suggestion was to review the literature and explore various technologies and guidelines that one SA SME can implement to assist with improving data quality and management. In the development phase, the researcher explored the literature and proposed guidelines and technologies that one SA SME can implement to improve their data quality and management. These guidelines can also be beneficial to other SMEs.

Sub-cycle 3

The awareness of the problem in this phase focused on the data management and quality needs of one SA SME. The suggestion was to perform an exploratory study and identify various needs of one SA SME. In the development phase, an exploratory study was completed to identify various needs and data management requirements of one SA SME.

Development

In the development phase, four circumscriptive phases were completed as a result of new awareness arising from evaluation while constructing the different phases of the artefact. An existing inefficient data storage facility and user interface had to be improved. A data warehouse had to be implemented, as well as a new reporting solution.

Evaluation

The functionality of the artefact was rigorously evaluated. In the evaluation phase of the main research cycle, a questionnaire was given to the IT administrator and data capturer at the SME to give feedback on the functionality of the artefact and other deliverables of the project as a whole, and interviews were conducted with the IT administrator and data capturers at the SME.

Conclusion

An artefact to improve the data quality and management in one SA SME was implemented. Furthermore, guidelines were proposed to improve data quality in SMEs. Also, in this phase it was necessary to reflect on outcomes achieved, contributions made, challenges encountered and lessons learned.

3.12 Rigour of the study

Qualitative research is often criticised for the following reasons: being too subjective, hard to replicate, researcher bias, lack of transparency and issues of generalisation (Bryman & Bell, 2011; Cypress, 2017:254). Additionally, Cypress (2017:253) mentions that there are persistent concerns about accomplishing rigour in qualitative research, which also leads to questions regarding reliability and validity. Evaluating design artefacts theories in DSR is, therefore, a very critical activity as it assures the rigour of the research study and provides feedback for additional development (Venable *et al.*, 2017:77). Furthermore, Venable *et al.* (2017:87) emphasise that evaluation in DSR does not only consider the utility aspect of the artefact in the environment, but also takes into consideration the quality of the knowledge contributed by the development of the artefact.

Bryman and Bell (2011:408) state that the qualitative findings depend too much on the researchers' unsystematic views regarding what is crucial, and also upon inspection the close relationship the researcher might have with the participants. By contrast, there are no statistical tests in qualitative research that can be used to evaluate the validity and reliability of the interpretation and representation of the participants' narratives (Sutton & Austin, 2015:229). Consequently, researchers conducting qualitative research studies need to conduct such studies with rigor and provide the proof.

However, Venable *et al.* (2017:82) advocate rigour as:

- Efficacy – establish that the artefact instantiation improves the situation or environment;
- Efficiency – establish that the artefact works in a real-world amidst resource constraints such as time and money. Furthermore, establish that the artefact works within the environment despite complications associated to people, processes and systems.
- Ethics – address the artefact's potential risks to the people, animals, the organisation or the public.

Questions regarding trustworthiness of qualitative research always emerge, and as a result the researcher needs to report on the validity and reliability of the study to ensure that it is conducted with rigor to ensure that such issues are addressed accordingly. Bryman and Bell (2011:43) and Kumar (2019:351) advocate that trustworthiness is associated with particular aspects that need to be taken into consideration:

- Credibility – Are the findings believable?
- Transferability – Are the findings applicable other contexts?
- Dependability – will the findings likely be applicable to other times?
- Confirmability – has the researcher permitted his or her values to interfere to a high degree?

However, Cypress (2017:253) advocates that strategies required to ensure rigor must be incorporated into the qualitative research rather than assessing rigor after the inquiry is completed. According to Neergaard *et al.* (2009:4) researchers should adopt the strategies outlined in Table 3-10 to enhance rigour in qualitative studies. Table 3-10 also portrays how the researcher of this study adapted research actions to adhere to the strategies and techniques.

Table 3-10: Strategies to enhance rigour of this study (Neergaard et al., 2009:4).

Technique	In the study
Authenticity	
The participants must be able to speak freely, Voice out their opinions and their voices be heard. Their perceptions should be portrayed correctly. Transcriptions must be presented accurately.	The participants signed consent forms to indicate that they were participating in this study at their own accord. The researcher explained to them that all data collected during the interview sessions and from questionnaires will form part of the research study, however, participant's details and company details were not disclosed. Anonymity was maintained in the report, using a coding system. Participants were able to speak freely, sharing their thoughts and experiences. Audio recording were used to collect data during interview sessions to ensure that participants thoughts, perceptions and experiences are documented accordingly without leaving out any important information. Data were transcribed and once more reviewed to ensure accuracy. All data treated as confidential.
Integrity	
Reflects on researcher bias and participants must be checked and validated.	The IT manager suggested that the IT administrator and the data capturer were the relevant individuals to participate in this study. The participants were selected based on their level of knowledge about the phenomena and understanding of the data issues within the organisation. One of the participants was responsible for capturing and managing data while the other was responsible for administering IT systems and processes within the organisation. The IT manager also confirmed that they were the right individuals to take part in this study as they have in-depth knowledge and understanding about the issues within the SME. Furthermore, the participants were deemed to have integrity, high principles and would provide honest information required for the success of the study as they would also benefit from the results. The researcher ensured that all participants were treated with respect and with integrity taking into consideration their experiences, thoughts, desires, views, and differences. No participants were subjected to any form of discrimination.
Credibility	
It must portray and capture a truly researchers' perspective.	The values of the researcher formed part of the study and might have influenced the outcome of the study because the researcher was involved throughout the study. The researcher also interacted with the participants on several occasion even during interview sessions. The researcher will ensure that all information needed for the success of the study is collected and analysed. Furthermore, contact details of the researcher were shared with the participants to communicate and share information with the researcher. The researcher also requested the participants to communicate any changes, processes and plans that my come up during the course of the study at all times telephonically or via email. This ensured a transparent and a trustworthy relationship with the participants. Once the audio recorded interviews were transcribed, the researcher shared the transcriptions with the participants to review and confirm if it's valid and portrays a true reflection of their words.
Criticality	
Reflect on the critical appraisal.	The evaluation phase of the artefact involved an iterative process of testing, developing and communicating the changes. Several test cases were used and various evaluation criteria were used to assess the functionality, reliability and relevance of the artefact. The evaluation process gave the participants an opportunity to interact with the artefact to prove that it meets the desired requirements, it works in the real-world, is user-friendly, and it will add value to the organisation.

No participant or organisation details were disclosed in the research study and the data collected from the questionnaires and interviews were kept confidential. The audio-recorded interviews were transcribed and reviewed to check for accuracy and authenticity.

3.13 Conclusion

This section of the study explored different research paradigms, approaches, and qualitative strategies of inquiry, data collection and analysis techniques. The research onion presented in this chapter provides a graphical representation of the concepts discussed in this chapter. Researchers indicated that paradigms are used to guide research inquiries and to enable researchers to gain better understanding of the reality of the world and the beliefs about the nature of reality. In this chapter, paradigms such as positivism, critical social research, interpretivism and design science research were explored. Design science research was chosen as an appropriate research paradigm for this study because the researcher wanted to develop an artefact as a solution to a socio-technical research problem in one SA SME. Researchers have indicated that Design science research (DSR) is regarded as a permissible information system research paradigm that involves solving a real world problem through construction of artefacts. Design science research (DSR) model developed by Vaishnavi *et al.* (2004/2019:12) was selected in this study.

Furthermore, philosophical assumptions such as ontological, epistemological, axiological and methodological we discussed in this section. The ontological stance of the study was that the SME in which the research study was conducted is recognised as a socio-technological environment and the reality was that the participants and the researcher would be involved throughout the study and interact frequently with each other. The epistemological stance of the study was addressed through knowing the nature of knowledge through constructing an artefact to improve data quality and management in one SA SME.

Subsequently, the researcher explored research approaches and strategies, data collection and analysis techniques. A case study was chosen to be the appropriate strategy to be used because an in-depth investigation was conducted on one phenomenon. Mixed mode research approach was selected as an appropriate research design because the researcher used semi-structured interviews and observation data collection techniques to collect data needed for understanding the needs, feeling, perceptions and beliefs of the participants. Furthermore, a questionnaire was also used to collect data during the evaluation of the artefact. Data gathered using interviews will be analysed using ATLAS.ti and questionnaire data will be analysed using the statistical software for the social sciences (SPSS). Various design science research models, guidelines for design science research, artefact evaluation techniques and the study plan were discussed.

The study plan discussed in Section 3.11 provided a roadmap the researcher followed in achieving the objectives of the study. The researcher took into consideration ethical principles such as privacy of participants and data, informed consent, public interest, honesty and accuracy, and the quality of the artefact. For ensuring and enhancing rigor of the study, guidelines such as authenticity, integrity, credibility, and criticality were followed.

CHAPTER 4: DESIGN AND DEVELOPMENT

4.1 Introduction

The study commenced when the researcher was approached by the information technology (IT) manager of one South African (SA) small medium-sized enterprise (SME) to assist with improvement of data quality and management. The primary objective of this study was to design and develop an artefact that can be used by one SA SME to better manage data and improve data quality. In this section of the study, the researcher outlines the roadmap followed in designing and developing the solution to the research problem. To achieve the objective, the design science research (DSR) model developed by Vaishnavi *et al.* (2004/2019:12) as outlined in Chapter 3 section 3.8.3 was followed. Furthermore, the researcher consulted the literature to gain background knowledge on the problem and to familiarise herself with the process and techniques for developing the artefact. A case study research approach was used to enable the researcher to conduct an in-depth investigation about the phenomenon to gain better understanding of the problem and gather sufficient information for determining a solution to the research problem.

Initially, for the researcher to gain more understanding and knowledge about the research problem and formulate the organisational requirements, an exploratory study was used to gather the information at the SA SME. The project initiation and planning session was held virtually using online platforms between the researcher, IT administrator and IT manager of the SME to collect information about the problem and the IT manager indicated that they needed a solution to assist them with capturing, storing and managing data, as well as improving data quality. The researcher followed the four phases of conducting case study research by Rashid *et al.* (2019:2) as outlined in Figure 3-3 section 3.4.4.

In the foundation phase the researcher discussed literature about data quality, data storage, and management, SMEs and data analytics in Chapter 3 to set the tone of the study. DSR was selected as an appropriate paradigm for this study because the researcher constructed an artefact to solve a real-world problem.

In the prefield phase, the researcher selected the case study as an appropriate research strategy. Based on information gathered, the researcher formulated the research questions (section 1.4.4) and research objective as explained in section 1.4.2, formulated the requirements for gathering and project scope, obtained ethical considerations in section 3.10 and determined the data analysis process in section 3.6.

In the field phase, the researcher followed a mixed method research approach as discussed in Chapter 3 section 3.3.3, where data collection techniques such as semi-structured interviews (section 3.5.1), observations (section 3.5.4), and questionnaires (section 3.5.2) were used to collect data. In this phase, the researcher contacted and interacted with participants through a workshop session. Additionally, interviews were conducted with the participants to obtain feedback on the functionality of the artefact. Observations were used to assess participants' interaction with the artefact. The data collection methods enabled the researcher to collect data required for this study.

The feedback on the outcome of the study was presented in a reporting phase where the results of the data analysis process were discussed, including the conclusions reached.

Section 4.2 will discuss the design of the artefact as suggested in the main DSR model for the study where research sub-cycle 1 will be explained in section 4.2.1, research sub-cycle 2 (section 4.2.2) and research sub-cycle 3 (4.2.4). A brief discussion of the SME involved in this study is presented in section 4.2.3. The research sub cycle 1 provides a brief discussion of the literature the researcher explored in relation to the concepts of the research study, which includes data storage, management, and data quality, benefits of high data quality, and challenges of poor data quality. Furthermore, research sub-cycle 2 will discuss the literature explored in determining the guidelines suitable for improving data quality and management in SMEs. Research sub-cycle 3 will focus on the exploratory study.

Section 4.3 will discuss the development of the artefact. The conclusion of this chapter follows in section 4.4.

The next two sub-sections that follow will revisit the awareness and suggestion phases of the main DSR cycle. Following sections then discuss the design and development phases.

4.1.1 Awareness phase

The IT manager of one SA SME indicated to the researcher that they were having challenges in managing and using data for reporting and decision making as a result of poor data quality within the organisation. In this phase, the researcher had identified the problem and was aware that one SA SME needed a solution to assist with improving data quality, data storage and management for analytics purposes.

4.1.2 Suggestion phase

Based on the suggestion phase of the main DSR model for the study, three research sub-cycles were followed to gain better understanding of data quality and management. The researcher used the literature to study concepts related to the research (sub-cycle 1). The literature was also explored to find guidelines on how organisations could exploit technology to solve data quality and management problems (sub-cycle 2). However, to achieve this, the researcher used an exploratory study to gain in-depth understanding and knowledge of the problem at one SA SME (sub-cycle 3).

4.2 Designing the artefact

The following sub-sections will discuss research sub-cycles 1 and 2, followed by a description of the context in which the research was conducted. Research sub-cycle 3, which explains the exploratory study performed to gather requirements for the development of the artefact, will conclude the design phase.

4.2.1 Research sub-cycle 1: Literature review

The researcher followed this research cycle to investigate what literature says in relation to this study and determine how other researchers addressed similar issues.

Awareness

In the awareness phase of research sub-cycle 1, the researcher analysed the problem presented by one SA SME. The researcher studied literature about concepts related to the research problem such as data storage, management, data quality and data analyses.

Suggestion

The researcher made a suggestion to review the literature to gain a thorough understanding about data storage and management, benefits of high data quality, and challenges of poor data quality in SMEs. Furthermore, the researcher also planned to review the literature on previous studies to determine how other organisations are dealing with data storage, data management and data quality, and find out which technology solutions are used by other organisations to improve data management and data quality for data analysis.

Development

The researcher consulted academic databases such as IEEE, Google scholar and ScienceDirect to review literature on the topics related to the research problem. Chapter 2 of this study presented the literature reviewed in relation to data storage, data management, and data quality, benefits of high data quality and challenges of poor data quality (section 2.6, 2.7, 2.8 and 2.9).

The researcher explored all these concepts together with various technologies, methods and procedures that other SMEs are using, the benefits and challenges they encountered and the results they achieved in order to gain better knowledge about the tools and procedures that one SA SME can adopt. This also assisted the researcher in developing the foundation for the research study, to better understand the concepts of data quality, storage and management in the SME and how decision making, analytics and business processes could benefit from improved technology processes.

The researcher reviewed the literature on benefits of high data quality in section 2.6 and discussed the causes and challenges of poor data quality in section 2.6.4. Furthermore, the cost and impact of poor data quality was reviewed in section 2.6.4. Section 2.7 in the study provided more insight and knowledge about data storage and management in SMEs. Applications and tools that SMEs can use to capture, store, manage and improve data quality were discussed in sections 2.6, 2.6.2, 2.6.3, 2.7 and 2.9.

Even though this study is based on a South African SME, the researcher explored articles about the significance of SMEs worldwide in relation to economic growth and development, poverty mitigation, innovation, and increased global competition and how they responded to digitalisation, technological advancement, customer requirements, increased global competition, innovation and large volumes of data (section 2.6, 2.8, 2.9).

4.2.2 Research sub-cycle 2: Literature review

Based on the literature reviewed and the knowledge discovered in sub-cycle 1 the researcher decided to explore literature further to gain more insight about data storage, management and data quality.

Awareness

Through reviewing the literature in the first sub-cycle, the researcher became aware of the advantages of principles and/or guidelines that could be implemented at SMEs to improve data quality and management.

Suggestion

The researcher's suggestion was to explore the literature for existing guidelines discovered by other researchers that could be suitable for improving data quality and management in one SA SME.

Development

In this sub-cycle, the researcher reviewed the literature to discover what other researchers are recommending in terms of improving data quality and management in the SMEs (Section 2.6 and 2.7). The researcher explored the literature to gain knowledge about appropriate guidelines that could be proposed to one SA SME to implement to mitigate the problem.

Table 4-1: Guidelines for improving data quality and management from the literature

Guidelines from the literature	
It is an organisation's responsibility to educate employees about the importance of data and its relationship with information, knowledge and wisdom.	Section 2.1
Senior management must develop a data-driven culture and mindset within the organisation needed to maintain and sustain data quality.	Section 2.6
Organisations must take responsibility to conduct data quality trainings and enterprise-wide data quality awareness programs.	Section 2.7
Organisations must make their employees aware that data are no longer regarded as "by-product", but are valuable enterprise-wide corporate asset that are used for operational record keeping and for operational, tactical and strategic level decision-making.	Section 2.6, 2.7
Organisations must understand that it is every employee's responsibility to protect the integrity of the data.	Section 2.6
Data quality refers to data that are "fit for use". Organisations must understand what "*fitness for use"* means to those who use the data and determine the dimensions that can be used to evaluate the quality of the data.	Section 2.6, 2.6.1
It is imperative for organisations to use well-founded metrics or dimensions to measure and improve the quality of the data.	Section 2.6.1
Data are not useful until they are transformed into information and insight to support a business or decision-making process. Organisations must ensure that data go through a data transformation or cleansing process before they are used for analytics and decision making.	Section 2.6

Guidelines from the literature	
Metadata pays a critical role in the process of measuring data quality.	Section 2.6.1
Organisations should not overlook or neglect existing data quality issues within the organisations. They must be addressed and fixed as early as possible.	Section 2.6.4
Organisations must implement standard data handling procedures and formal data governance programs.	Section 2.6.4
Data capturing systems must be equipped with adequate data validations and checks to prevent erroneous data from entering the system.	Section 2.6.4

Furthermore, technologies that could be proposed to one SA SME to improve data management and data quality were explored in the literature.

Table 4-2: Technologies for improving data quality and management from the literature.

Technologies from the literature	Literature
Data profiling tools can be used to assist with monitoring the current state of organisational data.	Section 2.7
SQL Server Management Studio can be used to manage data stored in SQL relational databases and data warehouses.	Section 2.7, 2.9.1
Six Sigma DMAIC approach (define, measure, analyse, improve and control) can be used to improve data quality.	Section 2.7
Microsoft SQL Server Data Tools (SSDT) can be used to improve data quality. SSIS can be used to develop ETL solution for transforming and cleansing data to improve data quality.	Section 2.7, 2.9.2
Apache Hadoop can assist with managing and storing big data.	Section 2.7
Cloud-based Business Intelligence.	Section 2.7
Business intelligent tools such as Talend Open Studio, Information Server for Data Quality can be implemented by SME to improve data quality.	Section 2.7
Kimball methodology can be used to develop the data warehouse.	Section 2.9.1

4.2.3 Case study environment

In this section the researcher will provide a brief discussion of the SME involved in the study. The case study environment associated with this study was a company specialising in engineering, casting and manufacturing of steel mill rolls and rings. To maintain confidentiality of the SME, it was referred to as “one SA SME” throughout the study. The organisation was established in the 20th century in SA and it is regarded as a highly successful organisation providing high quality services and products worldwide. The SME plays a critical role in economic growth and development of this country as it contributes to job creation, economic development and poverty alleviation. One SA SME employs less than 120 employees and its operations include managing customer orders, processing invoices, manufacturing steel mill rolls and rings, defining the materials needed for production, metal casting and quality assurance. To manage and keep record of the steel roll manufacturing process, the IT department is required to provide an efficient data management process by ensuring that all details from the order process until the roll is ready to be delivered to the customer, are captured and managed accordingly. This information is important to this SME as it is used in decision-making processes, process improvement and performance measurement.

However, due to the high-quality products the SME manufactures, the organisation has seen a large increase in its customer base worldwide which led to an increase in data. Currently, the SME is collecting large volumes of data that are difficult to manage and analyse; poor data quality is an issue and they are unable to capture data simultaneously. Furthermore, due to globalisation, customer demand, exponential growth of data, competition, and advancement in technology, the SME realised that a better solution was needed to assist with efficient capturing, managing and analysis of data. The SME required a technological solution that could also enable them to innovate, be competitive, provide efficient service to customers, increase business performance, make informed decisions and keep pace with changing technologies.

The organisation has a small IT department consisting of seven employees as outlined in Figure 4-1. The IT department oversees all IT technical issues, and is responsible for implementing and managing IT solutions, which include management of hardware, software and data.

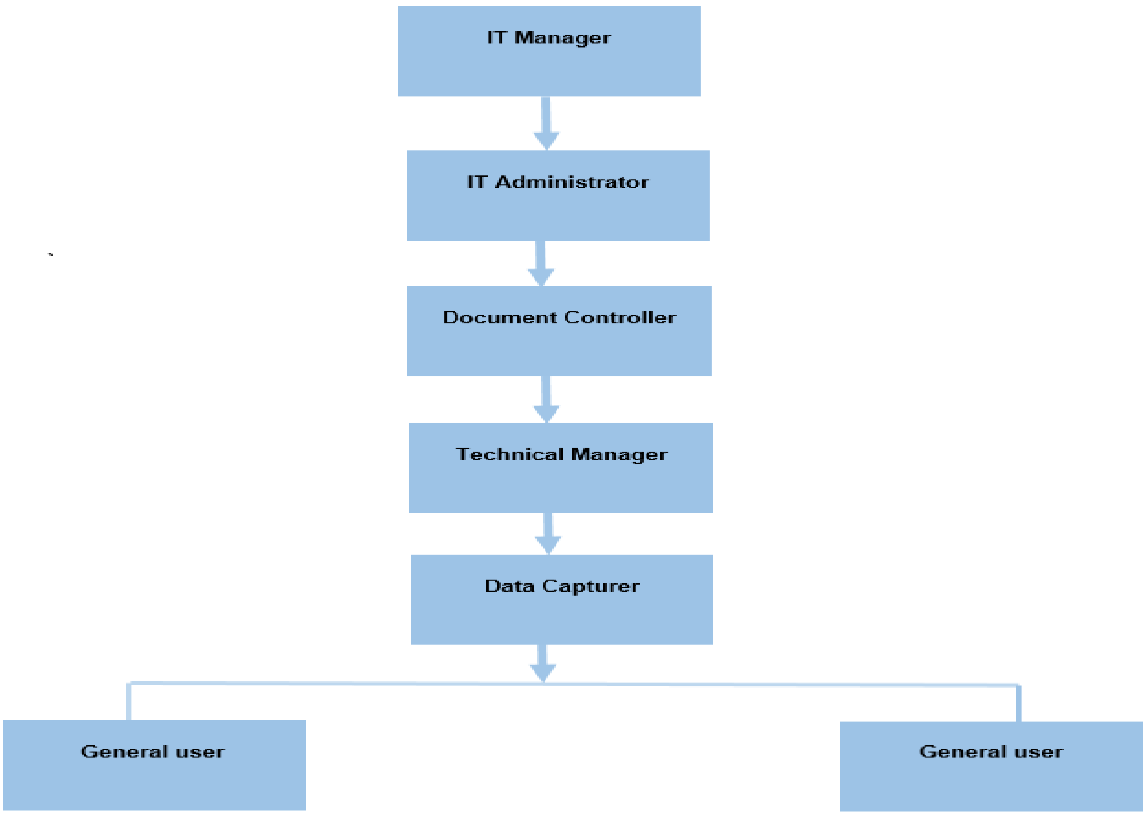


Figure 4-1: IT department structure of one SA SME

- The IT manager is responsible for initiating, planning, evaluating and managing IT functions within the organisation.
- The IT administrator is responsible for managing the IT infrastructure, upgrading and installing hardware and software, maintaining server and network and managing security procedures within the organisation.
- The document controller is responsible for managing the data and project plans, capture data and generate reports.
- The data capturer is responsible for capturing, retrieving and storing data into the databases.
- General users – two general users are responsible for handling day-to-day technical issues within the organisation.

At the start of the case study, the SME was using Microsoft Access 2016 as its primary application for capturing and storing data. Additionally, some of its data were kept in Excel spreadsheets and one SQL server relational database. Microsoft Excel was used as the main data reporting and analytics tool. To analyse the data stored in the databases, the data capturers either copied and pasted data to the Excel spreadsheet or linked the Excel spreadsheet to the database by defining a data source. Microsoft Excel pivot tables and analytical graphs were used for data visualisation. Figure 4-2 outlines the original entity relationship diagram of the MS Access database.

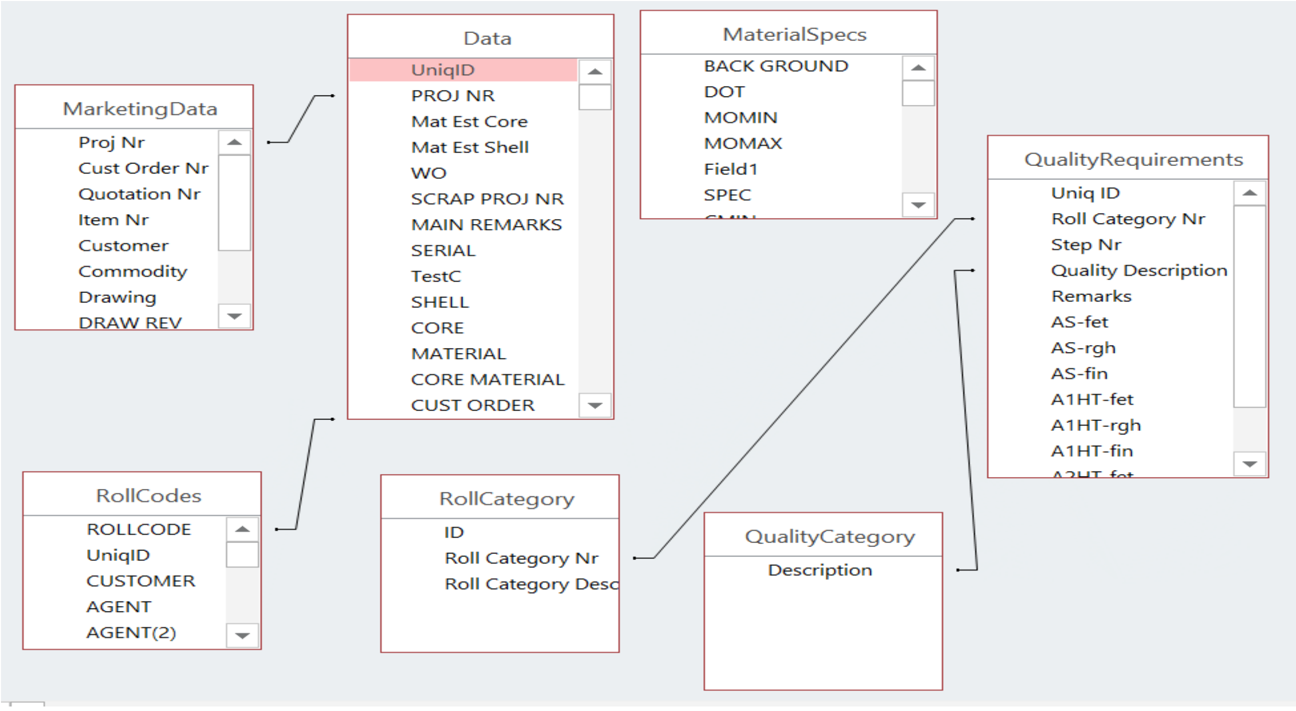


Figure 4-2: Original data management architecture in one SA SME.

The researcher conducted the study in order to develop an artefact for the SA SME to assist in improving data quality and data management for data analytics and decision-making processes. The artefact would be created to enable the SME to efficiently capture, retrieve, and store data in the database.

Before the artefact could be developed, the researcher conducted an exploratory study to gather requirements for improving the data management of the SME and confirm the feasibility of the research study.

4.2.4 Research sub-cycle 3: Exploratory study

Based on the knowledge acquired in sub-cycle 1 and sub-cycle 2, the researcher needed additional information about the phenomena, data quality problems and data management needs of the one SA SME.

Awareness

The researcher wanted to explore and better understand challenges with data capturing, storage, management and data quality that one SA SME was experiencing.

Suggestion

The researcher's suggestion was to determine the data management needs of the SME by conducting an exploratory study. In the process, the researcher organised an interview session with the IT administrator and senior manager of one SA SME to collect more information. The IT administrator and senior manager were selected for the interview session because they were the people knowledgeable about the data management issues within the SME. It was essential for the researcher to conduct the interview to gain in-depth knowledge and understanding about the phenomena and its problems.

Development

An exploratory study was conducted with the IT administrator and senior manager to answer questions outlined in Appendix 1. Due to work commitments, it was difficult to do face-to-face interviews, and therefore the researcher prepared the questions and emailed them to the IT administrator and senior manager to answer and provide more details about the problem, the current situation in the SME, the tools and applications currently used to manage and store data, and also to determine their needs for improvement of data management. The feedback received from the questions enabled the researcher to gain knowledge about the phenomenon, understand the current data management procedures that are used by the organisation, as well as the current state of the problem, needs and requirements of one SA SME.

Feedback from the IT administrator indicated that:

"Data are captured using Microsoft Access on all current databases and at times simultaneous logins are not permitted. We are currently experiencing errors in information on the database which makes it hard for management to have confidence in the data and trust their decision making processes. Data access is not efficient and poor data integrity is a problem. On a scale of 10, I can rate data integrity as 5."

More information was needed to assist with building the foundation of the study, to gain a better understanding of the phenomenon and to define the scope. The researcher requested a meeting with data capturers and the data manager to collect more information. A workshop was organised between the researcher, IT administrator and data capturers to provide them with a platform to speak openly about the situation and allow the researcher to gain an in-depth understanding of the problem from the users' experience. The researcher visited the organisation to assess the environment and the first face-to-face workshop was held, where the researcher interacted with participants to collect more information and to examine the current architecture within the organisation.

During the workshop, the data capturer demonstrated to the researcher how the data capturing and management process worked. The researcher observed the processes while taking notes about the issues and limitations presented. The system users shared their challenges and experiences with the researcher. One of the participants indicated that:

> "We need a technological solution to assist with efficient capturing and management of data. Currently data retrieval is slow and it is difficult to capture and manage roll production process. We need some sort of an interface to enable capturing roll order details, customer information, roll category, and roll codes, material specifications, quality requirements and heat treatment data into the database. We also need to analyse the data and generate reports. The current system does not check the data or perform any form of data validations which causes duplicate data, inconsistent, incomplete, and inaccurate data"

The workshop assisted the researcher:

- ❖ To gain a better understanding of the problems and needs of the SME from the users' perspective;
- ❖ To study the environment and explore the current data management architecture;
- ❖ To collect sufficient information needed to develop the foundation of the research study and gather requirements;
- ❖ To plan the scope of the project;
- ❖ To identify the target user audience;
- ❖ To meet the study participants; and
- ❖ To obtain ethical clearance from NWU to perform the research.

All information gathered in sub-cycle 1, sub-cycle 2 and sub-cycle 3 were used to develop the foundation of the study, develop a requirements document and to determine the most appropriate solution to the problem presented. Additionally, the information gathered assisted the researcher to formulate the research questions and the objectives of the study (section 1.4.4). The requirements gathered from the participants indicated that one SA SME needed a technological solution as indicated in section 4.2.1 that can assist with:

- ❖ Efficient capturing and retrieval of data;
- ❖ Enable multiple users to capture data simultaneously;

- ❖ Improving data quality for analytics and reporting;
- ❖ Improving decision-making processes;
- ❖ Enhancing data management processes;
- ❖ Provide efficient data storage;
- ❖ Minimise data capturing issues; and
- ❖ Providing data security.

The researcher suggested that an innovative artefact should be designed to assist one SA SME with capturing, managing, storing and retrieving data. Microsoft technologies would be used to develop the solution because the SME was already using Microsoft applications and one of their databases was using MS SQL server 2017. Additionally, the study by Raj *et al.* (2016:46) in section 2.1.4 proved that Microsoft technologies can assist SMEs to improve data quality, data capturing, management and data analytics.

4.3 Development of artefact

The researcher selected the DSR process model of Vaishnavi *et al.* (2004/2019:14) as outlined in Chapter 3, Figure 3-8 (section 3.8.3), which was further elaborated on in Figure 3-11 (section 3.11) as the appropriate model for the study. Figure 4-3 illustrates the timelines the researcher followed in delivering the artefact.

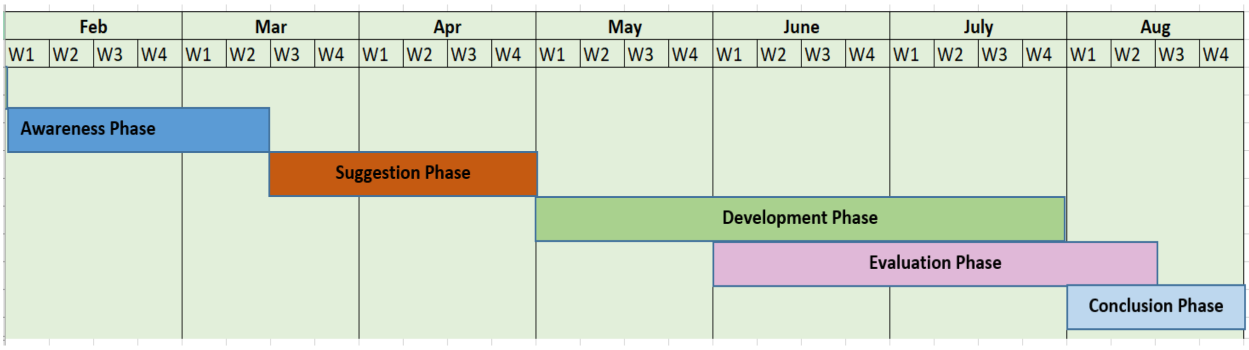


Figure 4-3: Gantt chart for artefact development.

In the next section, the researcher will discuss four development cycle circumscriptions of the main DSR cycle. Figure 4-4 illustrates the four-development cycle, the development of as SQL relational database, user interface for capturing and retrieving data, data warehouse and ETL solution, as well as the reporting and analytics.

Main DSR cycle – development phase circumscription

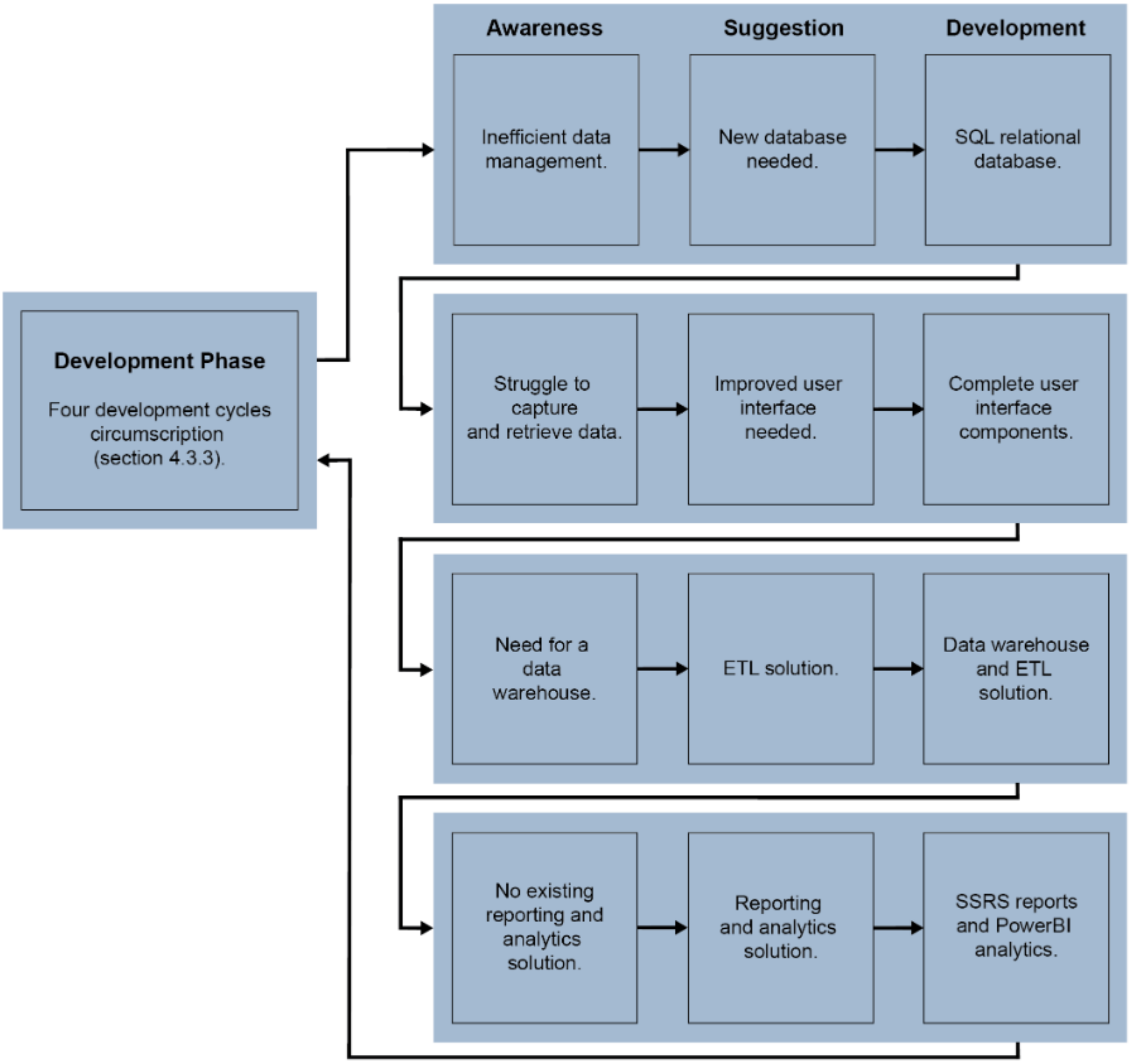


Figure 4-4: Four development cycles circumscription

A system requirements specification (SRS) document was developed based on the information gathered in the awareness phase and the proposal for the development of the artefact. Requirements gathering plays a critical role in software development, as it enables the researcher to understand user needs prior to embarking on system development (Ismail *et al.*, 2019:2). This process required the researcher to interact with stakeholders to obtain their wishes and expectations (Ismail *et al.*, 2019:2) and document all the critical information required for the development of the artefact such as the scope of the project, the business process, business rules, identify the stakeholders, the deliverables, business requirements, project risks, constraints, software and hardware to be used. Additionally, the SRS document was used as a reference guide throughout the entire artefact development process in ensuring that a high-quality artefact met the objectives and was delivered on time.

The entity relationship diagram (ERD) for the SQL relational database was designed during this phase to illustrate the relationship between the tables needed to store data. The output of this phase was the tentative design of the artefact. It took six weeks for the researcher to complete the initial development phase and produce a tentative design of the artefact. In this case, a high-level use case diagram was produced.

Figure 4-5 depicts the high-level use case diagram of users' interactions with the data management artefact. The data capturer would login to the artefact by providing a username and password to trigger the verify password use case. The login use case and verify password use case had an include relationship which indicated that every time the login use case was executed, the verify password was automatically executed. The relationship is illustrated by a dashed line from the login use case joining the verify password use case with the word "include" to indicate the type of relationship. Furthermore, display error message use case would sometimes happen, depending on the outcome of the base use case and this relationship is illustrated by a dashed line from the display error message use case, pointing to login use case with the word "extend" to indicate the type of relationship.

The artefact would also allow the data capturer to: search roll project details already stored in the database; capture customer information, order details, roll codes, roll material specification, heat treatment, and quality inspection details; and scrap information. For these use cases, validations were embedded in the code to check null values in mandatory fields, duplicate records and check if data formats are correct. These use cases had an include relationship that would automatically execute when data are captured to check data accuracy, consistency, availability and unique values. A dashed line with the word *include* indicated the relationship between the base use cases to the include use case. The error message use case illustrates an extend relationship which is dependent on the outcome of the base use cases.

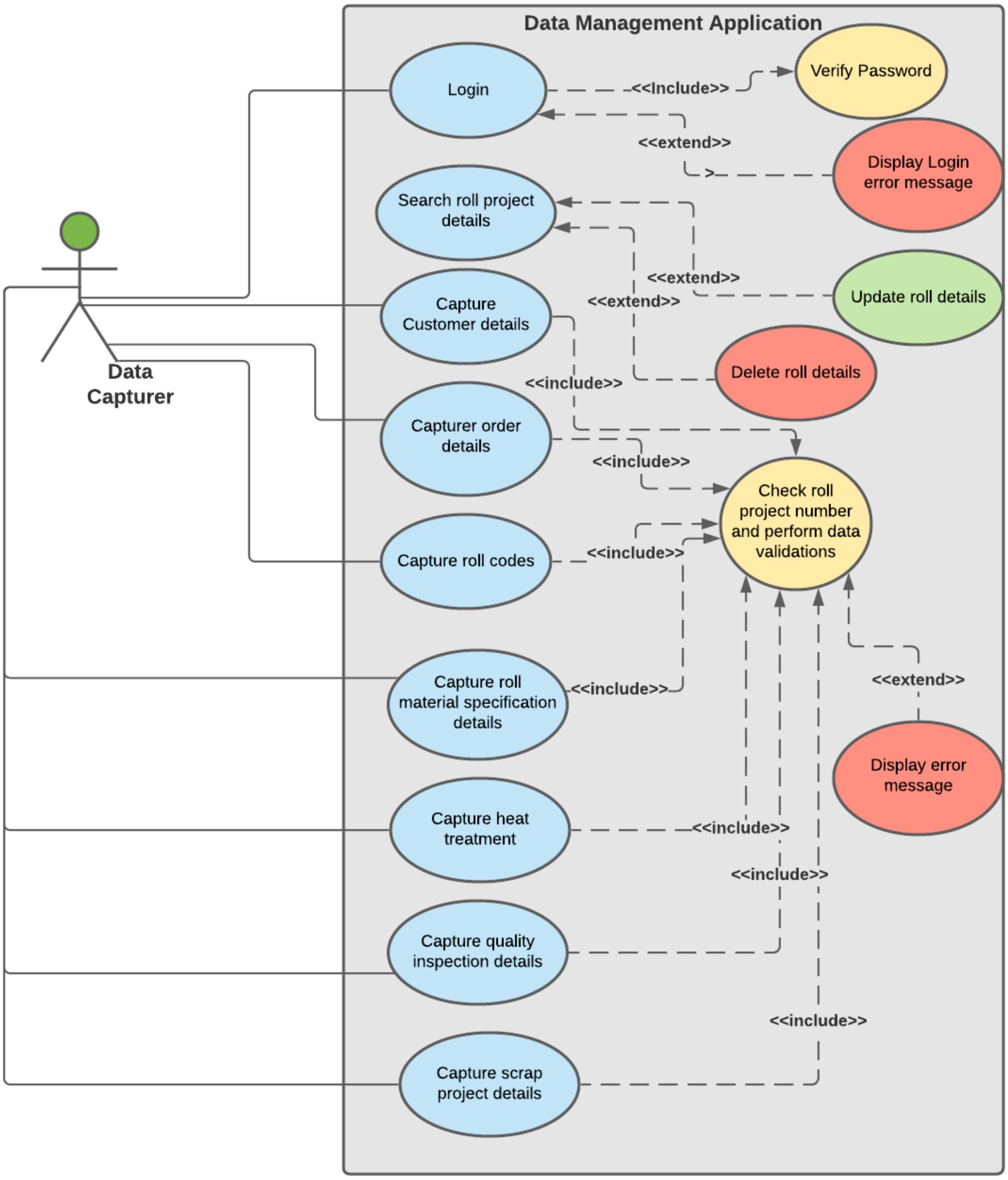


Figure 4-5: High-level use case diagram of user interaction with the artefact.

The researcher developed the use case diagram to paint a picture of how the user will interact with various elements of the artefact, and illustrated various functionalities or use cases associated with the artefact. The use case diagram described the expected functionality of the interface. The use case diagram was important in the artefact development phase, as it would drive the application development process by guiding the researcher in developing the anticipated system architecture and all functionalities required.

Artefact development cycles

This phase involves the development of the artefact where four circumscriptive cycles were performed in applying the DSR model while constructing the artefact.

Development cycle 1 – SQL relational database

Based on the tentative design produced in the initial phase, the researcher developed an SQL relational database using MS SQL Server 2019 to store and manage data. The ERD was followed in designing the SQL database to ensure that all entities, their relationships, data formats and constraints were developed accordingly. The researcher implemented not null constraints, unique, check, default and referential integrity constraints (based on section 2.7) to improve data quality (Gudivada *et al.*, 2017:14; Mullins, 2017:55; Nyaboga & Mwaura, 2009:20), and to be closer to achieving high quality data based on the data dimensions as outlined in section 2.6.1. Table 4-3 summarises the first database development cycle of the relational database needed for managing data.

Table 4-3: First development DSR cycle – relational database

Cycle	Step	Activity
Awareness	Problem	The old data management process was using Microsoft Access 2016. This tool presented the SME with multiple limitations and challenges. Multiple users were unable to efficiently capture data in the database simultaneously, there were poor data quality issues, poor data management, data security issues, data stored in various locations, slow performance on data retrieval and lack of trust in using the data for reporting and decision making.
Suggestion	Plan	To determine a solution to the identified problem, the researcher suggested the following actions: ❖ To send a list of questions through email to the IT administrator and senior manager to collect more information about the problems. ❖ To request access to the MS Access database and explore the database structure, relationship between the tables, functionality and performance. ❖ To send an email to request a workshop with the users in order to collect information about the problem from users' perspective based on their experiences, expectations and thoughts. ❖ Determine softwares required to develop the database and to develop the data capturing solution.
	Requirements	The following actions formed part of requirements gathering:

Cycle	Step	Activity
		❖ A workshop session was held with the users to gather more information about the problem. ❖ Understanding of the business processes and business rules. ❖ The researcher was granted access to the database and given sample data. ❖ The users gave the researcher permission to record the workshop session which formed part of the requirements gathering.
	Design	All the information gathered during the workshop session and current database structure were used to develop the logical or tentative design of the database. ❖ The table names, columns, data types and relationships were defined. ❖ Lucid chart diagramming application was used to develop the ERD diagram.
Development	Develop	The researcher installed MS SQL Server 2019 as an appropriate software for developing and managing the database. Based on the ERD diagram, the researcher developed the physical database using MS SQL Server 2019. Tables needed for capturing data using the artefact were defined including their foreign key primary key relationship. Constraints such as not null, unique, default, primary key, foreign key were developed to limit the type of data that can be stored in the database in achieving accuracy, consistency and reliability.
Evaluation	Test	The researcher evaluated the new database with the data capturer and the IT administrator. This process iterated several times to rectify column names, data types and formats, and relationships between the tables.
	Feedback	Documented the changes and revised the design. ❖ Decimal values in material specification tables were changed from two decimals place to three.

Once the physical database was developed, the researcher started planning the development of the interface for capturing data into the database. Table 4-4 provides a summary of the steps the researcher followed in the second development DSR cycle of developing the interface.

Development cycle 2 – User interface

Table 4-4: Second development DSR cycle – user interface.

Cycle	Step	Activity
Awareness	Problem:	Using Microsoft Access 2016, data capturers were struggling to capture and retrieve data efficiently. Additionally, this solution didn’t allow multiple users to efficiently capture and retrieve data in the database; it caused data inconsistencies, human error, poor data quality, poor data management, poor

Cycle	Step	Activity
		data security, no robust processes or methods to ensure data quality, and slow performance. ❖ Data capturers were unable to complete their tasks on time. ❖ Management were unable to make informed decisions using the data. ❖ Unreliable and inaccurate data affected decision making and production process. ❖ Management did not have confidence in the data and did not trust the reports generated using the data.
Suggestion	Plan	To determine a solution to the identified problem the researcher suggested the following actions: ❖ The researcher requested access to the current solution to analyse and gain in depth understanding of the problems. ❖ Requested a workshop session with employees relevant for this study. ❖ To install Microsoft Visual Studio 2019 (C#) on the researcher's computer.
	Requirements	The following actions formed part of requirements gathering: ❖ A workshop session was held with the users to gather more information about the problem. ❖ During the workshop, the data capturers demonstrated how the current system works and explained all the issues. ❖ The researcher gathered all the information needed for the study by taking notes during the discussions. ❖ The users also gave the researcher permission to record the workshop session which will form part of requirements gathering. ❖ IT administrator gave the researcher a requirements document outlining the functionality of the artefact. ❖ Based on all information gathered the researcher decided to develop a logical design of the artefact. ❖ Microsoft Visual Studio C# was installed on the researcher's laptop to be used for development.
	Design	The researcher used all the information gathered during the workshop session to develop the logical or tentative design of the user interface. ❖ Lucid chart diagramming application was used to develop the high-level use case diagram to demonstrate the functionality of the artefact. Actors and use cases were outlined in the use case diagram. The tentative design is an output in this phase and will be used in the development phase of the artefact.

From the suggestion phase, a tentative design was produced to guide the development phase. To construct the interface, the researcher used tools such as ASP.NET Microsoft Visual Studio 2019 (C#), cascading style sheet (CSS) and JQuery to develop the web-based solution.

To develop a high quality interface, the researcher took into consideration a set of conditions defined by Ali *et al.* (2020:5) for the development of an artefact (section 2.7). The researcher ensured that users will find the artefact easy to use, will find comfort and pleasure in using the artefact, that the artefact will enable efficient capturing and retrieval of data, kept fonts on all labels visible and readable, and all the interface components were kept in order and presentable. In addition to the set of conditions defined in section 2.7, the researcher also implemented error messages that are descriptive and easy to understand, as well as validation check messages to display whenever a particular condition is not fulfilled or when certain restrictions are violated.

Data validation was embedded within the source code and database constraints validated the data to check for accuracy, completeness, uniqueness and validity. Interface validations error messages displayed in a red colour to quickly indicate where issues are. The researcher used warm colours throughout the interface to ensure visibility, readability, the colour was not harmful to the eyes and to ensure ease of use. Furthermore, the researcher would apply proactive data quality management which involved revealing and eliminating particular data quality issues before they occur by embedding data validation rules or checks and strong data quality controls in data capturing systems and databases as described in section 2.7 by Mahanti (2018:319). A link creating communication between the newly created SQL relational database and the interface was defined to enable capturing and retrieval of data.

To manage the development process efficiently, the researcher followed the agile method of developing the artefact by breaking it down into small manageable components. The development of the artefact was split into five components: the development of the master page, customer page and marketing page (component 1), development of login and home page (component 2), development of roll codes page, material specification page and quality assurance page (component 3), and lastly, the development of the heat treatment page and scrap page.

In the subsequent sections, the researcher will elaborate on the development and evaluation of these components. The awareness and suggestion phases have already been discussed in Table 4-3 and Table 4-4. Evaluation of these components was conducted with the data capturer (P2) and the IT administrator (P1).

Development cycle 2: Component 1 (Customer, marketing and master page)

Based on the tentative design produced in the suggestion phase, the researcher commenced development with the master page. The purpose of the master page is to provide a template for all web pages in the artefact. The master page mainly includes the navigation links, texts, headers, footers, images and graphic elements which provide layout and functionalities that will be visible and accessible throughout all the subpages. The customer page was developed to assist with capturing basic customer information such as customer name, email address, telephone number and business address, and to provide efficient functionality to search and retrieve customer details. Table 4-5 provides a summary of the steps followed in the second development DSR cycle of component one of the user interface.

Table 4-5: Development cycle 2 – user interface (component 1).

Cycle	Step	Activity
Development	Develop	The researcher developed the master page with navigation links to access the home page, customer, marketing details, roll codes, material specification, quality, heat treatment and scrap information. Additionally, the company logo, logout and welcome links formed part of the master page. The customer page enabled the data capturer to capture, retrieve, update and delete customer details in a SQL database. The marketing page enabled the data capturer to capture order details per customer.
Evaluation	Testing	The first evaluation session was held virtually with P1 to evaluate the look, visibility, colours, pictures and layout used to design the master page. ❖ The criteria used for testing: simplicity, ease of use, functionality, consistency, relevancy, elegance, efficiency and accuracy. The second evaluation session was conducted in a workshop where the researcher demonstrated component 1 of the artefact to P1 and P2. The functionality, completeness, understandability, accuracy, efficiency, efficacy and consistency were used to evaluate component 1 of the artefact. ❖ Functional (black box) testing: The artefact was executed to determine defects and failures.
	Feedback	In the initial evaluation of component 1, P1 indicated that the layout and colours used by the researcher are satisfactory. During the workshop testing session, participants requested the following changes: ❖ Change email to be mandatory in the customer page and validate the format.

Cycle	Step	Activity
		❖ Telephone number should not be mandatory. ❖ Search functionality must use any value in the grid view to search for customer. ❖ P2 further indicated that the roll category field should be a textbox and not a dropdown list in the marketing web page.

Development cycle 2: Component 2 (Development of login and home page)

The login page controls access to the artefact by allowing only users who are registered and have valid usernames and passwords to access the artefact. The system administrator is responsible to register users. Furthermore, the home page is the landing page and enables users to search customer information stored in the database. The user can either use the project number or customer name to efficiently retrieve data from the database. Table 4-6 provides a summary of the steps followed in the second development DSR cycle of component two of the user interface.

Table 4-6: Development cycle 2 – user interface (component 2).

Cycle	Step	Activity
Development	Develop	The researcher developed the login page to enable only permitted users to access the artefact. Username and password are used to access the system. User credentials are stored in the database and are validated upon login. Password are hidden and can be changed by the user anytime. In a case where the user had forgotten his/her email, the user must send an email to the administrator requesting a password change. Change management process will be followed. The search functionality in the home page enabled the user to efficiently search and retrieve customer information from the database.
Evaluation	Testing	In the initial testing session, P1 tested access to the artefact. The researcher registered P1 as the system administrator responsible to manage the artefact. The username and password were used to connect to the artefact. During the workshop testing session the researcher demonstrated to P1 and P2 how to connect and access the artefact. P1 is responsible to register and grant users access to the artefact. Once the users were granted access, they were able to login into the artefact, change their passwords and update their user profiles. Users can send an email to P1 if any issues are encountered with login.

Cycle	Step	Activity
		Criteria used to evaluate component 2: Accessibility, functionality, consistency, relevancy, understandability, robustness, efficiency, accuracy, completeness and ease of use. ❖ Functional (black box) testing: The artefact was executed to determine defects and failures.
	Feedback	❖ No changes required. Usernames and passwords were authenticated successfully against credentials stored in the database. Error messages were displaying accordingly. ❖ The accuracy and completeness of the username and password were validated accordingly displaying understandable validation error messages. ❖ Correct credentials enabled participants to access the home page and the welcome user link displayed the participant's name. ❖ Password was hidden, not visible to other users.

Fig 4-6 represents the login page that the administrator and the data capturers will use to login and connect to the data capturing solution.

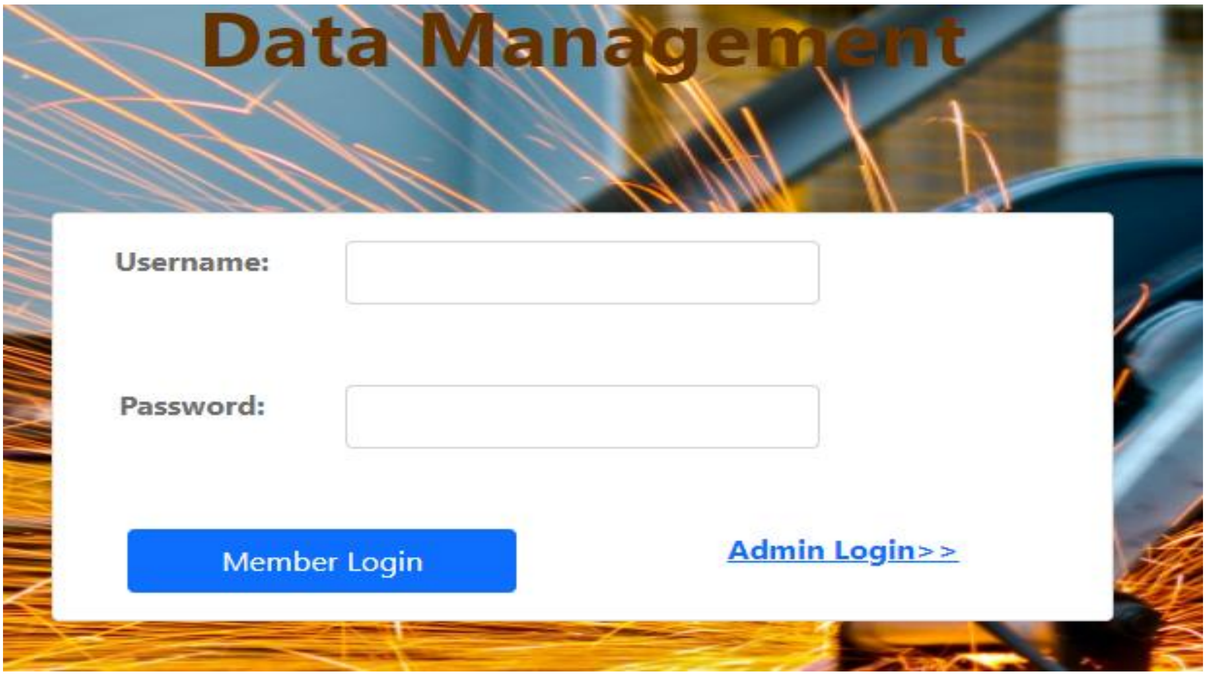


Figure 4-6: Login page of the artefact.

Once the users have been authenticated and granted access to the application, the home page with navigation links as outlined in Figure 4-7 are displayed. The home page enables users to efficiently search customer and project information stored in the database. The navigation links allow the users to access various functionalities of the application.

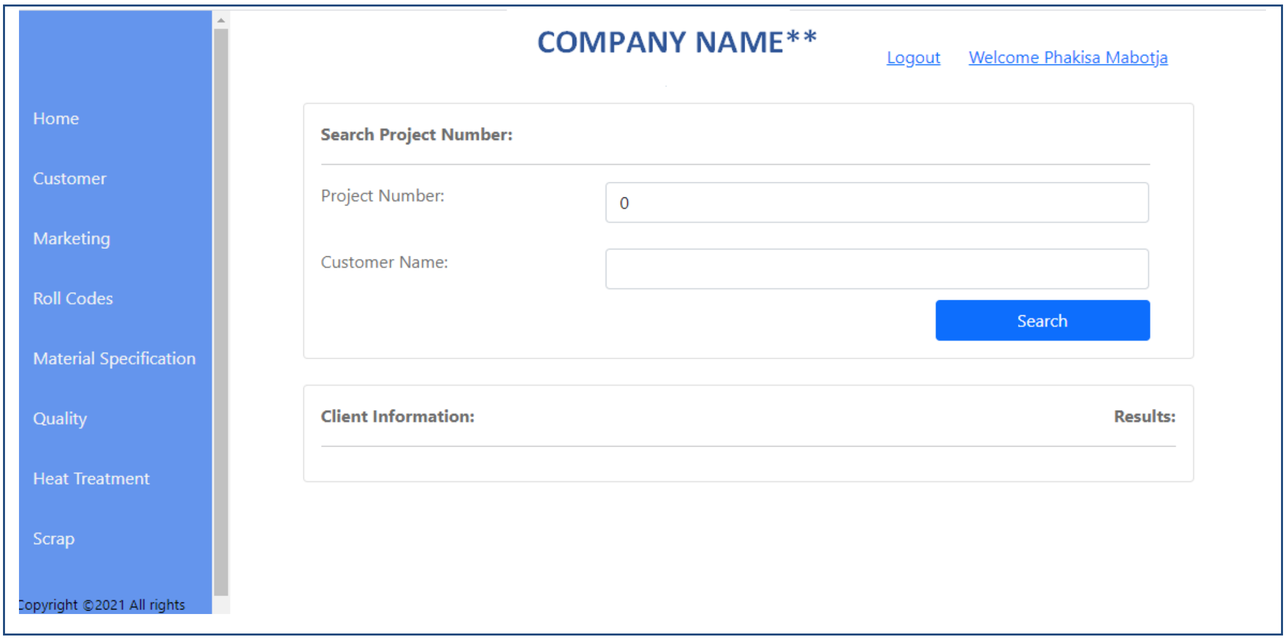


Figure 4-7: Home page and master page of the artefact.

Development cycle 2: Component 3 (Development of roll codes, material specification and quality assessment web pages)

Component 3 illustrates how the roll codes, material specification, quality requirements and inspection web pages were developed. Roll codes and material specification enabled the user to capture codes, measurements and elements needed for manufacturing the steel rolls. Furthermore, quality requirements and inspection web pages provided the functionalities for capturing, retrieving and deleting roll quality data. Table 4-7 provides a summary of the steps followed in the second development DSR cycle of component three of the user interface.

Table 4-7: Development cycle 2 – user interface (component 3).

Cycle	Step	Activity
Development	Develop	The search functionality to retrieve roll details stored in the database was developed. All fields required for capturing and updating roll codes, material specification quality requirements and quality inspection details were created. Dropdown lists were prepopulated with data in the database. Validation messages were implemented to validate mandatory fields and to check duplicates and null values. Data formats

Cycle	Step	Activity
		were checked (date, string, numeric) and appropriate error messages were implemented. The web pages should enable the data capturers to efficiently retrieve data using roll project no, capture data, and delete data in the database.
Evaluation	Testing	Criteria used for testing: Accessibility, functionality, consistency, relevancy, elegance, understandability, robustness, efficiency, accuracy, completeness and ease of use. ❖ Functional (black box) testing: The artefact was tested to determine defects and failures. ❖ P1 and P2 tested the component 3 and provided feedback.
	Feedback	❖ Text on the button search should be changed from search to retrieve data. ❖ Enlarge the size of the search button. ❖ Decimal values must be changed from two decimals to three.

Figure 4-8 represents the quality requirement web page developed in component 3 of the artefact design.

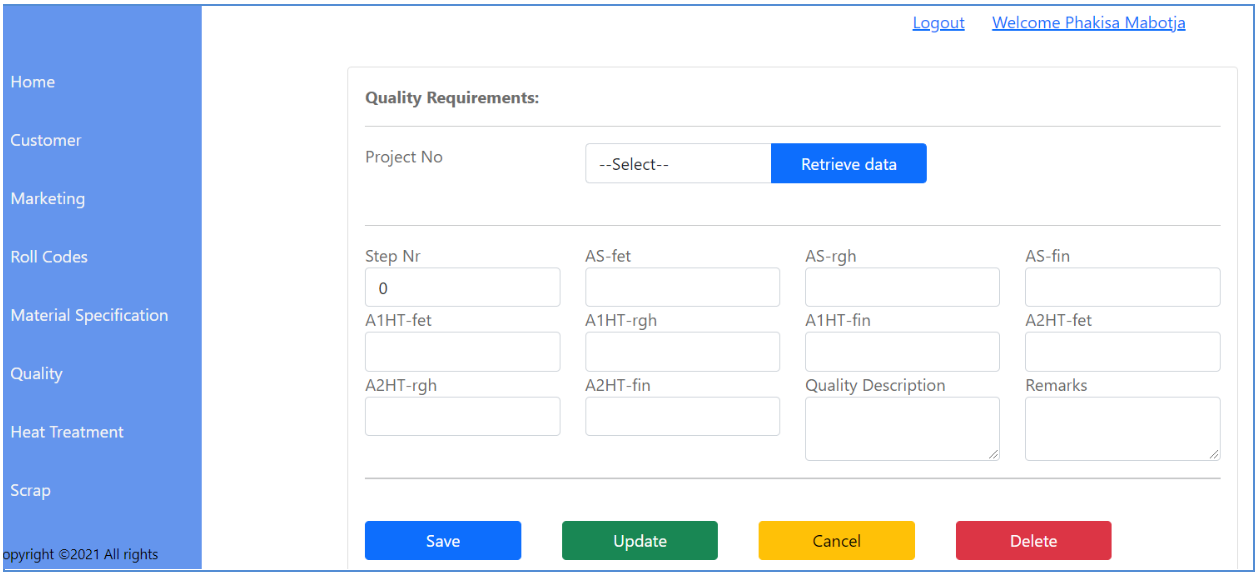


Figure 4-8: Quality requirements page.

Development cycle 2: Component 4 (Development of heat treatment page and scrap)

The following web pages enabled the user to capture details regarding the heat process of the roll into the database. The user can use the scrap page to capture rolls that failed the quality assessment stage. Table 4-8 provides a summary of the steps followed in the second development DSR cycle of component four of the user interface.

Table 4-8: Development cycle 2 – user interface (component 4).

Cycle	Step	Activity
Development	Develop	The researcher developed all fields required for capturing the heat treatment process and roll scrap details. Dropdown lists were prepopulated with data in the database. Validation messages were implemented to validate mandatory fields and to check duplicates and null values. Data formats were defined accordingly (date, string, numeric) and appropriate error messages were implemented. The web pages provide functionalities such as searching and retrieving data stored in the database using roll project no, capturing new data, update existing data and deleting unwanted information.
Evaluation	Testing	Criteria used for testing: Accessibility, functionality, consistency, relevancy, elegance, understandability, robustness, efficiency, accuracy, completeness and ease of use. ❖ Functional (black box) testing: The artefact was tested to determine defects and failures. ❖ P1 and P2 tested the component 4 and provided feedback.
	Feedback	Change the text on the search button from search to retrieve data.

Figure 4-9 outlines the heat treatment web page which is one of the web pages developed as part of component 4. This web page enabled the data capturers to efficiently search and retrieve data related to a particular project number in the database. Furthermore, it also allowed data capturers to capture new roll heat treatment details, update existing details in the database, and delete unwanted data.

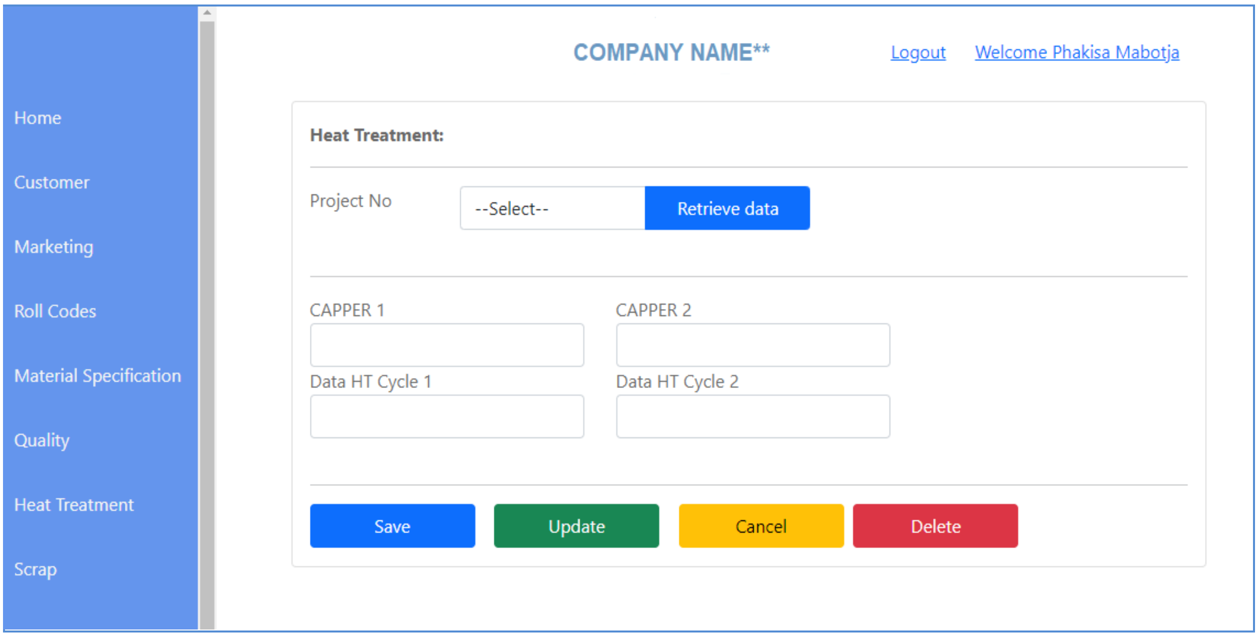


Figure 4-9: Heat treatment web page.

It took twelve weeks for the researcher to complete the development phase of the artefact because it involved an iterative process of developing and evaluating the artefact until all modules were functioning according to expectations. Feedback received during the evaluation phase was used to revise the design, to implement the changes and improve the functionality of the artefact. Once all components were completed, the researcher linked all the web pages to function together as one unit.

Development cycle 3 – Data warehouse and ETL solution

The purpose of this section is to discuss the development of the DW and data integration using the SQL server data tools. Due to large volumes of data, poor data management and data quality of the organisation, inefficient reporting and data analysis, as well as poor decision making, the SME decided to explore other technologies that can assist with integrating data stored in various locations into one trusted source database that can be used for querying, reporting and analysis. As discussed in section 2.9.1, organisations can use DW/BI technologies to extract data from various sources, clean and transform the data, and store it in a centralised database or data warehouse for querying, reporting and analysis. In this study, the researcher used the Kimball methodology to develop the DW/BI solution where the dimension modelling approach was used as stated in section 2.9.1.

Data warehousing

Data warehousing involved integrating a collection of data stored in various locations into one single trusted source. Due to data quality issues within the historical data, an ETL solution was required to clean the data and remove data inconsistencies before the data are loaded into the data warehouse (section 2.9.2). The ETL solution was developed using SQL server integration services (SSIS) in SQL Server Data Tools 2019. The researcher followed Figure 2-10 (section 2.9.2) to develop the ETL process. Microsoft management studio was used to develop the staging database and the data warehouse, which were hosted on the SQL database management server (DBMS). The data integration process was handled as follows:

Extract

As seen in Figure 2-10 (section 2.9.2), data were extracted from the live SQL operational database, which is a newly developed relational database used for capturing data using the artefact, MS Access legacy database which was the old database with historical production data, and lastly, the excel spreadsheet with customer information. The researcher developed the SQL stored procedures and the SSIS data integration packages to extract data from the three sources and integrate it into the landing environment. The landing database was designed to suit the live SQL latest database structure to maintain consistency. The landing database will be truncated and loaded with latest data at all times to ensure that the DW is populated with the latest information.

Once the data were loaded into the landing environment, the staging database was developed based on the Kimball dimensional modelling. The staging database dimensional structure consisted of the dimension tables to store the attributes and one-fact tables for storing the measurements. The dimension tables and fact table have a one-to-many relationship. An automated process of updating the staging database with latest data from the landing database was created.

Transform

In this stage, data quality issues such as data inconsistencies, duplicate data, data redundancy, and inaccurate values are transformed and cleansed. Business rules are implemented, null values are addressed, and incorrect values and incompatible formats are fixed, with surrogate keys assigned. All data originating from the source are converted and cleansed to suit the structure of the destination. An ETL process was followed in extracting, transforming and loading data into the central location as indicated in section 2.5.2. The researcher developed SSIS packages for extracting, transforming and loading data into the data warehouse as depicted in Appendixes 4 and 5.

Load

Once the data were transformed to suit the DW structure, data were loaded into the data warehouse, i.e., the fact and dimensions. The surrogate keys were defined as the primary keys in the dimension table. The primary key in the fact table is described by the combination of the surrogate key from the dimensions.

Development cycle 4 – SSRS reports and Power BI analytics solution

The next section will discuss the development of the reporting structure. Based on the requirements received from the participants, they indicated that the one SA SME requires a solution for generating reports efficiently. Test certificate reports are used by management for investigation whenever there are issues regarding the rolls, and they are also sent to customers to provide information on the material specifications used to create the roll. This report entails information about material specification of the roll, hardness, chemistry used, and quality assessment information. These reports are also sent to customers and are used by management for investigation and decision-making, so it is critical to ensure that the data that are used in these reports are of high quality.

Initial reporting solution

Participant 1 in this study explained that one SA SME didn't have reporting and analytics tools or applications to efficiently generate reports. MS access was used to generate the test certificate reports. However, based on the challenges and limitations presented by MS Access, the SME decided to implement a reporting solution.

New reporting solution project plan

Based on the requirements the researcher received from the participants, reports were developed using SQL Server Reporting Services (SSRS). The researcher designed four quality assessment reports also referred to as test certificates that would be sent to customers to indicate the details for the roll orders and the material specifications that were used to design the rolls.

Table 4-9: New reporting and analytics development

Cycle	Step	Activity
Awareness	Problem	One SA SME didn't have a reporting solution that management can use to efficiently generate reports needed for decision making and for analysis.
Suggestion	Plan	To determine a solution to the identified problem. the researcher performed the following actions: ❖ The researcher interviewed the participants to gather more information about the problem.

Cycle	Step	Activity
		❖ Assessed the infrastructure to determine if there were sufficient resources to host and manage the reporting solution.
	Requirements	The following actions formed part of requirements gathering: ❖ A meeting was held with the data capturer and IT administrator to gather more information about the problem. ❖ To manage the scope of the project, only four reports were designed. ❖ The IT administrator provided the researcher with sample reports that are needed. ❖ The researcher determined the individuals who will be using the reports. ❖ Information needed in the reports for management and customers were discussed. ❖ The number of the reports needed were determined. ❖ The layout, formats and colours, images, calculations that are needed in the reports were discussed. ❖ SSRS was selected as the appropriate reporting tool.
Development	Develop	The researcher used all the information gathered during the workshop session to develop the reports: ❖ The initial report was designed based on the requirements. ❖ The SSRS report was linked directly to the database to efficiently read data and display in the report. ❖ A sample report was sent via email to the data capturer for evaluation.
Evaluation	Testing	❖ The report was evaluated for accuracy, completeness, trustworthiness, elegance, completeness and efficiency. ❖ P1 and P2 evaluated the reports.
	Feedback	❖ The participant indicated that there are a few fields that must be added in the report. ❖ The name of the quality assurer had to be changed.

Based on the feedback received, the researcher revised the design and added additional fields that P1 provided. Furthermore, the incorrect quality assurer name was also changed. P1 and P2 evaluated the report for accuracy, efficient data retrieval, consistency, completeness, and functionality, ease of use, security and understandability. The participants confirmed that the reports are accurate and all information is visible.

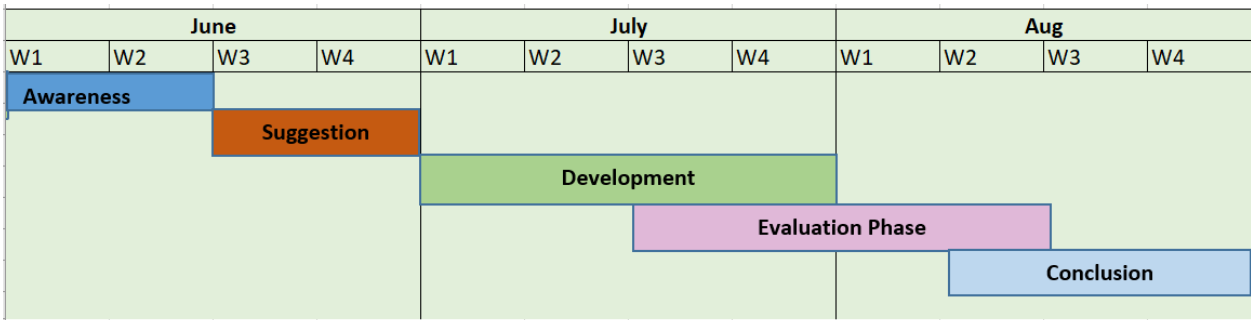


Figure 4-10: Gantt chart for reports development.

The researcher installed and configured the SQL report server on the one SA SME environment and deployed the reports. The reports were configured to link directly to the database, extract data and populate onto the reports. The researcher granted the IT administrator and the data capturer access to the reporting server. Active directory credentials were used to grant access to the users. Furthermore, Power BI was configured to read and analyse data stores in the data warehouse as indicated in Appendix 6.

4.4 Conclusion

The design and development process of the artefact for improvement of data quality and management in one South African (SA) small medium-sized enterprise (SME) was discussed. This chapter provided the process followed in designing and developing the solution as per the design science research model selected for this study. The main research cycle and three design sub-cycles preceding the development of the artefact were scrutinised. The awareness phase of the main research cycle explained the problem and challenges in one SA SME. Based on the issue presented, sub-cycle one was used to explore literature to gain in-depth understanding on data quality and management. Additionally, sub-cycle two was used to explore literature to investigate how organisations could exploit technology to solve data quality and management issues. Subsequently, an exploratory study was conducted to provide the researcher with more knowledge and understanding about the phenomenon and its data related issues. The feedback from the exploratory study was described. The awareness phase, suggestion phase and development phase of each sub-cycle were discussed.

An overview of the SME involved in this case study was provided in this section and figure 4-1 represent the information technology department structure. The original data management solution with its challenges and limitations was examined in section 4.2.3. Furthermore, the researcher discussed the artefact development processes followed based on the design science research model selected for this study. Figure 4-4 outlined the four development cycles circumscriptions. The high level use case diagram of the user interaction with the artefact was represented in Figure 4-5.

To develop the artefact, Microsoft Visual Studio 2019 (C#) was used to design the interface and MS SQL Server 2019 was used to design the relational database. The development process involved an iterative process of developing and evaluating the components of the artefact. The four development cycles circumscriptions involved the development of the SQL relational database as described in Table 4-3, user interface for capturing data described in (Tables 4-4, 4-5, 4-6, 4-7 and 4-8), followed by the development of the data warehouse and ETL (extract, transform and load) process for transforming and cleansing the data and, lastly the development of SQL Server reporting solution and analytics. In Chapter 5, the implementation and design science research evaluation phase of the artefact will be discussed.

CHAPTER 5: DEMONSTRATION AND EVALUATION

5.1 Introduction

In this chapter the researcher will discuss the artefact implementation process, the results and findings of the study. In the process, the researcher will address the practical objectives of the research study and the plan the researcher followed in answering the research questions in order to determine a solution to the research problem. A case study research strategy was followed to gain an in-depth understanding of the problem and to construct a solution to meet one SA SME's data quality and management needs. The design science research (DSR) process model of Vaishnavi *et al.* (2004/2019:14) as outlined in Chapter 3, Figure 3-8 (section 3.8.3) was followed to develop an innovative artefact for capturing and managing data in one SA SME.

Mixed method data collection techniques such as interviews, observation and a questionnaire were used to collect data from the participants. Participants' knowledge, personal experiences and involvement in the study contributed towards the success of the research study. Information acquired from the literature, as well as research study findings and results were used to compile the guidelines and technologies needed to improve data quality and management in one SA SME.

Section 5.2 will explain the participants involved in the research study.

Section 5.3 will discuss the evaluation phase of the main research cycle of the DSR methodology

The researcher will discuss the data collection techniques used in section 5.4, observation (section 5.4.1), questionnaire (section 5.4.2) and semi-structured interviews (section 5.4.3).

The researcher will discuss the results of the research study in section 5.5 where the themes and codes used to analyse data using ATLAS.ti are discussed. Additionally, results from the questionnaire analysed using SSPS are also discussed.

Section 5.6 presents the guidelines and technologies for improving data quality in one SA SME. Lastly, section 5.7 will discuss the conclusion of this chapter.

5.2 Participants

According to Yin (2018:279) the greatest mistake a researcher can make is to formulate a case study from an egocentric perspective by completing a case study without identifying a particular audience or without understanding the needs of the audience. Kumar (2019:363) emphasises that there are a number of factors that the researcher can take into consideration when selecting people to participate in a research study, such as: the researcher's judgement that the individual has immense knowledge about the problem, phenomena or situation of interest, the simplicity in accessing the prospective respondents, and how typical the case is of a category of people.

In this study the researcher selected a sample that is unbiased and that can provide in-depth and accurate information needed to understand the phenomena and to answer the research questions. The appropriate sampling technique to be followed in selecting a sample for this study is purposive sampling, because the researcher was already aware of where the participants would be selected from and believed that these participants would provide immense knowledge about the phenomena or situation. The senior manager recommended the IT administrator and the data capturer to participate in this study because the senior manager believed they would provide the researcher with critical information needed in this study. The SME has one IT manager, one IT administrator, document controller, technical manager and a data capturer as indicated in Figure 4-1 section 4.2.3. The IT administrator and data capturer agreed to participate in the research study to share their knowledge with the researcher. The data capturers, IT administrator and IT manager have extensive knowledge about the phenomena, the situation, current data management architecture, the technical architecture of the organisation and the data quality issues (section 4.2.3).

Data capturers

The data capturers are familiar with the current data management tools and technologies that the SME is currently using as they are responsible for capturing data on a daily basis. These participants can provide immense knowledge on the current data management infrastructure, the challenges and limitations presented and needs of the SME.

IT Administrator

The IT administrator was selected to provide in-depth knowledge about the SME's IT infrastructure and to share more insight about the phenomena and problems from a technical perspective in a broader context. This participant has a high-level understanding of different sources of data, which tools are used to capture, store and manage the data, which IT resources (hardware, software, servers) are in place, which database management systems and databases are used, and security and other features used to manage the environment.

IT Manager

The senior manager was selected to provide insight and vast knowledge about the phenomena and the role of data management architecture in business management, processes modelling and decision making. This participant's experience and knowledge provided the researcher with a better understanding of the problem in a wider context, the limitations and challenges from a business perspective, and in-depth knowledge about the SMEs business processes, functionalities and decision-making processes.

5.3 Evaluation – main research cycle

The evaluation phase of the main DSR model for the research enabled the researcher to rigorously test the functionality of the artefact. The researcher applied the criteria presented in Table 3-8 in section 3.9 to evaluate the artefact. The five components were evaluated separately and once the researcher was done with the construction of the entire artefact the criteria were reapplied to evaluate.

The artefact was deployed on one SA SME server to evaluate its functionality and stability in one SA SME environment. During the deployment phase, the connection between the database and the user interface failed. The user interface could not establish connection to the database to enable capturing and retrieval of data. To solve the issue, the researcher created a user in SQL database and granted the user sufficient rights (read and write) and updated the configuration manager of the user interface with the username and password of the created user in the SQL database to establish connection to the database. Once configured and tested successfully, the user interface was able to communicate with the SQL database.

Furthermore, participants were requested to access the user interface link to connect and start evaluating the artefact rigorously. Sonnenberg and Vom Brocke (2012:7) state that the evaluation of the artefact is an important step in DSR to rigorously prove that the artefact designed is functioning well in practice and it engrains the solution to the asserted problem. Reliability, consistency, completeness, usability, functionality, accuracy, performance and fit with the organisation are some of the attributes that researchers use to evaluate the IT artefacts (Hevner *et al.*, 2004:85). In this phase, the researcher applied a combination of attributes by Hevner *et al.* (2004:85) and attributes presented by Sonnenberg and Vom Brocke (2012:5) in section 3.9 to evaluate the artefact (Table 3-8).

The evaluation phase involved testing the artefact to discover defects and failures to improve the functionality of the artefact. The researcher created test cases and scenarios that were used to evaluate the artefact with P1 and P2. Table 5-1 describes some of the test cases used to test the artefact, expected results and feedback from the participants. The researcher designed these test cases to assess the artefact's functionality, accessibility, simplicity and elegance, ease of use, visibility, readability, efficiency, completeness, data accuracy, security, and understandability.

Table 5-1: Test case used for evaluating the artefact

Test description	**Expected results**	**Pass/Fail**	**Comment**
Enter valid username & password then click member login button.	Login successful message & redirected to home page.	Pass	
Incorrect username or password then click member login button.	Validations (enter username & password).	Pass	
Password is hidden.	*******	Pass	
Multiple users connect simultaneously.	Multiple users can connect simultaneously to capture and retrieve data efficiently.	Pass	
Click navigation link.	Links are visible, readable and when clicked redirect to relevant pages.	Pass	
Colours, font size, labels, textboxes, buttons are visible and text are readable.	Users are able to see and read all the details. Colours are not harmful for the eyes.	Pass	Some textboxes, dropdown lists, buttons must be enlarged to allow the text to fit properly. Colours are fine. Besides that, all looks good.
Email address not in the format (Tom@gmail.com).	Invalid email address error.	Pass	
Telephone number not accepting string values.	Error message displayed when string values are entered.	Pass	
Project number dropdown list prepopulated with project numbers in the database.	Dropdown list prepopulated with project numbers.	Pass	
Data formats, numeric fields, date and strings are defined accordingly.	Date format [dd/mm/yyyy].	Pass	
Search roll details using project number or customer name.	Efficient retrieval of roll and customer details.	Pass	
Search using incorrect project no or customer name.	Error message displaying project number does not exist).	Pass	
Duplicate project number & click save.	Error message (no duplicates allowed, update data).	Pass	
Save, update and delete buttons are functional.	User is able to capture, update and delete data in the database.	Pass	
Blank mandatory fields & click save.	Validation error messages.	Pass	
Data formats defined accordingly.	2 or 3 decimal, numeric, string, dates.	Pass	

Test description	Expected results	Pass/Fail	Comment
Search functionally.	Efficient retrieval of data.	Pass	
Consistent project numbers.	Project numbers are correct and consistent throughout.	Pass	

Based on the feedback received, the researcher revised the solution and implemented the changes. Only one issue was reported by the participants. Dropdown lists, textboxes and labels were enlarged to ensure that the text fits properly, are visible and readable.

5.4 Data collection

In this case study, a mixed method data collection technique was used to collect data from participants. Yin (2018:174) advocates that researchers conducting a case study should utilise diverse data collection techniques in ensuring that the case study is based on several sources of evidence, because without such sources, the purpose and worth of the case study will not be fully achieved. Semi-structured interviews, a questionnaire and observations were used to gather data. Initially, the researcher started gathering data in sub-cycle 3 as indicated in Figure 3-11 in section 3.11, where an exploratory study was conducted by sending interview questions as depicted in Appendix 1 via email to the IT administrator and IT manager to answer and provide the researcher with more information about the phenomenon and the problems. The purpose of the questions was to enable the researcher to gain an in-depth understanding of the situation, their challenges and needs, understand the tools and applications used and the existing data management issues.

5.4.1 Observation

Subsequently, a workshop testing session was held with the participants where the researcher collected information through observation. The researcher observed participants interacting with the artefact to see what impact it had on processes and people within the SME. The participants tested the functionality of the artefact. Furthermore, the accuracy, completeness, uniqueness, consistency, etc. of the data were tested. The researcher's results on observation are discussed in section 5.5.

5.4.2 Questionnaires

Once the artefact was developed, evaluated and implemented in the one SA SME environment, the researcher formulated a questionnaire through Google forms and sent them to the IT administrator and the data capturer to answer and give feedback regarding the functionality of the artefact solution. The purpose of the questionnaire was to gather data about the functionality of the artefact, and access participants' thoughts and perceptions about the artefact during post implementation. Appendix 2 depicts the questionnaire that participants were required to complete. When formulating the questionnaire questions, the researcher took into consideration the guidelines outlined in section 3.5.2 to ensure that the participants did not feel frustrated, embarrassed or uncomfortable in answering the questions.

5.4.3 Semi-structured interviews

Lastly, a follow-up interview session was conducted with the participants to gather more data needed for analysis and reaching the conclusion of the study. The researcher chose semi-structured interviews as the data collection technique because it provided a level of flexibility in the way the researcher posed the questions, and additionally it allowed participants to tell their story and discuss items with the researcher (section 3.5.1). Based on the data quality and management literature, the researcher formulated interview questions to ask two participants about the functionality of newly implemented data capturing and management of the artefact. The researcher ensured that the interview questions were formulated as outlined in Appendix 2 to obtain all required information needed. The researcher sent an email to the IT administrator and the data capturer requesting an interview session with them. They accepted the invitation and agreed to do a face-to-face interview session at one SA SME's premises.

To ensure that participants were not coerced in participating in the interview session, informed consent forms were signed before the interviews (section 3.10). Furthermore, the researcher requested to record the interview session and explained to the participants that all their personal details will be kept confidential and the data collected during the interview process will be transcribed and used only for analysis and research conclusions, and will be stored in a secure location for a period of five years as discussed in section 3.10. The researcher ensured that the interview was conducted professionally, in an ethical manner, ensuring that participants' interests were well protected from any form of harm.

The interview sessions between the researcher, IT administrator and the data capturer were recorded, then transcribed. The questionnaire was analysed using SSPS and the transcription of the interview was analysed using ATLAS.ti. The semi-structured interviews enabled the researcher to gain insight into one SA SME participants' experiences, thoughts, behaviour and understanding about the artefact. The results were analysed, interpreted and used for making recommendations and reaching the conclusions.

5.5 Results of the research

This section will discuss the results obtained from data collected from the IT administrator and data capturer. The objective of the researcher in conducting the semi-structured interviews was to collect data from participants in order to gain a better understanding of their thoughts, experiences, and challenges in using the new artefact. The results are based on feedback received from participants on the functionality of the new data management tool, user experience, complications with the new solution, the impact of the new solution on the business, and the new reporting solution. Feedback received from Interview questions outlined in Appendix 2 were used define the themes and code in Figure 5-1 and Figure 5-2.

The researcher used ATLAS.ti which is a qualitative data analytical tool to analyse the transcribed data. In the process, the researcher scanned through the transcribed data to identify important information and assigned labels to that information. This process is known as coding, as described in section 3.6. The codes and themes applied to the transcribed data are explained and represented in the tables below. The following themes were identified for the study:

- ❖ **The old data capturing solution** represents the comments from the participants about the old data capturing solution that was used by the SME. This theme is based on the comments from the data capturer and IT administrator.
- ❖ **New artefact** represents the comments from participants about the new artefact that was implemented in one SA SME. This theme is based on the comments from the data capturer and IT administrator.
- ❖ **New reporting solution** represents the comments from participants about the newly implemented reporting solution. This theme is based on the comments from the data capturer and IT administrator.

The theme represented by Table 5-2 is the old data capturing solution. Table 5-2 outlined the number of comments made by IT administrator (P1) and data capturer (P2) based on the old data capturing solution and the codes were illustrated in Appendix 7 using a network.

Table 5-2: Theme: old data capturing solution.

Old data capturing solution – related codes	IT Administrator	Data Capturer	Total
Inaccurate data	2	2	4
Difficult to use	3	4	7
Inaccurate data	2	2	4

Old data capturing solution – related codes	IT Administrator	Data Capturer	Total
Inconsistent data	3	2	5
Inefficient access and retrieval of data	4	2	6
Poor data management	2	5	7
Poor data security	1	2	3
Poor decision making	1	_	1
Poor performance	3	5	8
Poor productivity	4	2	6
Poor reporting	1	1	2
System malfunction	3	3	6
Tedious	5	3	8
Time consuming	2	2	4
Unable to access system simultaneously	1	2	3
Unavailable or inaccessible data	1	_	1
Grand Total	**38**	**37**	**75**

The theme presented by Table 5-3 is new artefact, which will focus on the number of comments made by participants about the new artefact and were illustrated in Appendix 8.

Table 5-3: Theme: New artefact.

New artefact: related codes	IT Administrator	Data Capturer	Total
Able to handle large volumes of data	1	1	2
Complications – System security	1	-	1
Complication – System updates	2	-	2
Complication – System upgrades	1	-	1
Data are more consistent	5	1	6
Data formats are accurate	5	-	5
Easy to use	4	1	5
Efficient data capturing	4	2	6
Efficient data retrieval	5	2	7
Efficient data update	2	1	3
Evaluation – Defects identified and resolved	2	1	3
Evaluation – Issues communicated to developer	3	1	4
Evaluation – Issues resolved on time	3	1	4

New artefact: related codes	IT Administrator	Data Capturer	Total
Evaluation – Minor issues detected	4	1	5
Evaluation – Tested accordingly	1	1	2
High level of efficiency	3	3	6
Improved data management	1	3	4
Improved data quality	4	3	7
Improved user performance	2	-	2
It is solving majority of our data issues	3	2	5
Meets expectations	2	3	5
Multiple users can connect and access the solution	1	1	2
No duplicate data	3	1	4
Quick turnaround time to decision making	3	2	5
Simplicity	2	-	2
The artefact adds value to the organisation	3	1	4
The requirements that we asked for it's exactly what we are getting	1	-	1
The ability to capture data simultaneously	2	1	3
The artefact is functional	6	2	8
The artefact is pleasant	3	-	3
The artefact provides clean data	2	-	2
The artefact provides error massages	2	-	2
The artefact provides level of data security	2	1	3
The artefact validates data before capturing into the database	1	-	1
Grand Total	**89**	**36**	**125**

Table 5-4 represents the new reports theme which will focus on the number of comments made by participants about the new reporting solution implemented at one SA SME, and the codes were further represented in a network as illustrated in Appendix 7.

Table 5-4: Theme: New reports

New Reports: related code	IT Administrator	Data Capturer	Total
Accessible	1	1	2
Accurate data	1	2	3
Add value to the organisation	1	1	2
Automated dates	1	-	1
Consistency	1	1	2
Convenient	1	-	1

New Reports: related code	IT Administrator	Data Capturer	Total
Decision making	2	1	3
Efficient	2	-	2
Everything that is needed is there	1	1	2
I'm happy	2	2	4
Pleasant	2	-	2
Printable in any format	1	-	1
Reading data from the database	1	-	1
Used for investigation and decision making	2	1	3
Grand Total	**19**	**10**	**29**

The old data capturing solution presented one SA SME with various challenges and limitations. Figure 5-1 outlines some of the comments made by the participants regarding the old data capturing solution.

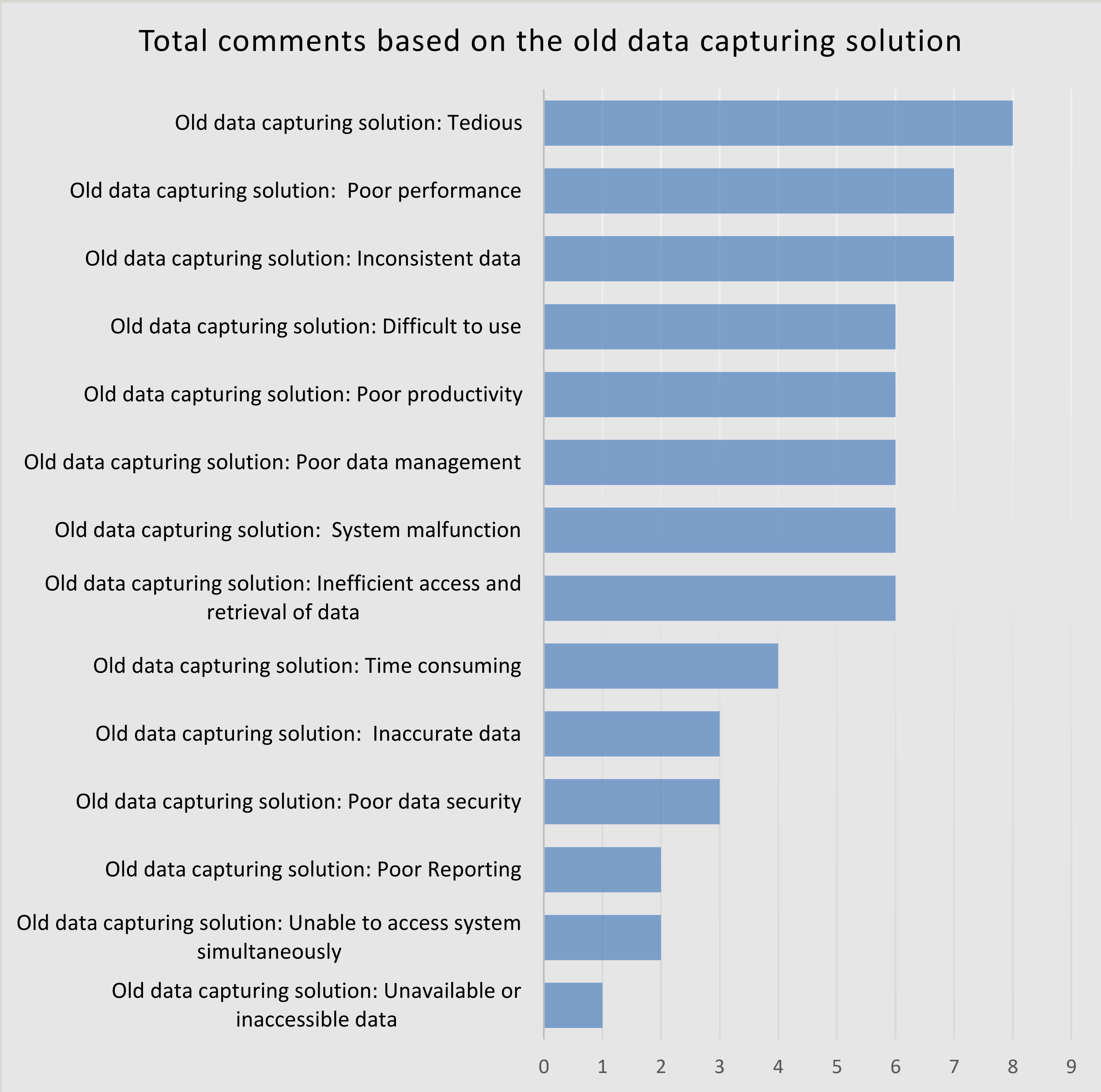


Figure 5-1: Total participant comments based on the old data capturing solution

The comments made by participants regarding the functionality and impacts of the new artefact are represented in Figure 5-2. .

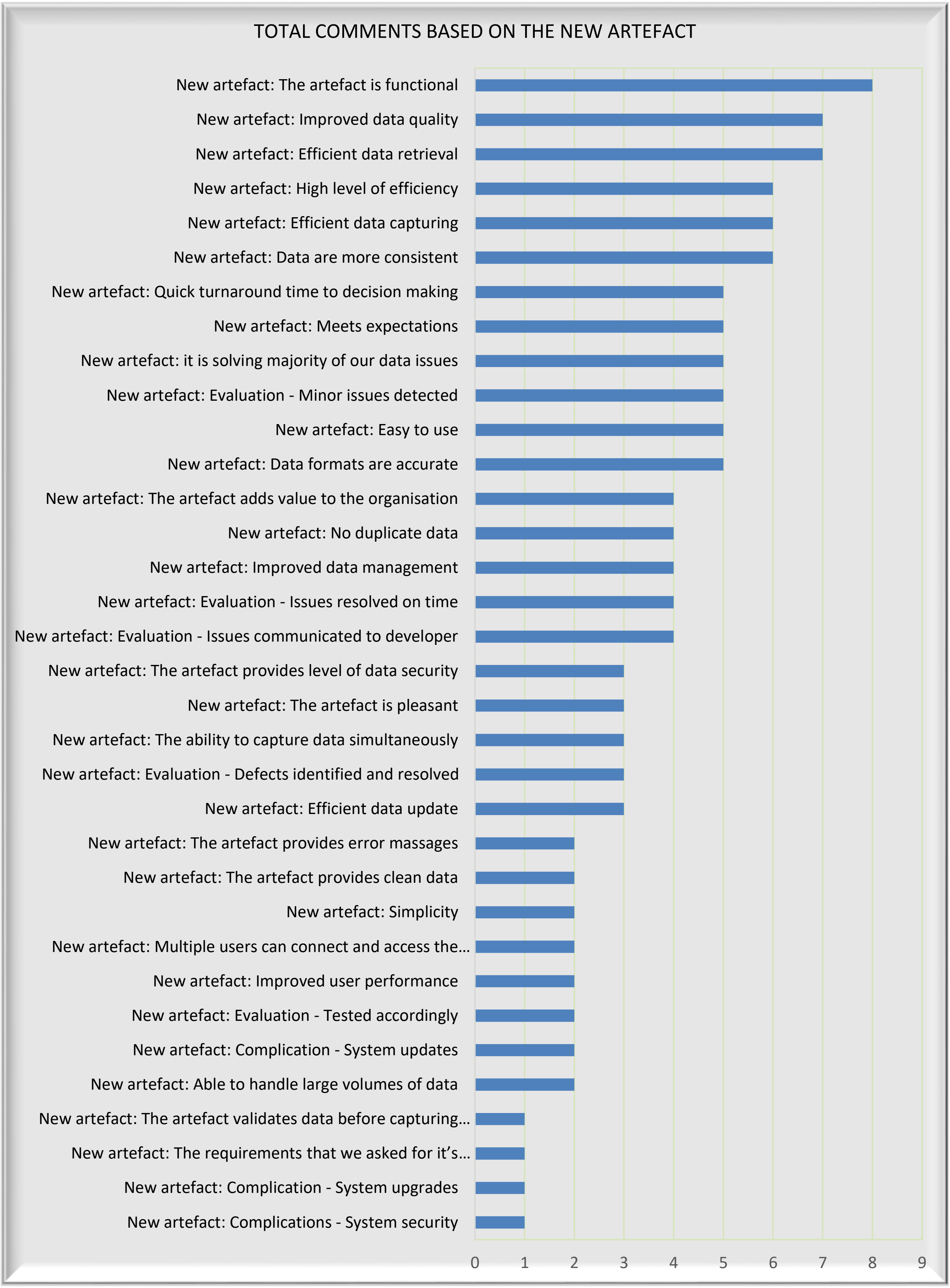


Figure 5-2: Total participant comments based on the new artefact.

The comments made by participants regarding the functionality of the new reporting solution are presented in Figure 5-3.

Figure 5-3: Total participants comments based on the new reporting solution.

Objective of implementing the data capturing and management artefact.

To understand the reasons why one SA SME decided to find solutions to their data capturing and management issues, the researcher collected data by interviewing the IT administrator (participant 1) and data capturer (participant 2). The following question was asked:

Interviewer: What motivated the decision to implement the user interface in the organisation?

Participant 1: Firstly, it was because of numerous data errors that we were getting and the inefficiency of data movement between departments. The old data management solution involved a tedious process of capturing and updating data while ensuring that other users had access to latest and relevant data. Also, extracting data was difficult and there was no proper data flow process. Additionally, data security was an issue, everyone had access to the data and could make changes. It was difficult to trust and have confidence in the data and use it for reporting and decision making because the accuracy of the data was not guaranteed.

However, for the artefact to satisfy all the requirements, it went through various test stages. The next section will address the defects and issues identified during the evaluation phase.

Defects and issues during artefact evaluation

In this section the participants gave feedback on the evaluation of the artefact to assess whether the evaluation stage was conducted ethically and that all issues were communicated to the developer to resolve. Several testing criteria such as accessibility, consistency, functionality, completeness, accuracy, elegance, understandability, efficiency, and ease of use were used to evaluate the artefact.

The participants' involvement in the evaluation stages as described in section 5.2 assisted the researcher to identify issues and complications in the artefact. This involved an iterative process of development, evaluation and communication until the final product was achieved.

Interviewer: Were you involved in testing the solution? Are there any defects and issues that were addressed?

Participant 1: Yes, I was involved and I can say standard testing was done and the solution was tested up to our expectations and to see if it satisfies our basic requirements.The were minor issues that were identified such as the project number value didn't fit properly in the dropdown list. Some of the textbox were small in size. But I think it was mainly caused by the resolution of some computers. These issues were not a serious setback.

Participant 2: Yes, I was involved in the testing process. The issues encountered were not major. It was textbox size issues and data format. Some values needed to be changes from two decimal place to three decimal place.

Both participants were involved in testing the artefact. The participants indicated that standard testing was conducted to see if the artefacts met their basic requirements. The participants indicated that there were no major issues, most of the issues only required the researcher to increase the dropdown list, textboxes and label sizes to accommodate long text while ensuring visibility. The project number was not displaying as expected. Participant 1 indicated that the value of the project number was not fitting properly within the dropdown list, the size needed to be adjusted to accommodate the text. The issues that were identified were communicated to the researcher to solve.

The participant explained that the initial timeframe of completion of the project was end of August. However, due to system upgrades within the organisation, things had not moved according to their expectations. However, after the evaluation phase was completed and all changes addressed, the researcher proceeded with artefact deployment. The next question addressed participant experience and thoughts on the artefact deployment phase.

Complications during deployment

Once the evaluation phase was completed and the artefact was deemed satisfactory, it was deployed to one SA SME environment. The developer had to deploy the artefact to their environment for further testing. The different environment comes with different challenges, so the following questions will focus on the complications encountered during the artefact deployment stage.

> **Interviewer:** If any, what complications were there with regards to the deployment of the artefact onto your environment?
>
> **Participant 1**: There were complications but not from the developer's side. They were due to the internal sytem updates that were performed. Additionally, security issues were encoutered when trying to publish the application through IIS, we could not establish a link between the interface and the SQL database. There were database and interface configuration issues but eventually apropriate rights were granted to the accounts that would be accessing the interface.

The artefact was successfully deployed to one SA SME server for further evaluation. The developer was granted access to the server to be able to connect remotely and attend to any requests or changes that may arise. Furthermore, the participants conducted another evaluation session to test the accessibility and functionality of the artefact after it was published in their environment. The researcher applied the criteria presented in Table 3-8 in section 3.9 to evaluate the artefact. A scale of 1 to 10 was used to measure the attributes of the artefact where 1 represented lowest and 10 represented highest. The overall feedback from participants based on the attributes defined in section 3.9 is discussed in Table 5-5:

Table 5-5: Overall feedback from participants based on attributes in section 3.9.

Criteria	Artefact Feedback	
	P1	P2
Accessibility	9	10
Accuracy	9	9
Completeness	9	10
Consistency	10	10
Ease of use	9	9
Effectiveness	10	10
Efficiency	10	10
Elegance	8	9
Fidelity with real world phenomena	9	10
Generality	10	9

Criteria	Artefact Feedback	
	P1	P2
Impact on the environment and on the artefact's users	10	9
Level of detail	10	10
Operationality	10	9
Relevance	9	10
Robustness	9	10
Simplicity	10	10
Understandability	9	10

The next question was used to obtain more feedback on the operation ability and ease of use of the artefact.

Interviewer: Are you able to complete tasks or work using the solution? What functionalities are you able to perform with the solution?

Participant 1: All I can say is that the requirements we asked for it's exactly what we are getting.

Participant 2: Well, we are already capturing and retrieving data. And we are already pulling test certificates.

The participants indicated that they were able to capture and retrieve data in the database. Furthermore, Participant 1 indicated that they have SQL skills in-house to assist with data-related issues and requests. Therefore, they will not struggle with managing the data stored in the database.

The next question was used to obtain feedback on user expression and thoughts about the functionality of the artefact.

Functionality of the artefact

In this section the researcher interviewed the participants about their overall impression regarding the functionality of the artefact. Both participants were involved in defining the requirements of the artefact and were also involved in various evaluation stages of the artefact.

Interviewer: Based on the requirements given to the developer, does the solution meet your expectations?

Participant 1: Yes, It is simple and quicker to navigate through various links in the menu, to capture and update the data. Additionally, multiple users are able to connect to the new data management solution simulteneously and perform various functionalities such as retrieving, capturing and updating data in the database. So, currently all users are able to connect and work together at the same time without any issues.

Participant 2: Yes, the solution is easier to use as compared to MS Access and very efficient when capturing, retrieving and updating data. The data are consistent.

In the following section, the interviewer asked participants questions about the impact of the artefact on the business and also their experience and thoughts about the solution.

User impressions and artefact impact on the business

The participants indicated that they find the artefact very useful and commented on the value added to the organisation.

Interviewer: So, are you confident that the solution adds value to the organisation?

Participant 1: Yes, definitely. It was difficult to use MS access to capture data because it was time consuming which eventually affected productivity. However, with the newly implemented solution it's a matter of click and go. Data capturing and retrieval has improved, it is very efficient. Data are consistent, data formats are defined accordingly and date are automated. It is easier to manage data. What business is looking for is clean data so that they can evaluate their processes, check material specifications being used and monitor roll dispatching processes. So far, turnaround time has improved, data are clean and consistent. So, it is adding value.

Participants indicated that since they have started using the artefact, even though it is not being used to full capacity, it is already adding value to the organisation. Based on the feedback, participants indicated that the data are more consistent, data formats are accurate, turnaround time has improved, data are clean, and users are able to capture and retrieve data efficiently.

Interviewer: Next question, do you think this solution will solve the problems that the company is experiencing with regards to poor data quality?

Participant 1: Yes, based on the data that we have captured so far, the data are consistent, data are clean, the solution is efficient, data formats are accurate and no duplicate data are stored in the database. If you capture incorrect information, validations and error messages are displayed and informing the user were the issue is.You cannot proceed and capture the data until you fix the issue and that on its own is cleaning the data before it is entered into the database. Already, this is saving us a lot of time and everything is in sync.

Participant 2: Yes, currently the data are more consistent and there are no duplicate data in the database. The solution validates the data before it is captured into the database.

Furthermore, participants indicated that the solution is solving the majority of the problems that the SME had. From the feedback, participant 1 emphasised that the error messages and validations embedded within the application were assisting with data cleaning at the data acquisition stage. Figure 2-6 in section 2.6.4 by Mahanti (2018:12) explained that it is advisable to address and solve any data quality issues while in the data acquisition before they creep through the data lifecycle stages. The researcher used the next question to access the thoughts and impressions of the participants.

Interviewer: Next question, what would you say was the overall impression or impact that the solution had on the business users using the new data management solution? Were there any implications to business operations?

Participant 1: The system is working wonderfully, it's pleasant to use. It gives a quicker turnaround time to decision making and also provides improved productivity in term of material specification and other functionalities. Also, data on test cerificate are consistent and accurate. The test certificates are coming out elegantly and data are retrieved efficiently from the database and populated accordingly where they are suppose to display. The changes that were communicated to the developer were developed and implemented promptly, it didn't take time. I'm quite impressed.

The participants found the solution wonderful and easy to use. In the next question, the researcher wanted to find out if there were any improvements that the users would have liked to see in the artefact.

Interviewer: What would you change about the new data capturing solution, if anything?

Participant 1: Honestly, the way the artefact is designed, it is very difficult to pinpoint where it needs improvements. It's still early to say, maybe we can improve it in such a way that we can be able to see historical information related to a specific roll on the interface.

Participant 2: I think we need to give it some time, use the artefact more then we can see if there are some information that needs to be added, changed or improved.

In addition to the interviews conducted, the researcher sent the questionnaire outlined in Appendix 3 to participant 1 and participant 2 to explore their thoughts and experiences of the newly implemented solution. The section below represents the statistical software for the social sciences (SPSS) results based on the questionnaire data collected from participant 1 and participant 2. Participant 1 and participant 2 were involved in the requirements gathering stage, development, evaluation, and deployment of the artefact. To gather information about their thoughts and expressions, the researcher asked the participants to rate the artefact, where a rating of 0 indicated that they are less satisfied with the solution and 10 indicated that they were highly satisfied.

User satisfaction on the functionality of the artefact interface.

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Satisfaction	2	1	8	9	17	8.50	.500	.707	.500
Valid N (listwise)	2								

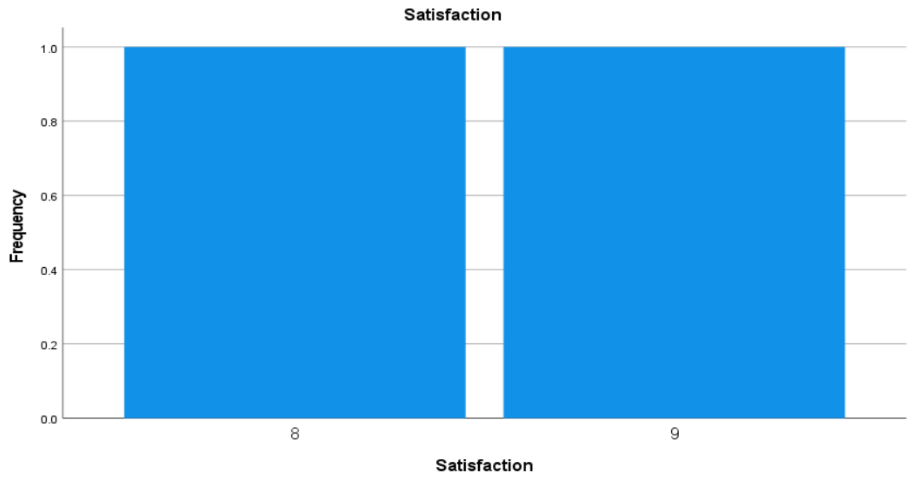


Figure 5-4: Satisfaction of the artefact

As outlined in Figure 5-4, the results based on user satisfaction on the functionalities of the artefact indicated that the participants were satisfied with the artefact functionalities. Figure 5-4 is based on user satisfaction; participant 1 rated the artefact and gave it a score of 9 out 10, whereas participant 2 rated it 8 out of 10.

From the feedback received from the SSPS results, the participants indicated that they are able to:

- efficiently capture and retrieve data using the artefact;
- the artefact is easy to navigate and use,
- multiple users are able to connect simultaneously and use the artefact;
- the data are consistent, accurate and there are no duplicate records; and
- they are able to complete data capturing tasks using the artefact.

Impression on the functionality of the artefact

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
The application is pleasant	2	8	9	17	8.50	.707	.500
Valid N (listwise)	2						

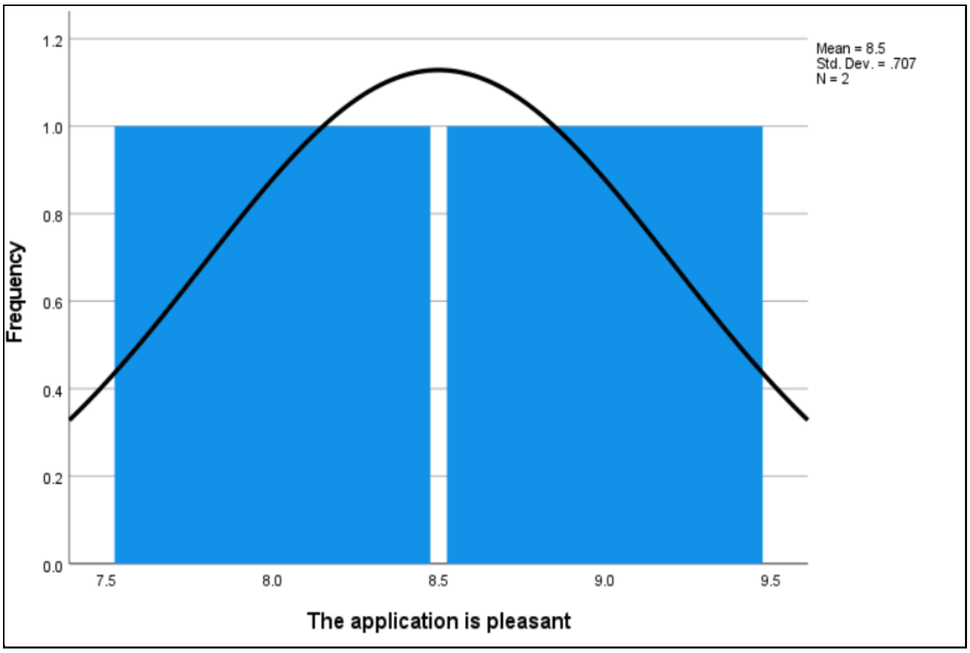


Figure 5-5: The application is pleasant.

The feedback received from participants was satisfactory. The results from the SSPS indicated that the participants were happy with the artefact and they find the artefact pleasant to use. As per Figure 5-5, participant 1 rated the artefact and gave it a score of 9 out 10, whereas participant 2 rated it 8 out of 10.

Results from the observation

The researcher observed participants' interaction with the artefact and the following were identified.

- The participants were happy about the layout and design of the artefact. They were able to view and read the information easily on the interface.
- The participants were able to connect to the artefact using the login credentials provided.
- As compared to the old data management solution, the artefact enabled users to connect simultaneously and capture data.
- The participants have a clear understanding about the organisation's business processes and the data flow processes, which made it easier for them to navigate and utilise various functionalities of the artefact without any issues.
- One of the main highlights was the efficient capturing and retrieval of data.
- The participants were happy with the layout of the reports.
- The participants were impressed with the artefact as a whole.

Section 5.6 will discuss the guidelines and technologies for improvement of data quality and management in one SA SME.

5.6 Guidelines and technologies for improvement of data quality in one South African SME (Section 4.2.2).

According to Mullins (2017:54) bad data quality can be avoided; however, many organisations still continue to ask "how". The literature explored by the researcher (Chapter 3), the DSR methodology applied in constructing the artefact (Chapter 4), and the results gathered from the participants in section 5.5 were used to compile guidelines for improving data quality in one SA SME. Table 5-6 outlines a list of guidelines that one SA SME can implement to improve data quality.

Guidelines extracted for one SA SME are highlighted in the table and the general description of the guideline is presented underneath that.

Table 5-6: Guidelines for improvement of data quality in one SA SME.

Number	Guidelines for improvement of data quality
Guideline 1	**Senior management in one SA SME must educate or train employees about the importance of data quality in the organisation especially its impact in decision making processes.** Employees must be trained on how to use data capturing systems, to handle and manage data.
Guideline 2	**Senior managers in one SA SME must develop a data-driven culture and mind set required to manage and preserve data quality.** Data-driven cultures require executive managers to view and acknowledge data as a strategic enterprise asset and resource rather than a tactical application by-product. By so doing, it will cascade down to the employees to adopt the same culture.
Guideline 3	**One SA SME senior management must create data quality awareness programs and ensure that the employees understand that data quality management is not a once off process but an ongoing activity.** It is senior management's responsibility to create enterprise-wide data quality awareness programs and data quality training.
Guideline 4	**It is important for executive managers in one SA SME to acknowledge the existing data quality issues within the organisation as early as possible and develop a plan to mitigate before they spread to other systems.** Data quality issues must be corrected as early as possible. There is no need to wait for a disaster to occur before actions are put in place to solve these issues.
Guideline 5	**One SA SME employees must know that it is every employee's responsibility to protect the integrity of the data.** Every employee needs to appreciate the value of data, understand the cost of poor data quality and view data as an organisational asset.
Guideline 6	**One SA SME must ensure that employees are comfortable with using the data capturing solution and understand how it works.** If employees do not find the data capturing easy to use, training must be provided. All information on the system must be visible and easy to read.
Guideline 7	**One SA SME need to ensure that their data capturing systems or interfaces are equipped with adequate validations and checks to prevent erroneous data from entering the system.** Human error was identified as the major contributor of poor data quality. Accurate procedures and measures must be implemented within the data capturing system.
Guideline 8	**Data capturing systems must use selection options such as checkboxes, dropdown lists and radio buttons as much as possible rather than free text capturing fields**. Free text capturing fields are major contributor to poor data quality.
Guideline 9	**One SA SME must understand what "*fitness for use*" means to customers and to management responsible for making key business decisions.** Dimensions such as completeness, uniqueness, timeliness, validity, accuracy and consistency can be used to evaluate the quality of the data.
Guideline 10	**The database or system administrator is required to perform data and system management duties.** A skilled individual in required to manage data security, perform data quality checks, monitor system functionality and provide data support when required.
Guideline 11	**The database administrator must perform data profiling as regularly as possible.** Data profiling involves examining data quality issues within existing data and determine corrective action to be performed.
Guideline 12	**The database or system administrator must develop constraints to protect the integrity of data.** Constraints such as referential integrity constraints to enforce primary key and foreign key relationships in databases, unique constraints to prohibit duplicate keys from being entered into the database, not null constraints to ensure completeness of data, check constraints and triggers can be used to enforce business rules onto data elements in the database.

Number	Guidelines for improvement of data quality
Guideline 13	**For effective data quality management, one SA SME must adopt a data stewardship approach.** A person or group of individuals must be appointed as data stewards to take care of a set of data within the organisation.
Guideline 14	**One SA SME must have standard methods of handling data.** Metadata are crucial to the business because they facilitate the understanding of data. So, to avoid data inconsistencies, different departments within one SA SME including IT must define and agree on common data definitions, common data standards, and standard metadata and have a standard data enterprise-wide dictionary.
Guideline 15	**One SA SME must develop a formal data governance program.** One SA SME must ensure that each department or business unit should not manage their data separately in silos as it will create data inconsistencies. So, one SA SME must develop a formal data governance program. Management must have understanding of how data governance is implemented in one SA SME.
Guideline 16	**Fragmented data systems, data silos and disparate departmental data stores must be avoided.** Disparate data stores lead to data discrepancies, data duplication and data inconsistencies.
Guideline 17	**One SA SME must determine data quality dimensions that the enterprise data must conform to and that are important to be fit for operational and analytical use.** Dimensions such as accuracy, validity, timeliness, consistency, uniqueness and completeness can be used to assess and measure data quality. Applying data quality dimensions may enhance the level of data quality and thus bring benefits to the organisation. In addition to the dimensions, the data must possess features such as; data must be usable and must come from a reputable source, must be accessible, interpretable, relevant, easy to read, etc.
Guideline 18	**Before data are used for reporting and analytics, they must be transformed and cleansed to ensure that they are of quality.** Data must go through the ETL process for transformation before they are loaded into a data warehouse or database that can be used for reporting and analytics.
Guideline 19	**Organisations must have a single trusted source for integrating high quality organisational data for reporting, querying and analytics**. The source must provide up-to-date, cleansed and reliable data required for decision making processes. There must be an automated process of extracting, transforming and loading data into the DW/BI solution.
Guideline 20	**One SME must have procedures to follow in managing and updating data that have reached their expiration date**. One SA SME can choose to use a date field to keep track of active records, expire the field by adding an indicator or a comment field to indicate which fields are inactive or updated.

Data quality tools play an important role in assisting SMEs to capture, store, manage, and analyse data. Furthermore, Mahanti (2018:412) indicates that data quality technologies should be selected based on what suits the business objectives best instead of only the IT department or based on their affordability. Furthermore, Van der Krogt *et al.* (2020:7) state that even though SMEs lack financial resources, lack qualified human resources, have limited knowledge of users, they shouldn't choose the BI technology based on how cheap the technology is, but it should be based on its functionality to support business and efficiency. It must be compatible with other systems, user friendly, and provide good solutions that will influence innovation and efficient decision making. Technologies that one SA SME can adopt to improve data quality and analytics are discussed next.

Table 5-7: Technologies for improvement of data quality.

Number	Technologies for improvement of data quality
Technology 1:	Data profiling tools can be used to evaluate the current state of organisational data by applying the appropriate dimensions to monitor data quality.
Technology 2:	SQL management studio can be used to develop, configure and manage SQL Server relational databases for storing day-to-day transactional data.
Technology 3:	SQL Server Data Tools proved to have the capability to handle data extraction, transformation and loading processes using SSIS. Additionally, SSRS can be used as a reporting tool for developing ad hoc reports.
Technology 4:	A SQL data warehouse designed using the Kimball methodology, the star schema can be used by SMEs to integrate data from multiple sources to provide clean and trustworthy data from reporting, analytics and reporting.
Technology 5:	Power BI proved to be a powerful tool for data visualisation. Organisations can adopt this tool to perform data analytics and to create dashboards.

Adopting data management and business intelligence tools should be perceived as an investment that will benefit the SMEs throughout if used accordingly. These tools can enable one SA SME to improve data quality and develop a data-driven business model that will enable the enterprise to gain a competitive advantage, increase revenue, and improve decision-making processes. Section 5.7 will give the conclusion of this chapter.

5.7 Conclusion

This section focused on the demonstration and evaluation of the artefact where various test cases where used to evaluate the functionality of the artefact. The IT administrator and the data capturer were involved in evaluating the artefact. The evaluation phase of the main research cycle followed the evaluation criteria presented in Table 3-8 in section 3.9. Evaluation attributes such as functionality, usability, completeness, consistency, accuracy, performance, reliability and fit for organisation were used to evaluate the artefact. Additionally, the participants evaluated the artefact accessibility, feasibility, look and feel, data formats, simplicity and elegance, ease of use, visibility, readability, efficiency and understandability. This involved an iterative process of testing and fixing the issues. The feedback received was satisfactory and minor issues were reported which the researcher fixed.

During post implementation, mixed method data collection technique was used to gather data from the participants. Semi-structured interviews were conducted with the IT administrator and data capturer to access their perceptions and thoughts about the functionality of the artefact, a questionnaire was used to collect data about the artefact and observation technique was used by the researcher to observe the users interaction with the artefact. Data analysis tools such as ATLAS.ti and statistical software for the social sciences (SPSS) were used to analyse the data. Codes and themes were applied to transcribed data.

The evaluation of the artefact was examined, and subsequently the results of the evaluation of the artefact were presented in section 5.5. Some of the feedback received from the participants indicated that they find the artefact pleasant to use, it was efficient and easy to use, it adds value to the organisation and users were able to connect to the artefact and capture and retrieve data. As compared to the old data capturing solution, the new artefact enabled multiple users to connect and capture data simultaneously. Finally, guidelines and technologies that one SA SME can implement to improve data quality were compiled and presented.

CHAPTER 6: RESEARCH CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

The research study was based on one SA SME which specialises in engineering casting and manufacturing of steel mill rolls and rings. The SME is important to the economic growth, poverty alleviation and job creation in SA. The organisation was using MS Access as their data management application. However, MS Access presented the SME with numerous limitations and challenges. The SME realised that they are in need of a technological solution that can enable multiple users to capture and retrieve data simultaneously, a solution that can improve data quality, enable users to efficiently capture, store and manage data, increase productivity, improve decision making processes, and data analytics.

The aim of this study was to develop an artefact to assist one SA SME to better manage data and improve data quality, and to propose guidelines and technologies that one SA SME could use to improve data quality. In order to achieve this, the study followed a design science research (DSR) methodology as discussed in Chapter 3. This chapter represents the final activity in the main research cycle, namely the conclusion phase.

In the conclusion phase of the design science research (DSR) model, the researcher discusses the outcomes of the research study, and reflects on which methods worked and did not work when solving the problem. The final phase of the design science research (DSR) model enables the researcher to communicate the results of the research study and contribute to the large knowledge base, enabling the researcher to reach significant conclusions based on the knowledge acquired from the research study.

The objective of this chapter is therefore to conclude and communicate the research. The chapter summarises the research problem in section 1.3, answers the research question in section 1.4.4, and addresses the research objectives of the study.

6.2 Research objectives addressed

In this section, the researcher will explain how this study addressed the theoretical and practical objectives defined in Chapter 1 of the study. The main objective of the research study was:

> To improve data quality and management in one South African SME for data analytics.

In order to reach the main objectives of the study, the following research questions were addressed. The main research question of this study is:

> How can SMEs exploit technology to better manage data and improve the quality of their data for data analytics?

The study aims to answer the following supporting research sub-questions indicated by SQ1 to SQ3:

- ❖ SQ1: Which applications are SMEs using to store and manage their data?
- ❖ SQ2: What are the impacts of poor data quality in SMEs?
- ❖ SQ3: What are the challenges of poor data quality and benefits of high data quality in SMEs?

The design requirements were formulated based on the information gathered from the participants and through the research questions. The design science research (DSR) model by Vaishnavi *et al.* (2004/2019:14) was followed to address the objectives of the study and to answer the research questions.

6.2.1 Theoretical objective

The sections that follow discuss the key findings in literature concerning the theoretical objectives in order to improve data quality and management in one South African SME for data analytics as the main objective of the study.

Theoretical objective 1: To explore technologies that SMEs can implement to better manage and store their data.

The researcher explored literature in Chapter 2 to discover what literature say about applications and technologies that SMEs can implement to improve data capturing, storage and manage data as outlined in section 2.3 and section 2.5. The researcher also explained technologies and applications that can be used to improve data quality, data storage and management.

Microsoft technologies proved to have capabilities to handle data quality and management issues. In this research study, Microsoft visual studio 2019 (C#) was used to develop the user interface, Microsoft SQL Server 2019 was used to develop a relational database and a data warehouse to store and manage data. Microsoft SQL Server Data tools (SSIS) was used to develop the ETL solution and SSRS to develop the reporting solution. Microsoft Power BI was used for analysis and visualisation of the data.

However, developing interface in-house for capturing data using Microsoft visual studio requires the interface to be designed in such a way that it is user-friendly, easy to use, with the information on the interface visible and readable. Furthermore, the interface must be equipped with data validations and checks to prohibit human error, ensure that data are free from defects, that data are validated in the data acquisition stage, and that incorrect data are prohibited from entering the system. Automated processes of capturing data must be used rather than using free text capturing fields, as human error was discovered to be the major contributor of poor data quality. The database system must be embedded with constraints and validations to filter out unwanted data and to ensure that data are accurate, consistent, complete, valid, current and not duplicated. Security must be implemented to authenticate authorised users.

Automated processes of extracting, transforming and loading data into a data warehouse proved to provide cleansed data stored in a trusted source. The solution provided a single version of the truth about the organisational data needed for reporting and analytics. One SA SME have access to clean updated data stored in a trusted source.

Theoretical objective 2: To investigate the impacts of poor data quality on SMEs.

Chapter 2 explored literature to determine the impact of poor data quality on organisations. The literature indicated that poor data quality can affect SMEs in various aspects. From the literature, section 2.6.4 indicated that poor data quality can affect an organisation’s productivity, financial state, as well as affect risk, compliance and an organisations’ confidence.

Poor data quality in one SA SME led to poor decision-making, mistrust in organisational data, poor performance, and put the SME at a competitive disadvantage.

Theoretical objective 3: To identify the challenges of poor data quality and benefits of high data quality in SMEs.

Section 2.6 provided an overview of the benefits of high data quality. From the literature, it is evident that high data quality enables organisations to increase revenue, provides excellent customer service, produces better analytics and reports, enables efficient decision-making processes, discovers business patterns, predicts future trends and competes with rivalries. High data quality enabled one SA SME to increase productivity and performance, improve decision making, increase insight and business opportunities, produce better analytics and reporting, increased confidence in the data and decision-making, and increase employee morale.

Furthermore, literature was explored in section 2.6.4 to determine the challenges of poor data quality. Poor data quality increases operational costs, lowers employee morale, and leads to poor analytics, reporting and poor decision making. In one SA SME poor data quality caused mistrust in organisational data, reporting and decision-making processes. Additionally, poor data quality contributed to poor productivity, burdensome business processes, and poor employee morale, hindered significant initiatives, and put the SME at a competitive disadvantage.

Theoretical objective 4: To explore different BI models, data warehousing and ETL concepts suitable for SMEs.

The researcher explored literature in section 2.9, 2.9.1, 2.9.2 on BI technologies, data warehousing and ETL concepts. In this study, Microsoft SQL Server 2019 and Microsoft data tools 2019 were used.

From the literature the researcher discovered that Microsoft business intelligence (BI) suites, open-source cloud based business intelligence (BI) tools are cheap and can be used by SMEs to improve data management while enhancing data quality for reporting and data analytics. Data profiling can also be used by one SA SME to frequently monitor the quality of the data in the organisation.

The data warehouse developed in this study has enabled the SME to integrate data from multiple sources into one trusted source. The ETL (extract, transform and load) solution developed using SQL server integration services (SSIS) was used to transform the data to improve quality before data were stored in the data warehouse. One SA SME has more timely access to information and reports.

The solution has improved the decision-making process, provided efficient access to accurate information and reports, improved employee satisfaction, reduced time to collect data, and reduced defects.

6.2.2 Practical objectives

The sections that follow discuss the key findings concerning the practical objectives in order to improve data quality and management in one South African SME for data analytics as the main objective of this study.

Practical objective 1: Design an artefact that one SA SME could use to capture, store and manage data for analytics.

The researcher followed the design science research paradigm in developing an innovative artefact to be used by the SME. The design science research model by Vaishnavi *et al.* (2004/2019:14) as discussed in section 3.8.3 was used in this study. The process of developing the artefact was discussed in Chapter 4.

The IT administrator (P1) and data capturer (P2) were involved in the development and testing of the artefact where various test cases were used to evaluate the functionality of the artefact. The evaluation criteria by Sonnenberg and Vom Brocke (2012:5) in section 3.9 was used to evaluate the artefact.

To obtain feedback from participants about the artefact, semi-structured interviews, observation and questionnaires were applied in this study as data collection methods. The researcher presented the results received from participants in section 5.5. The results from the data analysed using ATLAS.ti presented in section 5.5 indicated that P1 and P2 are happy with the functionality of the artefact and that the artefact adds value to the organisation, because they are able to efficiently capture, store, retrieve and manage data. Additionally, results presented in section 5.5 from the questionnaire analysed using SPSS indicated that they find the artefact pleasant to use and that they are satisfied with the overall functionality of the artefact.

Practical objective 2: to propose guidelines and technologies for improvement of data quality in one SA SME.

Subsequently, the literature explored in Chapter 3, the DSR methodology applied in constructing the artefact (Chapter 4) and the results gathered from the participants in section 5.5, were used to compile guidelines and technologies that could be suitable for improving data quality and management.

Table 6-1: Proposed guidelines for improvement of data quality in organisations.

Guidelines for improvement of data quality
Senior management in one SA SME must educate or train employees about the importance of data quality in the organisation especially its impact in decision making processes.
Senior managers in one SA SME must develop a data-driven culture and mind set required to manage and preserve data quality.
Senior managers in one SA SME must develop a data-driven culture and mind set required to manage and preserve data quality.
It is senior management's responsibility to create enterprise-wide data quality awareness programs and data quality training.
It is important for executive managers in one SA SME to acknowledge the existing data quality issues within the organisation as early as possible and develop a plan to mitigate before they spread to other systems.
One SA SME employees must know that it is every employee's responsibility to protect the integrity of the data.
One SA SME must ensure that employees are comfortable with using the data capturing solution and understand how it works.
One SA SME need to ensure that their data capturing systems or interfaces are equipped with adequate validations and checks to prevent erroneous data from entering the system.
Data capturing systems must use selection options such as checkboxes, dropdown lists and radio buttons as much as possible rather than free text capturing fields.
One SA SME must understand what "fitness for use" means to customers and to management responsible for making key business decisions.

Guidelines for improvement of data quality
The database or system administrator is required to perform data and system management duties.
The database administrator must perform data profiling as regularly as possible.
The database or system administrator must develop constraints to protect the integrity of data.
For effective data quality management, one SA SME must adopt a data stewardship approach.
One SA SME must have standard methods of handling data.
One SA SME must develop a formal data governance program.
Fragmented data systems, data silos and disparate departmental data stores must be avoided.
One SA SME must determine data quality dimensions that the enterprise data must conform to and that are important to be fit for operational and analytical use.
Before data are used for reporting and analytics, they must be transformed and cleansed to ensure that they are of quality.
Organisations must have a single trusted source for integrating high quality organisational data for reporting, querying and analytics.

Based on the feedback received, Microsoft technologies such as Microsoft SQL Server 2019 and SQL Server Data Tools 2019 proved to have capabilities of improving data quality and management in one SA SME. Microsoft visual studio 2019 enabled the researcher to develop a high quality artefact for capturing and retrieving data. Organisations can develop SQL relational databases to store their transactional data and develop data warehouses to integrate and store data from various sources for reporting, querying and analytics. These tools can be adopted by any organisations looking to develop a data management solution in-house.

Data profiling tools can be used by SMEs to analyse and assess the state and quality of the data.

For SMEs looking to leverage big data, they can use cloud-based open-source big data tools to manage and improve data quality. Cloud-based open-source tools can be accessed easily by any organisation looking to exploit big data.

Apache Hadoop is an open-source technology that can be implemented to assist with storing and analysing unstructured data.

ETL tools can be used to integrate, cleanse and transform data to improve data quality. Software vendors such as IBM, SAS, Oracle, Talend, Trillium software, etc. have cloud-based data management tools that can be adopted by SMEs to improve data quality and management.

Furthermore, organisations can follow the five phases of the Six Sigma DMAIC approach to improve data quality.

6.3 Theoretical and practical contributions

The theoretical contributions in this research study include:

- ❖ The guidelines and technologies compiled and proposed by the researcher for improving data quality and data management contribute to the knowledge base, as they can be implemented by other organisations seeking ways to improve data quality and management issues.
- ❖ The artefact development and DSR model followed contribute to the knowledge base and other SMEs in general.
- ❖ The case study contributes to the knowledge base of the significance of SMEs in economic development and poverty alleviation, and how data can be used in SMEs to gain competitive advantage, to innovate and to respond positively to changing markets and globalization.
- ❖ DSR study contributes to the knowledge base of data quality and data management concepts, reporting, data analytics and decision-making in SMEs.

The practical contributions of the study include:

- ❖ The study proposed guidelines and technologies that organisations can adopt and implement to assist with better management of data and to improve data quality (section 5.6).
- ❖ The study produced an artefact to assist the SME with capturing, storing and managing data while improving data quality (section 4.3).

6.4 Conclusion phase – Main research cycle

The conclusion phase marks the completion of the design science research project. The three sub-cycles that the researcher followed enabled the researcher to gain critical information required for solving the research problem. Through these cycles, the researcher reviewed the literature to acquire knowledge about the data quality, data storage, management, challenges of poor data quality, impacts of poor data quality and benefits of high-quality data. The interactions that the researcher had with the participants allowed the researcher to gain in-depth knowledge about the phenomenon and its challenges.

The participant's involvement throughout the research project contributed immensely in ensuring that an artefact was developed to address one SA SME's challenges. The participants' knowledge and experience provided guidance in enabling the researcher to understand the phenomenon in-depth, including its challenges, needs and expectations. The feedback received from the participants throughout the evaluation, development and suggestion phases added great value in ensuring that innovative feedback is constructed.

It is significant for any researcher conducting a case study to select participants that have sufficient knowledge and understanding of the phenomenon and its issues. This enabled the researcher to obtain critical and trustworthy information relevant to the research study and ensured that the DSR is completed on time with all objectives attained.

The feedback received in the evaluation phase by the participants indicated that the participants are able to access and use the artefact to capture, store and manage data. From the researcher's observation, the participants find the artefact easy to use. The participants' involvement in the study from the requirements gathering stage to development, their experience and knowledge about the business processes, and the data made it easy for them to find the artefact simple to use. Additionally, the researcher designed the artefact in such a way that it was understandable, friendly to use, information is visible, readable and easy to find, and the navigation links are easily accessible.

6.5 Rigour of the research

The DSR process model applied in this study and DSR guidelines followed enabled the researcher to develop a high-quality artefact. The methods applied in this study were selected to suit the research study, were justified from the literature and, therefore, able to withstand critical scrutiny. A mixed method approach where qualitative data collection techniques such as interviews, observation and a questionnaire were used proved to be valid and provided the researcher with the information needed for the success of this case study. Strategies of applying rigour as described in section 3.12 were applied in this study.

Informed consent forms were signed by participants within the SME who were involved in the research study to indicate that they were not coerced to participate. Ethical clearance was obtained from the North-West University to conduct the research study, and one SA SME also gave the researcher permission to conduct the study. Participants' names and the organisation will not be disclosed.

Qualitative data analysis involved content analysis as suggested by Zhang and Wildemuth (2009:3) using ATLAS.ti in order to analyse transcribed audio recordings and discover important information needed to make recommendations and make conclusions. Member checks were implemented to assure accuracy of interview transcriptions (Table 3-10).

Table 3-10 in section 3.12 discussed the strategies to enhance rigour in this study, as suggested by Neergaard *et al.* (2009:4), and are summarised here:

- ❖ Prior to conducting the semi-structured interview, the researcher ensured that the participants understood what the interview session was about, how the data were to be collected, and handled and ensured that consent forms were signed to indicate that they understood what the study was about and that they were not coerced to participate.
- ❖ Participants' details and company details were not disclosed in this study. The company was referred to as "one SA SME", the IT administrator "Participant 1 (P1)" and the data capturer "Participant 2 (P2)". Anonymity was maintained throughout the study.
- ❖ Participants in this study were recommended by the IT manager and were selected based on their level of knowledge about the phenomena and their in-depth knowledge and understanding about the issues within the SME. The manager also indicated that the participants are trustworthy employees with integrity and high principle, and will provide the researcher will correct information.
- ❖ The researcher ensured that no participants were subjected to any form of discrimination during artefact development, evaluation and data collection sessions. All participants were treated with respect throughout the study.
- ❖ Participants were able to communicate openly and freely about their thoughts, perceptions and experiences throughout the research study. All issues that were addressed during the artefact evaluation phase were attended to and addressed accordingly.
- ❖ The participants evaluated the artefact for efficacy, utility, quality, etc. The evaluation criteria for DSR artefacts in Table 3-8, section 3.9 and the framework for evaluation in design science (FEDS) as outlined in Figure 3-10 was followed in this study.
- ❖ The audio recorded interviews were transcribed and reviewed for accuracy and authenticity. Codes and themes were applied to the transcribed data where ATLAS.ti was used to analyse the data. The theme and codes were verified and corrections were made where necessary. The results were evaluated for accuracy and integrity.
- ❖ The researcher ensured that the integrity and authenticity of the data collected from the participants were not lost.

The case study involved an iterative process of artefact development, evaluation and communication of the outcomes to ensure that a high-quality artefact is delivered. Various test cases were used by the participants to test the data captured using the artefact for accuracy, uniqueness, consistency, completeness, validity, reliability and availability. Furthermore, the researcher observed the participants' interaction with the artefact, taking into consideration their thoughts, experiences and perceptions.

As discussed in section 3.9, the researcher ensured that the artefact is evaluated according to the evaluation criteria by Gregor and Hevner (2013:350) and Sonnenberg and Vom Brocke (2012:5), because evaluating the artefact is regarded as a vital activity as it assures rigour.

The artefact's efficacy and efficiency were evaluated according to Venable *et al.* (2017:82) list for rigour:

- Efficacy – Participants have indicated that the artefact added value to the organisation and it has improved the data capturing, retrieval and management processes (section 5.5).
- Efficiency – Multiple users are able to connect to the artefact and capture and retrieve data efficiently. Users were able to connect at the same time and utilise the artefact without any issues. Participants have efficient access to reports and accurate data (section 5.5).
- Ethics – Ethical clearance from NWU was obtained to conduct the study and ensured that participants were protected from any harm and exploitation (section 3.10).

A checklist developed by Hevner and Chatterjee (2010:20) was used to assess this DSR project and to ensure that this project addressed the primary aspects of DSR. Table 6-2 below indicates where the items in the DSR checklist were addressed in this study. An elaboration on some of the checklist items is provided after the table.

Table 6-2: Checklist used to assess the DSR study (section 1.5.6)

Questions	**Addressed in this study**
1. What is the research question (design requirements)?	Section 1.4, 6.2
2. What is the artefact? How is the artefact represented?	Section 4.2, 4.3
3. What design processes (search heuristic) will be used to build the artefact?	Section 3.11, 4.3
4. How are the artefact and the design processes grounded by the knowledge base? What, if any, theories support the artefact and the design process?	Section 3.8.3, 3.11, 4.3
5. What evaluations are performed during the internal design cycles? What design improvements are identified during each design cycle?	Section 4.3, 5.3
6. How is the artefact introduced into the application environment and how is it field tested? What metrics are used to demonstrate artefact utility and improvement over previous artefacts?	Section 5.5, 5.6
7. What new knowledge is added to the knowledge base and in what form (e.g., peer reviewed literature), meta-artefacts, new theory and new methods)?	Section 5.6, 6.3
8. Has the research question been satisfactory addressed?	Section 6.2

The design science research (DSR) model by Vaishnavi *et al.* (2004/2019:14) was followed and involved various phases such as awareness, suggestion, development, evaluation and conclusion in developing the artefact, as discussed in section 3.8.3. This study consisted of a main research cycle and three sub-cycles within the suggestion phase in order to achieve the objective of the study.

The design of the artefact involved a constant process of developing, evaluating and communicating the outcome of the solution. To ensure that the artefact is developed and evaluated efficiently, the researcher split the artefact design into five manageable components. The components were developed and tested individually by the researcher and participants to ensure functionality, consistency, accuracy, etc. (section 4.3). A workshop session was held with the participants to evaluate the functionality of the artefact. Various metrics by Hevner *et al.* (2004:85) and attributes presented by Sonnenberg and Vom Brocke (2012:5) in section 3.9 were used to evaluate the artefact (Table 3-8). All changes were communicated to the researcher to solve as discussed in section 4.3. Once all the components of the artefact had been tested, the artefact was deployed in one SA SME environment.

The knowledge contribution of this study is provided in section 6.3 and a reflection on addressing the research question is provided in section 6.2.

6.6 Limitations of the study

A first limitation to the case study was that the study was conducted in one organisation only. It is possible that outcomes would have been different in another context. The second limitation is that the sample size was small because the SME has less than 120 employees with a relatively small IT department. The artefact was tested with only the IT administrator and the data capturer because they were recommended by the IT manager as the most knowledgeable individuals to provide the researcher with accurate information about the data issues and challenges and the needs of the SME. The IT administrator has immense knowledge and understanding with regards to the organisation’s business processes, data management processes, IT operations and data quality. A third limitation occurred when the COVID-19 pandemic presented the researcher with several challenges, as it was difficult to conduct multiple face-to-face artefact testing workshop sessions with participants. Some of the workshops for evaluating the artefact were conducted using online platforms. The face-to-face workshop session would have enabled the researcher to gain proper observation of the users’ interaction with the artefact and feedback from the users.

6.7 Recommendations for future research

Future studies are possible with a proper investigation and/or comparison on how multiple SMEs are using technology to better manage data and improve data quality. Further studies can be conducted on applications and tools that SMEs can use to better manage data and improve data quality. Additionally, researchers can also explore the benefits of business intelligence and data analytics in SMEs. Due to the volume of data and cloud-based technologies, further research can be conducted to investigate the cloud-based technologies that SMEs could implement to better manage big data and improve quality.

6.8 Reflection and conclusion

The practical objective of this case study was to design an artefact that one SA SME could use to capture, store and manage data for analytics, and subsequently propose guidelines and technologies for improvement of data quality in order to improve data management and data quality in one SA SME. Through the design science research (DSR) paradigm and case study research methodology, the researcher addressed the theoretical and practical objectives of the study. The researcher explored existing literature to gain a thorough understanding of the research concepts and research methodologies. The design science research (DSR) model selected for this study used three sub-cycles and the main research cycle where five phases (awareness, suggestion, development, evaluation and conclusion) enabled the researcher to develop and produce a high quality artefact. The iterative process of developing, testing and communicating the changes ensured that the functional artefact meeting the design requirements was delivered.

The highlights of this case study were that design science research (DSR) was used in this research study to guide the researcher in developing an innovative artefact as a solution to a socio-technical problem. Based on the feedback received from participants in Chapter 5, it is evident that the solution is adding value to the organisation and solving the SME's data quality and management issues. From the feedback, participants indicated that data are more consistent and accurate and data capturing turnaround time has improved. The current reporting and business intelligent solutions have enabled the organisation to have access to accurate and organised data stored in a central location for reporting, analysis and decision-making purposes.

Knowledge contributions were achieved through utilising the design science research (DSR) methodology and validating the technologies and processes that were used to develop the artefact proved to have worked accordingly. Furthermore, contributions were made to the research body of knowledge by proposing guidelines and technologies that can be used by organisations to better manage data and improve data quality. Design science research (DSR) has proved to be a significant methodology for developing an artefact to address socio-technical problems. The proposed guidelines and technologies can be implemented and used by other SMEs seeking ways to better manage data and improve data quality. Based on the results outlined in section 5.5 and the problem statement presented in Chapter 1 (section 1.3), the artefact is adding value to one SA SME.

BIBLIOGRAPHY

Ahsan, S. & Shah, A. 2006. Data, Information, Knowledge, Wisdom: A Doubly Linked Chain? . (*In*. In the proceedings of the 2006 international conference on information knowledge engineering organised by. p. 270-278).

Ali, M.D.S., Khan, S. & Miah, S.J. 2017. Analysis of interaction between business intelligence and SME: Learn from each other. *Journal of Information Systems and Technology Management*, 14(2):151-168.

Ali, T.Z., Abdelaziz, T.M., Maatuk, A.M. & Elakeili, S.M. 2020. A Framework for Improving Data Quality in Data Warehouse: A Case Study. Paper presented at the 2020 21st Internatiotional Arab conference on Information Technology.

Ali, Z. & Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. *Indian J Anaesth*, 60(9):662-669.

Aljumaili, M. 2016. Data Quality Assessment: Applied in Maintenance. Luleå, Sweden: Luleå University of Technology. (Doctoral Thesis).

Assarlind, M. 2011. Antecedents for Quality Management in Small and Medium Enterprises. Sweden: Chalmers University of Technology.

Baharuden, A.F., Isaac, O. & Ameen, A. 2019. Factors Influencing Big Data & Analytics (BD&A) Learning Intentions with Transformational Leadership as Moderator Variable: Malaysian SME Perspective. *International Journal of Management and Human Science (IJMHS)*, 3(1):10-20.

Ballou, D.P. & Pazer, H.L. 1999. The impact of data quality information on decision making: An exploratory analysis. *IEEE Transactions on knowledge and data engineering*, 11(6):853-864.

Ballou, D.P. & Tayi, G.K. 1999. Enhancing data quality in data warehouse environments. *Communications of the ACM*, 42(1):73-78.

Batini, C., Cappiello, C., Francalanci, C. & Maurino, A. 2009. Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3):1-52.

Batini, C. & Pernici, B. 2006. Data Quality Management and Evolution of Information Systems. *n IFIP International Federation for Information Processing*, 214:51-62.

Bellinger, G., Castro, D. & Mills, A. 2004. Data, Information, Knowledge, and Wisdom. https://www.systems-thinking.org/dikw/dikw.htm Date of access: 23 Feb 2021.

Bernardino, J. 2015. Emerging Business Intelligence Technologies for SMEs. Open Source Technology. p. 1694-1721).

Bernus, P. & Noran, O. 2017. Data Rich – But Information Poor. Collaboration in a Data-Rich World. p. 206-214).

Bertossi, L., Rizzolo, F. & Jiang, L. 2011. Data Quality Is Context Dependent. Enabling Real-Time Business Intelligence. p. 52-67).

Bianchini, M. & Michalkova, V. 2019. Data Analytics in SMEs:Trends and Policies. (26164353).

Bryman, A. & Bell, E. 2011. Business Research Methods. 3rd Ed. US: Oxford university press.

Bulao, J. 2020. How Much Data Is Created Every Day in 2020? https://techjury.net/blog/how-much-data-is-created-every-day/ Date of access: 11 October 2020.

Bustos, E.S. & Vicuña, S.D. 2016. Decision making and adaptation processes to climate change. *Ambiente & Sociedade*, 19(4):215-234.

Cleven, A., Gubler, P. & Hüner, K.M. 2009. Design alternatives for the evaluation of design science research artifacts. Paper presented at the Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09.

Coghlan, D. & Brannick, T. 2019. Doing action research in your organisation. 4th Ed. Los Angeles: Sage.

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X. & Reis, M.S. 2016. How Can SMEs Benefit from Big Data? Challenges and a Path Forward. *Quality and Reliability Engineering International*, 32(6):2151-2164.

Cook, L.D. & Kamalodeen, V.J. 2019. Mixed Methods Case Study Research: (Mmira), M.M.I.R.A.

Creswell, J.W. 2013. Qualitative Inquiry and Research Design : Choosing among five approaches. 3rd: SAGE.

Creswell, J.W. & Plano Clark, V.L. 2018. Designing and Conducting Mixed Methods Research. 3rd Ed: Sage.

Curraj, E. 2018. Business digitilization of SMEs in Albania: Innovative Approaches and their impact on performance. European University of Tirana.

Cypress, B.S. 2017. Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. *Dimens Crit Care Nurs*, 36(4):253-263.

DAMA-UK. 2013. The Six primary dimensions for data quality assessment - Defining data quality dimentions. UK: The premiere organisation for data professionals in the UK.

Davenport, T.H. & Prusak, L. 1998. Working Knowledge: How Organizations Manage What They Know. *Harvard Business Press*.

Dawson, C. 2009. Introduction to Research Methods: A practical guide for anyone undertaking a research project. 4th Ed. UK: How to content.

De Silva, F. 2005. Data Warehousing & Business Intelligence ROI. (*In*. Proceedings of the International Conference on Computer and Industrial Management, ICIM, organised by Bangkok, Thailand.

Del Vecchio, P., Minin, A.D., Petruzzelli, A.M., Panniello, U. & Pirri, S. 2018. Big Data for Open Innovation in SMEs and Large Corporations: Trends, Opportunities, and Challenges. *Creativity & Innovation Management*, 27:6-22.

Denchev, E. 2017. The development of a model for athe consolidation of information in small and medium-seized enterprises (SME) clusters. 2.

Denscombe, M. 2010. A good research guide: For small-scale social research projects: McGraw Hill.

Department of small business development. 2019. Revised Shedule 1 of the National Definition of Small Enterprise in South Africa: Gazette, G.

Dominguez, A., Litchmore, K. & Willshire, M.J. 2019. Assessment of acceptance factors impacting adoptionand use of business intelligence and analytics systems among small and medium-sized US manufacturing organisations. ProQuest LLC: Capella University.

Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q. & Lin, Z. 2017. Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. Paper presented at the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA).

Eckerson, W. 2002. Data quality and the bottom line. United States: Llc, C.

Ehrlinger, L., Rusz, E. & Wöß, W. 2019. A survey of data quality measurement and monitoring tools. *ArXiv*, abs/1907.08138.

Experian. 2018. The 2018 global data management benchmark report.

Forbes Insight. 2017. The data differentiator: How improving data quality improves business: Bowes, P.

Gadenne, D. & Sharma, B. 2005. Quality Management Factors in SMEs: An Investigation of Their Effect on Organisational Performance.

Geiger, J.G. 2004. Data Quality Management:The Most Critical Initiative You Can Implement. *Data Warehousing, Management and Quality*.

Gesmann-Nuissl, D. & Kirchner, G. 2018. Responsibility for data quality. *Journal of Contemporary Management*, 15:245-264.

Gherghina, Ș.C., Botezatu, M.A., Hosszu, A. & Simionescu, L.N. 2020. Small and Medium-Sized Enterprises (SMEs): The Engine of Economic Growth through Investments and Innovation. *Sustainability*, 12(1).

Ghobadian, A. & Gallear, D.N. 1996. Total quality management in SMEs. *International Journal of managem,ent science*, 24(1):83-106.

Grabova, O., Darmont, J., Chauchat, J. & Zolotaryova, I. 2010. Business Intelligence for Small and Middle-Sized Entreprises. *SIGMOD*, 39(2):39-50.

Gray, J. 1996. Evolution of Data Management. *Computer*, 29(10):38-46.

Gregor, S. & Hevner, A.R. 2013. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2):337-355.

Gregor, S., Kruse, L. & Seidel, S. 2020. Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, 21:1622-1652.

Grillenberger, A. & Romeike, R. 2017. Key Concepts of Data Management – an Empirical Approach. *In Proceedings ofKoli Calling 2017. ACM,*:10.

Guarda, T., Santos, M., Pinto, F., Augusto, M. & Silva, C. 2013. Business Intelligence as a Competitive Advantage for SMEs. *International Journal of Trade, Economics and Finance*:187-190.

Guba, E.G. & Lincoln, Y.S. 1994. Competing paradigms in qualitative research.: Sage. p. 105-117).

Guba, E.G. & Lincoln, Y.S. 2005. Paradigmatic, controversies, contraditions and emerging confluences.

Gudfinnsson, K. 2019. Towards facilitating BI adoption in small and medium sized manufacturing companies. Sweden: University of Skövde. (Doctoral).

Gudivada, V., Apon, A. & Ding, J. 2017. Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software 10.1*:1-20.

Günthera, L.C., Colangelo, E., Wiendahla, H. & Bauer, C. 2019. Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises. Paper presented at the 18th International Conference on Sheet Metal, SHEMET 2019, Germany, 28-30 June 2017.

Hafeez, A., Kumar, A. & Ahmad, S. 2019. Importance of data integrity & its regulation in pharmaceutical industry. *The Pharma Innovation Journal 2019*, 8(1):306-313.

Hart, C. 1998. Doing a literature review: Releasing the social science research imagination. Los Angeles: Sage.

Haug, A., Zachariassen, F. & Liempd, D.v. 2011. The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2).

Heinrich, B., Hristova, D., Klier, M., Schiller, A. & Szubartowicz, M. 2018a. Requirements of data quality metrics. *Journal of data and Information quality*, 9(2).

Heinrich, B., Klier, M., Schiller, A. & Wagner, G. 2018b. Assessing data quality – A probability-based metric for semantic consistency. *Decision Support Systems*, 110:95-106.

Hevner, A. & Chatterjee, S. 2010. Design Research in Information Systems : Theory and Practice. Vol. 22. NY: Springer.

Hevner, A.R. 2018. Effective Design Science Research.

Hevner, A.R., March, S.T., Park, P. & Ram, S. 2004. Design science in information systems research. *MIS Quarterly*, 28(1):75-105.

IBM. 2016. The Four V's of Big Data. https://www.ibmbigdatahub.com/infographic/four-vs-big-data Date of access: 06 September 2020.

IBM. 2020. Extracting business value from the 4 V's of big data. https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data Date of access: 06 September 2020.

Informatica. 2006. Boosting Data Quality for Business Success: Company, I.T.D.I.

Inmon, W.H. 2002. Bulding the Data Warehouse. Third Edition. Canada: John Wiley & Sons, Inc.

Ismail, M.F., Tiun, S. & Bakar, N.H. 2019. Requirements Gathering for Specialized Information Systems in Public Administration.

ISO/IEC. 2008. ISO/IEC 25012:2008: Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model. https://www.iso.org/obp/ui/#iso:std:iso-iec:25012:ed-1:v1:en Date of access: 2 Mar 2021.

Jaya, M.I., Sidi, F., Ishak, I., Affendey, L.S. & Jabar, M.A. 2017. A review of data quality research in achieving high data quality within organisation. *Journal of Theoretical and Applied Information Technology*, 95(12).

Jesiļevska, S. 2017. Data Quality Dimensions to Ensure Optimal Data Quality. *The Romanian Economic Journal*(63):89-103.

Juran, J.M. & Godfrey, A.B. 1999. Juran's quality handbook. 5th ed. United States of America: The McGraw-Hill.

Kademeteme, E., Kalema, B., M & Pretorius, P. 2017. Managing and improving data quality through the adoption of data warehouse in the public sector. *African Journal ofScience, Technology, Innovation and Development*, 9(1):31-41.

Kalan, R.S. & Unalir, M.O. 2016. Leveraging big data technology for small and medium-sized enterprises (SMEs). Paper presented at the 2016 6th International Conference on Computer and Knowledge Engineering (ICCKE).

Kamensky, J. 2018. Data Rich, But Information Poor. http://www.businessofgovernment.org/blog/data-rich-information-poor Date of access: 16 December 2020.

Karanasios, S. 2011. New & Emergent ICTs and Climate Change in Developing Countries. .

Kfouri, G. & Skyrius, R. 2017. Factors influencing the implementation of business intelligence among small and medium enterprises in Lebanon. *Informacijos mokslai*, 76(76).

Kilani, M.A. & Kobziev, V. 2016. An Overview of Research Methodology in Information System (IS). *OALib*, 03(11):1-9.

Kilimis, P., Zou, W., Lehmann, M. & Berger, U. 2019. A Survey on Digitalization for SMEs in Brandenburg, Germany. *International Federation of Automatic Control (IFAC)*, 52(13):2140–2145.

Kimball, R. & Ross, M. 2013. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. Indianapolis, Indiana: John Wiley & Sons, Inc.

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. & Becker, B. 2008. The Data Warehouse Lifecycle Toolkit, Second Edition: Practical Techniques for Building Data Warehouse and Business Intelligence Systems. 2nd ed. Indianapolis, Indiana: Wiley Publishing.

Kothari, C.R. 2004. Research methodology: Methods and techniques: New Age International.

KPMG. 2016. Now or never: 2016 Global CEO Outlook.

Krauss, S. 2015. Research Paradigms and Meaning Making: A Primer. *The Qualitative Report*.

Kuechler, W. & Vaishnavi, V. 2012. A Framework for Theory Development in Design Science Research: Multiple Perspectives. *Journal of the Association of Information System*, 13(6):395-423.

Kumar, R. 2019. Research Methodology- Step-by-step guide for beginners. 5th. UK: SAGE.

Laranjeiro, N., Soydemir, S.N. & Bernardino, J. 2015. A Survey on Data Quality: Classifying Poor Data. Paper presented at the 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC).

Li, L., Peng, T. & Kennedy, S. 2011. A rule based taxonomy of dirty data. *GSTF International Journal on Computing*, 1(2):140-148.

Liew, A. 2013. DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their Interrelationships. *Business Management Dynamics*, 2(10):49-62.

Limpeeticharoenchot, S., Cooharojananone, N., Chanvanakul, T., Tuaycharoen, N. & Atchariyachanvanich, K. 2020. Innovative Mobile Application for Measuring Big Data Maturity: Case of SMEs in Thailand. *International Journal of Interactive Mobile Technologies (iJIM)*, 14(18).

Liu, Q., Feng, G., Zhao, X. & Wang, W. 2020. Minimizing the data quality problem of information systems: A process-based method. *Decision Support Systems*, 137.

Liu, Y., Han, H. & DeBello, J.E. 2018. The Challenges of Business Analytics: Successes and Failures. (*In*. Proceedings of the 51st Hawaii International Conference on System Sciences organised by.

Llave, M.R. 2019. A Review of Business Intelligence and Analytics in Small and Medium-Sized Enterprises. *International Journal of Business Intelligence Research*, 10(1):19-41.

Loshin, D. 2001. Enterprise Knowledge Management. The Data Quality Approach. US: Academic Press.

Loshin, D. 2011. Evaluating the Business Impacts of Poor Data Quality: International, I.

Mackenzie, N. & Knipe, S. 2006. Research dilemmas: Paradigms, methods and methodology. *Educational Research*, 16(2).

Mahanti, R. 2018. Data Quality : Dimensions, Measurement, Strategy, Management, and Governance: Milwaukee: ASQ Quality Press.

Mahanti, R. 2019. Data quality and data quality dimensions. *SQP*, 22(1).

Matthews, B. & Ross, L. 2010. Research methods: A practical guide for the social sciences. England: Pearson.

Mauch, J.E. & Park, N. 2003. Guide to a successful dissertation: A handbook for students and faculty. 5th. Unites States of America: Marcel Dekker.

Mehrad, A. & Tahriri, M. 2019. Comparison between Qualitative and Quantitative Research Approaches: Social Sciences. *International Journal For Research In Educational Studies*, 5(7).

Mertens, D.M. 2010. Research and Evaluation in Education and Psychology: Integrating diversity with qualitative, quantative and mixed methods. 3rd Ed. California: Sage.

Moore, S. 2018. How to Create a Business Case for Data Quality Improvement. https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/ Date of access: 06 September 2020.

Moyo, M. & Loock, M. 2016. South African small and medium-sized enterprises' reluctance to adopt and use cloud-based business intelligence systems: A literature review. Paper presented at the The 11th International Conference for Internet Technology and Secured Transactions (ICITST-2016).

Mullins, C.S. 2013. Database Administration:The Complete Guide to DBA Practices and Procedures. 2nd Ed. US: Addison-Wesley.

Mullins, C.S. 2017. The high cost of poor data quality. https://www.dbta.com/Columns/DBA-Corner/The-High-Cost-of-Poor-Data-Quality-122101.aspx Date of access: 01 February 2021.

Myers, M.D. & Klein, H.K. 2011. A set of principle for conducting critical research in Information Systems. *MIS Quarterly*, 35(1).

Myers, M.D. & Venable, J.R. 2014. A set of ethical principles for design science research in information systems. *Information & Management*, 51(6):801-809.

Neergaard, M.A., Olesen, F., Andersen, R.S. & Sondergaard, J. 2009. Qualitative description - the poor cousin of health research? *BMC Med Res Methodol*, 9:52.

Nenzhelele, T.E. & Pellissier, R. 2014. Competitive Intelligence Implementation Challenges of Small and Medium-Sized Enterprises. *Mediterranean Journal of Social Sciences*.

Ngek, N. 2014. Determining high quality SMEs that significantly contribute to SME growth: Regional evidence from South Africa. *Problems and Perspectives in Management*, 12(4).

Nkwe, N. 2012. Role of SMES in Botswana. *American International Journal of Contemporary Research*, 2(8):29-37.

Nyaboga, A.B. & Mwaura, M.F. 2009. Strategies For Gaining Competitive Advantage In A Dynamic Environment Thru Data Quality. *International Journal of Management & Information Systems*, 13(1).

Obeidat, M., North, M., Burgess, L., Parker, R. & North, S. 2015. DRIP – Data Rich, Information Poor: A Concise Synopsis of Data Mining. *Universal Journal of Management*, 3(1):29-35.

OECD. 2020. Financing SMEs and enterpreneurs 2020: An OECD Scorecard. https://www.oecd-ilibrary.org/sites/37b75ad0-en/index.html?itemId=/content/component/37b75ad0-en Date of access: 20 February 2021.

Olexová, C. 2014. Business intelligence adoption: a case study in the retail chain. *WSEAS transactions on business and economics*, 11:95-106.

Olszak, C. & Ziemba, E. 2012. Critical Success Factors for Implementing Business Intelligence Systems in Small and Medium Enterprises on the Example of Upper Silesia, Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*, 7.

Papachristodoulou, E., Koutsaki, M. & Kirkos, E. 2017. Business intelligence and SMEs: Bridging the gap. *Journal of Intelligence Studies in Business*, 7(1).

Parapi, J.M.O. & Masykuri, E.S. 2020. Data and its increasing importance among organisations and businesses. *English Departmental Journal*, 7(1).

Peffers, K., Tuunanen, T., Rothenberger, M.A. & Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45-78.

Pitney Bowes. 2015. The ROI of data quality: Realize lower costs, increase revenue and higher profits. https://www.pitneybowes.com/content/dam/pitneybowes/us/en/customer-information-management/data-quality/roi-of-data-quality-white-paper-may-2018.pdf Date of access: 2 February 2021.

Pitney Bowes. 2019. Data quality gets smart. https://www.pitneybowes.com/content/dam/pitneybowes/us/en/customer-information-management/data-quality/data-quality-gets-smart-whitepaper-us.pdf Date of access: 02 February 2021.

Ponis, S.T. & Christou, I.T. 2013. Competitive intelligence for SMEs: a web-based decision support system. *International Journal of Business Information Systems*, 12(3).

Popovič, A., Puklavec, B. & Oliveira, T. 2019. Justifying business intelligence systems adoption in SMEs: Impact of Systems Use on Firm Performance. *Industrial Management & Data Systems*, 119(1):210-228.

Rahi, S. 2017. Research Design and Methods: A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development. *International Journal of Economics & Management Sciences*, 06(02).

Raj, R., Wong, S. & Beaumont, A.J. 2016. Business Intelligence Solution for an SME: A Case Study. (*In*. In Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) organised by: SCITEPRESS. p. 41-50).

Ram, S. 2013. Benefits of Data Warehousing and Data Mining Tools. Auckland, New Zealand: Unitec Institute of Technology.

Ramakrishnan, R. & Gehrke, J. 2000. Database Management Systems. 2nd Edition. Boston: McGraw-Hill Higher Education.

Ramakrishnan, R. & Gehrke, J. 2003. Database Management Systems. 3rd ed: McGraw-Hill.

Ramukumba, T. 2016. Overcoming SMEs challenges through critical success factors: A case of SMES in the Western Cape Province, South Africa. *Economics and business review*, 16(1):19-36.

Randolph, J. 2009. A Guide to Writing the Dissertation Literature Review. *Practical Assessment, Research and Evaluation*, 14(13).

Rashid, Y., Rashid, A., Warraich, M.A., Sabir, S.S. & Waseem, A. 2019. Case Study Method: A Step-by-Step Guide for Business Researchers. *International Journal of Qualitative Methods*, 18.

Razbonyalı, C. & Güvenoğlu, E. 2016. Traditional Data Storage Methods and the Big Data Concepts. *International Research Journal of Engineering and Technology (IRJET)* 3(6):2556.

Redman, T.C. 1998. The impact of poor data quality on the typical enterprise *The typical communicationof the ACM*, 41(2).

Redman, T.C. 2004. Data: An Unfolding Quality Disaster: Review, D.

Rehman, A.A. & Alharthi, K. 2016. An introduction to research paradigms. *International Journal of Educational Investigations*, 3(8):51-59.

Runkler, T.A. 2012. Data Analytics: Models and Algorithms for Intelligent Data Analysis3rd ed: Springer Vieweg. p. 1-3).

Russom, P. 2006. Taking Data Quality to the Enterprise through Data Governance: Publication, A.C.

Sahoo, S. & Yadav, S. 2018. Total Quality Management in Indian Manufacturing SMEs. (*In*. 15th Global Conference on Sustainable Manufacturing organised by: Elsevier B.V. p. 541–548).

Saldaña, J. 2013. The coding manual for qualitative researchers. 2nd Ed: SAGE.

SAS. 2020. Data Management: What It Is and Why It Matters. https://www.sas.com/en_za/insights/data-management/data-management.html Date of access: 8 September 2020.

Saunders, C., Lewis, P. & Thornhill, A. 2019. Understanding research philosophy and approaches to theory development. Research Methods for Business Students 8th Ed ed: Pearson.

Scannapieco, M., Missier, P. & Batini, C. 2005. Data Quality at a Glance.

Schniederjans, M.J., Schniederjans, D.G. & Starkey, C.M. 2015. Business Analytics Principles, Concepts, and Applications with SAS What, Why, and How New Jersey Pearson Education.

Scholz, P., Schieder, C., Kurze, C., Gluchowski, P. & Boehringer, M. 2010. Benefits and challenges of business intelligence adoption in small and medium-sized enterprises Paper presented at the 18th European Conference on Information Systems. http://aisel.aisnet.org/ecis2010/32 Date of access: 05 September 2020.

Scotland, J. 2012. Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms. *English Language Teaching*, 5(9).

Seale, C. & Sillverman, D. 1997. Ensuring rigour in qualitative research. *Europian Journal of Public Health*, 7(4):379-384.

Sherman, R. 2014. Business intelligence guidebook: From data integration to analytics: Newnes.

Sidi, F., Shariat Panahy, P.H., Affendey, L.S., Jabar, M.A., Ibrahim, H. & Mustapha, A. 2012. Data quality: A survey of data quality dimensions. Paper presented at the 2012 International Conference on Information Retrieval & Knowledge Management.

Silverman, D. 2013. Doing qualitative research. 4th Ed: Sage.

Singh, J. & Vashishtha, S. 2015. Data quality tools for data warehouse models. *International Journal of Engineering Science and Technology*, 7(5):181-191.

Singh, R. & Singh, K. 2010. A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. *International Journal of Computer Science Issues*, 7(3).

Smithers, J. 2013. Top three reasons SMEs should takle big data, 24 Oct 2013. https://www.dbta.com/Editorial/News-Flashes/The-Top-3-Reasons-Small-to-Midsize-Enterprises-Should-Tackle-Big-Data-92688.aspx Date of access: 15 December 2020.

Sonnenberg, C. & Vom Brocke, J. 2012. Evaluation Patterns for Design Science Research Artefacts. Practical Aspects of Design Science. p. 71-83).

Sousa, S.D., Lopes, I. & Teixeira, H.N. 2012. A methodology for quality problems diagnosis in SMEs. Paper presented at the World Academy of Science, Engineering and Technology

Sutton, J. & Austin, Z. 2015. Qualitative Research: Data Collection, Analysis, and Management. *CJHP*, 68(3).

Tayi, G.K. & Ballou, D.P. 1998. Examining data quality. *communications of the ACM*, 41(2).

TDWI. 2020. TDWI Checklist Report : Maximizing Data Utility by Ensuring Trust in the Data.

Timmerman, Y. & Bronselaer, A. 2019. Measuring data quality in information systems research. *Decision Support Systems*, 126.

Vaishnavi, V., Kuechler, W. & Petter, S. 2004/2019. Design Science Research in Information Systems.

Van der Krogt, A., Coronel, R. & Núñez_Castillo, C. 2020. Adoption of business intelligence for SMEs. Paper presented at the 8th International Conference Innovation Management, Entrepreneurship and Sustainability, Prague.

Van Scheers, L. 2011. SME's Marketing Skills Challenges in South Africa. *African Journal of Business Management*, 5(13):5048-5056.

Vargas-Solar, G., Zechinelli-Martini, J.L. & Espinosa-Oviedo, J.A. 2017. Big Data Management: What to Keep from the Past to Face Future Challenges? *Data Science and Engineering*, 2(4):328-345.

Vaziri, R. & Mohsenzadeh, M. 2012. A Questionnaire-Based Data Quality Methodology. *International Journal of Database Management Systems*, 4(2):55-68.

Venable, J., Pries-Heje, J. & Baskerville, R. 2017. FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, 25(1):77-89.

Venter, M.I. 2005. Business Intelligence (BI) initiatives: Failure versus success.149-163.

Vom Brocke, J., Hevner, A. & Maedche, A. 2020. Introduction to Design Science Research. Design Science Research. Cases. p. 1-13).

Vosburg, J. & Kumar, A. 2001. Managing dirty data in organizations using ERP:Lessons from a case study. *Industrial Management & Data Systems*, 101/1:21±31.

Vuleta, B. 2020. How Much Data Is Created Every Day? [27 Powerful Stats]. https://seedscientific.com/how-much-data-is-created-every-day/ Date of access: 11 October 2020.

Wang, R.Y. 1998. A Product Perspective on Total Data Quality Management. *Communications of the ACM*, 41(2).

Wang, R.Y., Kon, H.B. & Madnick, S.E. 1992. Data Quality Requirements Analysis and Modeling. Paper presented at the Ninth International Conference of Data Engineering, Vienna, Austria.

Wang, R.Y., Lee, Y.W., Pipino, L.L. & String, D.M. 1998. Manage your information as a Product. *Sloan Management Review*, 39(4).

Wang, R.Y. & Strong, D.M. 1996. Beyond_Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4):5-33.

Williams, D. & Tang, H. 2020. Data Quality Management for Industry 4.0: A Survey: Quality, A.S.F.

Yeoh, W. & Popovič, A. 2016. Extending the understanding of critical success factors for implementing business intelligence systems. *Journal of the Association for Information Science and Technology*, 67(1):134-147.

Yin, R.K. 2018. Case study research and applications: Design and methods. LA: Sage Publications.

Zafary, F. 2020. Implementation of business intelligence considering the role of information systems integration and enterprise resource planning. *Journal of Intelligence Studies in Business*, 10(1):59-74.

Zainal, Z. 2007. Case study as a research method. *Jurnal Kemanusiaan*:1-6.

Zainun Tuanmat, T. & Smith, M. 2011. The effects of changes in competition, technology and strategy on organizational performance in small and medium manufacturing companies. *Asian Review of Accounting*, 19(3):208-220.

Zhang, Y. & Wildemuth, B. 2009. Qualitative Analysis of Content. *Applications of social Research methods to questions in information and library science*.

Zhu, Y. & Cai, L. 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14(2):1-10.

APPENDIX 1: EXPLORATORY STUDY – QUESTIONS

1. Current state of the IT infrastructure in the organisation

1.1 Data storage

1.1.1 Brief overview of your IT infrastructure in terms of the systems and hardware's used to manage data.

1.1.2 What does the existing data architecture look like?

1.1.3 How are data currently stored? In CSV/Spreadsheet/RDBMS/XML/GIS/Other?

1.1.4 If you are using databases to store data, are they all centralised?

1.1.5 What type of database applications are you using?

1.1.6 How are data validated? How is data integrity ensured?

1.1.7 Who is responsible for data management?

1.1.8 What type of data storage issues is the company experiencing?

1.2 Data capturing

1.2.1 Currently, how do you capture the data? What application/s are you using?

1.2.2 What type of data do you capture?

1.2.3 How many data captures do you have? Are they able to capture data simultaneously?

1.2.4 What is the daily volume of your data? How big is it?

1.2.5 On a scale of 1 to 10, how do you rate your data integrity?

1.2.6 Do you experience problems relating to capturing of data?

1.3 Data Analysis and reporting

1.3.1 Once the data are captured, what do you use it for?

1.3.2 Who is allowed to access and view it? Are there different types of permission levels such as read/write to protect the data?

1.3.3 How efficient is your database application in term of retrieving and accessing the data?

1.3.4 Do your data influence organisational decision-making processes?

2. The main goal of this project is to produce an artefact that will enable efficient capturing of data and provide a reliable data storage system.

2.1 What do you envisage about this phase of the project? What value will the solution add to the company?

2.2 What are the organisations goals for the project?

2.3 What issues would you like this project to address?

2.4 What type of data would you like to capture?

2.5 Who should be able to capture data?

2.6 Who should be able to access and view the data in the database?

2.7 Will organisation's confidential information also form part of the project?

2.8 Once the data are captured, what will they be used for?

2.9 Is the company familiar with business intelligence systems and concepts?

2.10 Would the company consider adopting a business intelligence solution in the near future?

2.11 Do we have sample data that can be used in the development of the artefact?

APPENDIX 2: POST IMPLEMENTATION INTERVIEW QUESTIONS

1. Which tools were you using for capturing and managing data?
2. What shortcomings did the old data management tool have?
3. What motivated the decision to implement a user interface in the organisation?
4. What was your role in this project?
5. Were you involved in defining the requirements of the solution?
6. Based on the requirements given to the researcher, does the solution meet your expectations?
7. Were you involved in testing the solution? Would you say the solution was rigorously tested?
8. Were defects and failures addressed and resolved accordingly? Were any issues that were identified during testing sessions communicated to the developer on time to solve?
9. What was the initial timeframe of completion regarding data management artefact development? Was the solution delivered on time?
10. What, if any, complications arose during the solution project completion?
11. If any, what complications where there with regards to the deployment of the solution to your environment?
12. Tell me about how you experienced the changeover and adjustment to the new data capturing and management solution? Explain
13. What would you say was the overall impression or impact that the solution had on the business users using the new solution? Where any implications to business operations?
14. Are you able to complete tasks using the solution? What functionalities are you able to perform with the solution?
15. Based on the issues that the old system presented, would you say the new solution solved them or is solving them?
16. Are you confident that the solution adds value to the organisation?
17. Do you think this solution will solve the problems that this company is experience with regards to data quality?
18. Do you have the skills in-house to extract and read data stored in the database?
19. What would you change about the new data capturing solution, if anything?
20. What would you say was the overall impression and impact that the solution had on the organisation and business users using the newly developed solution?

Reporting (Test certificates)

21. Which tools were you using for reporting?

22. What shortcomings did the old reporting structure have?

23. How is the new reporting solution?

24. How did you experience the transition from the old reporting tools to the new reporting tools?

25. What was the reaction from other users interacting with the newly implemented reporting solution?

26. Is the solution adding value to the organisation? Please elaborate

27. Is management using the reports (test certificates)? Please explain

28. Has the reporting solution made any improvements (provide accurate, consistent and reliable data) for decision making?

29. What would you change about the new reporting structure, if anything?

APPENDIX 3: QUESTIONNAIRE (POST IMPLEMENTATION)

DM Interface - Questionnaire

This questionnaire is based on the user interface and SQL relational database designed to assist with improving data quality and data management.

Interface display

This section will assess the look and feel of the application (colours, display and visibility)

The application is accessible. *

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

The application provides level of security (username and password required to connect) *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

The information on the application is visible and readable. *

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

The information on the interface is clearly organised. *

- Strongly agree
- Agree
- Disagree
- Strongly Disagree

The information on the application is easy to understand *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Are you satisfied with the interface display, colours and font size. *

- Very satisfied
- Satisfied
- Dissatisfied
- Very dissatisfied

It is easy to navigate through the application and find what I'm looking for. *

- Strongly agree
- Agree
- Disagree
- Strongly Disagree

I feel comfortable using the application *

- strongly Agree
- Agree
- Disagree
- Strongly Disagree

The error messages, validations and warnings are visible and readable. *

- Strongly agree
- Agree
- Disagree
- Strongly Disagree

The application gives errors that clearly states what the problems are. *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

The application is pleasant *

	1	2	3	4	5	6	7	8	9	10	
slightly	○	○	○	○	○	○	○	○	○	○	highly

I am satisfied with the application layout *

- Extremely satisfied
- Satisfied
- Dissatisfied
- Extremely dissatisfied

Overall comments or suggestions on the look and feel of the application *

Long answer text

Interface Functionality

This section will assess the functionality of the application

I am able to connect and use the application anytime I want *

- Strongly disagree
- Disagree
- Agree
- Strongly agree

Multiple users are able to connect to the application simultaneously without experiencing access limitation issues. *

- Strongly agree
- Agree
- Disagree
- Strongly Disagree

I can efficiently capture and store data in the database using the application. *

- Yes
- No

I can efficiently retrieve and update data in the database using the application. *

- Yes
- No

I can delete data stored in the database using the application *

- Yes
- No

Duplicate data cannot be inserted in the database *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Data are stored accurately *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Data are consistent throughout. *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Data validation checks are working accordingly (checking duplicates, unique values, no null or blank values accepted). *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

I am able to efficiently complete my data capturing tasks using the application *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Do you think the application can minimise data capturing issues caused by human error? *

Long answer text

The application was developed to assist with efficient capturing, retrieval and management of data and to improve data quality (accuracy, consistency, completeness, uniqueness, availability, eliminate duplicates, improve data format). Do you think the application meets its objectives? *

Long answer text

How did the data capturing process improve since the application was implemented? *

Long answer text

Overall, how will the application add value to the organisation? *

Long answer text

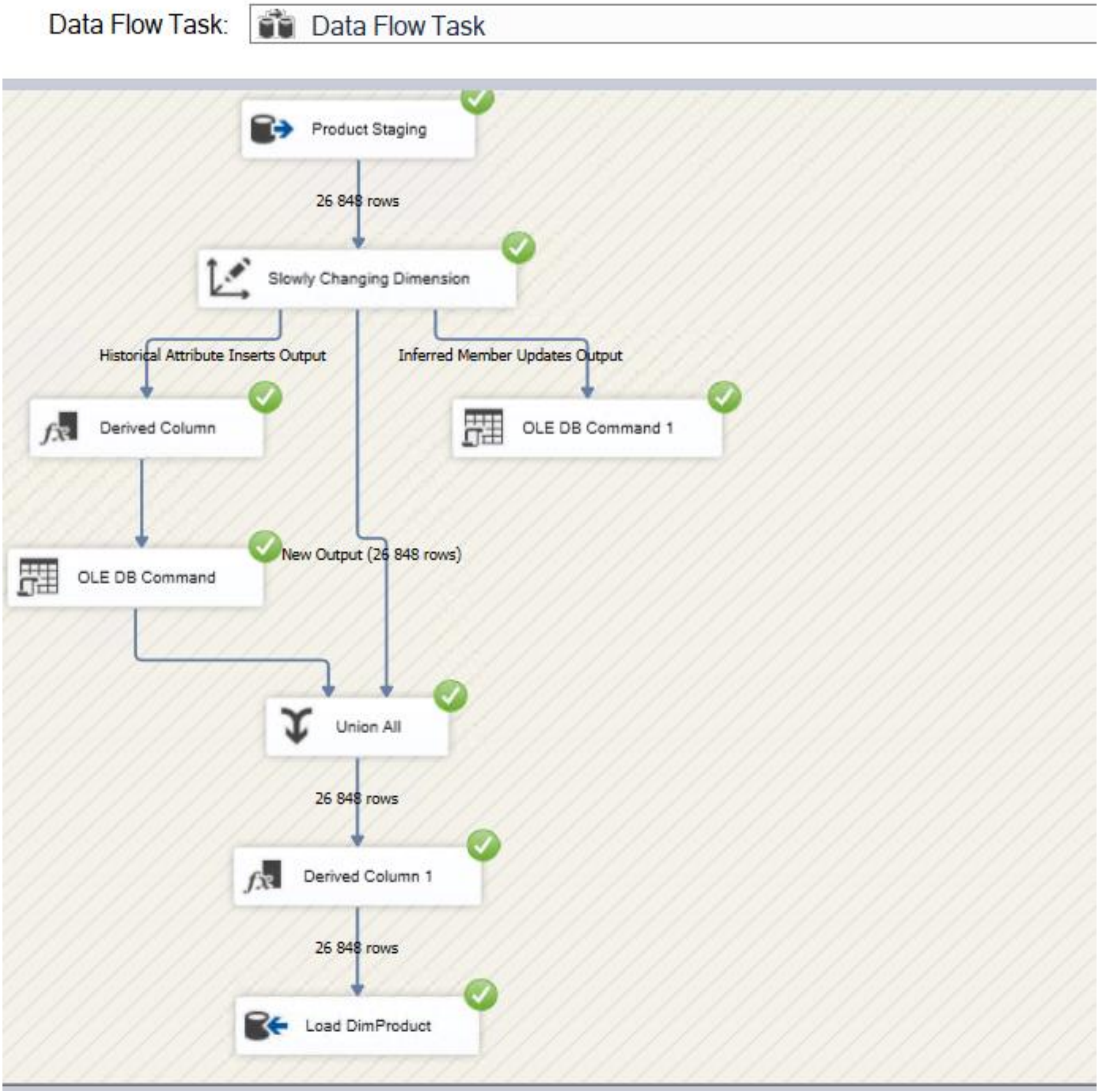
Overall, are you satisfied with the interface and its functionalities? *

	1	2	3	4	5	6	7	8	9	10	
Not satisfied	○	○	○	○	○	○	○	○	○	○	Extremely satisfied

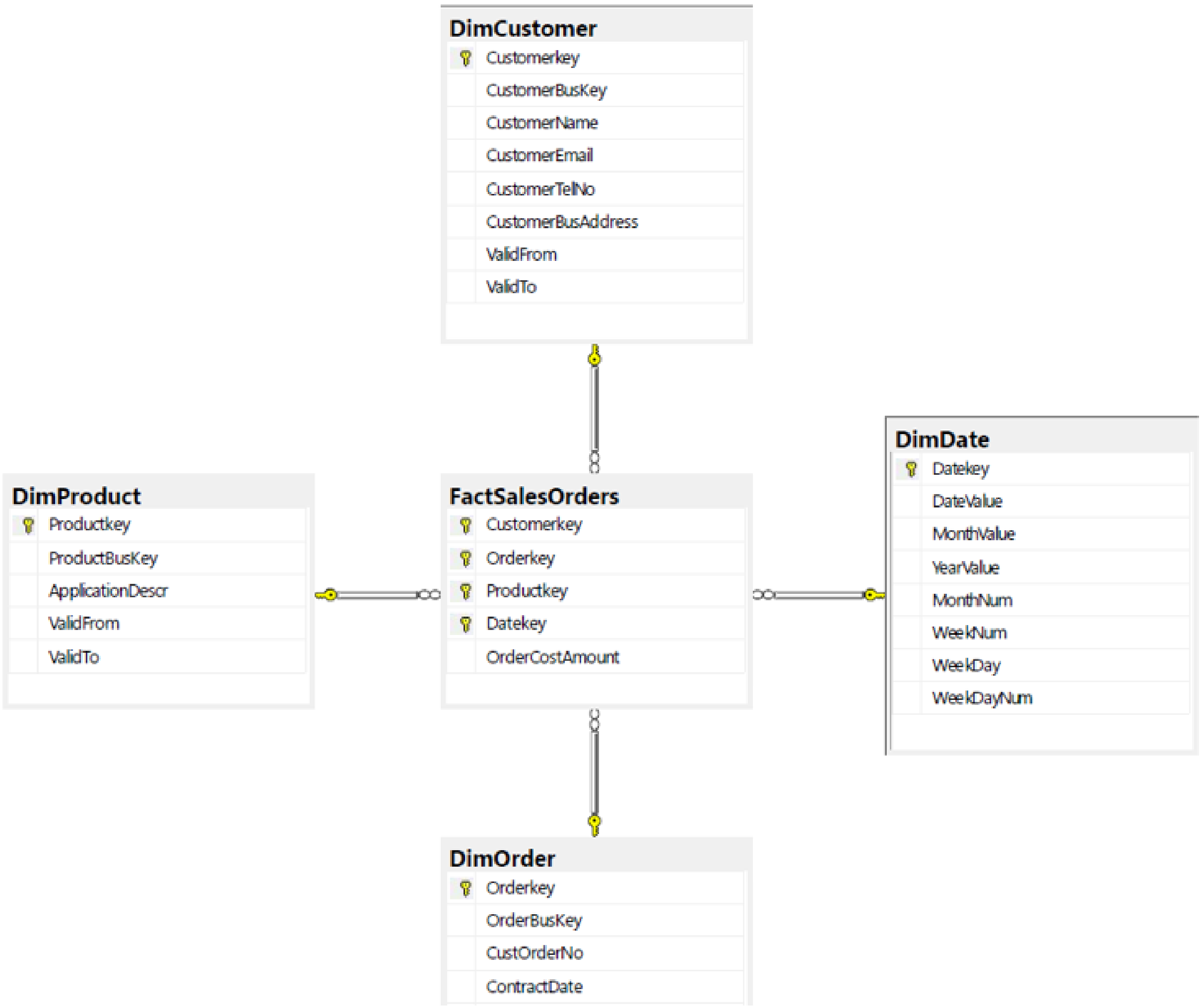
Any comments or recommendations

Long answer text

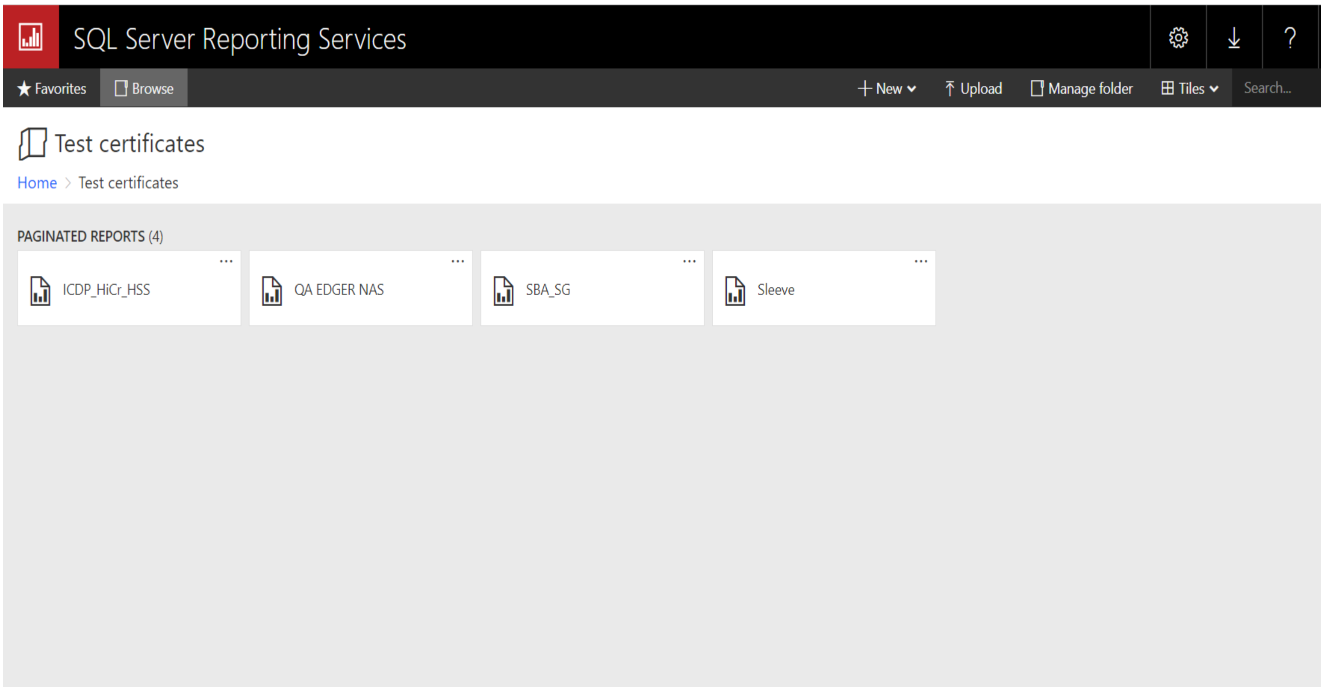
APPENDIX 4: DATA TRANSFORMATION SSIS PACKAGE

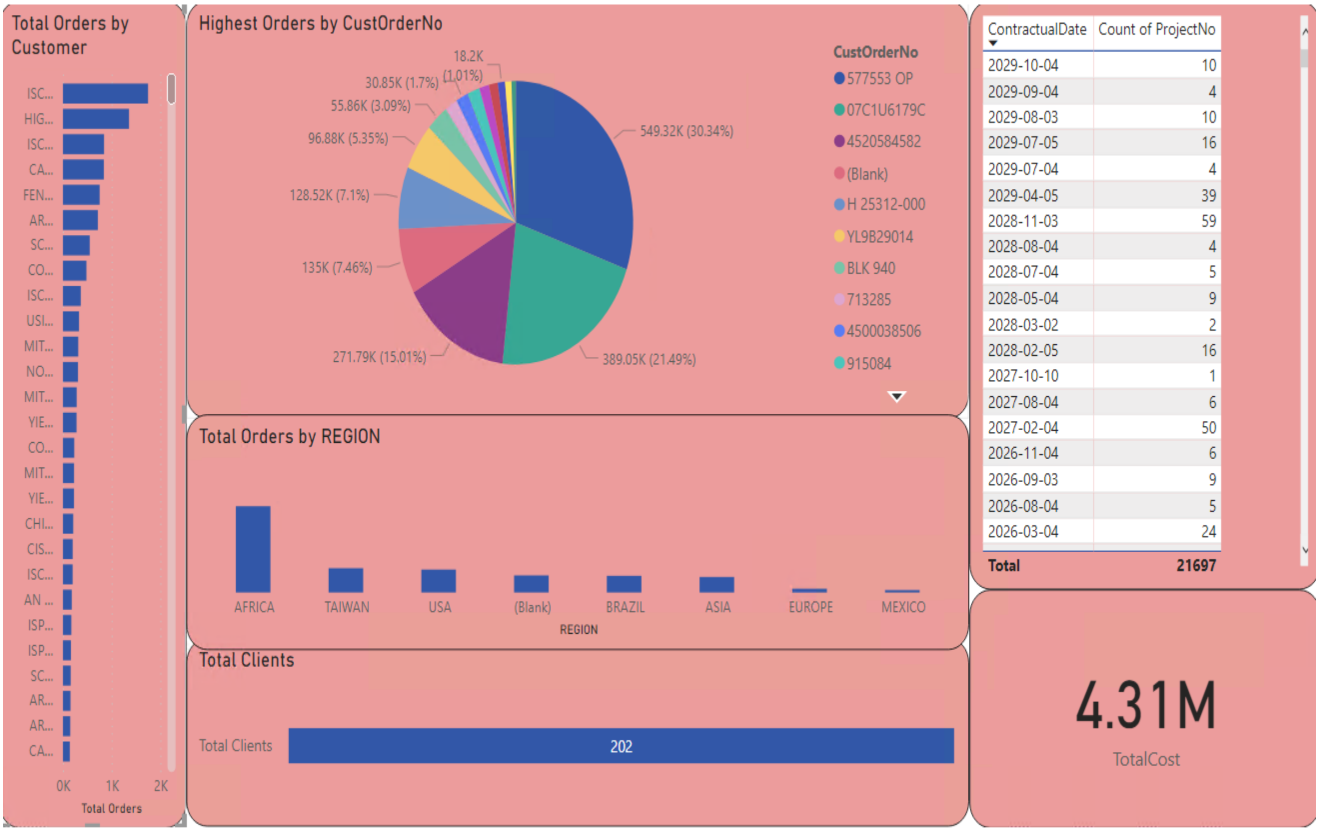


APPENDIX 5: DW STAR SCHEMA



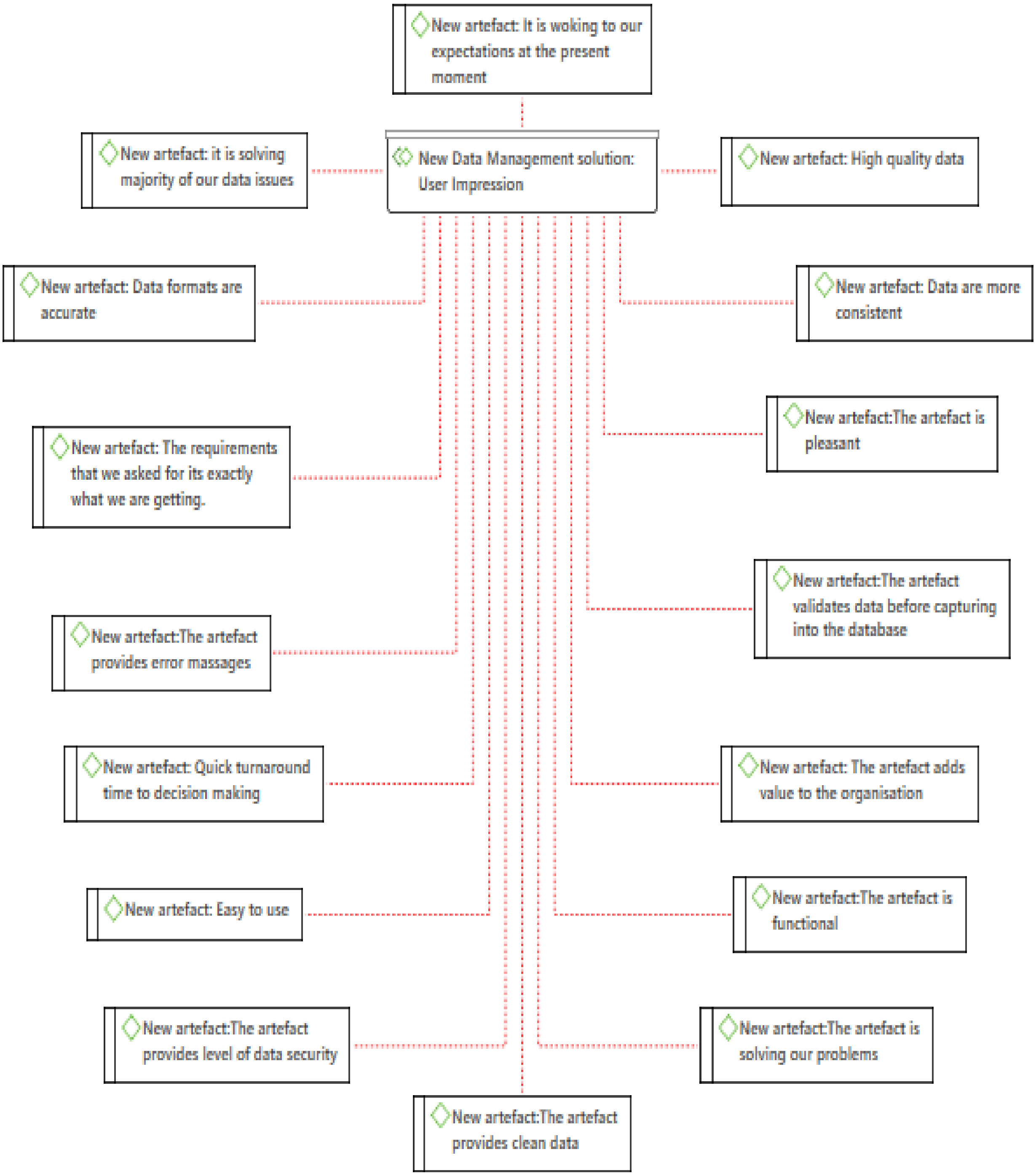
APPENDIX 6: SQL SERVER REPORTING AND ANALYTICS



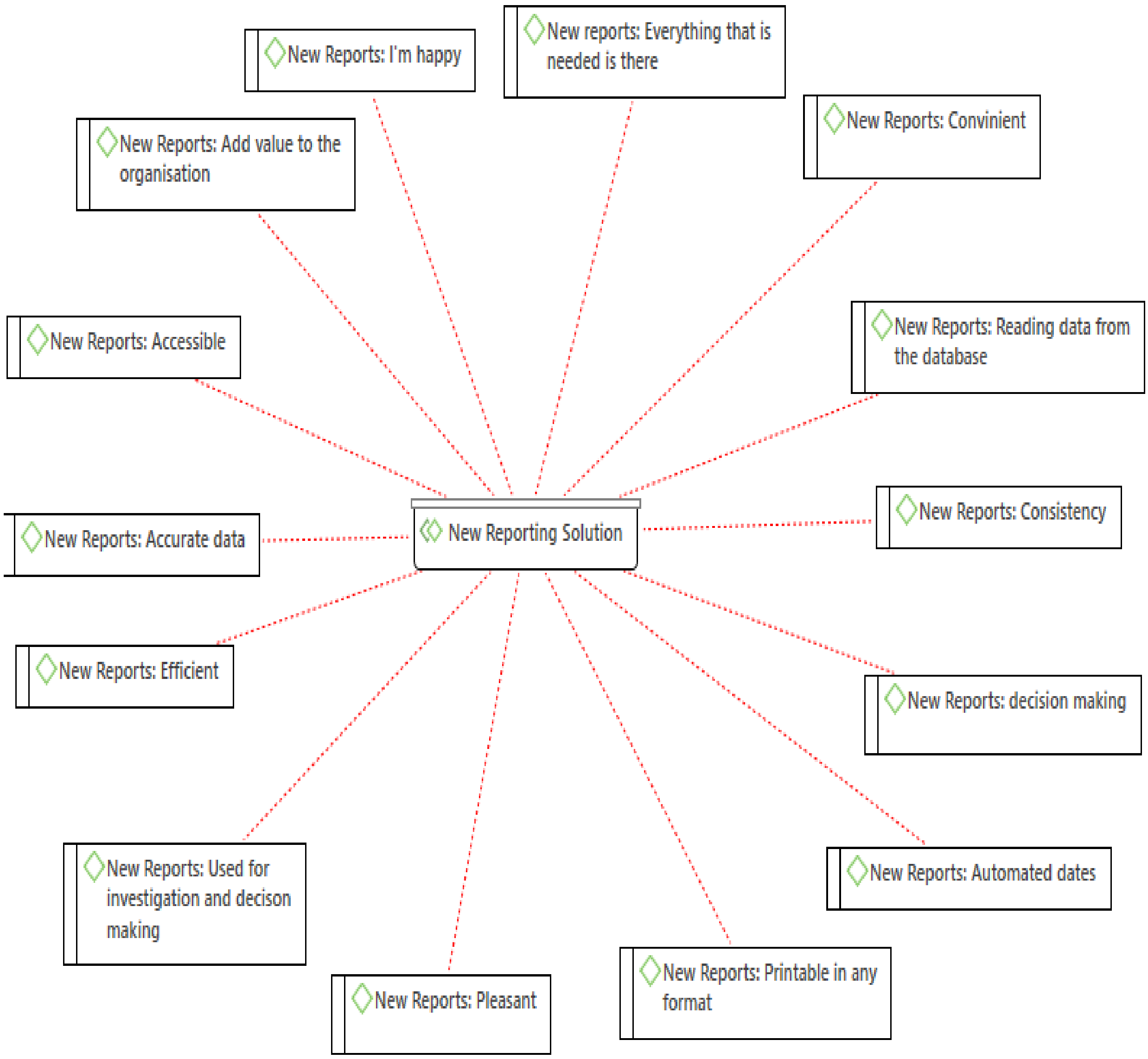


APPENDIX 7: OLD DATA MANAGEMENT CODE NETWORK

APPENDIX 8: NEW ARTEFACT CODE NETWORK



APPENDIX 9: NEW REPORTING SOLUTION CODE NETWORK



APPENDIX 10: SUPPORTING CODE

```
<asp:Content ID="Content1" ContentPlaceHolderID="head" runat="server">
</asp:Content>
<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" runat="server">

    <div class="container" id="searchCust">

            <div class="col-md-9 mx-auto">


                <div class="card" id="CardGridCustomer">
                    <div class="card-body">
                                <div class="row">
                                    <div class="col-md-5">
                                        <br />
                                        </div>
                                    </div>



                          <div class="row">
                              <div class="col-md-6">
                                   <input type='text' onkeyup='Filter(this);' id="txtSearch"
placeholder="Search customer" class="form-control" />
                              </div>
                          </div>
                        <br />


                        <asp:GridView ID="GrdCustomerDetails" class="table-striped table-bordered
alert-light" runat="server" AutoGenerateColumns="false"
OnRowDataBound="OnRowDataBound"
            DataKeyNames="CustomerId"  OnRowCancelingEdit="OnRowCancelingEdit" PageSize
= "10" AllowPaging ="true" OnRowEditing="OnRowEditing" OnPageIndexChanging =
"OnPaging"
            OnRowDeleting="OnRowDeleting"
OnRowUpdating="GrdCustomerDetails_RowUpdating"
            Width="800" OnRowCommand="GrdCustomerDetails_RowCommand">
            <Columns>
                <asp:TemplateField HeaderText="Customer" ItemStyle-Width="200">
                    <ItemTemplate>
                        <asp:Label ID="lblCustName"  runat="server" Text='<%# Eval("CustName")
%>'></asp:Label>
                    </ItemTemplate>
                    <EditItemTemplate>
                        <asp:TextBox ID="txtCustName" ReadOnly="true" runat="server" Text='<%#
Eval("CustName") %>' Width="200"></asp:TextBox>
                    </EditItemTemplate>
                </asp:TemplateField>
                <asp:TemplateField HeaderText="TelNo" ItemStyle-Width="200">
                    <ItemTemplate>
                        <asp:Label ID="lblTelNo" runat="server" Text='<%# Eval("TelNo")
%>'></asp:Label>
```

```
                </ItemTemplate>
                <EditItemTemplate>
                    <asp:TextBox ID="txtTelNo" runat="server" Text='<%# Eval("TelNo") %>'
Width="140"></asp:TextBox>
                </EditItemTemplate>
            </asp:TemplateField>

               <asp:TemplateField HeaderText="Email" ItemStyle-Width="150">
                <ItemTemplate>
                    <asp:Label ID="lblEmailAddr" runat="server" Text='<%# Eval("EmailAddr")
%>'></asp:Label>
                </ItemTemplate>
                <EditItemTemplate>
                    <asp:TextBox ID="txtEmailAddr" runat="server" Text='<%# Eval("EmailAddr")
%>' Width="140"></asp:TextBox>
                </EditItemTemplate>
            </asp:TemplateField>


               <asp:TemplateField HeaderText="Address" ItemStyle-Width="150">
                <ItemTemplate>
                    <asp:Label ID="lblBusinessAddr" runat="server" Text='<%#
Eval("BusinessAddr") %>'></asp:Label>
                </ItemTemplate>
                <EditItemTemplate>
                    <asp:TextBox ID="txtBusinessAddr" runat="server" Text='<%#
Eval("BusinessAddr") %>' Width="140"></asp:TextBox>
                </EditItemTemplate>
            </asp:TemplateField>

            <asp:CommandField ButtonType="Link" ShowEditButton="true"
ShowDeleteButton="true"
                ItemStyle-Width="150" />
        </Columns>
                         <RowStyle BackColor="#F7F6F3" VerticalAlign="Middle" />
                        <HeaderStyle BackColor="#99CCFF" />

    </asp:GridView>
                    </div>
                </div>
            </div>
    </div>
<br />
  <br />
 <div class="container" id="AddCust">

        <div class="col-md-9 mx-auto">
            <div class="card">
                <div class="card-body">
                    <div class="form-group">
                        <div class="row">
                            <div class="col-md-4">
                                <asp:Label ID="Label5" runat="server" Text="Add Customer
Information:" ForeColor="GrayText" Font-Bold="true"></asp:Label>
                            </div>
```

```
                </div>
                 
              <div class="row">

                 <hr />
              </div>

                <div class="row">
                   <div class="col-md-4">
                      <asp:Label ID="Label1" runat="server"  Text="Customer Name" 
ForeColor="GrayText"></asp:Label>
                    <asp:TextBox ID="txtCustName" runat="server" CssClass="form-control 
text-uppercase"   ToolTip="customer name"></asp:TextBox>



                      <asp:RequiredFieldValidator ID="ReqValCustname" 
ValidationGroup="ValCustData" CssClass="text text-danger" Font-Italic="true" 
ControlToValidate="txtCustName" runat="server" ErrorMessage="Customer name required 
"></asp:RequiredFieldValidator>
                   </div>

                      <div class="col-md-4">
                         <asp:Label ID="Label3" runat="server" Text="Telephone No:" 
ForeColor="GrayText"></asp:Label>


                         <asp:TextBox ID="txtTelNo" runat="server" CssClass="form-control" 
placeholder="0000000000" ToolTip="Telephone" MaxLength="10" 
TextMode="Number"></asp:TextBox>
                     <asp:RegularExpressionValidator ID="RegularExpressionValidator1" 
Font-Italic="true" ValidationGroup="ValCustData" runat="server" SetFocusOnError="true" 
Display="Dynamic" ControlToValidate="txtTelNo" ForeColor="Red" ValidationExpression="^\d$" 
ErrorMessage="10 digits telephone number allowed"></asp:RegularExpressionValidator>


                         </div>
                    </div>

                 <div class="row">

                      <div class="col-md-4">
                         <asp:Label ID="Label4" runat="server" Text="Email Address" 
ForeColor="GrayText"></asp:Label>

                         <asp:TextBox ID="txtEmail" runat="server" CssClass="form-control" 
ToolTip="Email Address" TextMode="Email"></asp:TextBox>
                         <asp:RequiredFieldValidator ID="RequiredFieldValidator1" Font-
Italic="true" ValidationGroup="ValCustData" CssClass="text text-danger" 
ControlToValidate="txtEmail" runat="server" ErrorMessage="Email address required 
"></asp:RequiredFieldValidator>
                      </div>
```

```
                     

                            <div class="col-md-4">
                                <asp:Label ID="Label6" runat="server" Text="Business Address"
ForeColor="GrayText"></asp:Label>

                                <asp:TextBox ID="txtBusAddress" runat="server" CssClass="form-
control" TextMode="MultiLine" ToolTip="Location Address"></asp:TextBox>
                             </div>

                        </div>
                          
                  <div class="row">

                     <hr />
                 </div>
                    <div class="row">
                        <div class="col-md-4">
                            <asp:Button ID="btnSaveCustInfo" CssColor="" runat="server"
ValidationGroup="ValCustData"  CssClass="btn btn-primary btn-block" Width="200"
Text="Save" OnClick="Button1_Click" />
                            <%--<button type="submit" runat="server" id="btnExit" class="btn btn-
default" data-complete-text="completed" causesvalidation="false" ><span class="glyphicon
glyphicon-backward"></span></button>--%>
                        </div>

                           
                        <div class="col-md-4">
                            <asp:Button ID="btnCancel" runat="server" CssClass="btn btn-warning
btn-block" Width="200" Text="Cancel" OnClick="btnCancel_Click" />
                        </div>
                    </div>
                    <asp:HiddenField ID="HiddenDisplayName" runat="server" />
                    <br />

                  </div>


                </div>
            </div>
        </div>
    </div>

    <script type="text/javascript">
    //for serching gridview on keyup
    function Filter(Obj) {
        var grid;
        grid = document.getElementById('<% =GrdCustomerDetails.ClientID%>');

        var terms = Obj.value.toUpperCase();
        var cellNr = 0; //your grid cellindex like name
        var ele;
        for (var r = 1; r < grid.rows.length; r++) {
            ele = grid.rows[r].cells[cellNr].innerHTML.replace(/<[^>]+>/g, "");
            for (var col = 1; col < grid.rows[r].cells.length - 1; col++) {
```

```
            ele += grid.rows[r].cells[col].innerHTML.replace(/<[^>]+>/g, "");
        }
        if (ele.toUpperCase().indexOf(terms) >= 0)
            grid.rows[r].style.display = '';
        else grid.rows[r].style.display = 'none';
    }
  }
  </script>
</asp:Content>
```

APPENDIX 11: CERTIFICATE OF EDITING

This certificate declares that the dissertation with the title,

IMPROVING DATA QUALITY AND MANAGEMENT IN ONE SOUTH AFRICAN SME FOR DATA ANALYTICS

by **P.P. MABOTJA**, was edited by:

Ann-Lize Grewar

BA in Language and Literature Studies

BA Hons in Translation Studies

SATI-membership number: 1002647

SATI Accreditation: APSInterp Afr-Eng

Chairperson of SATI North-West Chapter

Professional Editor's Guild membership number BOS008

Language Director at Language Matters PTY(Ltd)

annlizeboshoff@gmail.com / 072 758 5797

Signed on 09/12/2021

Ann-Lize Grewar
Language Practitioner
B.A. Language and Literature Studies
B.A. Hons. Translation Studies

Contact us:
info@languagematters.co.za
www.languagematters.co.za

LANGUAGE MATTERS

Ann-Lize Grewar
Taalpraktisyn
B.A. Taal- en Literatuurstudies
B.A. Hons. Vertaalkunde

Kontak ons:
info@languagematters.co.za
www.languagematters.co.za

Disclaimer: The editor cannot take responsibility for any changes made after the signed date on this certificate. It remains the responsibility of the client to incorporate all reasonable changes and/or comments as suggested by the editor. The editor can only take responsibility for quality work within reasonable request with regards to timeous delivery of the manuscript from the client, i.e., at least five (5) working days for a manuscript length of 80 pages.

Date of receipt: 04/12/2021

Date of delivery: 09/12/2021