

**PRIORITISING DATA QUALITY CHALLENGES IN ELECTRONIC
HEALTHCARE SYSTEMS IN SOUTH AFRICA**

by

MARNA BOTHA

**submitted in accordance with the requirements
for the degree of**

MASTER OF SCIENCE

in the subject

COMPUTING

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: DR A BOTHA

CO-SUPERVISOR: PROF M HERSELMAN

OCTOBER 2015

ABSTRACT

Data quality is one of many challenges experienced in electronic healthcare (e-health) services in South Africa. The collection of data with substandard data quality leads to inappropriate information for health and management purposes. Evidence of challenges with regard to data quality in e-health systems led to the purpose of this study, namely to prioritise data quality challenges experienced by data users of e-health systems in South Africa. The study followed a sequential QUAL-quant mixed method research design to realise the research purpose. After carrying out a literature review on the background of e-health and the current status of research on data quality challenges, a qualitative study was conducted to verify and extend the identified possible e-health data quality challenges. A quantitative study to prioritise data quality challenges experienced by data users of e-health systems followed. Data users of e-health systems in South Africa served as the unit of analysis in the study. The data collection process included interviews with four data quality experts to verify and extend the possible e-health data quality challenges identified from literature. This was followed by a survey targeting 100 data users of e-health systems in South Africa for which 82 responses were received.

A prioritised list of e-health data quality challenges has been compiled from the research results. This list can assist data users of e-health systems in South Africa to improve the quality of data in those systems. The most important e-health data quality challenge is a lack of training for e-health systems data users. The prioritised list of e-health data quality challenges allowed for evidence-based recommendations which can assist health institutions in South Africa to ensure future data quality in e-health systems.

DECLARATION

Student number: 55768458

I declare that **Prioritising data quality challenges in electronic healthcare systems in South Africa** is my own work and that all sources used or quoted have been indicated and acknowledged by means of complete references.

SIGNATURE

(Ms M. Botha)

DATE

ACKNOWLEDGEMENTS

My sincere gratitude and appreciation go to my supervisor, Dr Adele Botha, for her hard work, support, advice, guidance, constructive criticism and motivation during this study. Without you I would not have completed this study – thank you.

I also would like to thank my co-supervisor, Prof Marlien Herselman, for her guidance, support and motivation during this study. Without your gentle and encouraging smile, I probably would have cried a few times.

I would like to thank my Heavenly Father, for the strength He provided me throughout the course of this study and also for the opportunities He granted me in this regard.

I would like to express my gratitude to the following people:

- My parents (Salomon and Lienie Botha) for their loving support and kindness – not only during the course of this study, but throughout my life.
- My dearest friend, Chanica Viljoen, for her love, support and patience during the study.
- Stefan van Zyl for the language editing.
- Annecke Botha and Inge Odendaal for the distribution of research questionnaires.
- The interview and survey participants for their time and valuable contributions.

With heartfelt gratitude,

LIST OF PUBLICATIONS

Peer reviewed papers:

1. Botha, M., Botha, A., & Herselman, M. (2014, May). Compiling a prioritized list of health data quality challenges in public healthcare systems. In *IST-Africa Conference Proceedings, 2014* (pp. 1-8). IEEE.
2. Botha, M., Botha, A., & Herselman, M. (2014, December). Data quality challenges: A content analysis in the e-health domain. In *Information and Communication Technologies (WICT), 2014 Fourth World Congress on* (pp. 107-112). IEEE.
3. Botha, M., Botha, A., & Herselman, M. (2014, December). The Benefits and Challenges of e-Health Applications: A Content Analysis of the South African context. In *International Conference on Computer Science, Computer Engineering and Social Media (CSCESM2014)*. SDIWC.
4. Botha, M., Botha, A., & Herselman, M. (2015, September). On the prioritisation of data quality challenges in e-health systems in South Africa. SAICSIT.
5. Botha, M., Botha, A., & Herselman, M. (2015, November). Prioritized e-health data quality challenges: An application on the Tiered ART Monitoring Strategy. In *Proceedings of the International IDIA Development Informatics Conference*.

LIST OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
e-Health	Electronic health
ICT	Information and Communication Technology
CSIR	Council for Scientific and Industrial Research
CSET	College of Science, Engineering and Technology
CREC	Research ethics Committee
UNISA	University of South Africa
HIS	Health Information System
EHR	Electronic Health Record

CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 INTRODUCTION	2
1.2 BACKGROUND	3
1.2.1 BACKGROUND TO THE STUDY	3
1.2.2 PROBLEM STATEMENT	4
1.3 SIGNIFICANCE OF THE STUDY	4
1.4 RESEARCH QUESTIONS	5
1.4.1 MAIN RESEARCH QUESTION	5
1.4.2 SECONDARY RESEARCH QUESTIONS	5
1.5 RESEARCH OBJECTIVES	6
1.6 SCOPE AND CONTEXT OF THE STUDY	8
1.6.1 SCOPE	8
1.6.2 DELINEATION	8
1.7 RATIONALE OF RESEARCH	9
1.7.1 PERSONAL RATIONALE	9
1.7.2 SCIENTIFIC RATIONALE	9
1.8 RESEARCH APPROACH	9
1.9 ETHICAL CONSIDERATIONS	10
1.10 DISSERTATION STRUCTURE	10
CHAPTER 2: BACKGROUND TO THE RESEARCH DOMAIN – E-HEALTH	13
2.1 INTRODUCTION	15
2.2 E-HEALTH IN SOUTH AFRICA	17
2.3 PURPOSE OF E-HEALTH	18
2.4 DEFINITION OF E-HEALTH	21
2.4.1 HEALTH INFORMATION SYSTEMS	22
2.4.2 ELECTRONIC HEALTH RECORDS	24
2.5 BENEFITS AND BARRIERS TO E-HEALTH IMPLIMENTATION	27
2.5.2 E-HEALTH BENEFITS	27
2.5.3 E-HEALTH BARRIERS	29
2.6 DATA IN ELECTRONIC HEALTH SYSTEMS	32
2.7 IMPORTANCE OF DATA QUALITY IN E-HEALTH	34
2.8 CONCLUSION	35
CHAPTER 3: CURRENT STATUS OF RESEARCH ON DATA QUALITY CHALLENGES	36

3.1 INTRODUCTION	38
3.2 DATA QUALITY DEFINITION	39
3.3 DATA QUALITY CHALLENGES IN GENERAL INFORMATION SYSTEMS.....	40
3.3.1 THE INFLUENCE OF QUESTIONNAIRES ON DATA QUALITY	42
3.3.2 THE INFLUENCE OF MANAGEMENT ON DATA QUALITY	43
3.4 DATA QUALITY CHALLENGES IN HEALTH INFORMATION SYSTEMS	44
3.5 GROUPING OF DATA QUALITY CHALLENGES.....	48
3.5.1 DATA QUALITY CHALLENGES AT DATA COLLECTION.....	48
3.5.2 DATA QUALITY CHALLENGES IN DATA PROCESSING.....	51
3.5.3 DATA QUALITY CHALLENGES AT DATA DELIVERY	53
3.6 SIMILAR STUDIES	54
3.7 CONCLUSION	56
CHAPTER 4: POSSIBLE E-HEALTH DATA QUALITY CHALLENGES	58
4.1 INTRODUCTION	60
4.2 IDENTIFYING POSSIBLE E-HEALTH DATA QUALITY CHALLENGES	60
4.3 POSSIBLE E-HEALTH DATA QUALITY CHALLENGES INPUT	62
4.3.1 DATA QUALITY DIMENSIONS.....	62
4.3.2 DATA QUALITY CHALLENGES	62
4.3.3 E-HEALTH BARRIERS THAT COULD IMPACT ON DATA QUALITY	63
4.4 POSSIBLE E-HEALTH DATA QUALITY CHALLENGE OUTPUT	63
4.5 CONCLUSION	70
CHAPTER 5: RESEARCH METHODOLOGY.....	71
5.1 INTRODUCTION	73
5.2 RESEARCH PROCESS	73
5.2.1 RESEARCH PHILOSOPHY.....	75
5.2.2 RESEARCH APPROACH	80
5.2.3 RESEARCH STRATEGY	84
5.2.4 RESEARCH CHOICE	92
5.2.5 TIME HORIZON.....	99
5.2.6 TECHNIQUES AND PROCEDURES	100
5.3 APPLIED RESEARCH PROCESS SUMMARY	106
5.4 ETHICAL CONSIDERATIONS	108
5.5 CONCLUSION	108
CHAPTER 6: QUALITATIVE RESEARCH FINDINGS.....	109
6.1 INTRODUCTION	111

6.2 ISSUES ADDRESSED BY THE RESEARCH INSTRUMENT.....	113
6.2.1 SEMI-STRUCTURED INTERVIEWS	113
6.3 TRANSCRIPTION, CODING AND ANALYSIS PROCESSES	113
6.4 BACKGROUND OF INTERVIEWEES.....	114
6.5 DATA ANALYSIS AND RESULTS	115
6.5.1 DATA QUALITY CHALLENGES EXPERIENCED BY DATA QUALITY EXPERTS .	117
6.5.2 ADDITIONAL ASPECTS REGARDING E-HEALTH AND DATA QUALITY.....	120
6.5.3 REVIEW OF DATA QUALITY CHALLENGES.....	124
6.6 CONCLUSION	132
CHAPTER 7: QUANTITATIVE RESEARCH FINDINGS.....	136
7.1 INTRODUCTION	138
7.2 ISSUES ADDRESSED BY THE RESEARCH INSTRUMENT.....	140
7.2.1 QUESTIONNAIRES	140
7.3 DATA COLLECTION AND ANALYSIS PROCESS.....	141
7.4 DEMOGRAPHIC INFORMATION OF THE RESEARCH PARTICIPANTS	142
7.5 DATA ANALYSIS AND RESULTS	144
7.6 CONCLUSION	168
CHAPTER 8: CONCLUSION	169
8.1 INTRODUCTION	171
8.2 SUMMARY OF CHAPTERS	172
8.3 OVERVIEW OF THE STUDY.....	173
8.3.1 RESEARCH PROCESS	173
8.3.2 REFLECTION OF KEY FINDINGS.....	175
8.4 SIGNIFICANCE AND CONTRIBUTION OF RESEARCH	179
8.5 LIMITATIONS OF RESEARCH	181
8.6 REFLECTIONS	181
8.6.1 SCIENTIFIC REFLECTION	182
8.6.2 METHODOLOGICAL REFLECTION.....	182
8.6.3 PERSONAL REFLECTION.....	183
8.7 POSSIBLE FUTURE RESEARCH	184
8.8 CONCLUDING REMARKS.....	184
REFERENCES	186
APPENDIX A: INTERVIEW CONSENT FORM	197
APPENDIX B: OPEN INTERVIEW QUESTIONS.....	199
APPENDIX C: QUESTIONNAIRE CONSENT FORM.....	201

APPENDIX D: QUESTIONNAIRE 203
APPENDIX E: ETHICAL CLEARANCE 207

LIST OF TABLES

Table 2.1: Health information (Jha *et al.*, 2009)

Table 2.2: Application areas of e-health (Adebesin, 2014)

Table 2.3: E-health benefits

Table 2.4: Barriers of e-health implementation and use

Table 2.5: Classification of healthcare data (Adebesin, 2014)

Table 3.1: Data quality dimensions (Vaziri & Mohsenzadeh, 2012)

Table 3.2: Potholes on the road to data quality. Adapted from Strong *et al.*, (1997)

Table 3.3: Data quality challenges (Beacon Community Program, 2013)

Table 3.4: Data quality challenges at data collection

Table 3.5: Data quality challenges in data processing

Table 3.6: Data quality challenges at data delivery

Table 4.1: Possible e-health data quality challenges

Table 5.1: Deductive and inductive research approaches (Saunders *et al.*, 2015)

Table 5.2: Research strategies with research designs (Ahmadnezhad, 2009)

Table 5.3: Qualitative vs quantitative research (Mamhood, 2013)

Table 5.4: Fundamental principles for conducting and evaluating interpretive studies (Klein & Myers, 1999)

Table 6.1: Questionnaire themes

Table 6.2: Background data of respondents

Table 6.3: Data quality challenges inclusion criteria

Table 6.4: Data quality challenges review analysis

Table 6.5: Enhanced list of data quality challenges

Table 7.1: Themes and objectives of questionnaire

Table 7.2: Data quality challenges with 95% confidence intervals

Table 7.3: Ordinal values of questionnaire questions

Table 7.4: Ranked data quality challenges according to weighted total scores

Table 8.1: Summary of research process

Table 8.2: Prioritised data quality challenges

Table 8.3: Data quality challenges with regard to data users

Table 8.4: Data quality challenges with regard to IT

Table 8.5: General data quality challenges

LIST OF FIGURES

Figure 1.1: Structure of Chapter 1

Figure 1.2: Phases of research

Figure 1.3: Dissertation structure

Figure 2.1: Chapter 2 – stage in research

Figure 2.2: Structure of Chapter 2

Figure 2.3: Systematic literature review (Levy & Ellis, 2006)

Figure 2.4: Current phase in research

Figure 2.5: Information objectives for e-health (Strategy for the modern NHS (1998-2005))

Figure 3.1: Chapter 3 – stage in research

Figure 3.2: Structure of Chapter 3

Figure 3.3: Current phase in research

Figure 3.4: Data process (Rodrigues & Riveill, 2010)

Figure 3.5: Data quality audit cycle (Verma, 2012)

Figure 3.6: Total data quality management (Health information and quality authority, 2011)

Figure 4.1: Chapter 4 – stage in research

Figure 4.2: Chapter 4 structure

Figure 4.3: Process towards identifying possible e-health data quality challenges

Figure 4.4: Possible e-health data quality challenges

Figure 5.1: Chapter 5 – stage in research

Figure 5.2: Structure of Chapter 5

Figure 5.3: Research phases

Figure 5.4: Research process onion (Saunders *et al.*, 2015)

Figure 5.5: Research philosophy

Figure 5.6: Positivist vs. Interpretivist (De Villiers, 2005)

Figure 5.7: Applied research philosophy adapted from Saunders *et al.* (2015)

Figure 5.8: Deductive research approach (Donnelly & Trochim, 2007)

Figure 5.9: Inductive research approach (Donnelly & Trochim, 2007)

Figure 5.10: Research approach

Figure 5.11: Inductive research process (Creswell, 2009)

Figure 5.12: Applied research approach (Saunders *et al.*, 2015)

Figure 5.13: Research strategy

Figure 5.14: QUAL-quant process (Creswell, 2009)

Figure 5.15: Expert review sample size (Ouma, 2014)

Figure 5.16: Applied research strategy (Saunders *et al.*, 2015)

Figure 5.17: Research choices (Saunders *et al.*, 2015)

Figure 5.18: Research choice

Figure 5.19: Mixed method design matrix (Johnson & Onwuegbuzie, 2004)

Figure 5.20: Applied research choice (Saunders *et al.*, 2015)

Figure 5.21: Applied time horizon (Saunders *et al.*, 2015)

Figure 5.22: Techniques and procedures

Figure 5.23: Forms of data triangulation used in the study

Figure 5.24: Applied techniques and procedures (Saunders *et al.*, 2015)

Figure 5.25: Applied research process onion (Saunders *et al.*, 2015)

Figure 6.1: Chapter 6 – stage in research

Figure 6.2: Structure of Chapter 6

Figure 6.3: Current phase in research

Figure 6.4: The hermeneutical circle (Gadamer, 1998)

Figure 6.5: Qualitative phase processes

Figure 6.6: Challenges experienced but not in initial list

Figure 6.7: E-health barriers that can impact on the data quality

Figure 6.8: Data quality challenges review

Figure 6.9: Enhanced and practise-specific list of data quality challenges

Figure 7.1: Chapter 7 – Stage in research

Figure 7.2: Structure of Chapter 7

Figure 7.3: Current phase in research

Figure 7.4: Data triangulation

Figure 7.5: Participant job title

Figure 7.6: Years of experience in health domain

Figure 7.7: Weighted total scores of significant data quality challenges

Figure 7.8: Responses – need for more training for data users

Figure 7.9: Responses – data entry errors

Figure 7.10: Responses – inaccurate data is collected from the patient

Figure 7.11: Responses – the shortage of necessary infrastructure

Figure 7.12: Responses – inconsistent definitions and formats because of different systems used

Figure 7.13: Responses – physicians are hesitant to change existing processes

Figure 7.14: Responses – data governance

Figure 7.15: Responses – information is produced by subjectivity of the data capturer

Figure 7.16: Responses – lack of appropriate software for data management

Figure 7.17: Responses – the form is too long

Figure 7.18: Responses – nurses' notes may go unread by physicians

Figure 7.19: Responses – transformation from paper records to electronic records

Figure 7.20: Responses – Privacy, confidentiality and legal challenges

Figure 7.21: Responses – omissions of data

Figure 7.22: Responses – data not meeting all relevant needs

Figure 7.23: Responses – information not suitable for analytic purposes

Figure 7.24: Responses – financial barrier to purchase necessary hardware and cost challenges in general

Figure 7.25: Responses – outdated data

Figure 8.1: Chapter 8 – stage in research

Figure 8.2: Structure of Chapter 8

Figure 8.3: Current phase of research

CHAPTER 1: INTRODUCTION

Structure of Chapter 1

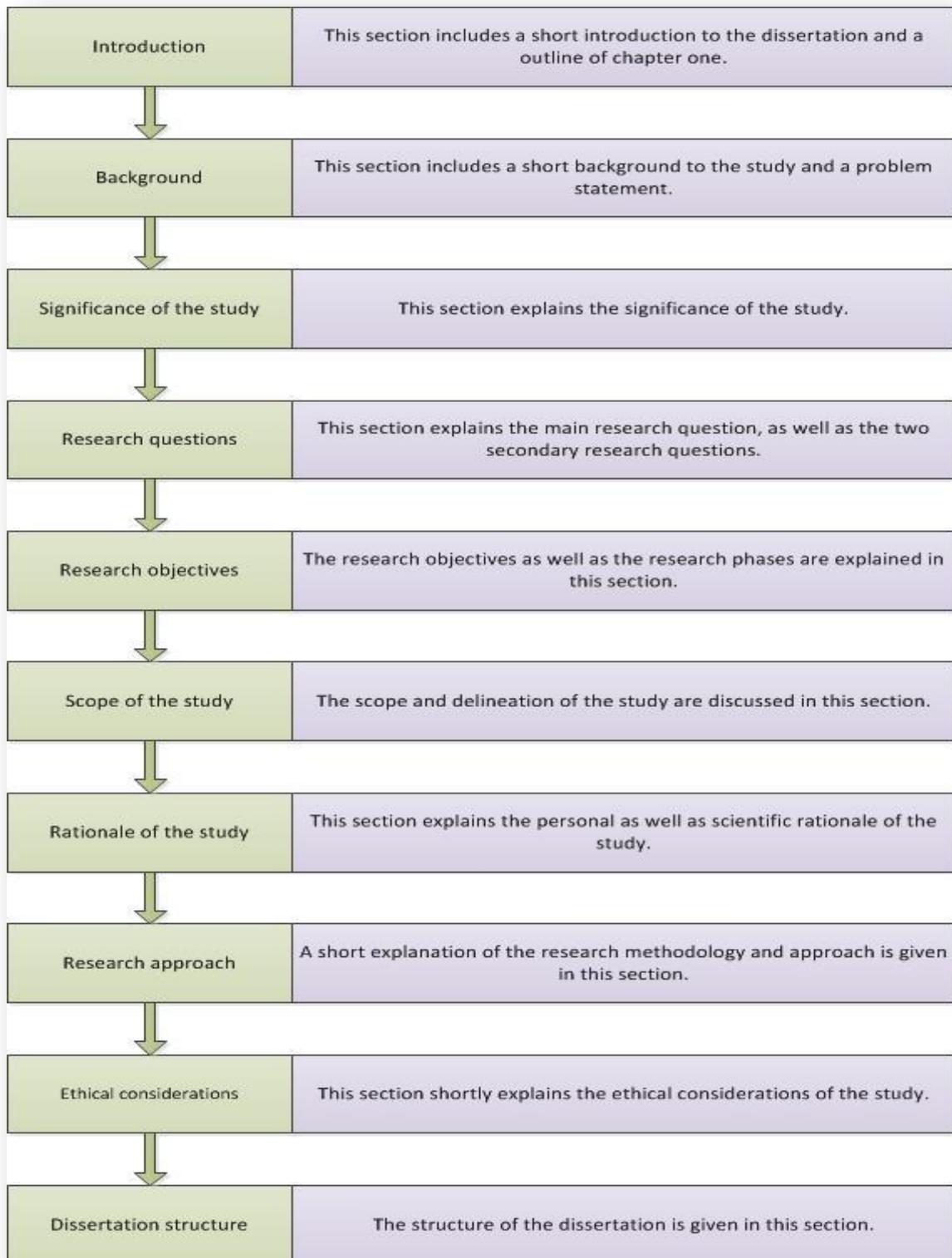


Figure 1.1: Structure of Chapter 1

1.1 INTRODUCTION

The purpose of this study is to prioritise data quality challenges experienced by data users of electronic healthcare (e-health) systems in South Africa in order to guide future interventions to improve the quality of health data. This will be done by exploring data quality challenges by means of a literature review and qualitative interviews, resulting in a description that will be prioritised through a quantitative study.

The National Department of Health adopts the World Health Organisation's definition of e-health as "the uses of information and communication technologies (ICTs) for health to, for example, treat patients, pursue research, educate students, track diseases and monitor public health" (Department of Health South Africa, 2012). E-health is mostly known as the use of various technologies to improve the status of health and quality of care for patients (Gerber, Olazabal, Brown & Pablos-Mendez, 2010).

The implementation of e-health has many benefits, including financial benefits (Francis, 2013; Iakovidis, 1998), health safety improvements (Francis 2013; Waegemann, 2002) and improved decision-making (Rodrigues, 2008; Waegemann, 2002), but it also poses many challenges. Amongst these challenges data quality proves to be a significant challenge (Francis, 2013; Accenture, 2010; Minear, 2009; Lobach & Detmer, 2007; Samyshkin & Timoshkin, 2004).

Data quality is defined as the extent to which i) data entries are aligned with actual values; ii) data is applicable or aligned with the needs of data users; iii) data is accessible; and iv) data is represented in an understandable manner (Wang & Strong, 1996). This definition of data quality serves as a standard to which data should conform to. All causes of substandard data quality can, for the purpose of the study, be regarded as data quality challenges.

There are many data quality challenges (Haug, Arlbjorn, Zachariassen & Schlichter, 2013) that may lead to insufficient data quality in various organisations, including health institutions in South Africa. These data quality challenges are explored in Chapter 3.

In this chapter the purpose and focus of the study are presented. Section 1.2 provides a background to the study and the problem statement is discussed. In Section 1.3 the significance of the study is explained. Sections 1.4 and 1.5 outline the research questions and objectives of the study. The scope and context of the study is discussed in Section 1.6. The researcher's personal rationale for the study and the scientific

rationale of the study are discussed in Section 1.7. A short introduction to the research approach used is given in Section 1.8 whilst Section 1.9 provides a description of the applicable ethical considerations. Section 1.10 provides the dissertation structure.

1.2 BACKGROUND

The background to the study, problem statement and significance of the study are discussed in this section.

1.2.1 BACKGROUND TO THE STUDY

Yakout, Elmagarmid and Neville (2010) identified that data quality is a core element for success in any organisation (including healthcare systems). Researchers and specialists agree that data quality is valuable to many real-world applications and that decisions, as well as organisational success, rely on solid data quality (Yakout, Elmagarmid & Neville, 2010). Together with the implications emerging from poor data quality in organisations, dependable healthcare relies on access to and usage of information that is correct, usable, consistent, appropriate, applicable, understandable and thorough (Health Information and Quality Authority, 2011).

The conversion from paper records to electronic data has led to expectations that data quality in healthcare systems will improve (Greiver, Barnsley, Aliarzadeh, Kreuger, Moineddin, Butt... & Kaplan, 2011). As electronic healthcare systems are systematically implemented in more facilities, it becomes more important to address data quality challenges (Chapter 3).

The *National eHealth Strategy* of the South African Department of Health (2012) outlines a comprehensive and integrated ICT health system to be implemented by 2017. While the South African Department of Health is in the process of implementing e-health in the public sector, health institutions in the private sector have been using different e-health technologies for many years (Health System Trust, 2008). South Africa faces many unique challenges such as connectivity, digital divide and monetary challenges. A set of prioritised data quality challenges experienced by data users of healthcare systems in South Africa could therefore lead to a better understanding and consideration of data quality and the importance thereof when implementing the *National eHealth Strategy*.

1.2.2 PROBLEM STATEMENT

Previous studies on data quality in health information systems focussed mainly on training for data users and data quality management interventions (Chapter 3). However, as a result of a rapid development in the implementation of e-health applications across the globe (including South Africa), there are many data quality challenges, other than training and management issues, which should be addressed (Nix & Rosenau, 2011). The importance of data quality in electronic healthcare systems has been emphasised by various authors (Bowen, 2012; Health Information and Quality Authority, 2011; Nix and Rosenau, 2011; The Connecting for Health Common Framework, 2006) and will be explored and discussed in this study.

The flow of health/patient data, used for the administering of the health facility, includes communication between the medical staff and patients. In South Africa there are nine provinces; each province is divided into several districts and each district is divided into sub-districts. The data used is collected by someone other than the data user – in most cases community health workers – thus the primary data can rather be regarded as secondary data (Wright & Odama, 2012). The data is used to calculate the value of authority, administration as well as medical and care functions for patients (Wright & Odama, 2012). Health data quality could have a significant effect on various factors, such as quality of care (Waegemann, 2002), workflow efficiency (Atreja, Gordon, Pollock, Olmsted & Brennan, 2008) and financial factors (Lobach & Detmer, 2007).

There have been various case studies with regard to possible causes of poor data quality, as well as data quality problems. These studies articulate a strong case for prioritising these challenges towards guiding future data quality interventions (Salati, Brunelli, Dahan, Rocco, Van Raemdonck & Varela, 2011; Verma 2012; Mphatswe, Mate, Bennett, Ngidi, Reddy, Barker & Rollins, 2012; Greiver *et al.*, 2011). Yakout (2010) states the importance of understanding data quality challenges in administering a national e-health system and this study builds on that understanding by exploring data quality challenges experienced by data users of electronic healthcare systems in South Africa.

1.3 SIGNIFICANCE OF THE STUDY

With the imminent large-scale conversion of paper records to electronic records it is envisaged that adequate data and information will be produced for the management,

monitoring and evaluation of the national health system's performance (Department of Health South Africa, 2012).

Greiver *et al.* (2011) identified problems with health data which include factors such as misspelled words, absent meta-data and data in unfit database fields. Substandard data quality can impact the quality of care in health facilities, lead to privacy and security issues, increase cost, create risks and decrease the benefits from using health information systems (The Markle Foundation, 2006). These data quality challenges, amongst various others explored in Chapter 3, should be addressed to administer data quality in e-health systems. The importance of data quality in health information systems is described in more detail in Section 2.6.

For the purpose of this study *data users* refer to any individual who uses or works with the healthcare data in the execution of their duties within the South African healthcare system. This may include nurses, doctors, community health workers, data entry clerks, data analysts, administration staff and other healthcare professionals.

1.4 RESEARCH QUESTIONS

From the background and problem statement one main and two secondary research questions are derived in order to conceptualise the final prioritised list of data quality challenges within the South African healthcare system. The two secondary research questions are used as supplementary questions to guide the study towards the answering of the main research question.

1.4.1 MAIN RESEARCH QUESTION

What are the prioritised data quality challenges in electronic healthcare systems in South Africa?

The objective of the main research question is to explore what data quality challenges are experienced by data users of healthcare systems in South Africa. The main research question is supported by two secondary research questions to guide the prioritisation of these data quality challenges.

1.4.2 SECONDARY RESEARCH QUESTIONS

Two secondary research questions are provided:

1.4.2.1 SECONDARY RESEARCH QUESTION 1

What is the current status of research on data quality challenges?

The objective of the first secondary question is to outline what is already known with regard to data quality challenges. Data quality challenges experienced globally and locally will be considered in the literature study. This will be done by means of a complete literature review, focussing on data quality challenges in healthcare systems.

The purpose is to find known data quality challenges from academic literature in order to explore challenges experienced in healthcare systems in South Africa.

1.4.2.2 SECONDARY RESEARCH QUESTION 2

What data quality challenges are experienced by data users of electronic healthcare systems in South Africa?

The objective of the second secondary question is to explore challenges experienced by data users in healthcare systems in South Africa. The purpose is to determine which experienced challenges from the literature (identified in secondary research question 1) are relevant in practise and within the South African healthcare system by means of a qualitative analysis to provide a description that will be prioritised through a quantitative study.

1.5 RESEARCH OBJECTIVES

To articulate the data quality challenges relevant to the South African health sector, this study follows a mixed-method approach, including qualitative as well as quantitative research. An exploratory mixed method QUAL-quan strategy is followed to answer the research questions. The objectives of the study are to:

1. define data quality as articulated by a literature review;
2. identify current data quality challenges by means of a literature review; as well as the compilation of an initial list of possible data quality challenges in e-health from the literature review to serve as a basis for the qualitative research in objective 3;
3. verify and expand data quality challenges by means of qualitative interviews with data quality experts;
4. conduct a quantitative study to prioritise the data quality challenges identified in

objectives 2 and 3;

5. prioritise the data quality challenges experienced by data users of healthcare systems in South Africa;
6. provide evidence-based recommendations with regard to the priority of data quality challenges in order to support healthcare system users with data usage based on a synthesis from the qualitative and quantitative phases of the research.

The objectives of the study (divided into four phases) are presented in Figure 1.2:

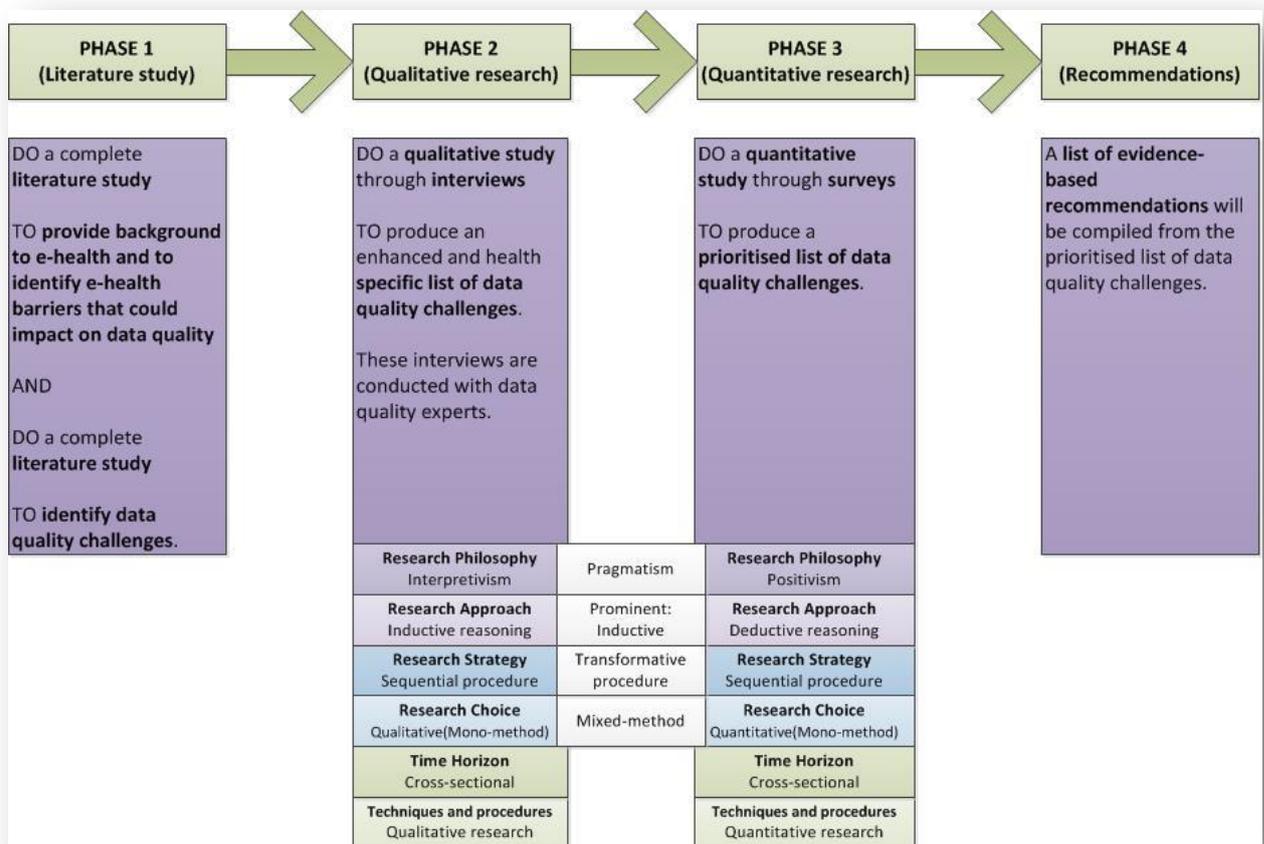


Figure 1.2: Phases of research

The research objectives outlined in Section 1.5 are illustrated in Figure 1.2. The research approach followed in the different phases of the study is also exemplified in Figure 1.2 and described in more detail in Section 1.8.

The scope and context of the study are described in Section 1.6.

1.6 SCOPE AND CONTEXT OF THE STUDY

The study will be limited to include only data users who make use of at least one electronic healthcare system in the execution of their daily medical tasks. The scope and delineation of the study are expounded on in 1.6.1 and 1.6.2 respectively.

1.6.1 SCOPE

This study is an exploratory short-term study and is not intended to be a longitudinal investigation on data quality challenges in healthcare systems. The purpose of the study is to explore data quality challenges experienced by data users of e-health systems in South Africa. Only data quality experts and data users of e-health systems in South Africa are targeted to participate in this study. This study does not include algorithms to measure data quality in these healthcare systems and is not a data quality intervention, but serves as a foundation for future data quality interventions. The priority of data quality interventions could be based on the prioritised list of data quality challenges.

1.6.2 DELINEATION

This research is conducted in various healthcare institutions in South Africa and includes feedback from doctors, pharmacists, dentists and pathologists. Only healthcare institutions that make use of at least one electronic healthcare system have been included in the study.

The study includes data users mainly from the private health sector. The researcher purposefully chose to investigate data quality challenges experienced in private health information systems as the public health sector of South Africa is not yet fully digitalised. The *National eHealth Strategy of South Africa* (2012) states that public e-health in South Africa is currently at Stage 3 of development (Department of Health South Africa, 2012). Stage 3 includes the migration of traditional district health information systems to electronic storage and reporting. Stage 4 and Stage 5 are still to be implemented in South Africa. Stage 4 includes the introduction of operational ICT systems as a source of data for the health information system (HIS) and Stage 5 is a comprehensive and integrated national HIS (Department of Health South Africa, 2012). As it develops, data quality challenges experienced in private health information systems can be identified and used as a guideline in the public sector (Department of Health South Africa, 2012).

1.7 RATIONALE OF RESEARCH

The research is motivated by both personal and scientific rationale. The rationale of the research is explained in 1.7.1 and 1.7.2.

1.7.1 PERSONAL RATIONALE

The researcher is an employee at the Meraka Institute, Council of Scientific and Industrial Research (CSIR), which is engaged in various projects regarding the computerisation of health records in South Africa. Data quality plays a significant part in these projects.

While working on several of these projects, the researcher became interested in data quality and ways to enhance the quality of data in healthcare systems. The researcher's work at the CSIR led to this study that prioritises data quality challenges experienced by data users of electronic healthcare systems in South Africa.

1.7.2 SCIENTIFIC RATIONALE

In the *National eHealth Strategy of South Africa*, the Department of Health of South Africa (2012) explains that data quality will remain substandard where systems are mainly paper-based or a combination of paper and computerised systems. In the midst of computerising health records, data quality becomes a significant part of the *National eHealth Strategy*.

A prioritised list of data quality challenges can be used as a guideline for data quality interventions to ensure an e-health system of sufficient quality in South Africa. The evidence-based recommendations provided in the study could advise data users about effective use of data to promote data quality in electronic healthcare systems.

1.8 RESEARCH APPROACH

This study follows a sequential exploratory mixed method approach towards answering the research questions. This implies that a qualitative design is followed by a quantitative design and analysis as described in Figure 1.2. Oates (2008) mentions that it is not unusual to combine qualitative and quantitative research methods. The data collection method for this study consist of a literature review (Chapter 2 and Chapter 3), which is used to compile an initial list of possible data quality challenges in e-health (Chapter 4). The initial list forms the basis for qualitative data collection (in the second phase) which is done by conducting interviews. The purpose of the interviews is to

verify and expand the data quality challenges identified in the literature. A quantitative study follows the qualitative study to prioritise the data quality challenges. The quantitative data is collected by means of questionnaires. The research methodology is explained in more detail in Chapter 5.

1.9 ETHICAL CONSIDERATIONS

All participants involved in the research were treated with respect and integrity as stipulated in the University of South Africa's (UNISA) *Research Ethics Policy*. The proposal, informed consent form and data collection instruments were presented to and approved by the College of Science, Engineering and Technology's (CSET) Research Ethics Committee (CREC).

No vulnerable community or data users were involved and the researcher guaranteed confidentiality. Each research participant had to sign an informed consent form.

More detailed information on ethical considerations is given in Section 5.4. The approval from CSET and CREC is attached as Appendix E.

1.10 DISSERTATION STRUCTURE

A graphical representation of the dissertation structure is presented in Figure 1.3. This representation indicates each chapter's title with a short description of its contents:

- *Chapter 1* is an introductory chapter that provides background to the study. This chapter explains the rationale, significance and objectives of the study.
- *Chapter 2* provides background to the research domain of the study, namely e-health. E-health in South Africa, its benefits, barriers and the importance of data quality within e-health systems are discussed. E-health barriers that could impact on data quality are identified.
- *Chapter 3* addresses the first secondary research question: *What is the current status of research on data quality challenges?* This chapter is an in-depth literature review on data quality challenges in general as well as in health information systems.
- *Chapter 4* provides an initial list of possible data quality challenges in e-health. This compiled initial list emerged from the literature review and is used to guide the interviews and questionnaires.

- *Chapter 5* addresses the research methodology followed. A background of various approaches is described shortly and the applied research approach is explained.
- *Chapter 6* partially addresses the second secondary question: *What data quality challenges are experienced by data users of electronic healthcare systems in South Africa?* This chapter explains the research findings that emerged from the qualitative data collection, namely the interviews with data quality experts.
- *Chapter 7* partially addresses the second secondary question: *What data quality challenges are experienced by data users of electronic healthcare systems in South Africa?* This chapter explains the research findings that emerged from the quantitative data collection, namely questionnaires. In this chapter the ranking or prioritisation of the data quality challenges is calculated.
- *Chapter 8* presents a conclusion of the entire study. The results and findings are summarised and discussed shortly. Recommendations for future studies are given in this chapter.

The structure of the dissertation, as well as the phase in which the research questions are addressed, is presented in Figure 1.3:

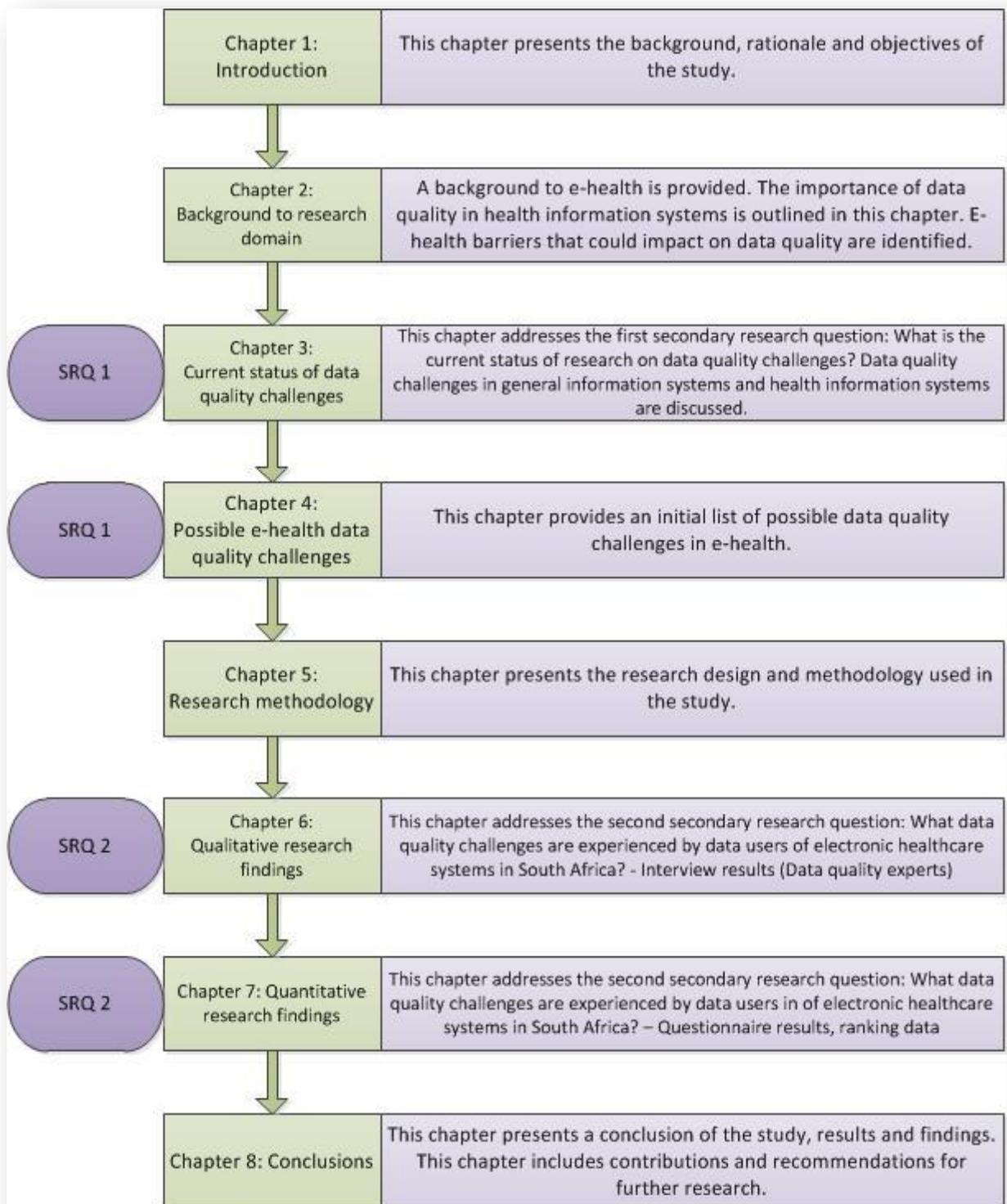


Figure 1.3: Dissertation structure

CHAPTER 2: BACKGROUND TO THE RESEARCH DOMAIN – E-HEALTH

Stage in research:

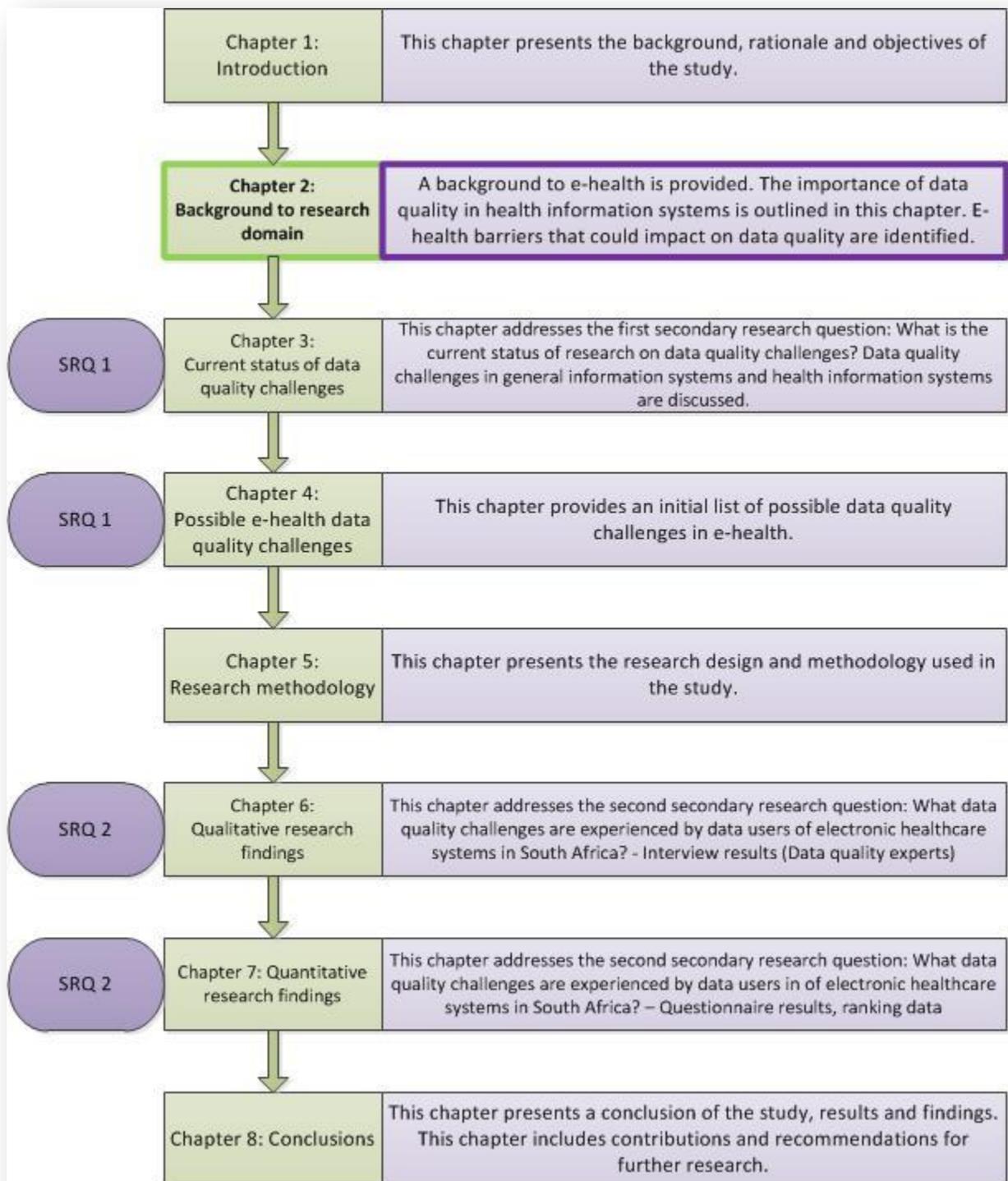


Figure 2.1: Chapter 2 – stage in research

Structure of Chapter 2:

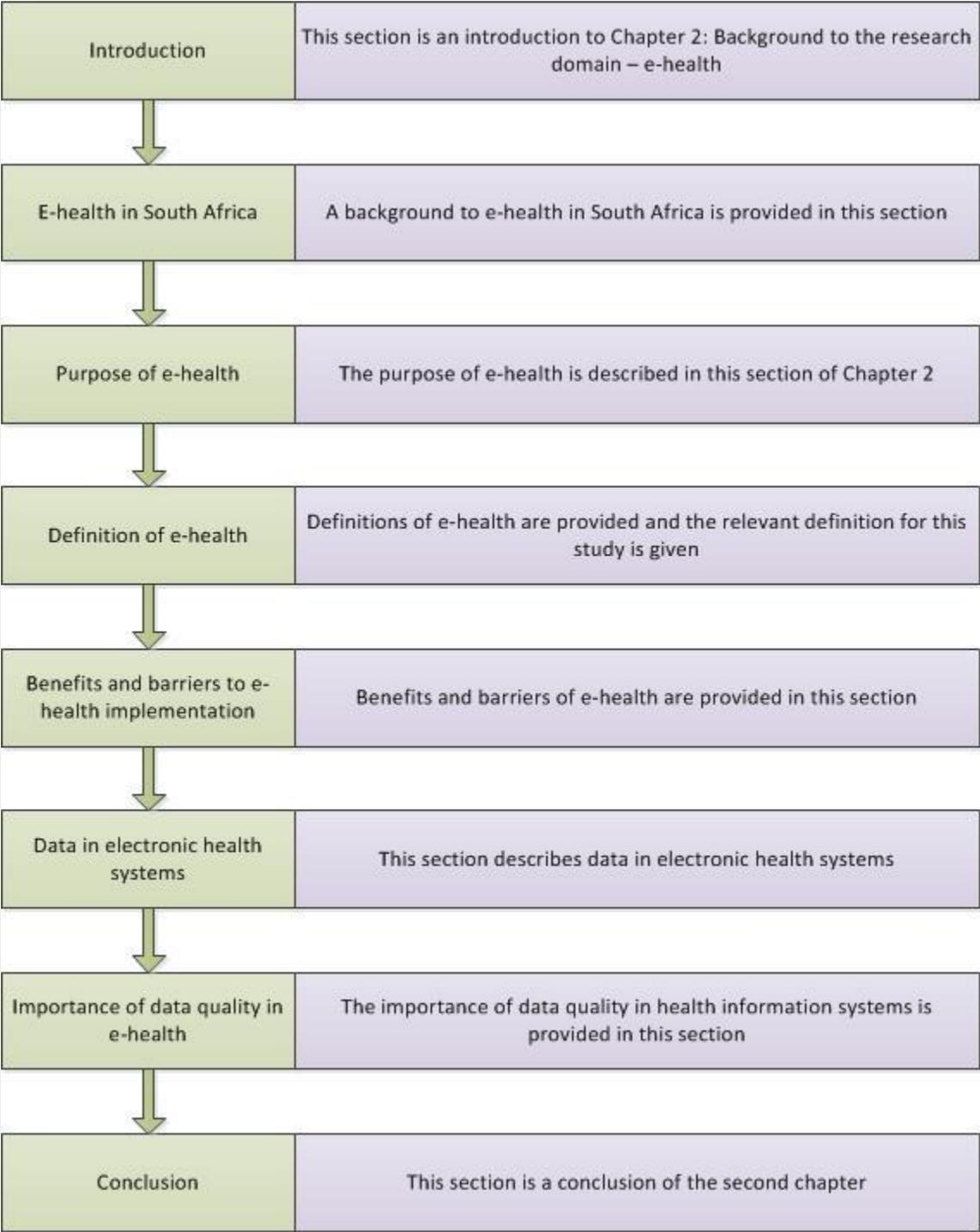


Figure 2.2: Structure of Chapter 2

2.1 INTRODUCTION

Chapter 1 presented an introduction and background to the study. The problem statement, as well as the research questions, was given in Chapter 1.

The purpose of this chapter is to contextualise the study within health and e-health in South Africa, to describe the importance of data quality in the e-health domain and to identify e-health barriers that could impact on data quality. This chapter serves as background to the domain of this study.

Kitchenham, Brereton, Budgen, Turner, Bailey and Linkman (2007) state that the most important reasons for conducting a systematic literature review in a research study are to:

- summarise existing research on technologies in order to find evidence of benefits and limitations of methods;
- find gaps in existing research in order to identify new areas for research; and
- compile a background or an initial list of possible data quality challenges in e-health to guide research questions and activities.

A process for conducting a systematic literature review is given by Levy and Ellis (2006):

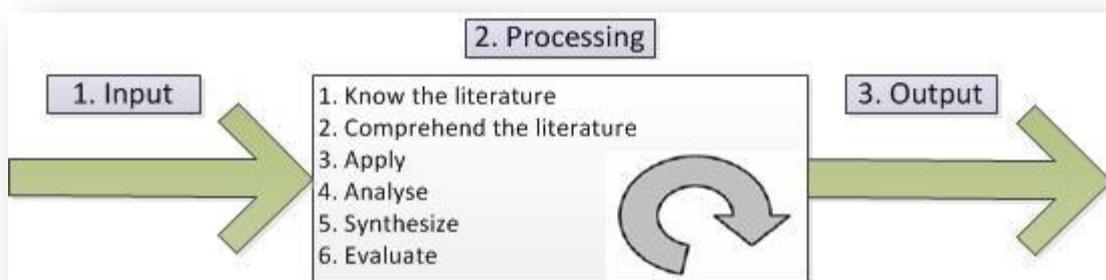


Figure 2.3: Systematic literature review (Levy & Ellis, 2006)

- *Input*: The input of a systematic literature review refers to the research in the literature review. The literature used should be relevant to the study.
 - *Know the literature*: The researcher should show that he/she has extracted relevant information from the articles that has been read.

- *Comprehend the literature:* The researcher should show that he/she understands the importance and meaning of the research.
 - *Apply:* To apply literature, the research should identify important concepts and theories and group them accordingly.
 - *Analyse:* The researcher should identify and explain why the literature presented is important.
 - *Synthesise:* The researcher should present the information as a whole. The literature should be assembled by concepts.
 - *Evaluate:* The researcher should differentiate between opinions, theories and facts
- *Output:* The output refers to the final, complete literature review.

Following the process presented by Levy and Ellis (2006), a systematic literature review on e-health in South Africa is conducted in Chapter 2 in order to provide background to the research domain and to identify e-health barriers that could affect data quality. Data quality challenges are explored in Chapter 3. The literature study was considered adequate when saturation was reached. Glaser and Strauss (1967, p65) defined saturation as the point at which “no additional data are being found whereby the [researcher] can develop properties of the category. As he sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated.”

Phase 1 of the research is partially presented in Chapter 2. Figure 2.4 illustrates the current phase in the research:

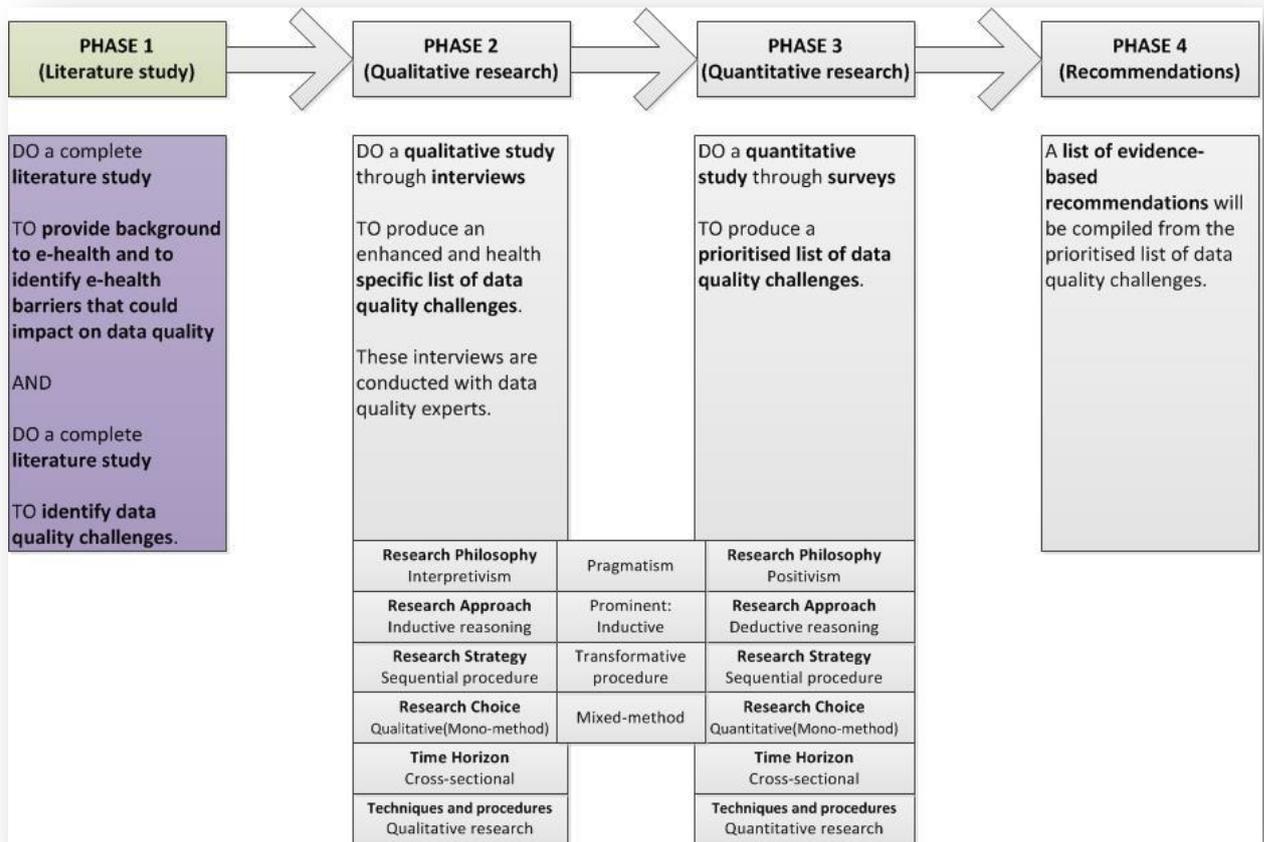


Figure 2.4: Current phase in research

To fully comprehend the need for data quality assurance in electronic healthcare systems in South Africa, it is important to understand electronic healthcare as the term has become broadly used (Soellner, Huber & Reder, 2014). This chapter consists of background on e-health in South Africa, the purpose of e-health, the definition of e-health, the benefits and barriers of e-health implementation and finally, the importance of data quality in e-health.

2.2 E-HEALTH IN SOUTH AFRICA

E-health is generally known to be the application of various technologies to improve the health status and quality of care for patients and has been globally implemented in cities and rural areas (Gerber *et al.*, 2010). This section describes the status of e-health in South Africa.

Since 1994 South Africa has made many improvements in the institution and application of a health information system (Health System Trust, 2011). Although many improvements have been made in this regard, the South African National Department of

Health is still facing many challenges (Department of Health South Africa, 2012). The *National eHealth Strategy of South Africa* (2012) outlines ten strategic priorities for the successful application of e-health to be implemented by 2017. The following is a list of the strategic priorities to be addressed to support the healthcare transformation in South Africa:

- strategy and leadership;
- stakeholder engagement;
- standards and interoperability;
- governance and regulation;
- investment, affordability and sustainability;
- benefits realisation;
- capacity and workforce;
- e-health foundations;
- applications and tools to ensure healthcare delivery; and
- monitoring and evaluation of the e-health strategy.

The Department of Health aims to realise their mission by applying these strategies (Department of Health South Africa, 2012). Their mission is *“to establish eHealth as an integral part of the transformation and improvement of healthcare services in South Africa, especially enabling delivery on the health sector’s Negotiated Service Delivery Agreement 2010-2014.”* (Department of Health, South Africa, 2012, p8).

Although there has been a concerted effort to update systems and data quality has improved during the process, many challenges still remain (Health System Trust, 2011). South Africa is in the process of implementing a national health information system and therefore encounters numerous challenges. The benefits and barriers of e-health are further discussed in Section 2.5.

The following section outlines the purpose of e-health.

2.3 PURPOSE OF E-HEALTH

Various countries are doing research on the implementation of e-health. These countries include Canada, Australia, USA, England, South Africa and Finland amongst

many others (Hayrinen, Saranto & Nykanen, 2008). The purpose of implementing e-health is discussed in this section.

Although the *e* in e-health generally refers to *electronic* (Schiavo, 2013; Fry & Neff, 2009), Eysenbach (2001) provides ten e-components that serve as a definition of e-health:

- *Efficiency* is one of the primary objectives of e-health. Healthcare should become more efficient through the use of e-health and therefore the cost of health should reduce.
- *Enhancing quality* refers to the enhancement of the quality of care. By improving communication and interoperability in healthcare systems and patient information the quality of care should improve.
- *Evidence-based* entails that the effectiveness of e-health should never be assumed, but rather be scientifically proven before engaging in an e-health intervention.
- *Empowerment* of both the consumers and patients. By making patient records accessible and enhancing the availability of medical information, patient-centred medicine is enhanced. This empowers patients to make informed decisions.
- *Encouragement* refers to encouragement of an improved relationship between patients and health workers. This is established by the easy sharing of information and transparency of patient data and medical information.
- *Education* of health professionals should be improved by making use of internet sources. The health education of patients and consumers should also be enhanced or their knowledge should be broadened through information on the internet. This can be regarded as preventative health information.
- *Enabling* communication and a standardised information exchange between different healthcare facilities is a key outcome of e-health.
- *Extending* e-health beyond the original borders of health. This refers to physical as well as metaphorical boundaries. With the use of communications technologies, it becomes possible to extend healthcare in terms of location.

- *Ethics*: in the midst of possibilities regarding information exchange, communication technologies and health information systems, it remains important to adhere to privacy and equity laws and ethical considerations.
- *Equity* refers to the equal access and use of e-health. The digital divide, especially in developing countries, remains a challenge in the successful implementation of e-health.

The information objectives of e-health, as given by the *Information Strategy for the Modern NHS (1998-2005)*, are illustrated in Figure 2.5:

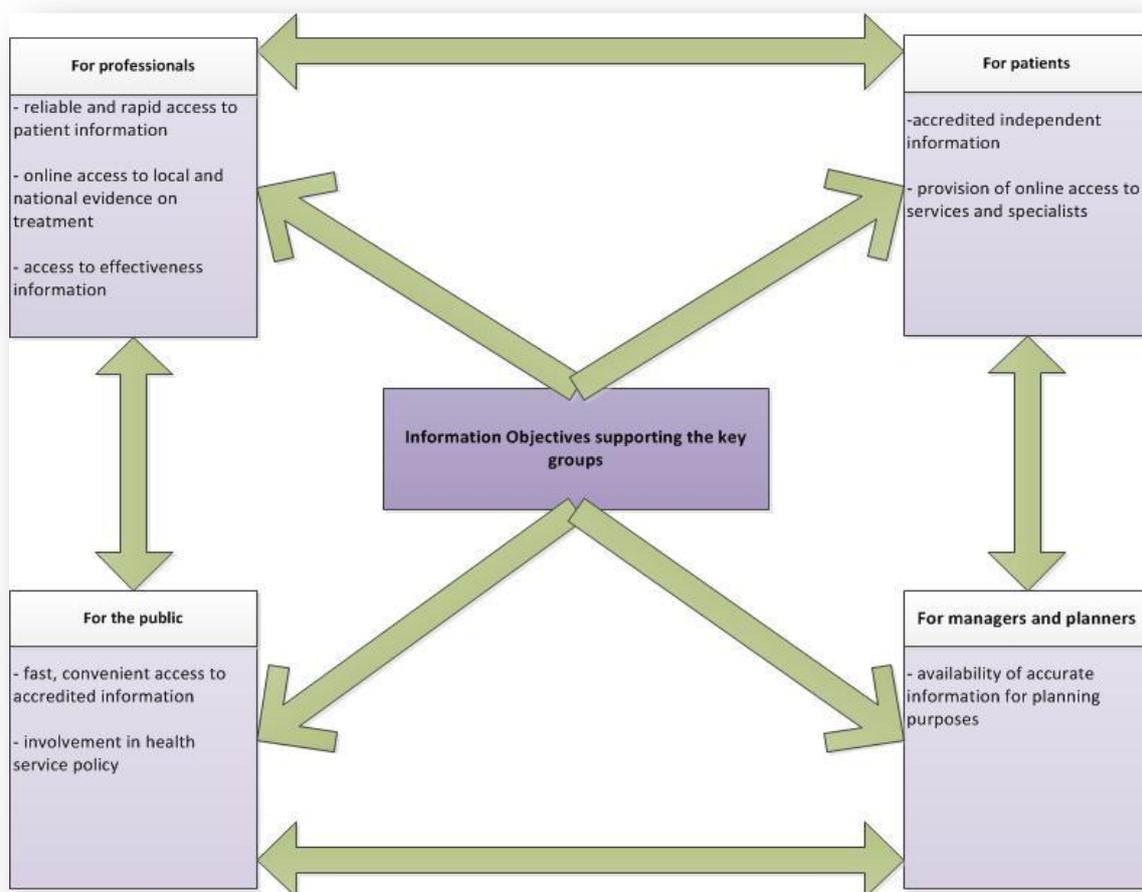


Figure 2.5: Information objectives for e-health (Strategy for the modern NHS (1998-2005))

The figure illustrates that professionals, patients, management and the public should benefit from health information in different ways.

Regardless of whether a system is electronic or paper-based, the data will not be useful if it is of substandard data quality (World Health Organisation, 2003). According to a guideline provided by the World Health Organisation (2003) health data should:

- determine the continuing and future care of a patient at all levels of healthcare;
- serve medical legal purposes for the patient, doctor and healthcare service;
- maintain accurate and reliable information about diseases treated and surgical procedures performed in a hospital and within a community, as well as immunization and screening programmes including the number and type of participants;
- help with clinical and health service research and outcomes of healthcare intervention, if required;
- provide accurate, reliable and complete statistical information about the uses of healthcare service within a community;
- teach healthcare professionals; and
- determine staffing requirements and plan healthcare services.

In this section, the researcher explains various reasons for implementing e-health. The purpose of e-health is essentially to enhance all aspects of healthcare and the communication thereof. Various countries, including South Africa should benefit from the implementation of e-health.

The definition of e-health for this study is formulated in Section 2.4.

2.4 DEFINITION OF E-HEALTH

The term *e-health* may include various applications of healthcare, internet resources, electronic patient information, data analysis tools, communication between health professionals as well as communication to patients, electronic health devices and administrative data (Ouma, 2008).

E-health can be defined as a conversion to electronic health records to benefit from internet use and health agreements (Khalifehsoltani & Gerami, 2010) in order to develop sufficient communication between healthcare professionals (De Clercq, 2008).

Nuq and Aubert (2013) define e-health as the use of information in electronic format and communication technologies in the health domain. The term *e-health* is thus a broad term that refers to everything that applies to the combination of computing or electronic devices and healthcare or medicine (Kwankam, 2004; Eysenbach, 2001).

More specifically, in the hospital, home and primary care settings, e-health could include (Ouma, 2008):

- electronic patient administration tools and technologies, laboratory information systems, electronic communication systems in the hospital setting;
- telephone consults, diabetes and asthma monitoring systems in the home setting; and
- patient records, electronic prescribing, medical records and patient management in the primary care setting.

For the purpose of this study Nuq and Aubert's (2013) definition will be applied: "*E-health is the use of information in electronic format and communication technologies in the health domain.*"

Data users of at least one electronic healthcare system in South Africa were included in the sample of this study. The data users are from different settings, such as hospitals, doctors' offices, pharmacies and pathologies. The data users included in this study do not make use of only one or the same electronic healthcare system, therefore the chosen definition of e-health is applicable to the study.

E-health is administered by health information systems (HIS), which are described in Section 2.4.1. Electronic health records (described in Section 2.4.2) are recorded in these health information systems.

2.4.1 HEALTH INFORMATION SYSTEMS

Canada, Australia, USA, England, South Africa and Finland (amongst many others) have the same prospects for their national health information system (Hayrinen *et al.*, 2008), namely to:

- include patients in the use of their personal health information records;
- define the essential information in these health records;
- standardise the system in terms of codes, languages and vocabularies to ease interoperability to provide open EHR systems; and
- implement data security, in terms of the system's infrastructure as well as security policies.

A guideline for the content of a successful health information system includes (Jha, DesRoches, Campbell, Donelan, Sowmaya, Ferris... & Blumenthal, 2009):

Table 2.1: Health information (Jha *et al.*, 2009)

Clinical documentation	Test and imaging results	Computerised provider-order entry	Decision support
Demographic characteristics of patients	Laboratory reports	Laboratory tests	Clinical guidelines
Physicians' notes	Radiologic reports	Radiologic tests	Clinical reminders
Nursing assessments	Radiologic images	Medications	Drug-allergy alerts
Problem lists	Diagnostic-test results	Consultation requests	Drug-drug interaction alerts
Medication lists	Diagnostic-test images	Nursing orders	Drug-laboratory interaction alerts
Discharge summaries	Consultant reports		Drug-dose support
Advanced directives			

From the content of a health information system depicted in Table 2.1, the use of data in health information systems is outlined. Data in electronic health systems is discussed in more detail in Section 2.6.

There are various health information systems currently implemented in health facilities in South Africa. The *National Health Normative Standards Framework for Interoperability in eHealth in South Africa* (2014) lists all the health information systems used in healthcare facilities in South Africa. The health information systems currently implemented each have various functionalities. Currently different healthcare facilities make use of different health information systems. *The National Health Normative Standards Framework for Interoperability in eHealth in South Africa* (2014) aims to standardise these systems so that all healthcare facilities in South Africa can use the same health information system in the future.

The *National eHealth Strategy of South Africa* states that a comprehensive and integrated ICT health system should be implemented by 2017 (Department of Health South Africa, 2012).

For the purpose of this study, the researcher is aware that various health information systems are currently implemented in healthcare facilities in South Africa. The data quality challenges will therefore be prioritised in general across various health information systems.

The definition and uses of electronic health records are described in Section 2.4.2.

2.4.2 ELECTRONIC HEALTH RECORDS

Electronic health records have been researched, tried, implemented and de-implemented since the 1990s. Studies now show that a fully interoperable, working e-health system could still be seven years away (Francis, 2013). The process of fully implementing electronic health records is continuous by nature.

This section defines electronic health records and discusses some of the uses of electronic health records in healthcare facilities.

2.4.2.1 DEFINITION OF ELECTRONIC HEALTH RECORDS

Electronic health records (EHRs) can formally be defined as a database or warehouse containing digital patient data which is stored in a secure manner, can be exchanged and is available to various authorised users (Hayrinen, *et al.*, 2008). Another definition of electronic health records is: an individual's healthcare data of his entire lifetime, with the aim of continuity of care, supporting teaching and research and the sharing of this information with confidentiality and security ensured (Iakovidis, 1998).

The definition of EHR has not changed in the past decade (Francis, 2013) and as seen in the mentioned definitions, an electronic health record is a patient's health information as a whole, stored in electronic format.

Along with the formal definitions of EHR, the content of a typical electronic health record serves as part of the definition and can be very specific.

2.4.2.2 USES OF ELECTRONIC HEALTH RECORDS

Electronic health records are used to facilitate continuity of care in healthcare facilities (Calman, Hauser, Lurio, Wu & Pichardo, 2012; Jha *et al.*, 2009; Atreja *et al.*, 2008). Continuity of care means that patient care should be continuous in all the phases of healthcare. These phases are (in chronological order) as follows (Iakovidis, 1998):

- Preventative care.
- Diagnosis of the patient's current condition.
- Treatment of the current condition.
- Rehabilitation.

Patient records are mainly used to record patients' treatment and to ease communication with the patient but patient/health records are now used to serve as

evidence in law cases, for general research and mainly to get an overview of the patient’s entire health profile and history that includes (Penoyer, Kendall, Noblin, Bullard, Talbert, Wilson... & Briscoe, 2014):

- the patient’s condition;
- care given;
- measurements taken;
- medication administered;
- human body systems; and
- common patient problems.

The patient’s health profile and history can guide clinicians to guarantee continuity and stability of care. They mainly make use of the patient profile and history to (Penoyer *et al.*, 2014):

- understand the patient’s overall condition or health status;
- make clinical choices; and
- communicate with other health workers.

Aside from the uses of electronic health records depicted here, there are several application areas of e-health. The application areas of e-health and their descriptions are presented in Table 2.2 adapted from Adebessin (2014):

Table 2.2: Application areas of e-health (Adebessin, 2014)

E- health application area	Description
Electronic medical record	Refers to a patient’s electronic record of health-related information. Supports the capturing, storing and sharing of a patient’s information between healthcare providers within an organisation.
Electronic health record	Refers to a longitudinal record of a patient with information from more than one event over a period of time and across different healthcare settings.
Personal health record	An electronic record of a patient, containing information generated through one or more encounter from one healthcare setting over a period of time. The personal health record may be generated by the patient or a

E- health application area	Description
	medical professional. The personal health record is a patient-centric record.
Computerised provider order entry system	A computerised provider order entry system that allows physicians to place an order for medication, radiology or laboratory tests electronically.
Electronic prescribing and pharmacy information system	Supports inventory management, medication administration, and patient dispensing history and could in some cases provide decision support. The electronic prescribing and pharmacy information system is usually implemented in pharmacies.
Clinical decision support system	Such systems provide medical physicians with relevant patient information at the right time. It is an intelligent system that can enhance decision support and proficient healthcare delivery.
Chronic disease management system	Chronic disease management systems manage and track the health status and test results of chronic disease patients in order to improve and support the management of healthcare for patients with chronic diseases.
Radiology information system	This is a management system, typically implemented in radiology departments, to support the workflow and business process of a hospital's radiology department.
Laboratory information system	This management system is department specific and supports the management of a hospital's laboratory department. It supports the recording, organising and archiving of laboratory test results.
Mobile health	Mobile health (m-health) refers to the use of mobile communication devices to support the delivery of healthcare
Telemedicine	Telemedicine refers to the provision of healthcare regardless of geographical location. Telemedicine includes communication between physicians, specialists and patients.
Practice, patient and clinical management system	Supports the capturing, storing and exchange of a patient's health information during the course of their care. Diagnostics management, appointment

E- health application area	Description
	scheduling, the generation of reports and resource management are also supported by this management system.
Healthcare information system for planning	Practice, patient and clinical management systems support collection, aggregation and analysis of health data from various sources in order to report on a country's health, health trends and burden of diseases.

The main uses of the electronic health records discussed in this subsection, underline the significance of the usage of electronic health records. In order to provide continuity of care and base clinical choices on electronic health records, one could argue that the quality of the electronic health records should be up to standard. The uses outlined in the table are an important part of healthcare and therefore it is clear from the narrative that *the quality of electronic health records is an important factor to consider when implementing e-health*. The importance of data quality in health information systems is described in more detail in Section 2.7.

The benefits and challenges of e-health are described in the following section.

2.5 BENEFITS AND BARRIERS TO E-HEALTH IMPLIMENTATION

The purpose of this section is to explore the benefits and barriers to e-health implementation. Separate lists of both e-health benefits and identified barriers are compiled and presented.

2.5.2 E-HEALTH BENEFITS

Table 2.5 provides a list of e-health benefits identified in literature. These benefits are given together with the sources from which they were extracted.

Table 2.3: E-health benefits

E-health benefits	Sources
Cost savings, financial benefits in general	Francis (2013); Iakovidis (1998); NHS Executive (1998)
Health safety improvements	Francis (2013); Atreja <i>et al.</i> (2008); Waegemann (2002); Iakovidis (1998)
Improvement of efficiency and effectiveness of healthcare	Kern, Edwards and Kaushal (2014); Calman <i>et al.</i> (2012); Jha

E-health benefits	Sources
	<i>et al.</i> (2009); Atreja <i>et al.</i> (2008); Waegemann (2002)
Improved decision-making	Penoyer <i>et al.</i> (2014); Atreja <i>et al.</i> (2008); Rodrigues (2008); Waegemann (2002); Iakovidis (1998)
Access to physicians remotely	Fontenot (2013); Lobach & Detmer (2007); Waegemann (2002)
Reduce medical errors	Rodrigues (2008); Lobach & Detmer (2007); Waegemann (2002)
Sharing of information	Waegemann (2002); Iakovidis (1998);
Medical science and research	Fontenot (2013); Lobach & Detmer (2007)
Workflow efficiency	Atreja <i>et al.</i> (2008); Waegemann (2002);
Employee satisfaction	Rodrigues (2008); Waegemann (2002);
Patient satisfaction	Rodrigues (2008); Waegemann (2002);
Reduces paperwork	Kern <i>et al.</i> (2014); Atreja <i>et al.</i> (2008);
Better data for management purposes	Iakovidis (1998)
Quality assurance for forecasting	Iakovidis (1998)
Comparative effectiveness	Fontenot (2013)
Improved diagnosis process	Iakovidis (1998)
Improved communication	Calman <i>et al.</i> (2012)
Standardisation of healthcare	Lobach & Detmer (2007)
Safety regarding drug dispensing	Lobach & Detmer (2007)
Enhancement of self-managing chronic diseases	Lobach & Detmer (2007)
Management improvements	Atreja <i>et al.</i> (2008)
Prevention support	Atreja <i>et al.</i> (2008)

Table 2.4 presents 22 e-health benefits identified from literature. E-health holds many promises for healthcare but there remain many barriers to successful implementation. These barriers regarding e-health implementation and use are tabulated in Section 2.5.3.

2.5.3 E-HEALTH BARRIERS

Table 2.6 provides a list of e-health implementation and use barriers. The table consists of e-health barriers that have been identified from literature as well as their relevant literature sources.

Table 2.4: Barriers of e-health implementation and use

E-health barriers	Sources
The financial barrier to purchase necessary hardware and cost challenges in general	Theobald (2014); Jaroslawski & Saberwal (2014); Jha <i>et al.</i> (2009); Minear (2009); Atreja <i>et al.</i> (2008); Anderson (2007); RTI International (2001); Iakovidis (1998)
Lack of IT and clinical resources	Theobald (2014); Jaroslawski & Saberwal (2014); Jha <i>et al.</i> (2009); Anderson (2007)
Difficulty learning and using the software	Theobald (2014); Anderson (2007); Braa, Monteiro and Sahay (2004); Ammenwerth, Graber, Herrmann, Burkle and Konig (2003)
Personnel costs	Jaroslawski & Saberwal (2014); Atreja <i>et al.</i> (2008); Iakovidis (1998)
Standardisation of all health information systems, since the content and structure of all health information systems should be standardised	Jaroslawski & Saberwal (2014); Hayrinen <i>et al.</i> (2008); Atreja <i>et al.</i> (2008); Rodrigues (2008); Ammenwerth <i>et al.</i> (2003)
It might be time consuming to update the EHR thoroughly	Penoyer <i>et al.</i> (2014); Wentzel, van Velsen, van Limburg, de Jong, Karreman, Hendrix and van Gemert-Pijnen (2014);

E-health barriers	Sources
	Anderson (2007)
The implementation of health information systems in small and rural facilities, and not only big hospitals, remains a challenge	
Data privacy	Accenture (2010); Minear (2009); Rodrigues (2008); Anderson (2007); Samyshkin & Timoshkin (2004); Waegemann (2002); RTI International (2001)
Interoperability	Accenture (2010); Minear (2009) Atreja <i>et al.</i> (2008); Westra <i>et al.</i> (2008); Anderson (2007); Waegemann (2002)
Sustainability	Jaroslowski & Saberwal (2014); Minear (2009); Braa <i>et al.</i> (2004)
Data quality	Francis (2013); Accenture (2010); Minear (2009); Atreja <i>et al.</i> (2008) Lobach & Detmer (2007); Samyshkin & Timoshkin (2004);
Usability	Car, Black, anandan, Cresswell, Pagliari, McKinstry, Procter, Majeed and Skeih (2008); Lobach & Detmer (2007); Anderson (2007); Ammenwerth <i>et al.</i> (2003); RTI International (2001)
Transferring data from paper to electronic records	Atreja <i>et al.</i> (2008); Braa <i>et al.</i> (2004); Waegemann (2002);
Physicians are hesitant to change existing processes	Jaroslowski & Saberwal (2014); Jha <i>et al.</i> (2009);
Forming electronic health records as part of the facilities daily routine	Car <i>et al.</i> (2008); Braa <i>et al.</i> (2004);
Meeting needs at each provider level, reaching goals	Heeks (2006); Samyshkin &

E-health barriers	Sources
	Timoshkin (2004);
Data access	Accenture (2010); Bowling, Rimer, Lyons, Golin, Frydman and Ribisl (2006)
Government	Jaroslawski & Saberwal (2014); Rodrigues (2008)
Logistics	Jaroslawski & Saberwal (2014), Rodrigues (2008)
The shortage or absence of the necessary infrastructure, such as internet connections	Theobald (2014); Anderson (2007)
The patient's medical history is not always available	Theobald (2014)
Although it saves time, the drop-down menus in health information systems may lack detailed information.	Penoyer <i>et al.</i> (2014)
Nurses' notes may go unread by physicians	Penoyer <i>et al.</i> (2014)
Scalability	Braa <i>et al.</i> (2004)
Modernizing existing systems	Samyshkin & Timoshkin (2004)
Extracting knowledge out of information	Westra <i>et al.</i> (2008)
Software	Minear (2009)
Patient consent	Accenture (2010)
Donors	Jaroslawski & Saberwal (2014)
Fragmentation programmes	Jaroslawski & Saberwal (2014)
Lack of appropriate software	Theobald (2014)

Table 2.6 presents 31 barriers to the implementation and use of e-health. Financial barriers such as purchasing of hardware, the lack of IT and clinical resources, difficulty of learning using software, personnel costs, standardisation of health information systems, time challenges, implementations in rural areas, data privacy, interoperability, sustainability, data quality, usability and the transfer from paper records to electronic records are frequently mentioned.

Data quality is frequently sighted as a significant barrier to the implementation and meaningful use of e-health (Francis, 2013; Accenture, 2010; Minear, 2009; Atreja *et al.*, 2008; Lobach & Detmer, 2007; Samyshkin & Timoshkin, 2004).

A background to data in electronic health systems is provided in Section 2.6 and the importance of data quality in e-health is discussed in Section 2.7.

2.6 DATA IN ELECTRONIC HEALTH SYSTEMS

The field of healthcare is very data intensive and involves the exchange of large amounts of data, such as diagnostic images, patient records, administrative data and test results forms an essential part of healthcare (Adebesin, 2014). A classification of healthcare data is presented in Table 2.7. The role and importance of data in the healthcare domain supports the purpose of the study by stressing the importance of data quality.

Table 2.5: Classification of healthcare data (Adebesin, 2014)

Healthcare information category	Description
Patient-specific data	This is information that could be linked to a specific patient by using a unique identifier, such as a hospital registration number. Examples of patient-specific data include diagnosis, medical history and progress notes.
Aggregated data	Aggregated data is a summary derived from patient-specific, administrative and personnel data. Examples of aggregated data include disease indices, statistical reports and trend analyses.
Comparative data	This data is generally used for performance assessments against an internal or external standard. One example of comparative data is patient waiting time in the facility compared to the average waiting time in the region.
Knowledge-based information	Knowledge-based information is compiled with facts and models, to support process, problem solving and decision support. This type of information can be found in literature, such as medical journals.

Medical records can also be divided into administrative and clinical data:

Administrative data mainly supports organisational processes and include (Reynolds & Bowman, 2010):

- *Demographic data* serves to identify a patient and to link a patient to a specific medical record. Demographic data typically include the patient's name, surname, address, date of birth and phone number.

- *Data for billing and reimbursement* consists of the necessary information to bill a patient. Examples of data include medical aid name and number, diagnosis and procedure codes.

Clinical data relate to patient care and include (Reynolds & Bowman, 2010):

- *Allergies* that list all substances to which a patient is known to be allergic to.
- *Care plan* that indicates instructions to the healthcare team about the patient's treatments and medications.
- *Consents and authorisations* represent the patient's consent or agreement to treatments.
- *Discharge summary* summarises a patient's visit or stay at a healthcare facility. This includes the reason for admission, treatments and guidelines for following care.
- *Imaging and radiology reports* represent report results from radiology. This includes X-rays, ultrasounds and scans.
- *Immunisation record* documents all vaccinations received by the patient.
- *Laboratory reports* record all blood tests.
- *Medical history* of a patient includes previous surgeries, illnesses and other interventions regarding the patients or their relatives.
- *Physical examination* stores a patient's physical condition based on inspection of the patient's body.
- *Progress notes* documents a patient's progress after treatment and admission and includes a reflection of changes in the patient's condition.
- *Referral notes* are used to refer a patient from a general practitioner to a specialist.
- *Referral reports* document the specialist's opinion regarding the patient.
- *Treatments and interventions* document all medical and surgical procedures of the patient.

As e-health is an information intensive arena, with many different uses for data, the importance of data quality becomes of paramount concern. The importance of data quality in e-health is discussed in Section 2.7.

2.7 IMPORTANCE OF DATA QUALITY IN E-HEALTH

Data quality is frequently sighted as a concern in literature. As the purpose of the study is to prioritise data quality challenges experienced by data users of electronic healthcare systems in South Africa, the researcher aims to state the importance of data quality in e-health.

Nix and Rosenau (2011) state that data quality challenges in electronic health records are still in the process of being understood. They acknowledge that time and patience is required to ensure that health information systems are functioning at full potential while meeting the needs of healthcare providers. They argue that accurate and complete data is compulsory when making use of electronic health records in healthcare systems. Conceding that the use of health information systems does not guarantee that data quality will improve, they acknowledge that quality of health records remains a significant challenge (Nix & Rosenau, 2011). The data collected and stored in health information systems is used for a variety of health related functions and the quality of the data determines the extent to which actions are reliable (Bowen, 2012).

Not only does data quality have an impact on effectiveness and advantage in the business sector (Redman, 1995), but it can also have significant effects on the health sector. The *Connecting for Health Common Framework* (2006) states that substandard data quality could:

- impact the quality of care in health facilities;
- lead to privacy and security issues;
- increase costs;
- create risks; and
- decrease the benefits from using health information systems.

The World Health Organisation (2003) states that healthcare data is imperative for the current and future care of a patient, irrespective of the level of care at which the health service is provided. Data must be accurate, complete, reliable, comprehensible and accessible to users if it is to meet the requirements of the patient, doctor and other

health professionals, the healthcare facility, legal authorities, as well as state, province and national government health authorities (World Health Organisation 2003). Data quality is a significant factor in the usability as well as the sustainability of e-health (Ure *et al.*, 2006).

The collection of data with substandard data quality leads to inappropriate information for health and management purposes (Samyshkin & Timoshkin, 2004); therefore the quality of that data is essential, not only for use in patient care, but also for monitoring the performance of the health service and employees (World Health Organisation, 2003).

E-health barriers that can impact on the quality of data are identified in section 2.5.3 and will be presented to experts to validate their impact. This process and outcome is documented in Chapter 6.

2.8 CONCLUSION

The purpose of this chapter was to contextualise health and e-health in South Africa and to describe the importance of data quality in the e-health domain. The chapter served as background to the domain of this study.

The chapter outlined the current status of e-health in South Africa. South Africa is one of many countries doing research on the implementation of a national health information system. The implementation of such a system is however only in the early phases.

The initiative of implementing e-health was described, followed by a definition of e-health relevant to this study. E-health for the purpose of this study is regarded as the use of information in electronic format and communication technologies in the health domain. No specific health information system was selected for the study and generic e-health barriers, across various health information systems, were explored.

There are many e-health benefits and barriers to implementation and use. One of the most frequently sighted e-health barriers is *data quality*. The importance of data quality in e-health was articulated in this chapter and various e-health barriers that could impact on data quality was identified.

The following chapter explores the current status of research on data quality challenges and aims to identified data quality challenges from literature in general and e-health specifically.

CHAPTER 3: CURRENT STATUS OF RESEARCH ON DATA QUALITY CHALLENGES

Stage in research:

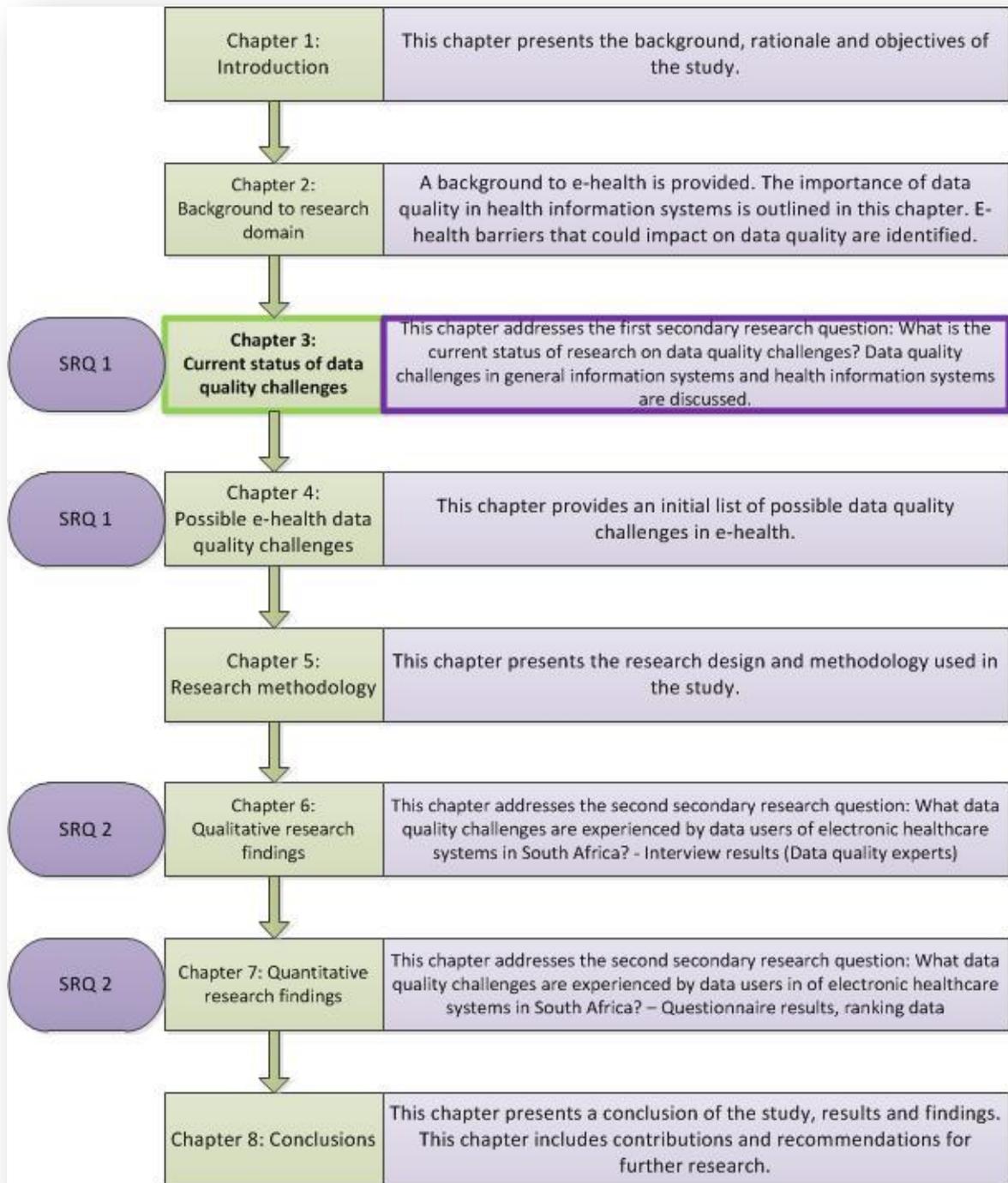


Figure 3.1: Chapter 3 – stage in research

Structure of Chapter 3:

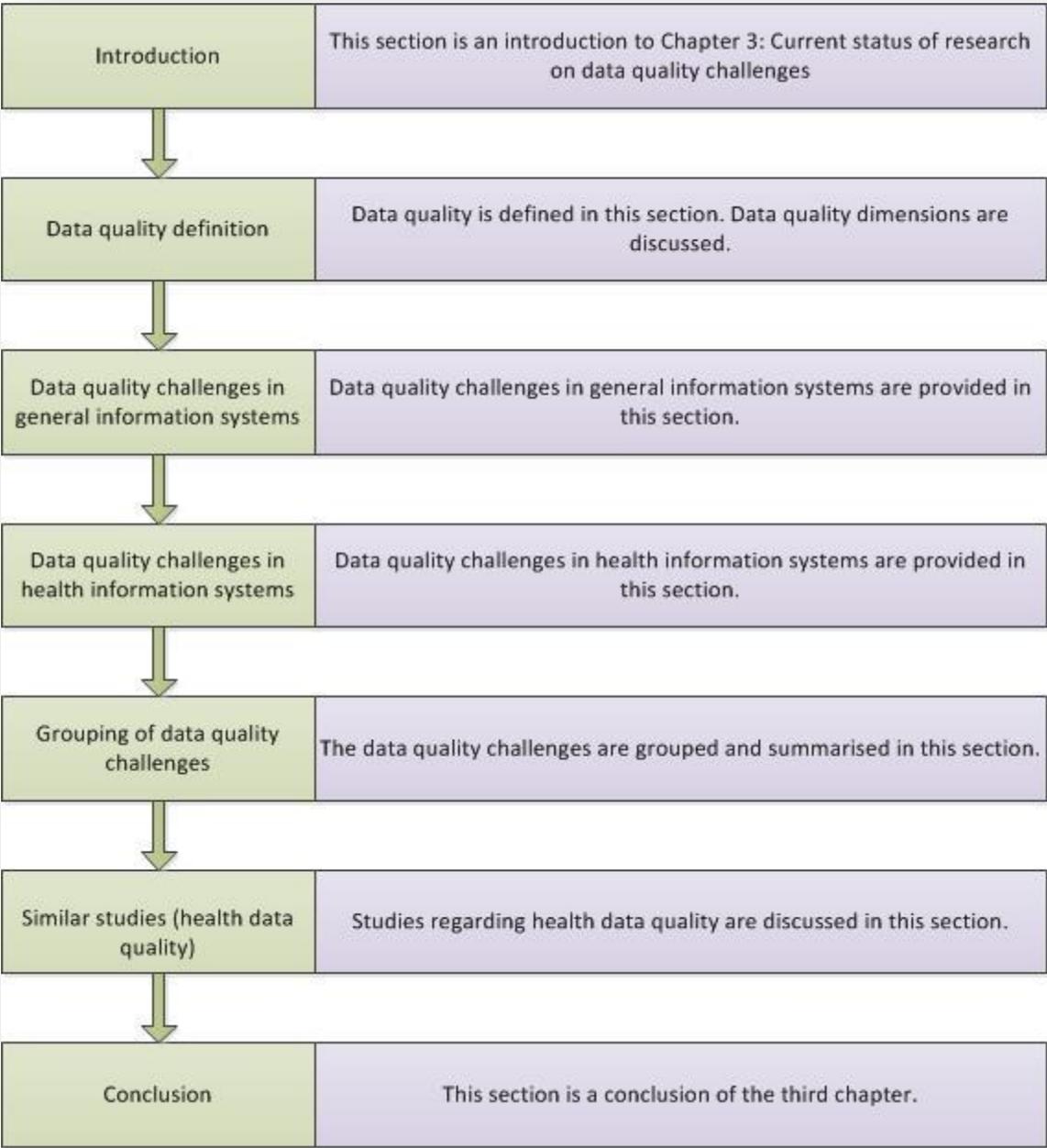


Figure 3.2: Structure of Chapter 3

3.1 INTRODUCTION

Chapter 2 contextualised e-health in South Africa and highlighted some e-health barriers that could impact on data quality. Chapter 3 explores the current status of research on data quality challenges and aims to identify data quality challenges from literature in general and e-health specifically.

A problem statement, as well as the research questions, was given in Chapter 1.

The first secondary research question is: *What is the current status of research on data quality challenges?*

This chapter aims to answer the first secondary research question towards identifying additional possible data quality challenges that could impact on e-health to guide the qualitative and quantitative phases of the research.

Phase 1 of the research is partially presented in Chapter 3. Figure 3.3 illustrates the current phase in the research:

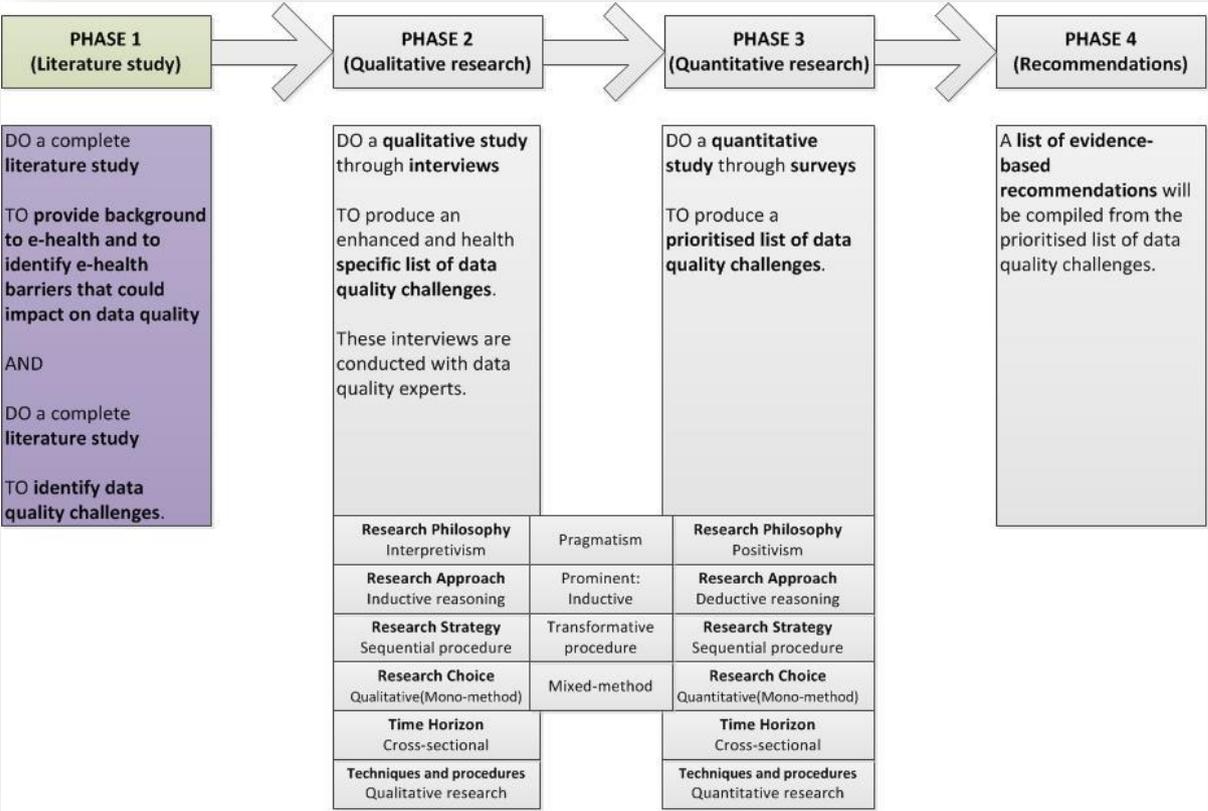


Figure 3.3: Current phase in research

Chapter 3 is a comprehensive literature review to identify data quality challenges. The chapter presents a definition of data quality for this study; challenges to data quality in both general and in health information systems; a summary of known data quality challenges; and similar studies with regard to data quality in health information systems.

A definition of data quality is provided in Section 3.2.

3.2 DATA QUALITY DEFINITION

The definition of data quality serves as a standard of what data should adhere to. In order to identify data quality challenges, a standard of data quality should be set. The researcher explores the definition of data quality in this section.

The research done in most cases tends to use information technologists' opinions of what data quality actually means (Wang & Strong, 1996). Wang and Strong (1996) proposed to define what data quality means to data users. The aim of their research was to identify aspects of data quality that is important from the consumer's point of view. Their research provided four categories for twenty dimensions of data. The four categories are (Wang, 1998; Strong, Yang & Wang, 1997; Wang & Strong, 1996):

- *Accuracy of data*: the extent to which the data entries are aligned with actual values.
- *Relevancy of data*: the extent to which data is applicable or aligned with the needs of the data user.
- *Representation of data*: the extent to which the data is represented in an understandable way.
- *Accessibility of data*: the extent to which data is accessible.

Vaziri and Mohsenzadeh (2012) displayed Wang's data quality dimensions in a descriptive table:

Table 3.1: Data quality dimensions (Vaziri & Mohsenzadeh, 2012)

Dimension	Explanatory question
Ease of access	Is the data easy to retrieve?
Suitable amount	Is the volume of data suitable for the specific task?
Believability	Is the data plausible?
Completeness	Is all or most of the data available and in enough depth?
Brief representation	Is the data compact?
Constant representation	Does the format of the data stay consistent?

Dimension	Explanatory question
Ease of manipulation	Is it easy to change the data to another format?
Free of mistakes	Is the data accurate and trustworthy?
Interpretability	Are the data definitions, symbols and languages comprehensible?
Impartiality	Is the data unbiased?
Relevancy	Is the data applicable or aligned with the needs of the data user?
Status	Is the data highly regarded in terms of its basis or source?
Timeliness	Is the data up to date?
Understandability	Is the data easy to comprehend?
Value added	Does the data provide return in terms of use?
Security	Is the data controlled correctly?

In 2010 Khan added *believability* to the four categories mentioned (Khan, Singh & Dey, 2010). This additional category can be seen as part of accuracy. Believability is an important attribute; the real or actual values aren't always available and data quality should be measured by only the data one currently has (Khan *et al.*, 2010).

The data quality dimensions could differ in each case, because the need and use for the data varies (Vaziri & Mohsenzadeh, 2012). Data quality in electronic health systems should adhere to a specific set of data quality dimensions, including completeness, correctness, concordance, comprehensibility, value added, consistency in capture, consistency in form, currency/timeliness, believability and security (Beacon Community Program, 2013; Bowen, 2012; Carter, 2012; Rodriguez & Riveill, 2010).

Data should conform to the dimensions of data discussed in this section. The data quality dimensions serves as a standard for data. The purpose of the study is to prioritise data quality challenges experienced by data users of electronic healthcare systems in South Africa. Data quality challenges refer to any hurdle or difficulties which could lead to data not conforming to the data quality dimensions discussed in this section.

The researcher explores data quality challenges in general information systems in Section 3.3.

3.3 DATA QUALITY CHALLENGES IN GENERAL INFORMATION SYSTEMS

The previous section provides a description of what is implied by data quality. It

focuses on the definition of data quality, which serves as a standard for data quality. This section's goal is to describe data quality challenges in general information systems, as stated by previous research in this field.

Since information technology has developed rapidly over the past few years, the use of data in all sectors became more crucial (Haug *et al.*, 2013). Haug *et al.* (2013) supports this argument by stressing the importance for companies or organisations to understand what the reason for poor data quality is in order to enhance the quality of data. The most significant data quality challenges according to Haug *et al.* (2013) are:

- absent assignment of responsibilities regarding data;
- ambiguity of roles with regard to data construction, use and maintenance;
- ineffective organisational procedures;
- lack of adequate administration focus with regard to data quality;
- lack of data quality assessments;
- lack of adequate rewards in terms of data quality;
- lack of adequate data training for data consumers;
- a need for written data quality policies and procedures;
- managers' lack of thorough emphasis on the importance of data quality;
- lack of appropriate software for data management;
- lack of adequate input options in data quality systems; and
- unfortunate usability of information technology systems.

According to Haug (2013) 67% of these data quality challenges are caused or influenced by human error. Singh and Singh (2010) propose causes of poor data quality, where the cause lies at the data source. Through literature reviews 52 causes of data quality at source were identified (Singh & Singh, 2010). These causes include factors such as: absent columns, insufficient candidate data sources, insufficient familiarity of inter-data dependencies, altering timeliness of data sources, unforeseen alterations in source systems, numerous sources for the same data, misspelled data, extra columns, conflicting use of special characters, diverse data types for the same columns, data values drifting from their field descriptions and business rules, unsuitable data relationships and an occurrence of outliers.

Literature suggests that the questionnaires used for data collection could have a

significant effect on the quality of data. Various factors with regard to questionnaires could have an effect on data quality. Section 3.3.1 explains the influence of questionnaires on data quality.

3.3.1 THE INFLUENCE OF QUESTIONNAIRES ON DATA QUALITY

Groves (1987) states that the mode of the questionnaire can contribute to the quality of data. Although questionnaires are used by professionals in many fields, questionnaires are not recognised as a field of study on its own. It is important to recognise the different role players in questionnaires – there are data collectors, whose main purpose is to collect the data and there are analysts, who are interested in analysing the results of the questionnaire (Groves, 1987). Groves (1987) also states that it is important to choose an applicable questionnaire for its specific use which will lead to improvement of data quality. The purpose of this study is to evaluate how the different role players in surveys (data collectors, analysts, describers and modellers) move towards different error sources. It is found that these differences cause researches to focus on different aspects of the questionnaire (Groves, 1987).

Based on prior research there are three significant aspects of questionnaires that might have an impact on data quality:

- *Structure of the questions:* A person-level questionnaire focuses on individuals, for instance: *Is John sick?* A household-level questionnaire asks general questions about the entire household. An example of this would be: *Is anyone in the household sick?* (Hess, Moore, Pascale, Rothgeb & Keeley, 2001). By using a quantitative methodology on both types of questionnaires, it is found that the person-level questionnaire might enhance the completeness of the data, but that the trustworthiness might reduce. Household-level questionnaires tend to have a lower completeness (Hess *et al.*, 2001).
- *Length of the questionnaire:* The length of a questionnaire might also influence the completeness and quality of data. It is found that shorter questionnaires are usually more complete, but that people are willing to answer a longer questionnaire when the questions are clear and understandable (Subar, Ziegler, Thompson, Johnson, Weissfeld, Reding, Kavounis & Hayes 2001).
- *Data collection modes:* One of the main tasks with regard to survey research is to decide on a data collection method. It is found that using a mixed method for

data collection is very efficient and that it might produce the most complete outcome (De Leeuw, 2005).

Three aspects of the data collection questionnaire could have an effect on the quality of data. The subsection emphasised the importance of questionnaires.

Another data quality challenge emphasised in literature is data quality management. The influence of management is described in Section 3.3.2.

3.3.2 THE INFLUENCE OF MANAGEMENT ON DATA QUALITY

Research on ways to improve data quality has been conducted by examining management strategies (Soares, 2015; Eckerson, 2014). This section of the literature review will include a few methodologies, guidelines or strategies to improve data quality.

Substandard data quality may have a major effect on the finances of corporations (Lacas, 2010). Therefore, there a lot of research is done on the development of methodologies to improve or measure data quality.

The *Total Data Quality Management* process is a cycle used to assess and improve data quality. It consists mainly of four steps, which is shown at a later stage. The four steps in *Total Data Quality Management* are (Health Information and Quality Authority, 2011):

- *Define*: This step entails scoping of the problem, identification of goals, identification and review of documentation and development quality measures.
- *Measure*: In this step business rules should be applied and suspect data should be flagged.
- *Analyse*: This step entails the identification of conformance problems, providing of references, highlighting of problems and validation of problems.
- *Improve*: This includes the development of opportunities, implementation of improvements, documentation of improvements and updating of data standards.

Like these data quality improvement guidelines, other methodologies and guidelines to improve data quality exist.

Section 3.3 described data quality challenges in general information systems. These challenges are experienced in general information systems but are applicable to health information systems. The data quality barriers identified in Section 3.3 will be included as possible data quality challenges in e-health (Chapter 4) to be verified by data quality

experts.

Although data quality challenges can be generic across any information system, specific research has been conducted on data quality challenges in health information systems. The researcher explores data quality challenges in health information systems in Section 3.4.

3.4 DATA QUALITY CHALLENGES IN HEALTH INFORMATION SYSTEMS

Since data quality remains a significant challenge in electronic health records (Francis, 2014; Accenture, 2010; Lobach & Detmer, 2007) it is valuable to identify data quality challenges in this regard. Various authors have done research on difficulties of data quality in electronic health records (Wentzel *et al.*, 2014; Beacon Community Program, 2013; Horbatuk, 2011; Nix & Rosenau, 2011; Braa & Hedberg, 2002; Rodriguez & Riveill, 2010; Minear, 2009; Ure *et al.*, 2006; Samyshkin & Timoshkin, 2004; Strong, Lee & Wang, 1997). This research will be discussed in this section.

Strong, Lee and Wang (1997) provide ten what they call *potholes* on the road to data quality. The term *potholes* refer to challenges regarding data quality and Table 3.2 shows which data quality dimension is impacted by each challenge.

Table 3.2: Potholes on the road to data quality. Adapted from Strong *et al.*, (1997)

Data quality challenge	Impacted data quality dimension	Example in health domain
More than one source of data, each producing different values.	Consistency, believability	A hospital makes use of more than one information system to guide decision support.
Information is produced by subjectivity	Objectivity, believability	Medical coders (data entry clerks) use their own judgement when entering disease codes.
Lost information because of systematic errors	Correctness, completeness, relevancy	Correct data is rejected because of faulty error checks and incorrect data gets accepted because of a faulty error check.

Data quality challenge	Impacted data quality dimension	Example in health domain
Difficulty in accessing data because of large amounts of stored data	Concise representation, timeliness, value-added, accessibility	Runtime on database queries is long because of a large amount of health records in the database.
Inconsistent definitions and formats because of different systems used	Concise representation, timeliness, value-added	Different systems are used in different departments of the health facility, all using different standards.
Indexing problems	Concise representation, timeliness, accessibility	Medical images may be difficult to access.
Automated cross-system analysis issues (interoperability)	Relevance, value-added, consistent representation	There might be issues in querying information across health information systems.
Useful and relevant information changes due to organisational changes.	Relevance, value-added, completeness	Health insurance policies may change, effecting current and previous health information.
Privacy and confidentiality issues	Security, accessibility, value-added	Although patient information should be kept confidential, researchers and analysts need access to the data.
Lack of sufficient computer resources	Accessibility, value-added, completeness	Unreliable communication lines, such as connectivity, may lead to incomplete data.

From the data quality challenges identified by Strong *et al.* (1997) it can be deduced that the causes of poor data quality are many and varied. The challenge is to understand the cause of the problem in order to improve the quality of data.

The Beacon Community Program (2013) shows how the use of electronic health records could administer the improvement of healthcare. This community states the importance of high quality data to ground the successful implementation of electronic health records. Data quality challenges, specifically with regard to electronic health records, are identified by the Beacon Community Program (2013):

Table 3.3: Data quality challenges (Beacon Community Program, 2013)

Data quality challenge	Description
Unavailable queries (database-level, table-level, item-level)	A certain database or table may not be included in the reporting tool. This leads to loss of value, since all relevant queries cannot be made.
Inconsistency across data elements	Systems in different departments of the health facility are inconsistent with a patient's health information.
Timeframe restrictions	A query may for example ask for information in the last 12 months, but the system only allows for queries in the current calendar year.
Data segmentation	It may not be possible to group data into certain clusters, since the correct data is not stored.
Tracking completed tasks	The system may lack the function to indicate whether a task is completed or not.
Information not stored	The system does not provide for certain relevant information to be stored.
Data stored in multiple places	The system might have more than one section for one assessment, but queries only extracts data from one source.

Horbatuk (2011) provides data quality challenges that may occur in health information systems:

- Data quality challenges at data capture:
 - A certain data element is not captured.
 - A certain data element is captured, but not in the required format.
 - A certain data element is in the correct format for one system or part of a system, but not for another.
 - A certain data element is captured differently in different parts of the electronic health system.
 - Transformation from paper-based records.
- Technical challenges:
 - Uncertainties in legal aspects of medical records.

- Interfaced systems may display data incorrectly.
- Process challenges:
 - Balancing patient care with data capture and electronic health record tasks.
 - Information needed for required fields.
 - A lack of necessary resources.

In addition to these data quality challenges, Rodriguez and Riveill (2010) identified three significant data quality challenges. The importance of improving data quality in electronic health records should be clear since important medical decisions are made by making use of the data in health information systems (Rodriguez & Riveill, 2010). The three main causes of substandard data quality from their study are:

- management of heterogeneous data (for example; medical images, numerical data, free text);
- management of big amounts of data; and
- human error such as wrong input, accidental deletions and uncertain and bias inputs.

According to Nix and Rosenau (2011) human error could be divided into two types:

- *Active error*: a person entered the wrong value.
- *Passive error*: the default value in the system was not corrected.

Along with the data quality challenges in electronic health systems described here, there are many more identified data quality challenges in e-health. These causes include data collection only meeting management needs (collection of data that is not usable for specific needs); limitation of the amount of data that is collected; information that is not usable for analytic purposes; lack of incentives/encouragement to collect high quality data; inaccurate data at collection; data not structured in a manner that supports the relevant clinical tasks; information that is difficult to access; dated data (lack of timeliness); information that is not always available (incomplete data); lack of adequate training for data users; incorrect data; incorrectly formatted data; data in the wrong field; data in the wrong sequence; inconsistent data within a single file; inconsistent data for the same patient but on different visits; omissions of data; contradictions between data ambiguity of data captured; authenticity of data elements; incomplete data because it is

not located in the logical location or not clustered correctly; the same data in multiple locations may differ from one another; captured data may not include all relevant elements; and uncertainty regarding data elements could lead to bias or subjectivity (Wentzel *et al.*, 2014; Nix & Rosenau, 2011; Braa & Hedberg, 2002; Minear, 2009; Ure *et al.*, 2006; Samyshkin & Timoshkin, 2004).

Section 3.4 described data quality challenges in health information systems. These are data quality challenges experienced specifically in e-health. The data quality challenges identified in Section 3.4 will be included in the initial list of possible data quality challenges in e-health (Chapter 4) to be verified by data quality experts.

3.5 GROUPING OF DATA QUALITY CHALLENGES

Rodriguez and Riveill (2010) suggest that data quality challenges could emerge in any of the three data stages:

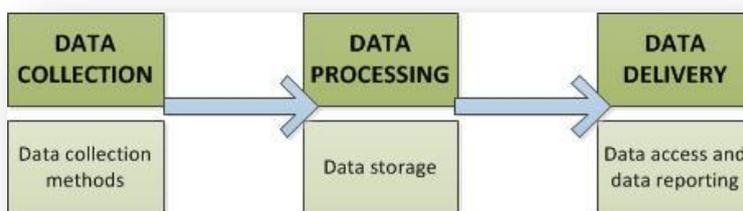


Figure 3.4: Data process (Rodrigues & Riveill, 2010)

The data quality challenges, discussed in Section 3.3 and 3.4, will be grouped into these three stages of data. As seen in Section 3.3 and 3.4, there are various data quality challenges in general as well as in the e-health domain. The following section provides a tabulated summary of the data quality challenges identified in literature, using Rodrigues and Riveill's (2010) outline. These are presented as follows:

- Data quality challenges at data collection (Section 3.5.1).
- Data quality challenges in data processing (Section 3.5.2).
- Data quality challenges in data delivery (Section 3.5.3).

3.5.1 DATA QUALITY CHALLENGES AT DATA COLLECTION

The table below lists data quality challenges that occur during data collection. This may also include data quality challenges before data collection, such as management responsibilities. The data quality challenges in this table are a summary of the data quality challenges presented in Section 3.3 and 3.4.

Table 3.4: Data quality challenges at data collection

Data quality challenges at data collection	Sources
Absent assignment of responsibilities regarding data	Soares (2015); Eckerson (2014); Haug <i>et al.</i> (2013); Singh & Singh (2010)
Ambiguity of roles with regard to data construction, use and maintenance	Eckerson (2014); Haug <i>et al.</i> (2013); Samyshkin & Timoshkin (2004);
Lack of adequate administration focus with regard to data quality	Eckerson (2014); Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Lack of adequate rewards in terms of data quality	Haug <i>et al.</i> (2013); Samyshkin & Timoshkin (2004)
Lack of adequate training for data consumers	Haug <i>et al.</i> (2013); Singh & Singh (2010); Minear (2009); World Health Organisation (2003);
Need for written data quality policies and procedures	Soares (2015); Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011); Singh & Singh (2010)
Not enough emphasis on the importance of data quality by managers	Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Misspelled data	Nix & Rosenau (2011); Singh & Singh (2010); Ure <i>et al.</i> (2006)
Conflicting use of special characters	Singh & Singh (2010)
More than one source of data, each producing different values	Horbatuk (2011), Nix & Rosenau (2011); Singh & Singh (2010); Strong

Data quality challenges at data collection	Sources
	<i>et al.</i> (1997)
Information is produced by subjectivity	Nix & Rosenau (2011); Strong <i>et al.</i> (1997)
A certain data element is not captured	Horbatuk (2011); Nix & Rosenau (2011)
A certain data element is captured, but not in the required format	Horbatuk (2011); Nix & Rosenau (2011)
A certain data element is captured in the correct format for one system or part of a system, but not for another	Horbatuk (2011); Singh & Singh (2010)
Transformation from paper records	Beacon Community Program (2013); Horbatuk (2011); Greiver <i>et al.</i> (2011); World Health Organisation (2003)
Accidental deletions	Rodriguez & Riveill (2010)
Accidental edits	Rodriguez & Riveill (2010)
The default value in the system was never corrected	Singh & Singh (2010)
Inaccurate data at collection	Nix & Rosenau (2011); De Leeuw (2005); Samyshkin & Timoshkin (2004); World Health Organisation (2003)
Questionnaire mode	De Leeuw (2005); Groves (1987)
Questionnaire length	Subar <i>et al.</i> (2001); Hess <i>et al.</i> (2001); Groves (1987)
Structure of questions	World Health Organisation (2003); Hess <i>et al.</i> (2001); Subar <i>et al.</i> (2001); Groves (1987)

Data quality challenges at data collection	Sources
Data entry errors	Nix & Rosenau (2011); Singh & Singh (2010); Health Organisation (2003)

3.5.2 DATA QUALITY CHALLENGES IN DATA PROCESSING

The table below lists data quality challenges that occur in the data processing phase. The data quality challenges in this table are a summary of the data quality challenges presented in Section 3.3 and 3.4.

Table 3.5: Data quality challenges in data processing

Data quality challenges in data processing	Sources
Ineffective organisational procedures	Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Lack of appropriate software for data management	Haug <i>et al.</i> (2013); Singh & Singh (2010)
Lack of adequate input options in data quality systems	Haug <i>et al.</i> (2013)
Unfortunate usability of information technology systems	Haug <i>et al.</i> (2013)
Absent columns	Singh & Singh (2010)
Insufficient candidate data sources	Singh & Singh (2010); World Health Organisation (2003)
Insufficient familiarity of inter-data dependencies	Singh & Singh (2010)
Changing timeliness of data sources	Wentzel <i>et al.</i> (2014); Vaziri & Mohsenzadeh (2012); Singh & Singh (2010)
Unforeseen alterations in source systems	Singh & Singh (2010)
Numerous sources for the same data	Beacon Community Program (2013); Singh & Singh (2010)
Extra columns	Singh & Singh (2010)
Data values drift from their field descriptions and business rules	Singh & Singh (2010)
Unsuitable data relationships	Singh & Singh (2010)
Lost information because of systematic errors	Strong <i>et al.</i> (1997)
Inconsistent definitions and formats because of different	Strong <i>et al.</i> (1997)

Data quality challenges in data processing	Sources
systems used	
Interoperability issues	Samyshkin & Timoshkin (2004); Strong <i>et al.</i> (1997)
Lack of sufficient computer and other resources	Horbatuk (2011); Strong <i>et al.</i> (1997);
Unavailable queries	Beacon Community Program (2013)
Information not stored	Beacon Community Program (2013); Nix & Rosenau (2011)
Uncertainties in legal aspects of medical records	Horbatuk (2011); Minear (2009)
Balancing patient care with electronic health record tasks	Horbatuk (2011)
Information needed for required fields	Horbatuk (2011), Singh & Singh (2010)
Managing large amounts of data	Rodriguez & Riveill (2010)
Managing heterogeneous data	Rodriguez & Riveill (2010)
Data structure problems	Wentzel <i>et al.</i> (2014); Singh and Singh (2010)
Outdated data	Wentzel <i>et al.</i> (2014); Singh & Singh (2010)
Data in the wrong field	Singh & Singh (2010); Ure <i>et al.</i> (2006)
Data in the wrong sequence	Ure <i>et al.</i> (2006)
Omissions of data	Beacon Community Program (2013); Nix & Rosenau (2011)
Data clustered incorrectly	Nix & Rosenau (2011)
Inconsistent data within a single file	Ure <i>et al.</i> (2006)
Authenticity of data elements	Nix & Rosenau (2011); Samyshkin & Timoshkin (2004)

3.5.3 DATA QUALITY CHALLENGES AT DATA DELIVERY

The table below lists data quality challenges that occur in the data processing phase. The data quality challenges in this table, is a summary of the data quality challenges presented in Section 3.3 and 3.4.

Table 3.6: Data quality challenges at data delivery

Data quality challenges at data delivery	Sources
Lack of adequate assessments	Haug <i>et al.</i> (2013); Singh & Singh (2010)
Occurrence of outliers	Singh & Singh (2010)
Timeframe restrictions	Beacon Community Program (2013)
Privacy and confidentiality issues	Samyshkin & Timoshkin (2004); Strong <i>et al.</i> (1997)
Data segmentation	Beacon Community Program (2013)
Tracking completed tasks	Beacon Community Program (2013)
Interfaced systems may not display data correctly	Horbatuk (2011)
Data not meeting all relevant needs	Vaziri & Mohsenzadeh (2012); Horbatuk (2011); Samyshkin & Timoshkin (2004)
Information is not usable for analytic purposes	Samyshkin & Timoshkin (2004)
Data is difficult to access	Wentzel <i>et al.</i> (2014); Samyshkin & Timoshkin (2004)

The data quality challenges explored in Sections 3.3 and 3.4 were grouped and summarised according to Rodrigues and Riveill’s (2010) outline. These challenges could serve as a reference to guide future health data quality interventions. Data quality interventions could be simplified by handling data quality challenges at the data stage where it occurs (Botha, Botha & Herselman, 2014). The data quality challenges presented in Section 3.5 will be verified and expanded by data quality experts in the qualitative phase of this research.

The researcher explores similar studies in Section 3.6.

3.6 SIMILAR STUDIES

Worldwide there has been research on data quality in health information systems. Section 3.6 provides background with regard to six similar studies. The similar studies provide a background to the researcher with regard to data quality in e-health.

- *Using a data entry clerk to improve data quality in primary care electronic medical records: a pilot study* (Greiver et al., 2011). The goal of this study was to evaluate the influence of using a data entry clerk for medical records in an intervention base. The study included a before and after design. It was found that by using a data manager for data queries and a data entry clerk to enter and check the data, health data quality will improve significantly. The study only focussed on the improvement of data quality by improving one data quality challenge, whereas this study focusses on exploring and prioritising all data quality challenges.
- *Data quality and clinical audit* (Verma, 2012). Verma states that valid conclusions can only be made as a result of good data quality. Data are used to simplify the clinical audit process. The importance of good data quality is illustrated in the Audit Cycle:

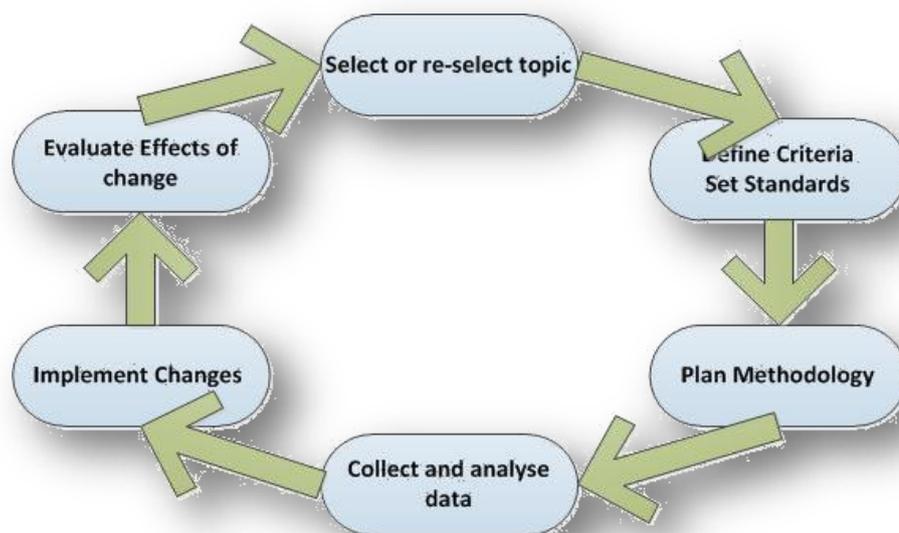


Figure 3.5: Data quality audit cycle (Verma, 2012)

- *Health data ownership and data quality: clinics in the Nyandeni District, Eastern Cape, South Africa* (Wright & Odama, 2012). The purpose of this study was to

evaluate the correlation between data ownership and data quality. The method used was a data audit as well as structured interviews with clinicians. It was found that secondary collected data are not used for decision-making in these clinics. Data quality is found to be very poor and it is important for clinicians to understand the importance of data quality. The study only focussed on the improvement of data quality by improving one data quality challenge, whereas this study focusses on exploring and prioritising all data quality challenges.

- *Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa* (Mphatswe *et al.*, 2012). The purpose of this study was to test the data quality improvement by implementing an intervention in KwaZulu-Natal. The intervention included training on data gathering, response for health information staff, regular data evaluations and data audits at healthcare facilities. A before and after study was used to evaluate the results. It was found that these simple, practical implementations have a major effect on the overall data quality of health-care facilities in KwaZulu-Natal. The study only focussed on the improvement of data quality by improving a few data quality challenges, whereas this study focusses on exploring and prioritising all data quality challenges.
- *Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) database* (Salati *et al.*, 2011). The purpose of the study was to develop a methodology to assess the quality of the ESTS database. The researchers used completeness, correctness, consistency and believability as metrics. The study includes a template and thresholds to assess data quality in medical databases. The study focussed on measuring data quality, whereas this study focusses on exploring and prioritising data quality challenges as experienced by data users.
- *Health information and quality authority (2011)*. It is found that up to 30% of the total American health budget is spent on data. Therefore it is important to have good data quality in health facilities. The health information and quality authority defines TDQM (Total Data Quality Management):

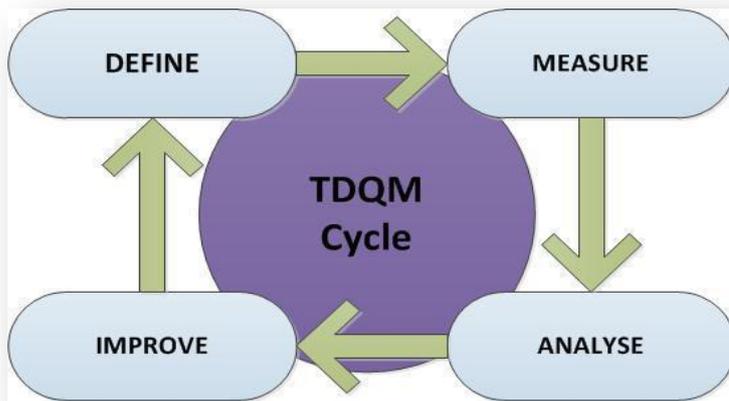


Figure 3.6: Total data quality management (Health information and quality authority, 2011)

The similar studies described in Section 3.6 all aimed to improve data quality within health information systems. The studies consisted of data quality interventions, including training, the use of a data entry clerk and management interventions. As summarised in Section 3.5, the researcher found that there are many data quality challenges. Data quality interventions should be focussed on more than only one data quality challenge, as seen in the cases described. The purpose of this study is to prioritise data quality challenges experienced by data users of electronic healthcare systems in South Africa. The prioritised list of data quality challenges could be used as a guideline for future data quality interventions in health information systems.

Section 3.7 provides a conclusion to Chapter 3.

3.7 CONCLUSION

The purpose of Chapter 3 was to explore data quality challenges which could lead to substandard data quality in electronic healthcare systems. Inductive content analysis was used to explore the data quality challenges presented in this chapter.

The definition of data quality, which consists of a number of data quality dimensions, can be regarded as a standard to which any data should adhere. Data quality challenges in both general and health information systems were provided. The data quality challenges explored in this chapter, were grouped into Rodrigues and Riveill's (2010) three data processes.

In Chapter 2, which provided a background to e-health and the importance of data quality in health information systems, the researcher identified e-health barriers that can impact on data quality. In Chapter 3, known data quality challenges were identified.

The first secondary research question: *What is the current status of research on data quality challenges?* was answered in the summarised data quality challenges tables, provided in Section 3.5.

The findings of this chapter will be used as an additional input to the possible data quality challenges in e-health presented in Chapter 4.

CHAPTER 4: POSSIBLE E-HEALTH DATA QUALITY CHALLENGES

Stage in research:

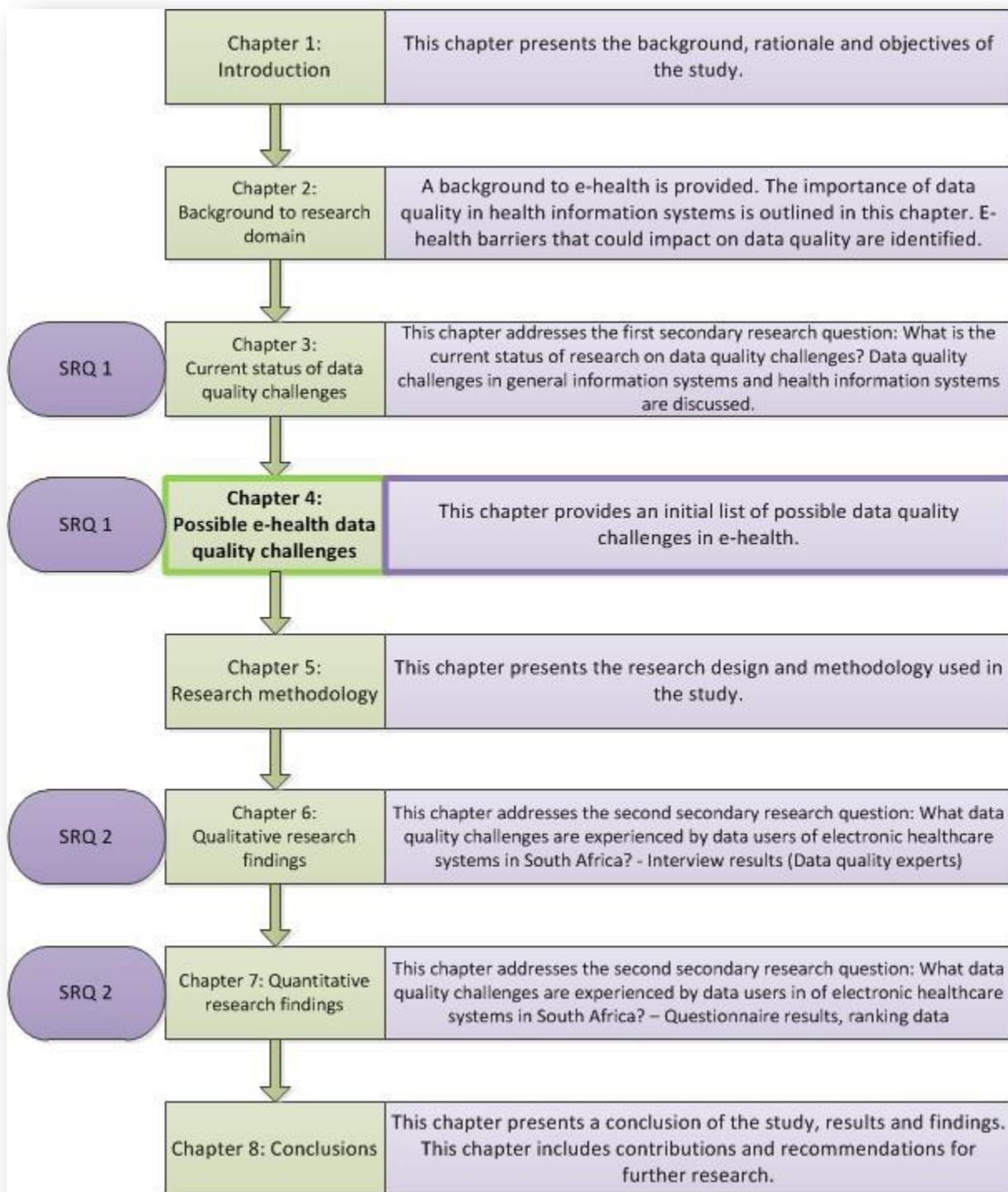


Figure 4.1: Chapter 4 – stage in research

Structure of Chapter 4:

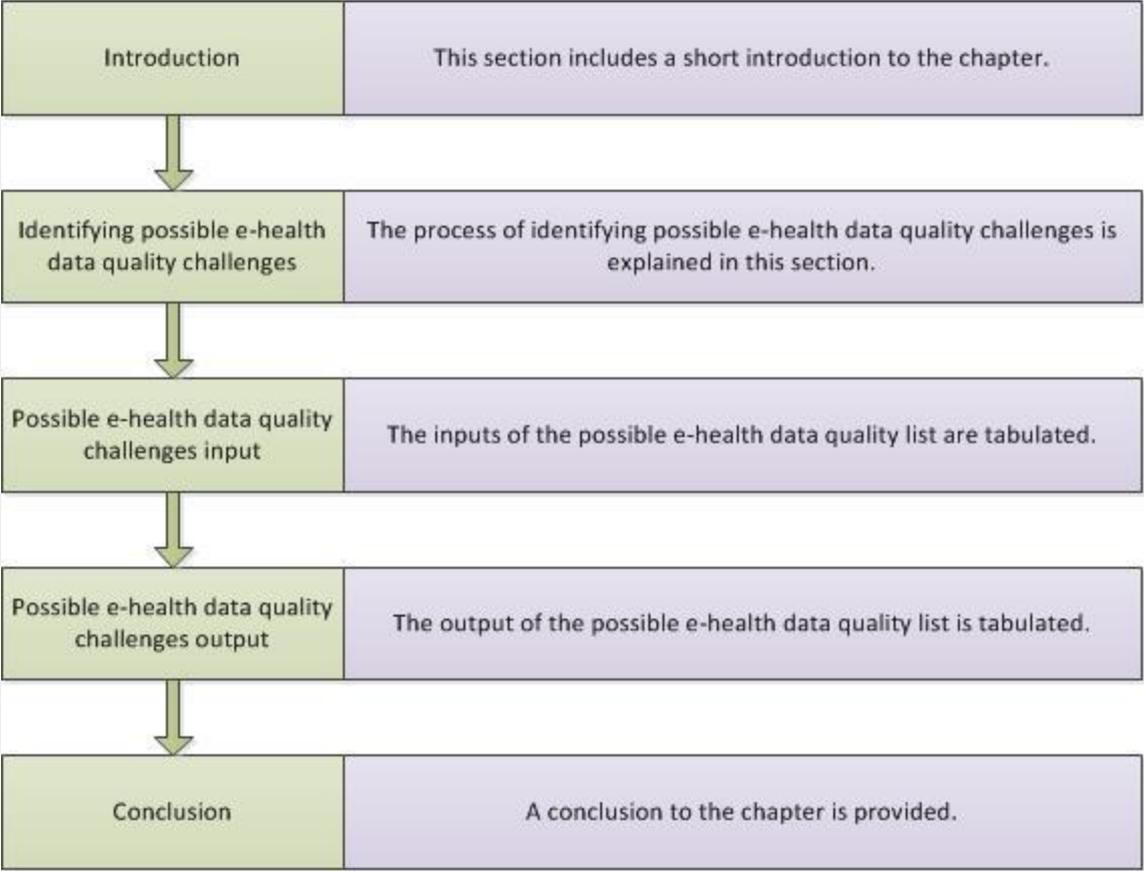


Figure 4.2: Chapter 4 structure

4.1 INTRODUCTION

Chapter 4, possible data quality challenges in e-health, describes the possible data quality challenges that emerged from the two literature chapters. Chapter 2 presented literature on e-health and the uses thereof and identified e-health barriers that could impact on data quality. Chapter 3 presented literature on data quality in general and e-health data quality specifically. The chapter articulated the definition of data quality used in this study and explored the causes of substandard data quality. Chapter 2 and Chapter 3 serve as underpinning for possible data quality challenges presented in this chapter. Chapter 2 and Chapter 3 were completed through a literature study and new literature was sought till data saturation point was reached (Glaser & Strauss, 2009).

4.2 IDENTIFYING POSSIBLE E-HEALTH DATA QUALITY CHALLENGES

The identified possible data quality challenges in e-health will inform and guide the qualitative (Chapter 6) and quantitative (Chapter 7) phases of the study. The possible e-health data quality challenges were identified through literature reviews documented in Chapter 2 and 3. The funnel in Figure 4.3 illustrates the process followed.

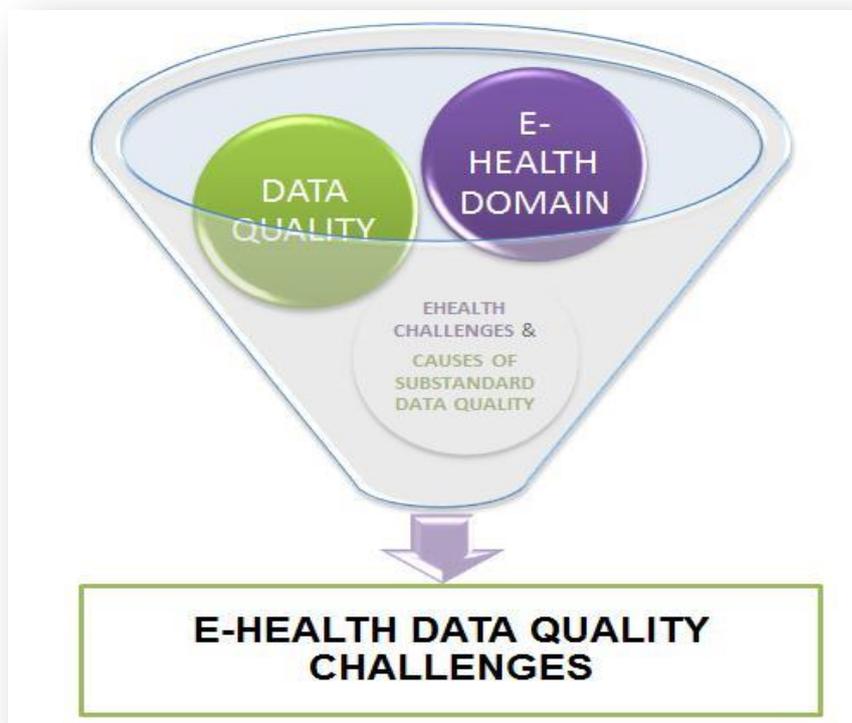


Figure 4.3: Process towards identifying possible e-health data quality challenges

The study covers two domains, namely, data quality and e-health. Literature reviews on both domains were given in the previous two chapters.

The definition of data quality (Section 3.2) is needed, since it serves as a standard for data quality. Data that does not conform to the defined data quality dimensions can be regarded as substandard data quality. As shown in Chapter 3, there are certain causes of substandard data quality.

The definitions and uses of e-health were given in Section 2.4. Although there are many e-health benefits, there are barriers to the implementation as well as the sustainability thereof.

Figure 4.4 is a graphical representation of the compilation of possible data quality challenges in e-health. Data quality challenges were identified and extracted from literature in the data quality domain. E-health barriers that could impact on data quality were identified from literature on the e-health domain. These identified data quality challenges and e-health barriers that could impact on the data quality are combined to present possible e-health data quality challenges.

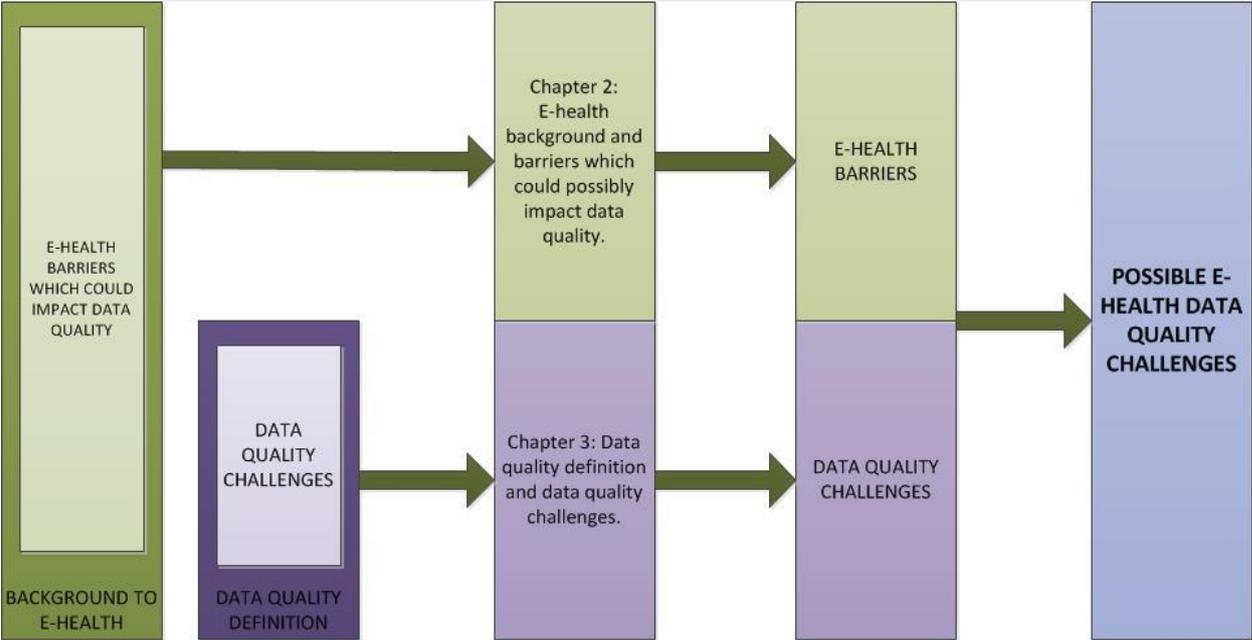


Figure 4.4: Possible e-health data quality challenges

The combination of the two domains leads to e-health data quality challenges. The inputs to the possible e-health data quality challenges are given in Section 4.3. The e-health data quality challenges are tabulated in Section 4.4.

4.3 POSSIBLE E-HEALTH DATA QUALITY CHALLENGES INPUT

The possible e-health data quality challenge inputs are discussed in this section. Data quality dimensions, data quality challenges and e-health barriers that could impact on data quality all serve as the input to the possible e-health data quality challenges. Data quality dimensions are discussed in Section 4.3.1.

4.3.1 DATA QUALITY DIMENSIONS

As seen in Chapter 3, there are many data quality dimensions that serve as a definition or standard for data quality. Data quality dimensions are factors to be considered with regard to data.

Data should be accurate, relevant, accessible and it should be a true representation (Wang, 1998; Strong, Yang & Wang, 1997; Wang & Strong, 1996). Furthermore, data quality dimensions include factors such as ease of access, believability, completeness, brief and suitable representation, free of mistakes, relevance and value (Vaziri & Mohsenzadeh, 2012).

Health data that does not conform to these data quality dimensions can be regarded as substandard data (Bowen, 2012; Carter, 2012). Therefore data quality dimensions play an important part in identifying the possible e-health data quality challenges for this study.

4.3.2 DATA QUALITY CHALLENGES

As seen in Chapter 3 there are various causes of substandard data quality. Causes of substandard data quality are seen as data quality challenges. The purpose of this study is to prioritise data quality challenges in electronic healthcare systems in South Africa. To realise the purpose of the study, an initial foundation of existing e-health data quality challenges is needed to support both the qualitative and quantitative phases of this study.

There exist various known causes of substandard data quality, including data quality challenges at data collection, at data processing and at data delivery (Rodriguez & Riveill, 2010).

Data quality challenges include factors such as (see Table 3.4, 3.5 and 3.6): management challenges, data entry errors, data structure problems, transformation from paper records, ineffective organisational procedures, data values drifting from field

descriptions and business rules, data that does not meet relevant needs, data that is difficult to access, privacy and confidentiality issues, interoperability issues amongst many other (Wentzel, 2014; Haug, 2013; Horbatuk, 2011; Singh & Singh 2010; Samyshkin & Timoshkin 2004; Strong *et al.* 1997).

The data quality challenges, presented in Chapter 3, are a key part of the initial identified possible e-health data quality challenges for this study.

4.3.3 E-HEALTH BARRIERS THAT COULD IMPACT ON DATA QUALITY

Chapter 2 presented a background of e-health as research domain. The definition and purpose of e-health were provided. The background included benefits and barriers of e-health implementation and use, as well as the importance of data quality in health information systems.

The barriers to e-health implementation that could affect the quality of data were presented in Table 2.5 and extends to include financial barriers, lack of IT and clinical resources, the difficulty of learning and using e-health software, personnel costs, standardisation of Health Information Systems, time challenges, the implementation of e-health in rural areas (connectivity), data privacy, interoperability, sustainability, data quality, usability and the transition from paper to electronic health records.

E-health barriers that could impact on data quality should be considered when prioritising data quality challenges in electronic healthcare systems in South Africa.

4.4 POSSIBLE E-HEALTH DATA QUALITY CHALLENGE OUTPUT

The possible e-health data quality challenge output, presented in this section, serves as the foundation for the qualitative phase (Chapter 6) and the quantitative phase (Chapter 7) of this study.

The possible e-health data quality challenges presented in Table 4.1 consists of the data quality challenges presented in Chapter 3 (which includes data quality dimensions) and the e-health barriers that could impact on data quality presented in Chapter 2.

Table 4.1: Possible e-health data quality challenges

Possible e-health data quality challenges	Sources
Absent assignment of responsibilities regarding data	Soares (2015); Eckerson (2014); Haug <i>et al.</i> (2013); Singh & Singh

Possible e-health data quality challenges	Sources
	(2010)
Ambiguity of roles with regard to data construction, use and maintenance	Eckerson (2014); Haug <i>et al.</i> (2013); Samyshkin & Timoshkin (2004);
Lack of adequate administration focus with regard to data quality	Eckerson (2014); Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Lack of adequate rewards in terms of data quality	Haug <i>et al.</i> (2013); Samyshkin & Timoshkin (2004)
Lack of adequate training for data consumers	Haug <i>et al.</i> (2013); Singh & Singh (2010); Minear (2009); World Health Organisation (2003);
Need for written data quality policies and procedures	Soares (2015); Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011); Singh & Singh (2010)
Managers' lack of thorough emphasis on the importance of data quality	Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Misspelled data	Nix & Rosenau (2011); Singh & Singh (2010); Ure <i>et al.</i> (2006)
Conflicting use of special characters More than one source of data, each producing different values	Singh & Singh (2010) Horbatuk (2011), Nix & Rosenau (2011); Singh & Singh (2010); Strong <i>et al.</i> (1997)
Information is produced by subjectivity	Nix & Rosenau (2011); Strong <i>et al.</i> (1997)
A certain data element is not captured	Horbatuk (2011); Nix & Rosenau (2011)
A certain data element is captured, but not in the required format	Horbatuk (2011); Nix & Rosenau (2011)
A certain data element is captured in the correct format for one system or part of a system, but not for another	Horbatuk (2011); Singh & Singh (2010)
Transformation from paper records	Beacon Community

Possible e-health data quality challenges	Sources
Accidental deletions	Program (2013); Horbatuk (2011); Greiver <i>et al.</i> (2011); World Health Organisation (2003)
Accidental edits	Rodriguez & Riveill (2010)
The default value in the system was never corrected	Rodriguez & Riveill (2010)
Inaccurate data at collection	Singh & Singh (2010)
Questionnaire mode	Nix & Rosenau (2011); De Leeuw (2005); Samyshkin & Timoshkin (2004); World Health Organisation (2003)
Questionnaire length	De Leeuw (2005); Groves (1987)
Structure of questions	Subar <i>et al.</i> (2001); Hess <i>et al.</i> (2001); Groves (1987)
Data entry errors	World Health Organisation (2003); Hess <i>et al.</i> (2001); Subar <i>et al.</i> (2001); Groves (1987)
Ineffective organisational procedures	Nix & Rosenau (2011); Singh & Singh (2010); Health Organisation (2003)
Lack of appropriate software for data management	Haug <i>et al.</i> (2013); Health Information and Quality Authority (2011)
Lack of adequate input options in data quality systems	Haug <i>et al.</i> (2013); Singh & Singh (2010)
Unfortunate usability of information technology systems	Haug <i>et al.</i> (2013)
Absent columns	Singh & Singh (2010)
Insufficient candidate data sources	Singh & Singh (2010); World Health Organisation (2003)
Insufficient familiarity of inter-data dependencies	Singh & Singh (2010)
Changing timeliness of data sources	Wentzel <i>et al.</i> (2014); Vaziri & Mohsenzadeh (2012); Singh & Singh

Possible e-health data quality challenges	Sources
	(2010)
Unforeseen alterations in source systems	Singh & Singh (2010)
Numerous sources for the same data	Beacon Community Program (2013); Singh & Singh (2010)
Extra columns	Singh & Singh (2010)
Data values drift from their field descriptions and business rules	Singh & Singh (2010)
Unsuitable data relationships	Singh & Singh (2010)
Lost information because of systematic errors	Strong <i>et al.</i> (1997)
Inconsistent definitions and formats because of different systems used	Strong <i>et al.</i> (1997)
Interoperability issues	Samyshkin & Timoshkin (2004); Strong <i>et al.</i> (1997)
Lack of sufficient computer and other resources	Horbatuk (2011); Strong <i>et al.</i> (1997);
Unavailable queries	Beacon Community Program (2013)
Information not stored	Beacon Community Program (2013); Nix & Rosenau (2011)
Uncertainties in legal aspects of medical records	Horbatuk (2011); Minear (2009)
Balancing patient care with electronic health record tasks	Horbatuk (2011)
Information needed for required fields	Horbatuk (2011), Singh & Singh (2010)
Managing large amounts of data	Rodriguez & Riveill (2010)
Managing heterogeneous data	Rodriguez & Riveill (2010)
Data structure problems	Wentzel <i>et al.</i> (2014); Singh and Singh (2010)
Outdated data	Wentzel <i>et al.</i> (2014); Singh & Singh (2010)
Data in the wrong field	Singh & Singh (2010); Ure <i>et al.</i> (2006)
Data in the wrong sequence	Ure <i>et al.</i> (2006)
Omissions of data	Beacon Community Program (2013); Nix & Rosenau (2011)
Data clustered incorrectly	Nix & Rosenau (2011)
Inconsistent data within a single file	Ure <i>et al.</i> (2006)

Possible e-health data quality challenges	Sources
Authenticity of data elements	Nix & Rosenau (2011); Samyshkin & Timoshkin (2004)
Lack of data quality assessments	Haug <i>et al.</i> (2013); Singh & Singh (2010)
Occurrence of outliers	Singh & Singh (2010)
Timeframe restrictions	Beacon Community Program (2013)
Privacy and confidentiality issues	Samyshkin & Timoshkin (2004); Strong <i>et al.</i> (1997)
Data segmentation	Beacon Community Program (2013)
Tracking completed tasks	Beacon Community Program (2013)
Interfaced systems may not display data correctly	Horbatuk (2011)
Data not meeting all relevant needs	Vaziri & Mohsenzadeh (2012); Horbatuk (2011); Samyshkin & Timoshkin (2004)
Information is not usable for analytic purposes	Samyshkin & Timoshkin (2004)
Data is difficult to access	Wentzel <i>et al.</i> (2014); Samyshkin & Timoshkin (2004)
The financial barrier to purchase necessary hardware and cost challenges in general	Theobald (2014); Jaroslowski & Saberwal (2014); Jha <i>et al.</i> (2009); Minear (2009); Atreja <i>et al.</i> (2008); Anderson (2007); RTI International (2001); Iakovidis (1998)
Lack of IT and clinical resources	Theobald (2014); Jaroslowski & Saberwal (2014); Jha <i>et al.</i> (2009); Anderson (2007)
Difficulty learning and using the software	Theobald (2014); Anderson (2007); Braa <i>et al.</i> (2004); Ammenwerth <i>et al.</i> (2003)

Possible e-health data quality challenges	Sources
Personnel costs	Jaroslowski & Saberwal (2014); Atreja <i>et al.</i> (2008); Iakovidis (1998)
Standardisation of all health information systems, since the content and structure of all health information systems should be standardised	Jaroslowski & Saberwal (2014); Hayrinen <i>et al.</i> (2008); Atreja <i>et al.</i> (2008); Rodrigues (2008); Ammenwerth <i>et al.</i> (2003)
It might be time consuming to update the EHR thoroughly	Penoyer <i>et al.</i> (2014); Wentzel <i>et al.</i> (2014); Anderson (2007)
The implementation of health information systems in small and rural facilities, and not only big hospitals, remains a challenge	Francis (2013); Rodrigues (2008); Braa <i>et al.</i> (2004)
Data privacy	Accenture (2010); Minear (2009); Rodrigues (2008); Anderson (2007); Samyshkin & Timoshkin (2004); Waegemann (2002); RTI International (2001)
Interoperability	Accenture (2010); Minear (2009) Atreja <i>et al.</i> (2008); Westra <i>et al.</i> (2008); Anderson (2007); Waegemann (2002)
Sustainability	Jaroslowski & Saberwal (2014); Minear (2009); Braa <i>et al.</i> (2004)
Data quality	Francis (2013); Accenture (2010); Minear (2009); Atreja <i>et al.</i> (2008); Lobach & Detmer (2007); Samyshkin & Timoshkin (2004);
Usability	Car <i>et al.</i> (2008); Lobach & Detmer (2007); Anderson (2007); Ammenwerth <i>et al.</i> (2003); RTI International (2001)

Possible e-health data quality challenges	Sources
Transferring data from paper to electronic records	Atreja <i>et al.</i> (2008); Braa <i>et al.</i> (2004); Waegemann (2002);
Physicians are hesitant to change existing processes	Jaroslawski & Saberwal (2014); Jha <i>et al.</i> (2009);
Forming electronic health records as part of the facilities daily routine	Car <i>et al.</i> (2008); Braa <i>et al.</i> (2004);
Meeting needs at each provider level, reaching goals	Heeks (2006); Samyshkin & Timoshkin (2004);
Data access	Accenture (2010); Bowling <i>et al.</i> (2006)
Government	Jaroslawski & Saberwal (2014); Rodrigues (2008)
Logistics	Jaroslawski & Saberwal (2014), Rodrigues (2008)
The shortage or absence of the necessary infrastructure, such as internet connections	Theobald (2014); Anderson (2007)
The patient's medical history is not always available	Theobald (2014)
Although it saves time, the drop-down menus in health information systems may lack detailed information.	Penoyer <i>et al.</i> (2014)
Nurses' notes may go unread by physicians	Penoyer <i>et al.</i> (2014)
Scalability	Braa <i>et al.</i> (2004)
Modernizing existing systems	Samyshkin & Timoshkin (2004)
Extracting knowledge out of information	Westra <i>et al.</i> (2008)
Software	Minear (2009)
Patient consent	Accenture (2010)
Donors	Jaroslawski & Saberwal (2014)
Fragmentation programmes	Jaroslawski & Saberwal (2014)
Lack of appropriate software	Theobald (2014)

The identified possible data quality challenges are presented in Table 4.1. Section 4.5 provides a conclusion to this chapter.

4.5 CONCLUSION

The purpose of this chapter was to articulate the possible e-health data quality challenges to serve as a foundation for the qualitative and quantitative phases of the study.

As explained in more detail in Chapter 5, this study adopted a transformative procedure. This research strategy is based on a theoretical perspective, which is used as an initial list of possible e-health data quality challenges for the rest of the research. Mixed methods are often used to validate the theoretical background. Either a sequential or concurrent approach could be used in transformative procedures (Creswell & Clark, 2011; Creswell, 2009).

Chapter 2 and Chapter 3 provided the inputs for the possible data quality challenges presented in this chapter. The possible data quality challenges are presented as a combination of data quality challenges and e-health barriers that could impact data quality. As such these e-health barriers that could impact on data quality will be considered as data quality challenges in the remainder of this study. These factors are taken into consideration in the qualitative and quantitative phases of the research towards prioritising data quality challenges in electronic healthcare systems in South Africa.

Chapter 5 presents the design and methodology followed in this study.

CHAPTER 5: RESEARCH METHODOLOGY

Stage in research:

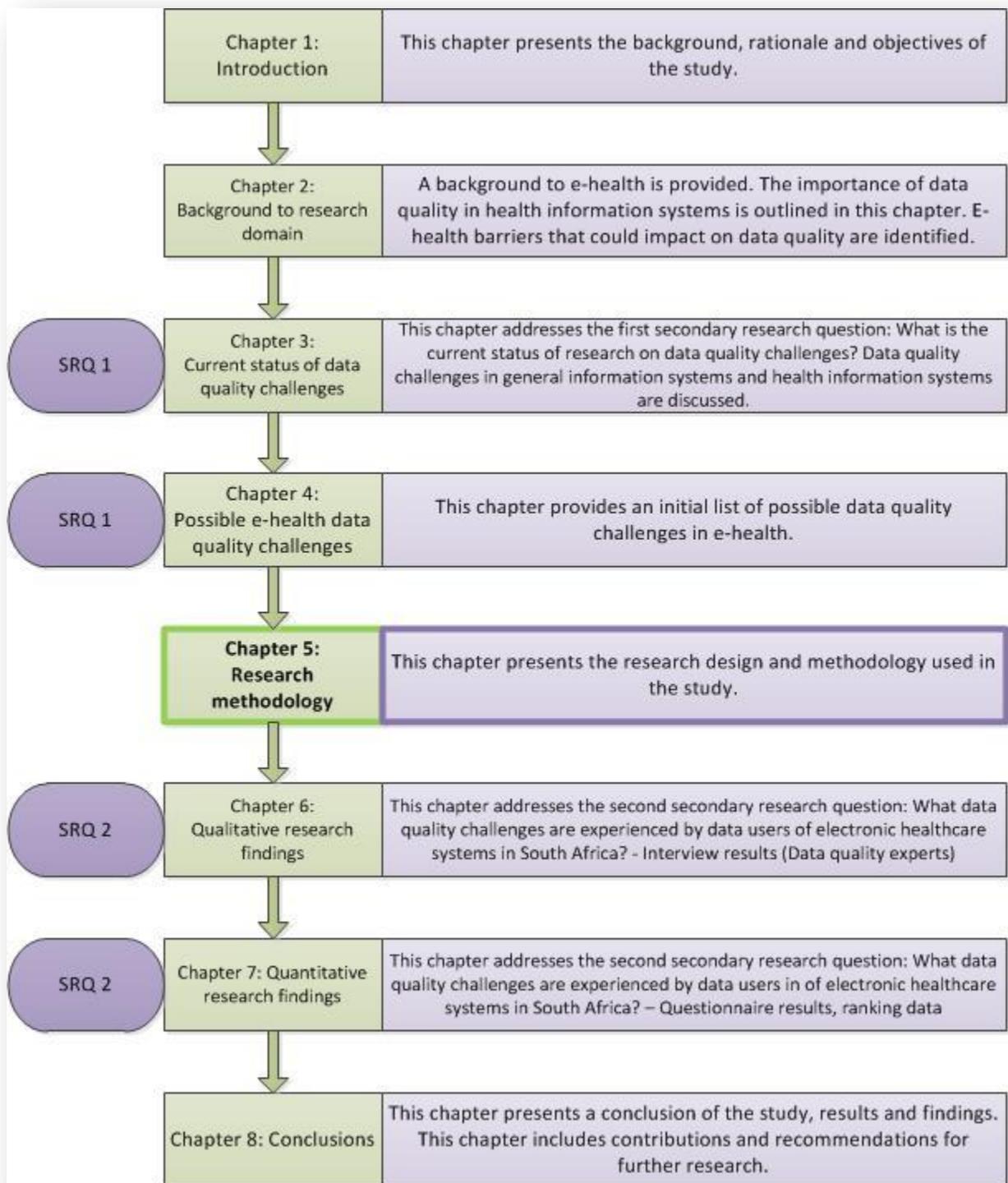


Figure 5.1: Chapter 5 – stage in research

Structure of Chapter 5:

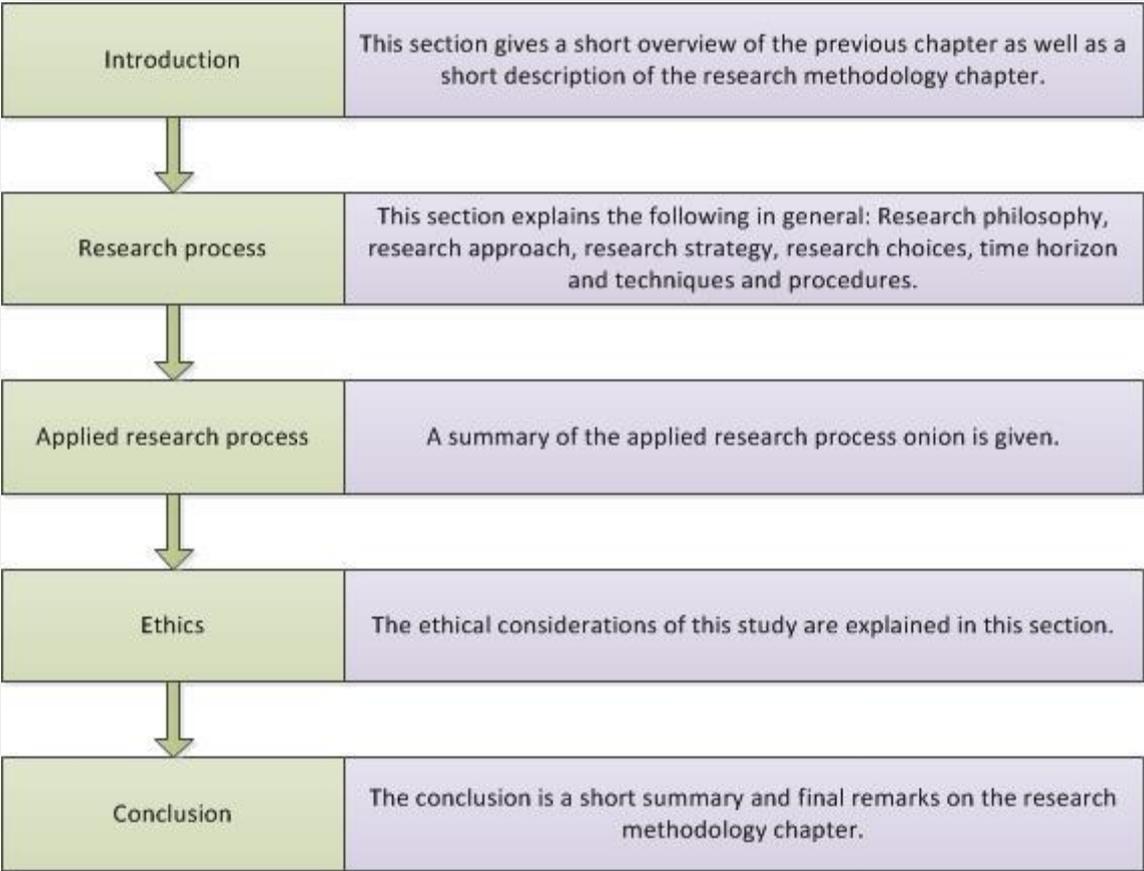


Figure 5.2: Structure of Chapter 5

5.1 INTRODUCTION

Researchers approach research questions by following particular research methodologies which are appropriate to their specific studies (Leedy & Ormrod, 2005). This chapter explains the research methodology followed in this study.

The previous chapter provided an overview of the possible e-health data quality challenges, which emerged from the literature review in Chapter 2 and Chapter 3. These possible e-health data quality challenges are used to compile the interviews as well as questionnaires.

This chapter consists of general explanations of the research philosophy, research design, research strategy, research approach, data collection methods and data analysis techniques in Section 5.2. The applied research process is summarised in Section 5.3. The ethical considerations are discussed in Section 5.4. Finally, a conclusion of the chapter is given in 5.5.

5.2 RESEARCH PROCESS

This study took place in four separate phases towards answering the research question. The four separate phases of the research plays a significant role in the applied research methodology. The study followed an exploratory sequential QUAL-quant research design.

The research phases are as follows:

1. Literature reviews on the background of e-health and data quality challenges (Chapter 2 and Chapter 3) were provided. Identified possible e-health data quality challenges (Chapter 4) emerged from the literature. These identified challenges are used in the following phases.
2. A qualitative study is done in order to verify and expand the data quality challenges identified in literature. The qualitative interviews are conducted with data quality experts.
3. A quantitative study was used to produce a prioritised list of the data quality challenges that emerged from the second phase of the research.
4. Evidence-based recommendations are provided to mitigate the risks identified in data quality challenges in order to support healthcare system users about the

use of data.

The research phases are illustrated in Figure 5.3.

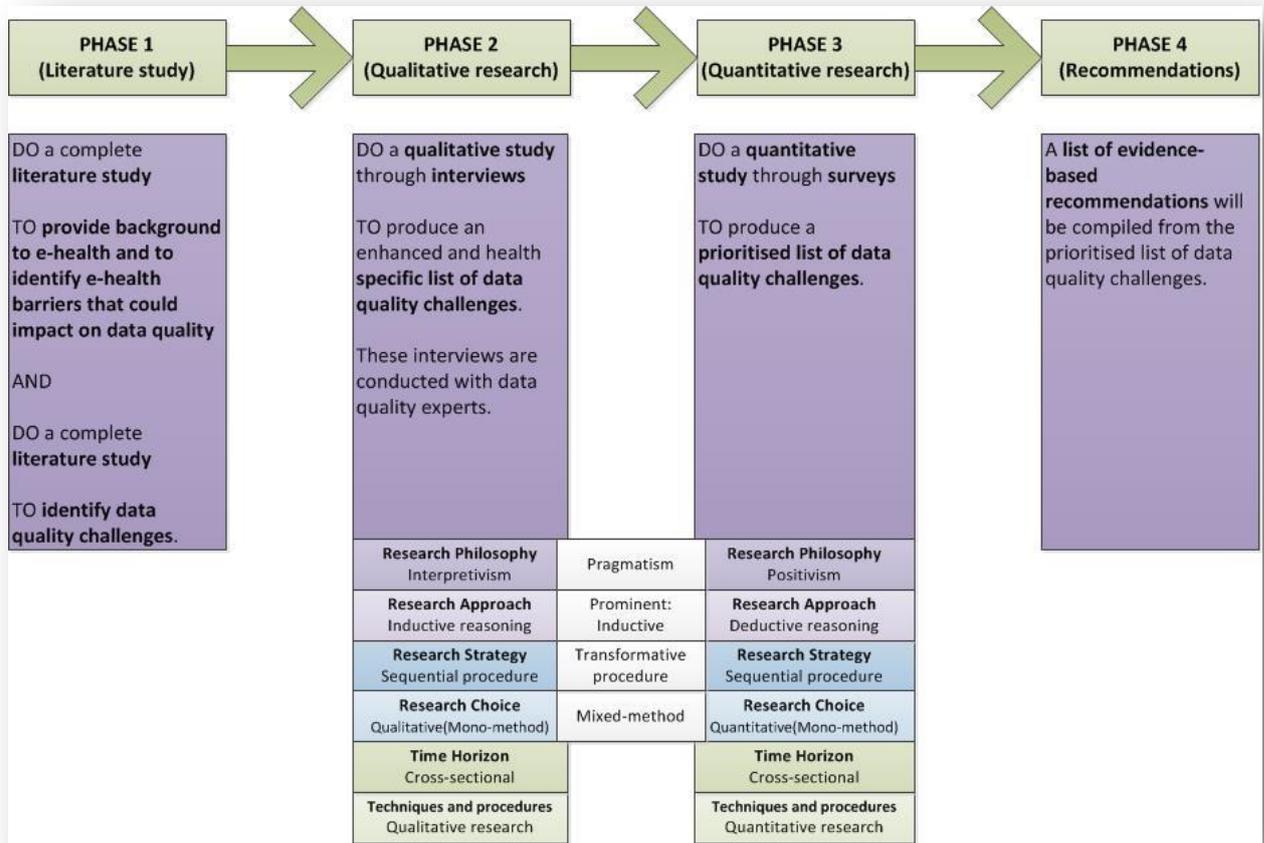


Figure 5.3: Research phases

The figure illustrates that the research took place in four sequential phases. The implication of this research process is that Phase 3 and Phase 4 of the research are based on different research foundations.

The research process on by Saunders, Lewis and Thornhill (2015) is used to describe the methodological approach followed in the research:

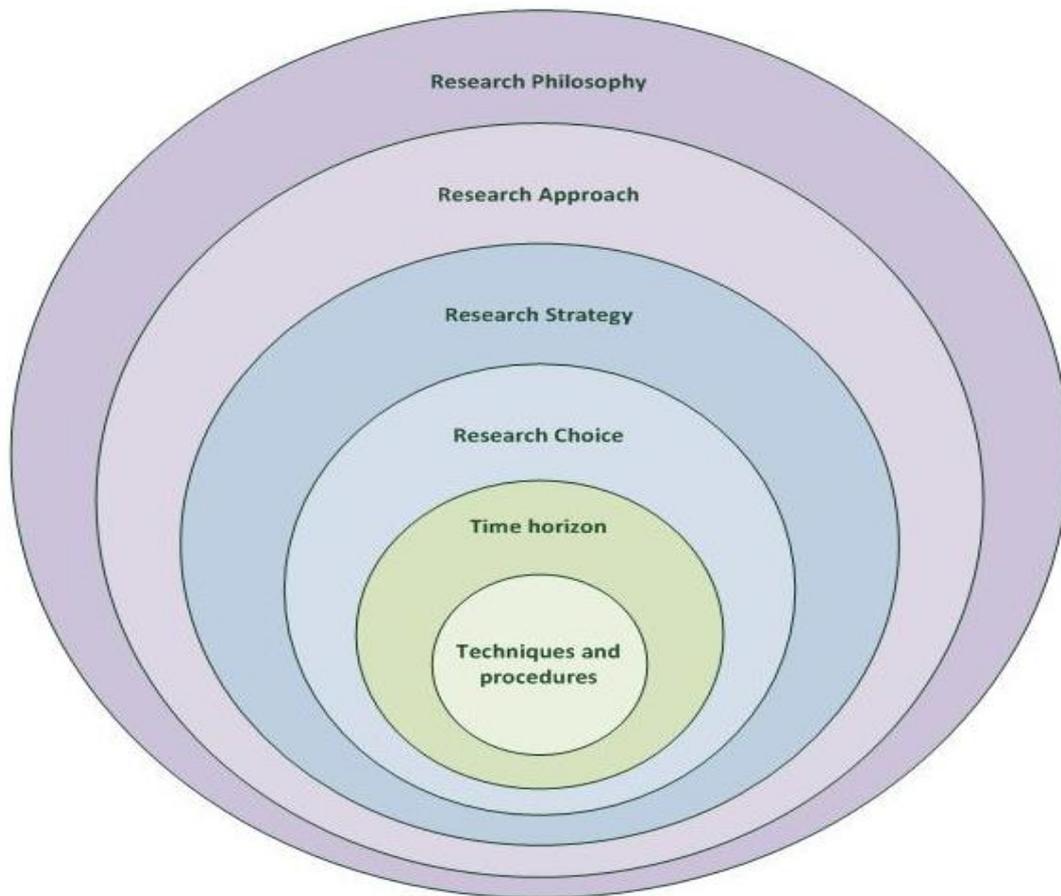


Figure 5.4: Research process onion (Saunders *et al.*, 2015)

The research process onion is a guideline to approach the research methodology by defining the research philosophy, research approach, research strategy, research choice, time horizon and techniques and procedures used in the study.

Each of these layers is discussed in sections 5.2.1 to 5.2.6. For each layer of the research process onion, an overview of the relevant layer is given followed by the applied philosophy, approach, strategy, choice, time horizon and techniques and procedures. The applied processes of the research process onion are motivated by the researcher.

5.2.1 RESEARCH PHILOSOPHY

Research is grounded on a fundamental philosophical paradigm (Oates, 2008; Myers, 1997). The first layer of the research process onion is discussed in this section. Research approaches, research strategies, research designs and data collection methods may correspond better with certain research philosophies (Oates, 2008). Lee

(1989) states that these philosophies are not always restricted or clear. Depending on the study the philosophies can be combined or bent.

5.2.1.1 RESEARCH PHILOSOPHIES OF INQUIRY

The research philosophy consists of different paradigms, of which positivism, realism and interpretivism are the most prominent.

- *Positivism* relies on structured and ordered views (Oates, 2008). In positivism it is assumed that reality can be defined independently of the researcher (Myers, 1997), thus positivism can be regarded as an objective research philosophy (Cornford & Smithson, 1996). Oates (2008) outlines the features of the positivistic research philosophy:
 - *The world is independent of people:* Social and physical aspects in the world may exist without dependency on the human mind. These aspects can be captured and measured.
 - *Measurement and modelling:* By making use of observations, models and procedures, the researcher discovers and quantifies the world or scenario.
 - *Objectivity:* Facts regarding the world is not based on the researcher's personal beliefs.
 - *Hypothesis testing:* Experimental testing of theories and hypotheses are the foundation of the research.
 - *Quantitative data analysis:* Mathematical and statistical methods are most often used in this philosophy – the results of the research are obtained in a logical and objective manner.
 - *Universal laws:* Patterns, simplifications and facts are used to conclude the findings of the research.
- *Critical research philosophy:* The critical research philosophy assumes that reality is established historically, in other words, what we observe to be real now, can be changed in the future (Donnelly & Trochim, 2007; Myers, 2004).

- *Interpretivism*: Social factors are believed to influence reality in the interpretivist philosophy (Kaplan & Maxwell, 1994). In this philosophical paradigm, the researcher attempts to understand the world through meanings and perceptions of people (Deetz, 1996).

Besides the three main research philosophies explained here, *pragmatism* is a philosophy applied in some studies. The main focus of pragmatism lies in practical implementations of theories (Goldkuhl, 2004). Pragmatism is a method often used for research that does not fit perfectly into quantitative or qualitative approaches (Darlington & Scott, 2003).

5.2.1.2 APPLIED RESEARCH PHILOSOPHY

The applied research philosophy is explained by referring to the different phases of research.

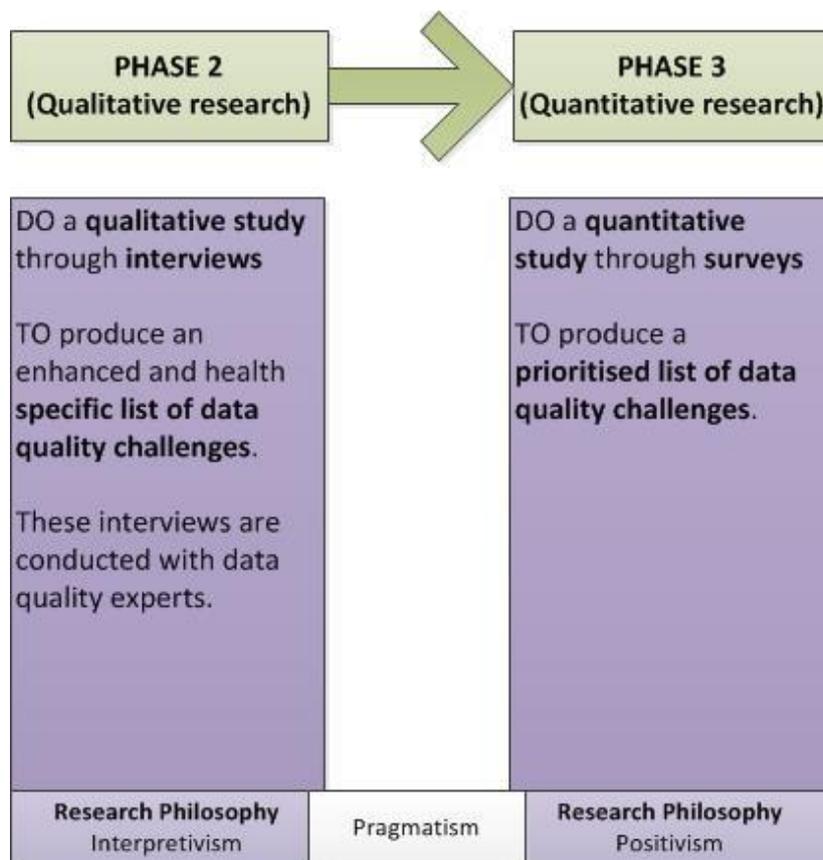


Figure 5.5: Research philosophy

- **Interpretive philosophy**

The second phase (qualitative) of the research follows the interpretive research philosophy.

The philosophy underlying the interpretative approach assumes that reality is gained through social factors, such as meanings and documents. Interpretive research focusses on the opinions of people (Kaplan & Maxwell, 1994). The research question is answered through the opinions that people have with regard to the research problem (Deetz, 1996).

This philosophy fits the purpose of the study, since the *perceived* data quality challenges in electronic healthcare systems in South Africa are investigated. The study focusses on the opinions and perceptions of data users of these systems, rather than actual, tested facts.

- **Positivistic philosophy**

The third phase (quantitative) of the research follows the positivistic research philosophy.

Positivistic philosophers assume that actuality can be defined by measurable factors (Myers, 1997). Quantitative data analysis, objectivity, measurements and patterns are all properties of the positivistic research philosophy (Oates, 2008). This philosophy is suitable for the prioritizing or measuring of the perceived data quality challenges in electronic healthcare systems in South Africa.

Figure 4.6 illustrates that positivistic and interpretivistic research philosophies are not exclusive (De Villiers, 2005).

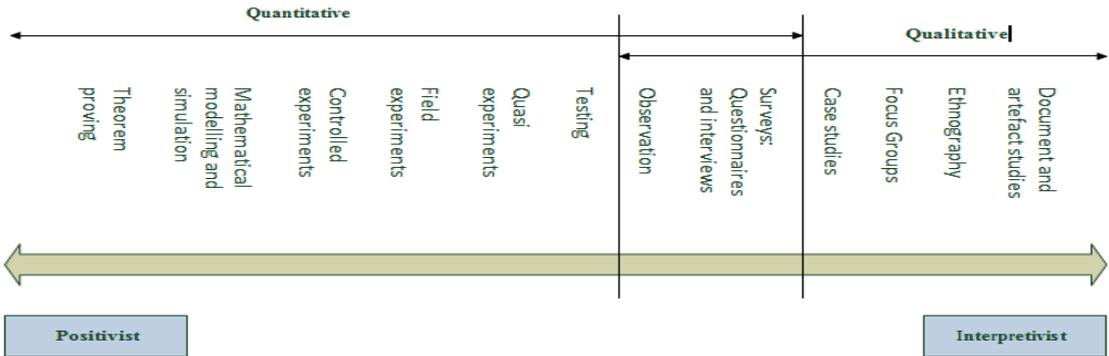


Figure 5.6: Positivist vs. Interpretivist (De Villiers, 2005)

The mixed method research approach can be beneficial, since different conventions may supplement each other (Trauth & Jessup, 2000).

- **Pragmatism**

Since Phase 2 and Phase 3 of the research adapt to different research philosophies, the appropriate research philosophy for this study is pragmatism.

Pragmatism focusses on practical implementation or application or significances of the theory (Goldkuhl, 2004). Pragmatist researchers usually make use of mixed method data collection methods (Creswell, 2007). The pragmatist researcher will make use of an appropriate approach to answer questions that is not suitable for qualitative or quantitative approaches (Darlington & Scott, 2002). In this case, a literature review is used along with the qualitative and quantitative data collection methods. Pragmatism is a good fit for this research, since the research requires a practical as well as theoretical approach, thus both qualitative and quantitative approaches are suitable for the research.

The applied first layer of the research process onion is as follows:

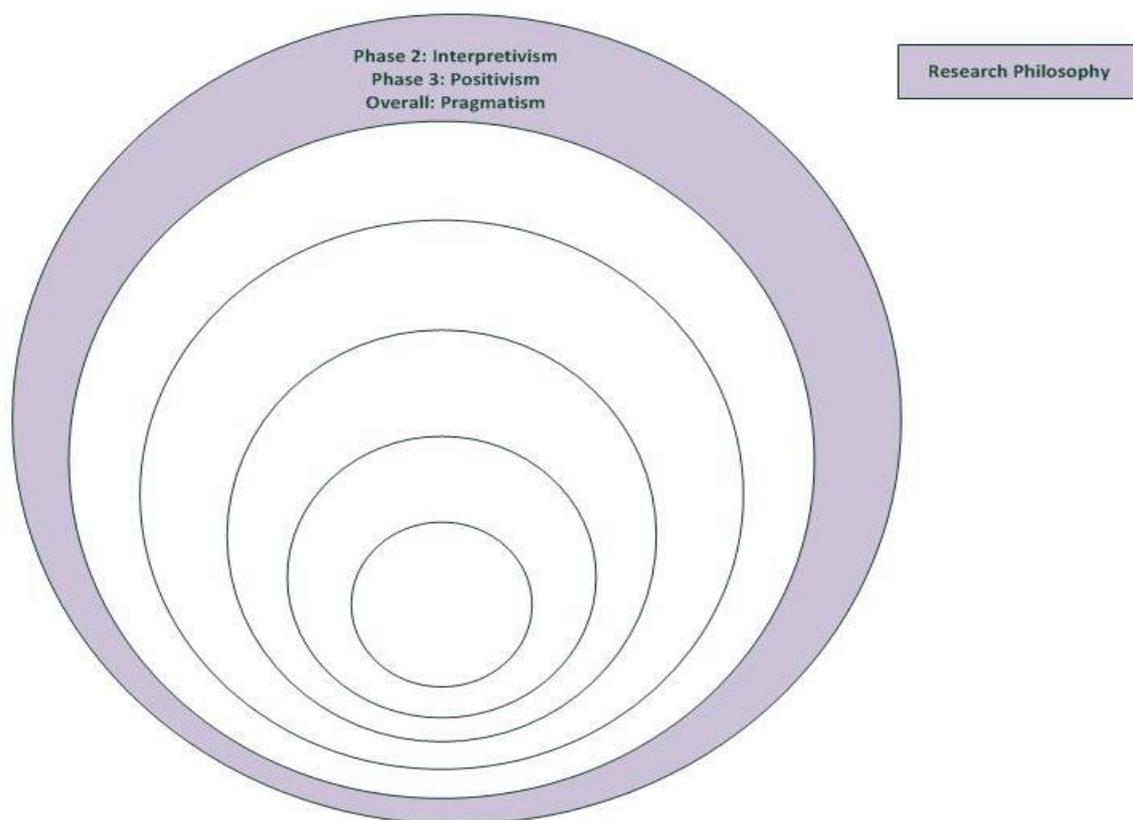


Figure 5.7: Applied research philosophy adapted from Saunders *et al.* (2015)

The following layer of the research process onion, the research approach, is discussed in Section 5.2.2.

5.2.2 RESEARCH APPROACH

The research approach refers to the way the research is conducted; from a specific or a general point of view.

5.2.2.1 RESEARCH APPROACHES OF INQUIRY

- *Deductive reasoning* begins at an understanding of theory and then conceptualises a certain subject of focus (Wills, 2007). This reasoning is also referred to as the top-down research approach (Donnelly & Trochim, 2007). Deductive reasoning transfers from wide-ranging or general theory to particular observations (Babbie, 2005). The figure below illustrates deductive reasoning:

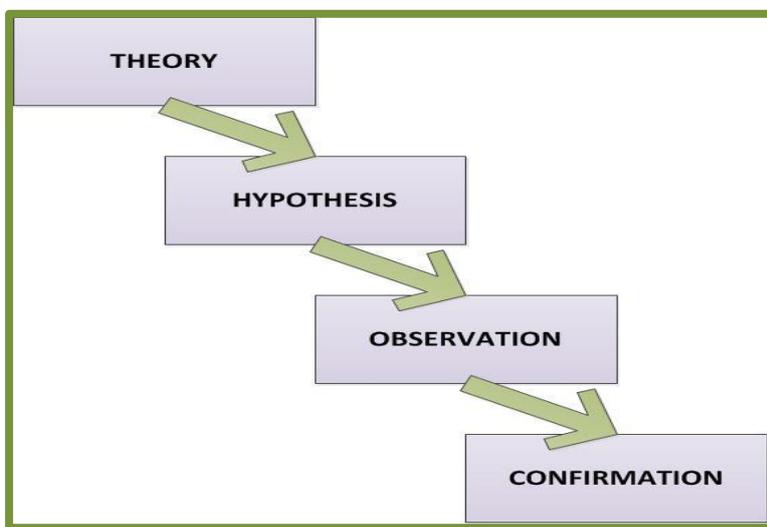


Figure 5.8: Deductive research approach (Donnelly & Trochim, 2007)

- *Inductive reasoning* is also referred to as the bottom-up research approach (Donnelly & Trochim, 2007). Inductive reasoning begins its focus on specific observations and broadens it to a more general theory (Wills, 2007). Inductive reasoning transfers from particular observations to general observations (Babbie, 2005). The figure below illustrates inductive reasoning.

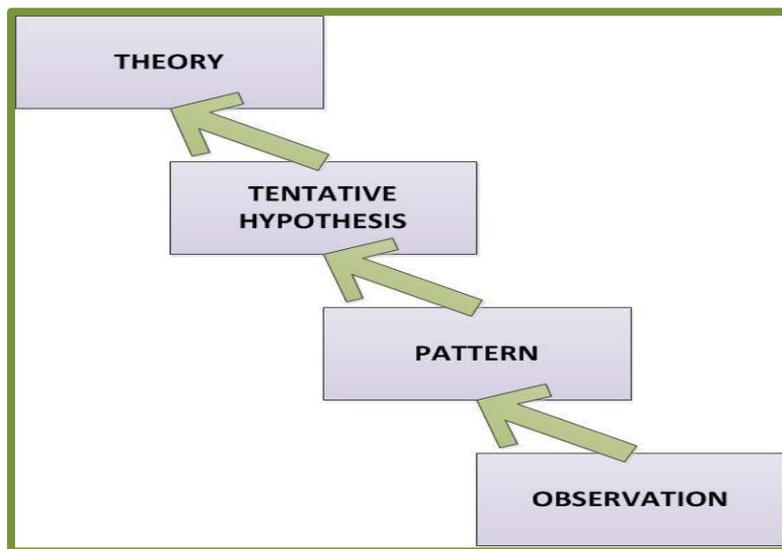


Figure 5.9: Inductive research approach (Donnelly & Trochm, 2007)

The table below highlights the important differences between the deductive and inductive research approaches:

Table 5.1: Deductive and inductive research approaches (Saunders *et al.*, 2015)

Deductive approach	Inductive approach
<ul style="list-style-type: none"> Scientific principles 	
<ul style="list-style-type: none"> Moving from theory to data 	<ul style="list-style-type: none"> Gaining an understanding of the meanings humans attach to events
<ul style="list-style-type: none"> Need to explain relationships between data 	<ul style="list-style-type: none"> Close understanding of the research context
<ul style="list-style-type: none"> Collection of quantitative data 	<ul style="list-style-type: none"> Collection of qualitative data
<ul style="list-style-type: none"> Application of controls to ensure validity of data 	<ul style="list-style-type: none"> More flexible structure to permit changes of research emphasis as the research progresses
<ul style="list-style-type: none"> Operationalisation of concepts to ensure clarity of definition 	
<ul style="list-style-type: none"> Highly structured approach 	
<ul style="list-style-type: none"> Researcher independence of what is being researched 	<ul style="list-style-type: none"> Realization that the researcher is part of the research process
<ul style="list-style-type: none"> Necessity to select samples of sufficient size in order to generalise conclusions 	<ul style="list-style-type: none"> Less concern with the need to generalise

5.2.2.2 APPLIED RESEARCH APPROACH

The applied research approach is explained by referring to the research phases.

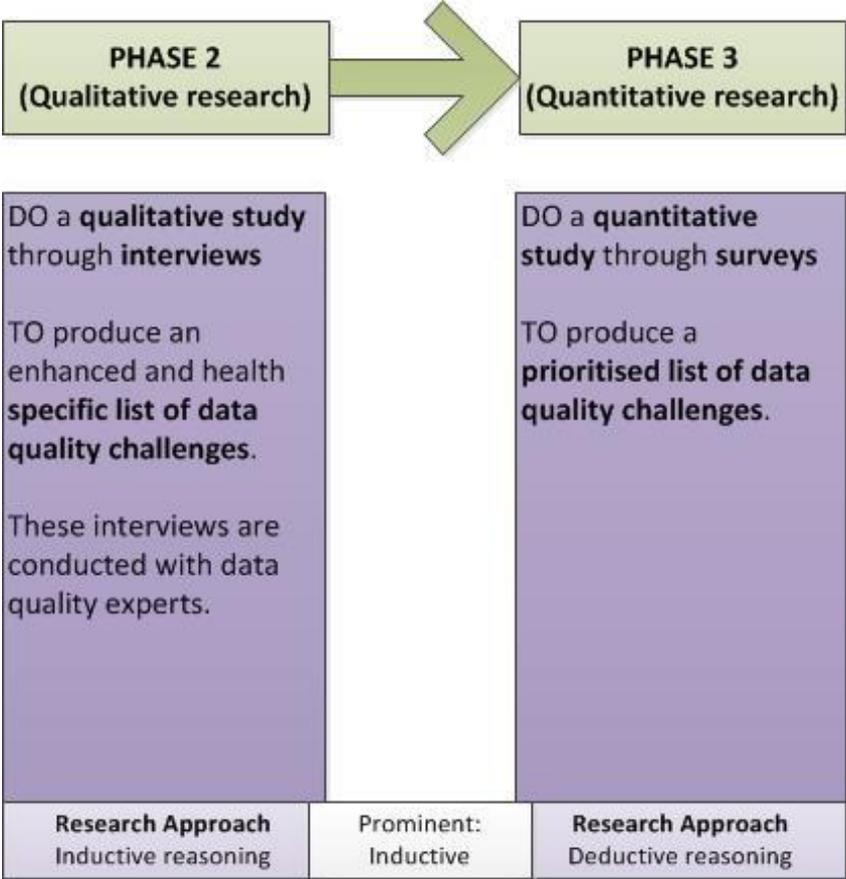


Figure 5.10: Research approach

- **Inductive reasoning**

Phase 2 of this study follows the inductive research approach, since it is the qualitative phase of this study. This approach will be explained by means of the adapted graphical presentation below.

To draw a conclusion on an entire class or events, based on separate as well as individual facts, is known as inductive reasoning (Leedy & Ormrod, 2005). Inductive reasoning is be implemented in this study, where the purpose is to explore and describe data quality challenges in electronic healthcare systems in South Africa.

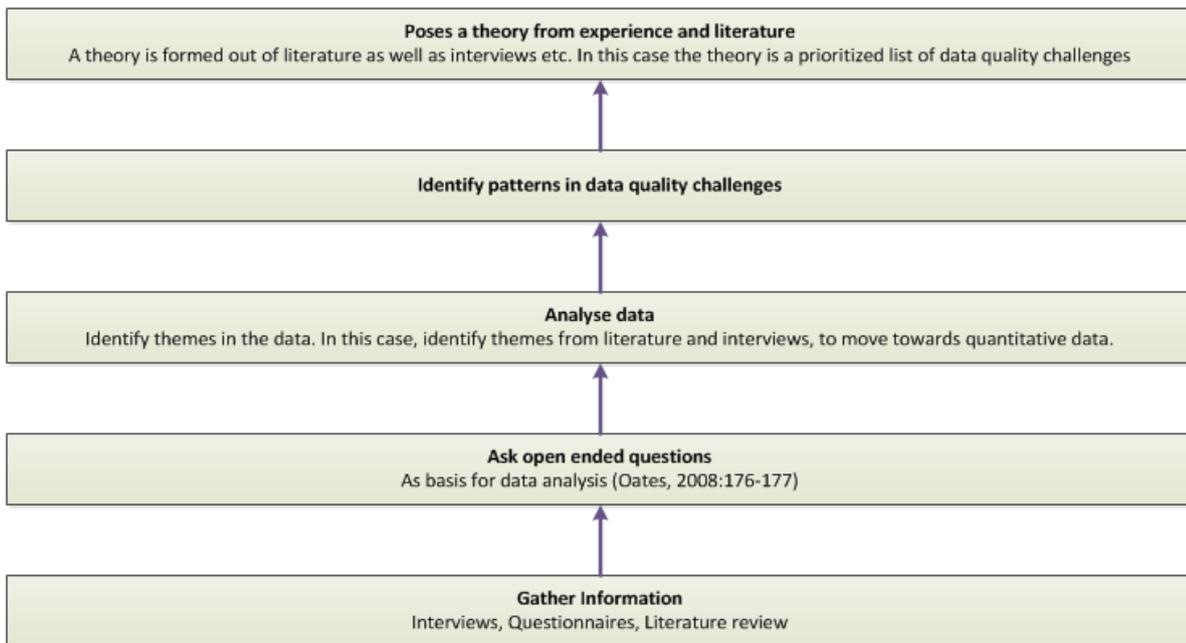


Figure 5.11: Inductive research process (Creswell, 2009)

- **Deductive reasoning**

Phase 3 of this study, follows the deductive research approach. This phase is the quantitative phase of the research, and therefore lends itself to deductive reasoning.

Since the qualitative approach of this study is dominant, this study mainly follows an inductive approach. The quantitative approach in this study is only used to rank or prioritise the data quality challenges in electronic healthcare systems in South Africa. Although numeric data emerged from the quantitative research, the data will not be used for anything other than ranking the challenges; therefore a deductive approach is less prevalent in this study.

The second layer of the research process onion is illustrated as follows in Figure 5.12:



Figure 5.12: Applied research approach (Saunders *et al.*, 2015)

The following layer of the research process onion, the research strategy, is discussed in Section 5.2.3.

5.2.3 RESEARCH STRATEGY

A research strategy is a defined process or technique that should guide the entire research process, in order to answer the research questions (Oates, 2008; Yin, 2003).

5.2.3.1 RESEARCH STRATEGIES OF INQUIRY

Certain research strategies could be more appropriate for certain research designs. Table 5.2 pairs research strategies to the research designs (Ahmadnezhad, 2009) and shows which research strategies are used in the quantitative, qualitative and mixed method research designs:

Table 5.2: Research strategies with research designs (Ahmadnezhad, 2009)

Quantitative research design	Qualitative research design	Mixed method research design
<ul style="list-style-type: none"> Experiments 	<ul style="list-style-type: none"> Narrative research 	<ul style="list-style-type: none"> Sequential procedure
<ul style="list-style-type: none"> Surveying 	<ul style="list-style-type: none"> Phenomenological research 	<ul style="list-style-type: none"> Concurrent procedure
	<ul style="list-style-type: none"> Ethnography 	<ul style="list-style-type: none"> Transformative procedure
	<ul style="list-style-type: none"> Grounded theory 	
	<ul style="list-style-type: none"> Case studies 	
	<ul style="list-style-type: none"> Action research 	

The key research strategies will be discussed below (Creswell, 2009; Oates, 2008):

- Experiments* can be done either in a laboratory or in the field (Oates, 2008). An experiment is based on cause and effect relations and usually includes variables and measurements (Oates, 2008). There are two types of experiments, namely true experiments and quasi-experiments. In true experiments the subjects are allocated randomly to treatments. In quasi-experiments, the allocation of the subjects' treatments is not done in a random manner (Keppel, 1991). For an experiment to be of value, a substantial amount of subjects or participants are required (Oates, 2008).
- Surveying* is a research strategy which provides the same type of data from a large group of participants in a consistent manner (Oates, 2008). Questionnaires and/or interviews are used for data collection in order to generalise findings from a sample to give meaning to a population (Oates, 2008; Olivier, 2004). In order to support the generalisation of the findings, the participants or sample should be of a significant size (Oates, 2008).
- Narrative research* is a research strategy where the researcher tells the story of a research participant. The researcher usually gains the information by asking the participants to tell a story about their lives (Clandinin & Connelly, 2000). Narrative research is mainly used in fields such as behavioural studies and social sciences (Rosaldo, 1989).

- *Phenomenological research* is a research strategy where a certain phenomenon is identified by exploring the perceptions of the participants. Phenomenology is focussed on experiences of humans or participants of the research (Measor, 1985). The data gathering is done through personal engagement and relationships with the participants (Moustakas, 1994).
- *Ethnography* is the study of a cultural group in their natural location over a certain period of time and data is mainly collected by means of observation (Creswell, 2009; Fetterman, 2010). Ethnography provides information in a specific context and may usually not be generalised (Oates, 2008). This research process is flexible and may differ between different contexts (LeCompte & Schensul, 1999).
- *Grounded theory* is a specific approach to qualitative research and includes field research and the analysis of data, in order to ensure that a certain theory in the field is grounded or true (Oates, 2008). Grounded theory has two main characteristics: continuous assessment of data with developing classifications and theoretical sampling of different clusters to optimise the similarities and differences of the data (Creswell & Clark, 2011).
- *A case study* is a research strategy where a certain event, action, programmes, procedures, a group of people or individuals are being explored by the researcher. Data collection methods may include various forms, such as interviews, questionnaires and observations over a certain period of time (Yin, 2003). Yin (2003) further defines a case study as the investigation of a certain phenomenon within its real/actual setting and context.
- *Action research* is a research strategy where a problem or a challenge in a certain area is recognised by the researcher. The process of action research comprises the identification of a problem and the implementation of certain actions to solve identified problem (Oates, 2008). The applied actions should be validated in terms of effectiveness and this is done by observing the outcomes. The process continues until a significant outcome is reached (Cooper & Schindler, 2003).
- *Sequential procedure* is a research strategy where the researcher makes use of joint procedures or research designs to answer the research questions. With a sequential procedure, two methods follow on each other. It may be that a

qualitative method is done first, followed by a quantitative method to generalise the results. A quantitative method may also be done first, followed by a qualitative method, to gain more detail on the initial findings (Creswell & Clark, 2011; Creswell, 2009).

- *Concurrent procedure* is a research strategy where the researcher makes use of joint procedures or research designs to answer the research questions. With a concurrent procedure, two methods are followed at the same time. The results are presented based on overall findings that emerged from both methods. Qualitative and quantitative methods are used concurrently, but may be used to answer different research questions (Creswell & Clark, 2011; Creswell, 2009).
- *Transformative procedure* is a research strategy based on a theoretical perspective, which is used as a framework for the rest of the research. Mixed methods are often used to support the theoretical background. Either a sequential or concurrent approach can be used in transformative procedures (Creswell & Clark, 2011; Creswell, 2009).

5.2.3.2 APPLIED RESEARCH STRATEGY

The research strategies applied in this study are described by referring to the research phases.

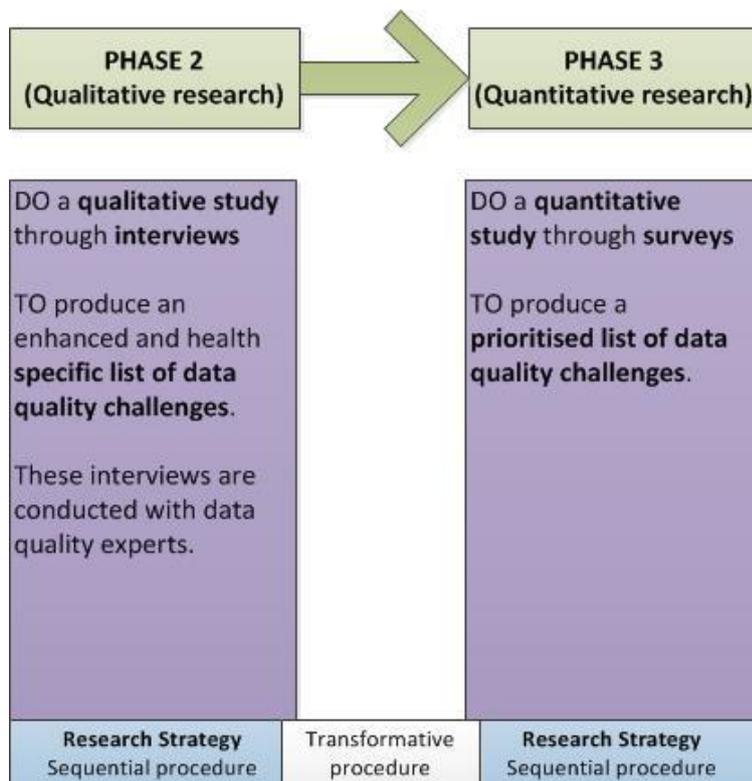


Figure 5.13: Research strategy

Since Phase 2 and Phase 3 follow on each other, the appropriate research strategy for this study is a sequential procedure.

Sequential procedure is a research strategy where the researcher makes use of joint procedures or research designs to answer the research questions. With a sequential procedure, two methods follow on each other. It may be that a qualitative method is done first, followed by a quantitative method to generalise the results. A quantitative method may also be done first, followed by a qualitative method, to gain more detail on the initial findings (Creswell & Clark, 2011; Creswell, 2009).

The research is done by following a sequential QUAL-quan design. The qualitative research included interviews to gather in depth data regarding data quality challenges in electronic healthcare systems. Thereafter, a quantitative study was conducted in order to prioritise these data quality challenges. The qualitative research is dominant, since the main data emerged from literature surveys and interviews. The quantitative research is less prevalent, since it is only used for the prioritisation of the data.

Sequential procedures, as depicted above, are a form of the mixed method approach. In this case, the research will begin with a qualitative approach, for exploratory reasons, and will be followed up by a quantitative approach with a large sample, in order to

generalise the results (Creswell & Clark, 2011; Creswell, 2009). In this case, generalisation of results refers to the prioritisation of the data quality challenges that emerged from the qualitative approach.

The sequential exploratory design is explained in Figure 5.14:

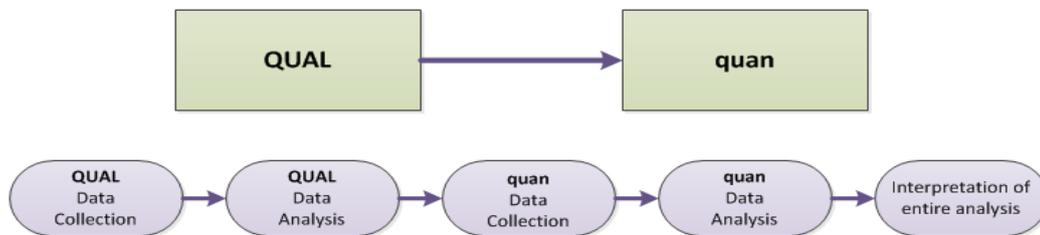


Figure 5.14: QUAL-quan process (Creswell, 2009)

This figure explains that the first phase of the research is to collect qualitative data. The second phase consists of analysis of the qualitative data. The third phase is the collection of quantitative data, whereas the fourth phase entails the analysis of the quantitative data. The final phase of the research is the interpretation of the entire analysis, resulting into the final outcome which in this case is a prioritised list of data quality challenges in electronic healthcare systems in South Africa. A set of evidence-based recommendations is provided.

Transformative procedure is a research strategy based on a theoretical perspective, which is used as a framework for the rest of the research. Mixed methods are often used to support the theoretical background. Either a sequential or concurrent approach could be used in transformative procedures (Creswell & Clark, 2011; Creswell, 2009).

This study is a transformative procedure, since an in depth literature review is used as an initial list of possible e-health data quality challenges for the study (Chapter 4). Both the qualitative and quantitative phases of the research is based on theory from literature.

- **Sampling frame**

Qualitative sampling frame: The interviewed data quality experts were identified using LinkedIn. People who indicated data quality as an expertise on LinkedIn were asked to participate in the study by agreeing to be interviewed on data quality and data quality

challenges. The participants in this study were purposefully chosen to include only those who have experience with data quality and data quality challenges.

Quantitative sampling frame: The data users of electronic healthcare systems (questionnaire respondents) were selected by visiting medical facilities in South Africa. The managers of these facilities were asked whether they make use of e-health systems and the questionnaires were given to medical facilities that indicated that they make use of electronic healthcare systems.

- **Sampling technique**

The qualitative interview used a non-probability purposeful sampling technique. The interviewees had to be data quality experts. The respondents had to be data users of electronic healthcare systems and therefore not all medical facilities could be used in this study. The sample was purposefully chosen to include only data users of electronic healthcare systems.

- **Response rate and non-response**

The researcher targeted four data quality experts on LinkedIn. The response rate for the qualitative interviews with data quality experts was 100%.

The researcher made physical visits to medical facilities to obtain research participants and mainly gave a few questionnaires to a facility manager. There were three main reasons for non-participation in the questionnaires:

1. Potential respondents did not have time to complete the questionnaire.
2. Potential respondents could not relate to any of the data quality challenges.
3. Potential respondents were not interested in participating in the study.

A total of 100 questionnaires were handed out for completion, with a total of 82 returned. Thus, the response rate for the questionnaires was 82%.

- **Sample size**

Qualitative sample size: Nielson and Launder (2003) suggest that three to five experts are appropriate in the domain of human-computer interaction. The percentages of errors that can be detected through expert reviews are displayed in the figure below (Ouma, 2014):

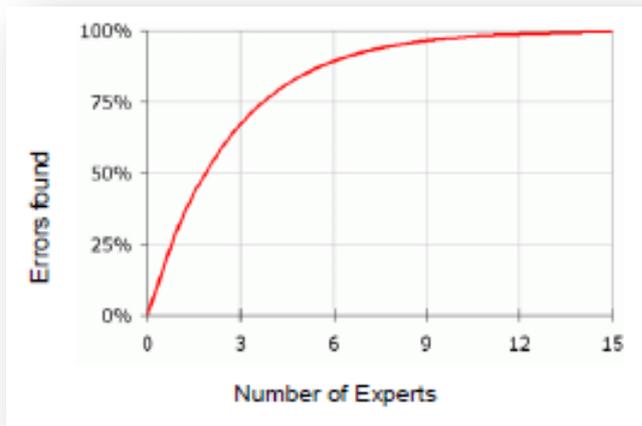


Figure 5.15: Expert review sample size (Ouma, 2014)

As depicted in the figure above, using three to five experts can detect approximately 85% of errors. The benefits of using more than five experts might not meet the costs. This study therefore uses four data quality experts for the qualitative phase of research.

Quantitative sample size: The final sample size of data users of electronic healthcare systems in South Africa that participated in the study was 82. Although the sample size may not practically be representative of all data users of electronic healthcare systems in South Africa, the sample size is statistically significant according to the following sample size formula (Falk, Marohn & Tewes, 2002): $n = 2 \times m + 1$, where n is the sample size and m is the number of questions on the questionnaire. The questionnaire consisted of 39 data quality challenge-related questions. Thus, for this study, a sample size of 79 is statistically significant. This study aims not to generalise, but rather to explore data quality challenges experienced by data users of electronic healthcare systems in South Africa.

The research strategy was discussed in this section and the applied third layer of the research process onion is as follows in Figure 5.16:

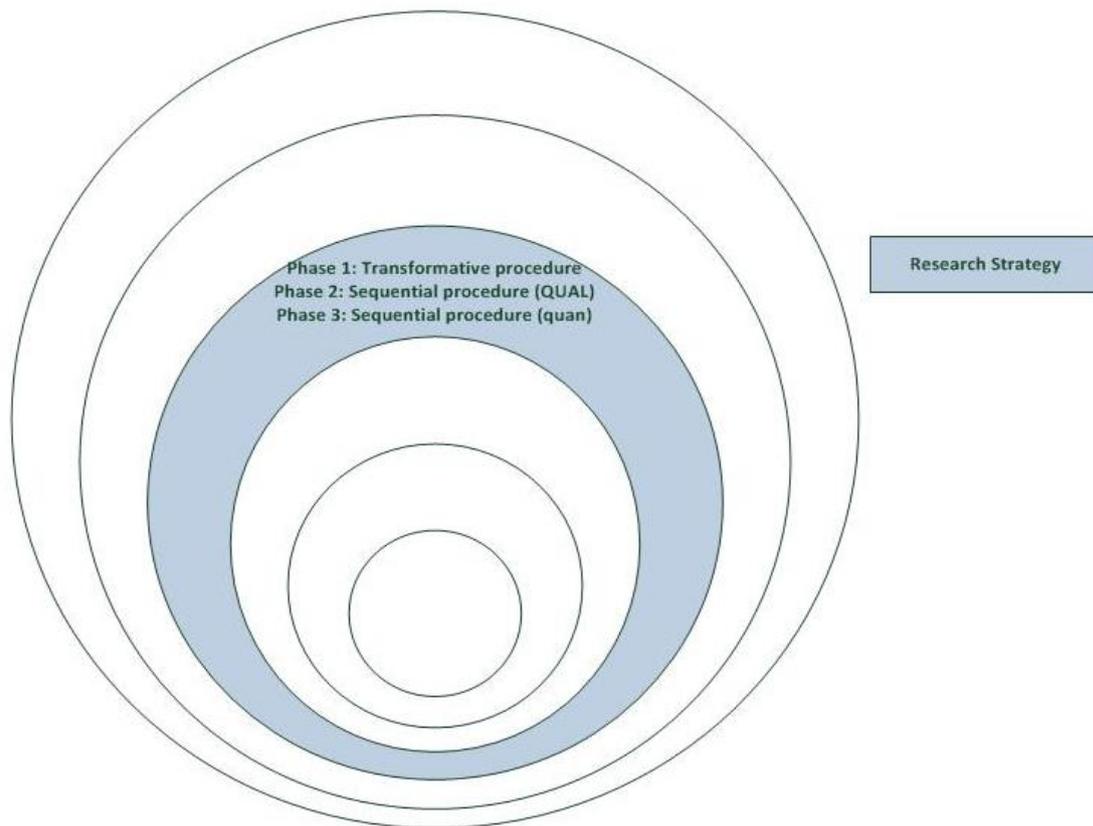


Figure 5.16: Applied research strategy (Saunders *et al.*, 2015)

The following layer of the research process onion, the research choice, is discussed in section 5.2.4.

5.2.4 RESEARCH CHOICE

Section 5.2.4 explains the research choices that were of inquiry, along with the applied research choice.

5.2.4.1 RESEARCH CHOICES OF INQUIRY

The different research choices are discussed in this section. Figure 5.17 illustrates the hierarchy of research choices.

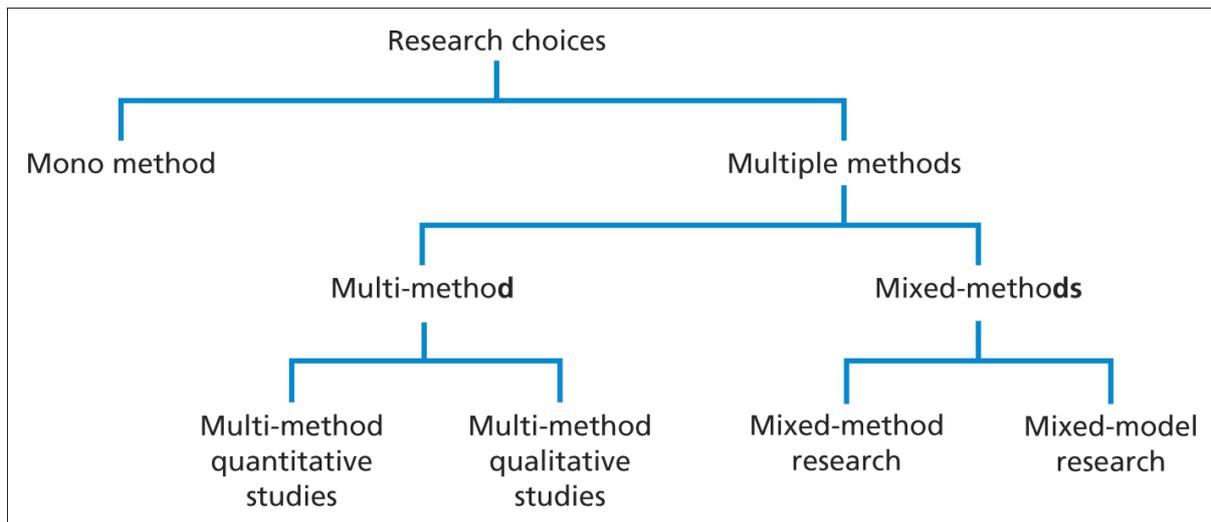


Figure 5.17: Research choices (Saunders *et al.*, 2015)

- *Mono-method*: Only one research method is used in the study.
- *Multi-method*: More than one qualitative method or more than one quantitative method is used in the study.
- *Mixed method*: Both qualitative and quantitative methods are used in the study.

The research choice consists of both qualitative and quantitative research methods:

- *Qualitative research*: Qualitative research is used for exploring and understanding a social or human problem through the perspective of individuals or groups. The research process includes the combining of processes and questions, inductive analysis, building from facts to more general themes and interpreting the data. A flexible structure is suitable for the final report (Creswell, 2009).
- *Quantitative research*: the researcher explains and describes general outcomes in a numerical manner (Leedy & Ormrod, 2005). Leedy and Ormrod (2005) further state that the process for quantitative research may be the same as qualitative research. Variables and statistical methods are used to analyse quantitative data. Patterns in data are used to conclude findings (Oates, 2008).

Table 5.3 briefly articulates the main differences between qualitative and quantitative research:

Table 5.3: Qualitative vs quantitative research (Mamhood, 2013)

Criteria	Qualitative research	Quantitative research
Purpose	To understand and interpret social interactions	To test hypotheses, look at cause and effect, and make predictions
Group studied	Smaller and not randomly selected	Larger and randomly selected
Variables	Study of the whole, not variables	Specific variables studied
Type of data collected	Words, images or objects	Numbers and statistics
Form of data collected	Qualitative data, such as open-ended responses, interviews, observations, field notes and reflections	Quantitative data based on precise measurements using structured and validated data-collection instruments
Type of data analysis	Identify patterns, features and themes	Identify statistical relationships
Objectivity and subjectivity	Subjectivity is expected	Objectivity is critical
Role of the researcher	Researcher and their biases may be known to participants and participant characteristics may be known to the researcher	Researcher and their biases are not known to participants and participant characteristics are deliberately hidden from the researcher (double blind studies)
Results	Particular or specialised findings that are less generalizable	Generalizable findings that can be applied to other populations
Scientific method	Exploratory or bottom-up; the researcher generates a new theory from the data collected	Confirmatory or top-down; the researcher tests the hypothesis and theory with the data
View of human behaviour	Dynamic, situational, social and personal	Regular and predictive
Most common research objectives	Explore, discover and construct	Describe, explain and predict
Focus	Wide-angle lens; examines	Narrow-angle lens; tests a

Criteria	Qualitative research	Quantitative research
	the breadth and depth of phenomena	specific hypothesis
Nature of observation	Study behaviour in a natural environment	Study behaviour under controlled conditions; isolate casual effects
Nature of reality	Multiple realities, subjective	Single reality, objective
Final report	Narrative report with contextual description and direct quotations from research participants	Statistical report with correlations, comparisons of means and statistical significance of findings

In the late 1980s and early 1990s the mixed method research design was adopted for the first time (Creswell & Clark, 2011). The mixed method research design is discussed in the section below.

The mixed method design is divided into the following designs (Mahmood, 2013):

- sequential exploratory design;
- sequential explanatory design;
- sequential transformative design;
- concurrent triangulation design;
- concurrent embedded design; and
- concurrent transformative design.

The applied research choice and design is discussed in Section 5.2.4.2.

5.2.4.2 APPLIED RESEARCH CHOICE

The applied research choice is discussed by referring to the research phases.

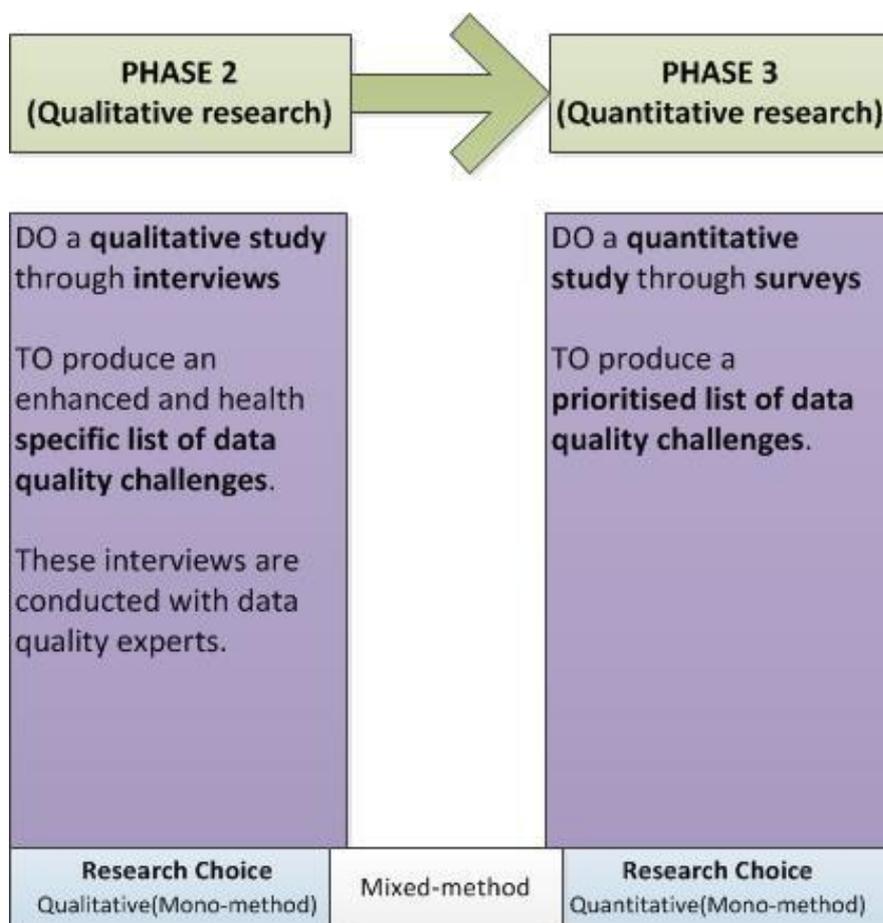


Figure 5.18: Research choice

- **Qualitative research**

Phase 2 of the research follows a qualitative method. Qualitative research focuses on day by day actions of people; social and administrative settings; and uses facts in the form of words rather than numerical data (Kaplan & Maxwell, 1994).

Since this study focusses on the perceived data quality challenges by data users of electronic healthcare systems in South Africa, this design is suitable for the second phase of research. The first stage of the research is to find out what data quality challenges data users experience.

- **Quantitative research**

Phase 3 of the research follows quantitative method. In addition to the first qualitative research phase, the researcher will also quantitatively expand on the experienced data quality problems and possible causes thereof. The second phase of this research is therefore positivistic (Maree & Van der Westhuizen, 2007).

The positivistic methodology states that hypotheses should be tested systematically and controlled. Experiments should inspect each possible casual factor to observe the effects, while at the same time it should exclude all other factors (Chapman, 2011).

The qualitative research results are used to prioritise the experienced data quality issues as well as the possible causes.

- **Mixed methods**

A mixed method research design is used to answer the research question as well as the secondary research questions. Qualitative as well as quantitative methods are used in this study. The combination of these two designs allows the researcher to provide comprehensive evidence of the research problem (Johnson, Onwuegbuzie & Turner, 2007).

Figure 5.19 illustrates the mixed method design in terms of time order and paradigm emphasis:

Note: "qual" = qualitative "quan = quantitative "+" = concurrent "-." = sequential Capital letters = high priority Lower case letters = low priority		Time order decision	
		Concurrent	Sequential
Paradigm Emphasis Decision	Equal Status	QUAL + QUAN	QUAL – QUAN QUAN - QUAL
	Dominant Status	QUAL + quan QUAN + qual	QUAL – quan qual – QUAN QUAN – qual Quan -QUAL

Figure 5.19: Mixed method design matrix (Johnson & Onwuegbuzie, 2004)

Johnson and Onwuegbuzie (2004) explain that a researcher should decide on a mixed method design by using the matrix above. The researcher should firstly choose a time order, which is concurrent or sequential. The researcher should also choose a paradigm emphasis, which is needed to decide whether the qualitative or quantitative approach will be more dominant, or if they will have equal weight. This study follows a sequential QUAL-quant research design and the fourth layer of the research process is illustrated in Figure 5.20:

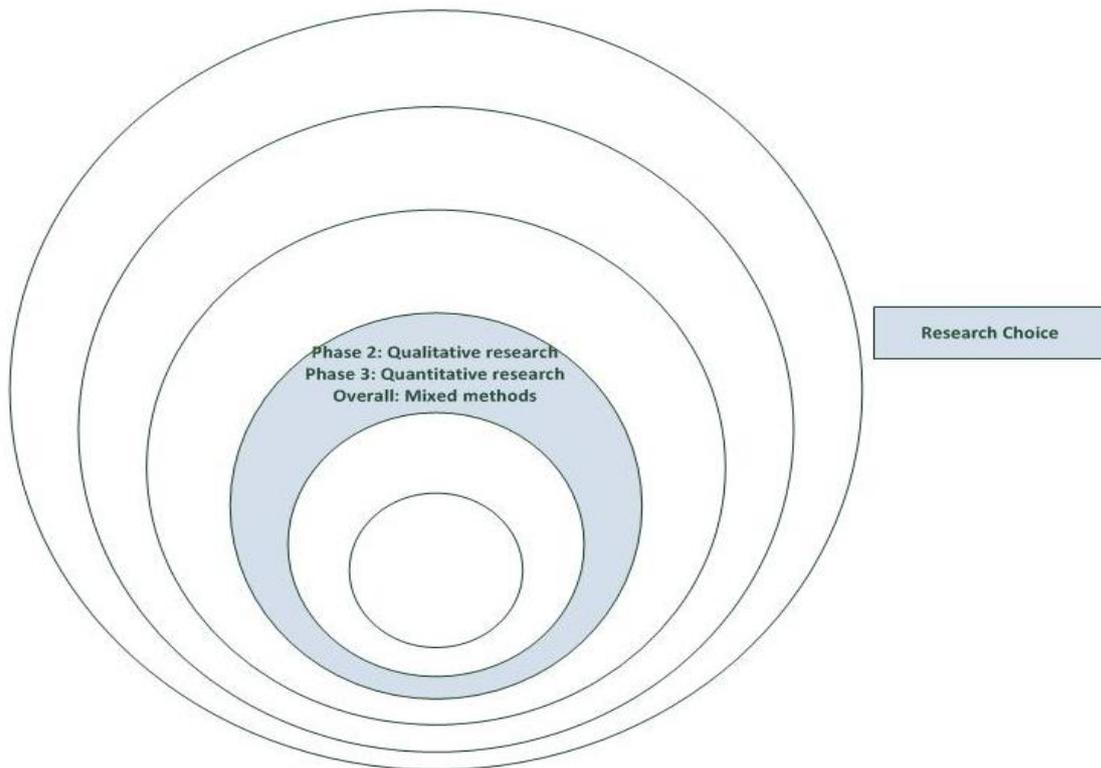


Figure 5.20: Applied research choice (Saunders *et al.*, 2015)

The next layer of the research process onion, the time horizon, is discussed in Section 4.2.5.

5.2.5 TIME HORIZON

The time horizon layer in the research process onion refers to when the research will be conducted.

5.2.5.1 TIME HORIZONS OF INQUIRY

- *Cross-sectional research* is conducted at a specific point in time. This data can be from the past, present or future (Donnelly & Trochim 2007).
- *Longitudinal research* is conducted over a certain time period (Donnelly & Trochim, 2007).

5.2.5.2 APPLIED TIME HORIZON

This study can be regarded as cross-sectional research. The study took place in a specific point in time, since data quality challenges as perceived by data users of electronic healthcare systems in South Africa were collected. The data quality challenges experienced in the present is of importance to the study.

The fifth layer of the applied research process onion is given below:

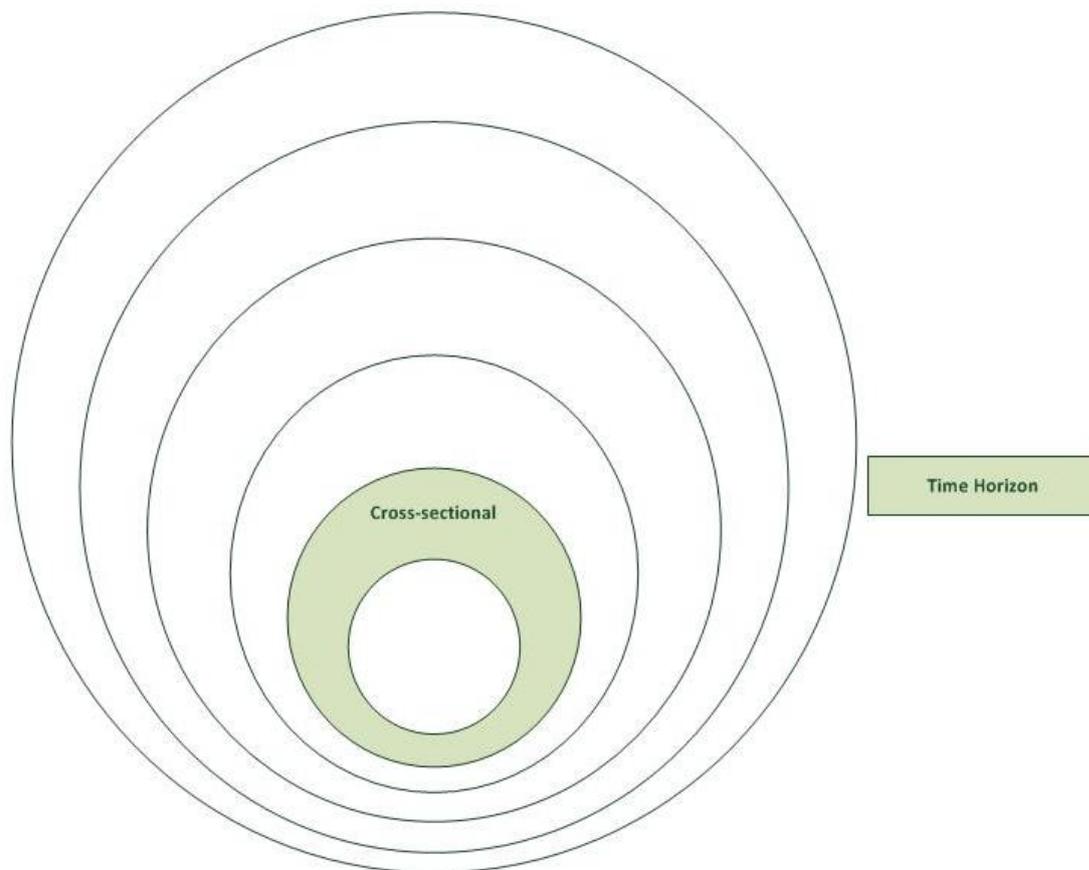


Figure 5.21: Applied time horizon (Saunders *et al.*, 2015)

The final layer of the research process onion, techniques and procedures, is discussed in Section 5.2.6.

5.2.6 TECHNIQUES AND PROCEDURES

The techniques and procedures layer of the research process onion refers to the data collection and data analysis techniques used in a study.

5.2.6.1 DATA COLLECTION METHODS AND ANALYSIS TECHNIQUES OF INQUIRY

A data collection method refers to the manner in which the researcher gathers relevant data in order to answer the research questions (Oates, 2008). There are various data collection methods, but the chosen method should adhere to the requirements of the research question (McNiff, 2002). There exist various data collection methods for qualitative as well as quantitative studies, namely questionnaires, interviews, document analysis, observations, literature reviews and focus group discussions.

5.2.6.2 APPLIED DATA COLLECTION AND ANALYSIS TECHNIQUES

By using more than one data collection method, the researcher is empowered to investigate the case in numerous ways (Oates, 2008). Therefore, this study uses two dominant data collection methods. Since the researcher is following an interpretivistic and positivistic research philosophy, and makes use of qualitative as well as quantitative research designs, the research lends itself to a mixed data collection approach. The data collection methods in this case include:

1. interviews; and
2. questionnaires.

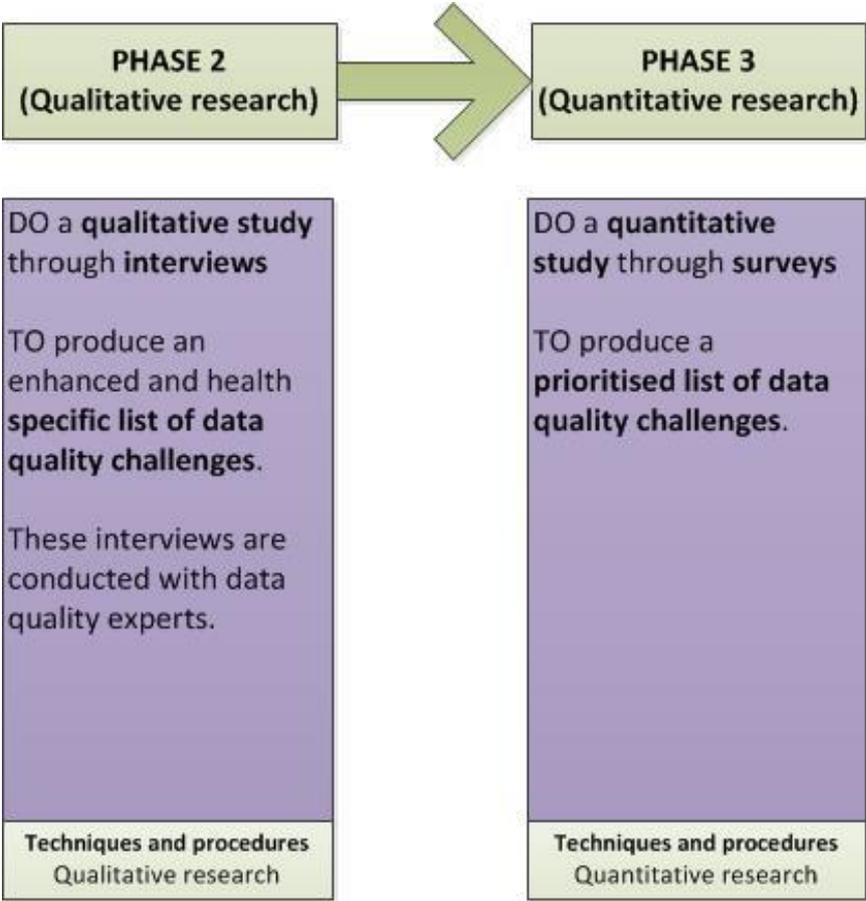


Figure 5.22: Techniques and procedures

- **Interviews**

Phase 2 of the research adapted to qualitative techniques and procedures. Interviews are a data collection method that contains oral questioning of respondents (Denscombe, 2001). In this case interviews involve the oral questioning of data quality experts in various countries and domains.

The interviews include discussions on the experts' perceived data quality challenges. The questions in the interview are conducted by using information obtained from the literature reviews. The interviews are semi-structured. A list of specific questions is conducted, but the structure was open to change with regard to the direction of the conversation. See Appendix B for the open interview questions.

- **Qualitative data analysis**

Qualitative data analysis uses themes and classifications in words to analyse data in order to provide meaningful results (Oates, 2008).

The data that results from the qualitative investigation is studied, manufactured, programmed and grouped. The data is analysed and grouped by using the open source software system *Hyper Research 3.7.1*.

The qualitative responses from the interviews with four data quality experts were transcribed and grouped by coding responses of the same questions and with regard to the same data quality challenges. Although software has been used as a tool for the qualitative data analysis process, the researcher applied a manual process of coding the qualitative data.

The interviews resulted in an enhanced and specific list of data quality challenges used in the quantitative questionnaires. The qualitative data analysis process is described in more detail in Section 6.3.

- **Hermeneutics**

According to Crotty (1998) hermeneutics is based on the interpretive research philosophy. Hermeneutics analyses the numerous sections of the text while taking the whole picture into account. It also analyses the whole picture while considering the numerous separate texts (Gadamer, 1998).

Phase 2 of the research is based on the interpretive research philosophy and therefore uses hermeneutics as a data analysis technique.

Klein and Myers (1999) provide the fundamental principles for conducting and evaluating interpretive studies. Table 5.4 summarises Klein and Myers' (1999) principles and indicates where in the study these principles are applied.

Table 5.4: Fundamental principles for conducting and evaluating interpretive studies (Klein & Myers, 1999)

Fundamental principle for conducting and evaluating interpretive studies	How and where applied in this study
<p>The fundamental principle of the hermeneutic circle This principle suggests that all human understanding is achieved by iterating between considering the interdependent meaning of parts and the whole that they form. This principle of human understanding is fundamental to all the other principles.</p>	<p>Applied in the qualitative data analysis in Chapter 6. The data collected from the four data quality experts were analysed in parts and as a whole by using themes and classifications to analyse the data, as prescribed by Oates (2006).</p>
<p>The principle of contextualization Requires critical reflection of the social and historical background of the research setting, so that the intended audience can see how the current situation under investigation emerged.</p>	<p>Applied in Section 6.4. The backgrounds of the interviewees are provided. This ensures that the research question is positioned within the context of the research setting.</p>
<p>The principle of interaction between the researchers and the subjects Requires critical reflection on how the research materials (or “data”) were socially constructed through the interaction between the researchers and participants.</p>	<p>The researcher served as interviewee. The data were socially constructed by asking questions and allowing the interviewees to answer the questions according to their background and opinions.</p>
<p>The principle of abstraction and generalization Requires relating the idiographic details revealed by the data interpretation through the application of principles one and two to general theoretical concepts that describe the nature of human understanding and social action.</p>	<p>No generalization applied to the study. The details revealed by the data were content specific. Questions specific to data quality challenges were proposed to the interviewees.</p>
<p>The principle of dialogical reasoning Requires sensitivity to possible contradictions between the theoretical preconceptions guiding the research design and actual findings (“the story which the data tell”) with subsequent cycles of revision.</p>	<p>The interpretations of the data, provided in Chapter 6, were done based on the literature reviews provided in Chapter 2 and Chapter 3.</p>
<p>The principle of multiple interpretations Requires sensitivity to possible differences in interpretations among the participants as are typically expressed in multiple narratives or stories of the same sequence of events being studied. Similar to multiple witness accounts even if all tell it as they saw it.</p>	<p>Interviews were sought from data quality experts in various countries (South Africa, England and Canada) and work settings (medical IT development and general data quality experts). The different settings were taken into account during analysis of the results (Chapter 6).</p>
<p>The principle of suspicion Requires sensitivity to possible “biases” and systematic “distortions” in the narratives collected from the participants.</p>	<p>The data were collected via interviews. The questions asked differed slightly between interviewees, according to their different settings and experiences.</p>

Hermeneutics as applied in this study is discussed in more detail in Section 6.5.

- **Questionnaires**

Phase 3 of the research entailed quantitative techniques and procedures. Questionnaires are used to obtain numerical data in order to compile a prioritised list of data quality challenges experienced by data users of electronic healthcare systems in South Africa. The questions in the questionnaires were closed-ended and derived from the information obtained from the literature review as well as the qualitative results. The questions were conducted to make it possible for the researcher to prioritise the identified data quality challenges. See Appendix D for the questionnaire used in the study.

- **Quantitative data analysis**

Quantitative data analysis uses statistical and mathematical methods to analyse the data and to get meaningful numeric results (Oates, 2008).

The data that results from the quantitative study is analysed by using statistical methods using Microsoft Excel. The questionnaire prompted respondents to provide ordinal data. Data quality challenges were listed on the questionnaire and respondents had the option to rate the challenge.

Statistically significant data quality challenges were identified by determining an appropriate level C confidence interval (Altman, Machin, Bryant and Gardner, 2014).

An appropriate level C confidence interval for p is $P \pm z^* \sqrt{\frac{p(1-p)}{n}}$ where z^* is the upper $(1-C)/2$ critical value from the standard normal distribution (Altman *et al.*, 2014).

The data quality challenges which proved to be regarded as challenges by more than 50% of the participants are included in the list of prioritised data quality challenges.

To prove that each item is regarded as a data quality challenge by more than 50% of the participants, a 95% confidence interval has been determined for the proportions.

The statistically significant data quality challenges were prioritised by determining the weighted total score for each item. The weighted total score is determined by summing the ordinal values multiplied by the frequency of responses.

The quantitative data analysis process is described in more detail in Section 7.3.

- **Triangulation**

The combination of more than one data source, methodological approaches or analytical methods is known as triangulation (Kimchi, Polivka & Stevenson, 1991). Triangulation is done to strengthen data and to provide reliability to the research findings (Denzin, 1970).

In this study, triangulation is used to generalise the findings from the qualitative research (Phase 2) and quantitative findings (Phase 3).

According to Thurmond (2001) there are various types of triangulation, namely:

- data sources triangulation;
- investigator triangulation;
- methodological triangulation;
- theoretical triangulation; and
- data analysis triangulation.

This study applies various types of triangulation. The triangulation applied in this study is illustrated in Figure 5.23:

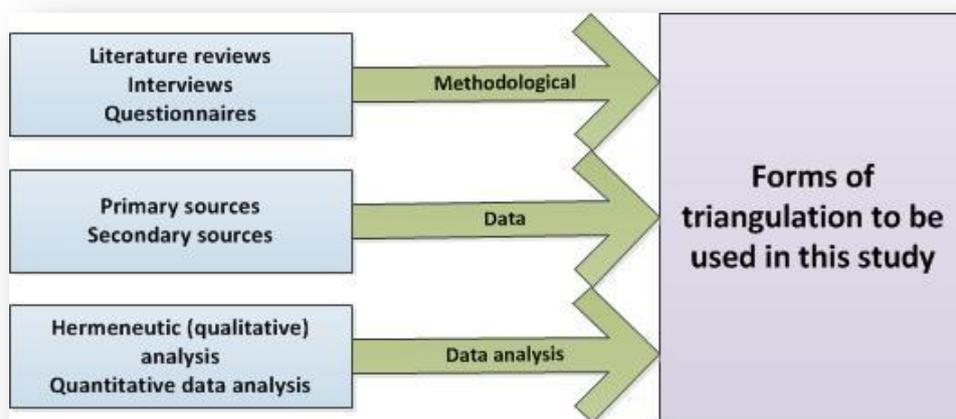


Figure 5.23: Forms of data triangulation used in the study

The data triangulation applied in this study, is described in more detail in Section 7.3. The techniques and procedures of this study were described in this section. The final layer of the applied research process onion is displayed in Figure 5.24:

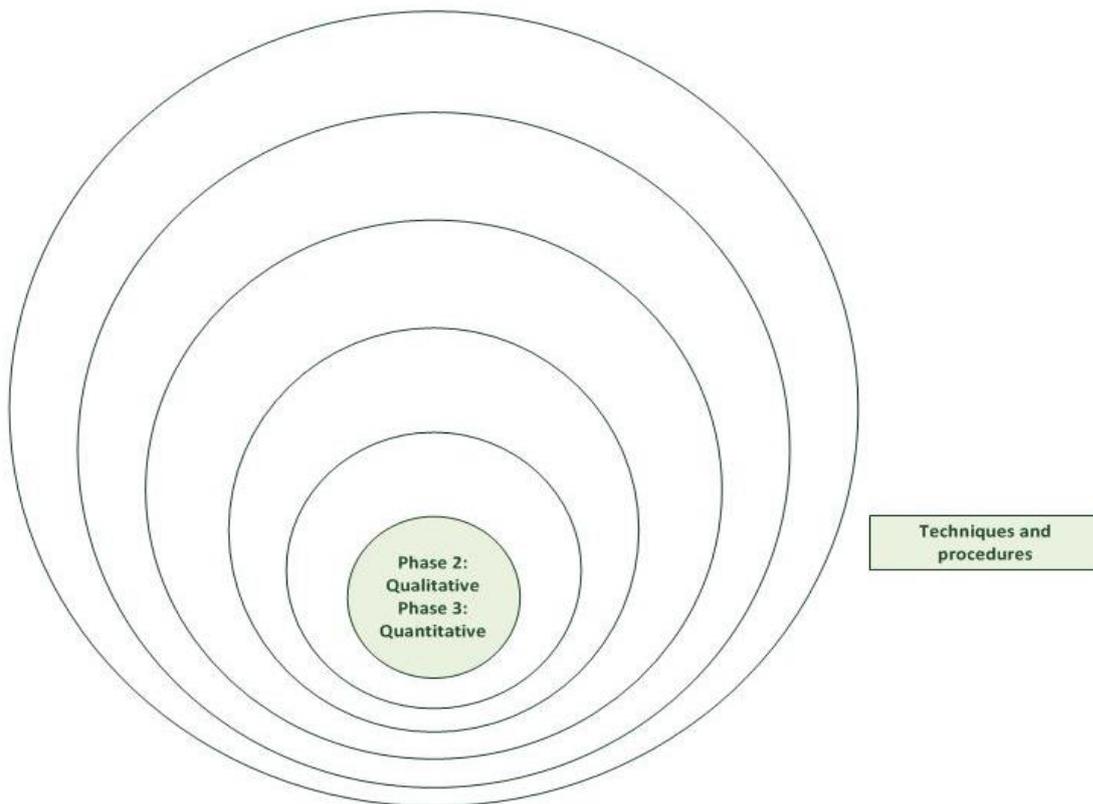


Figure 5.24: Applied techniques and procedures (Saunders *et al.*, 2015)

The complete applied research process onion is given in Section 5.3.

5.3 APPLIED RESEARCH PROCESS SUMMARY

The applied research process, as discussed in Section 5.2 can now be summarised in the form of the research process onion.

Figure 5.23 illustrates the applied research process onion of this study:

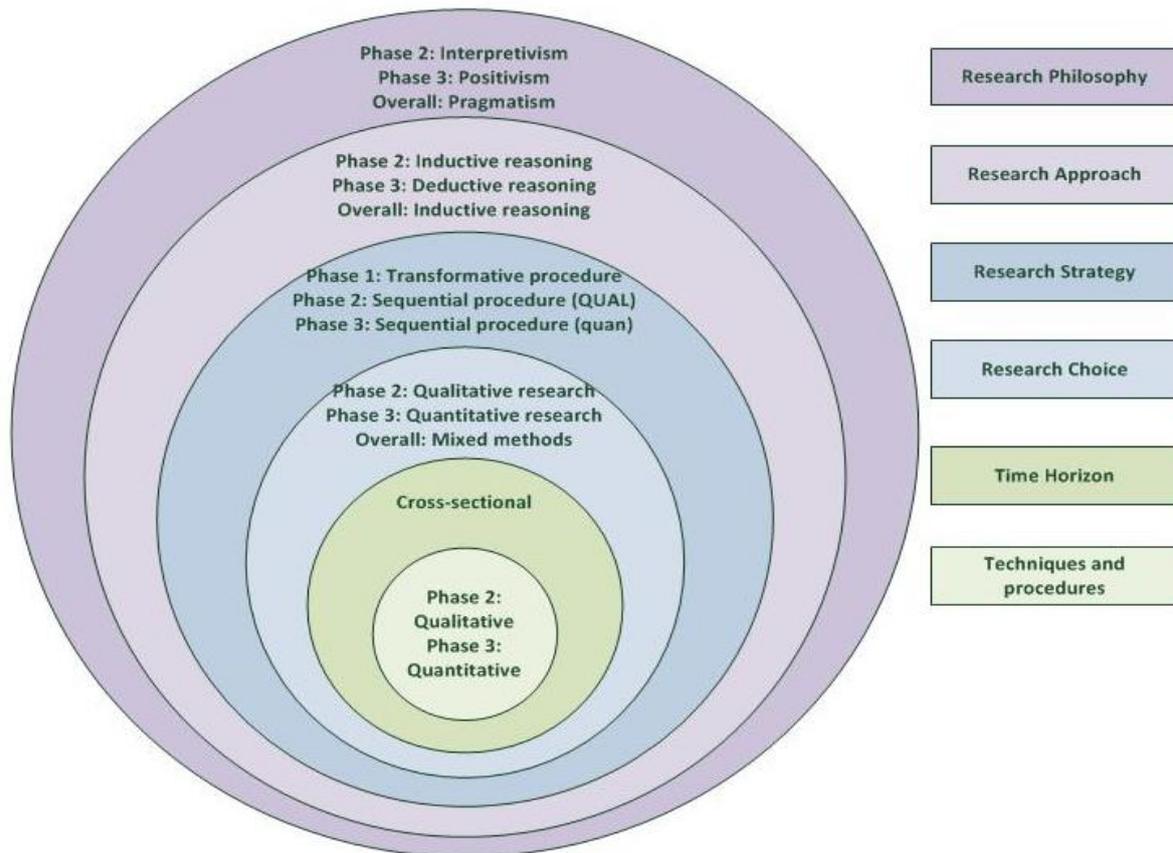


Figure 5.25: Applied research process onion (Saunders et al., 2015)

Pragmatism is the philosophy followed in this study. The pragmatist philosophy is suitable for this study, since both qualitative and quantitative philosophies were followed respectively in Phase 2 and 3 of the research. Although inductive as well as deductive reasoning were followed in this research, the quantitative study was less prevalent; therefore the inductive research approach is more suitable for the entire study. This study follows a sequential exploratory mixed method QUAL-quant research design. The qualitative study took place in the second phase of research, whereas the quantitative study only took place after the qualitative study has been completed; therefore sequential procedure is a suitable research strategy for this study. The qualitative study emerged from the identified possible e-health data quality challenges, which makes a transformative procedure a suitable research strategy for this study. Phase 2 and Phase 3 distinctly follow mono-methods (qualitative and quantitative). The final research choice is an exploratory sequential mixed method procedure. The study had a cross-sectional time horizon, since it took place in a specific point in time. The data collection and analysis techniques and procedures were applied respectively in Phase 2 and Phase 3

of the study. Qualitative techniques applied to Phase 2, whereas quantitative techniques applied to Phase 3 of the research.

5.4 ETHICAL CONSIDERATIONS

It is essential that the researcher conducts research in a direct and open way, in all studies (Olivier, 2009). The researcher explained the significance and objectives of the study to the participants and read the following rights to the participants of the study (Strode, Slack, Grant, Bamjee & Mushariwa, 2009):

1. The participant has a right to not take part in the study.
2. The participant has the right to withdrawn at any time during the study.
3. The participant has the right to agreement before the study starts.
4. The participant has the right to request that any information should be kept private or not to be used in the study.

The participants included data quality experts and data users of electronic healthcare systems. The researcher collected information regarding data quality challenges experienced by data users of e-health systems. These data users included nurses, doctors, community health workers, data entry clerks, administrative staff and any other staff members that work with data. No patients were used as participants in the study.

No patient data was requested. The content of the electronic data is irrelevant in this study; the researcher focussed only on the data quality challenges experienced by data users. See Appendix A for the interview consent form and Appendix C for the questionnaire consent form used in the study.

5.5 CONCLUSION

This chapter discussed general explanations of the research philosophy, research design, research strategy, research approach, data collection methods and data analysis techniques.

This study follows a positivist, interpretivist and pragmatist research philosophy. An exploratory sequential mixed method design is adopted in this study. Since the research adapts to a bottom-to-top approach, the inductive research approach is followed. Qualitative as well as quantitative data collection methods and data analysis techniques are used in this study. Chapter 6 describes the qualitative research findings.

CHAPTER 6: QUALITATIVE RESEARCH FINDINGS

Stage in research:

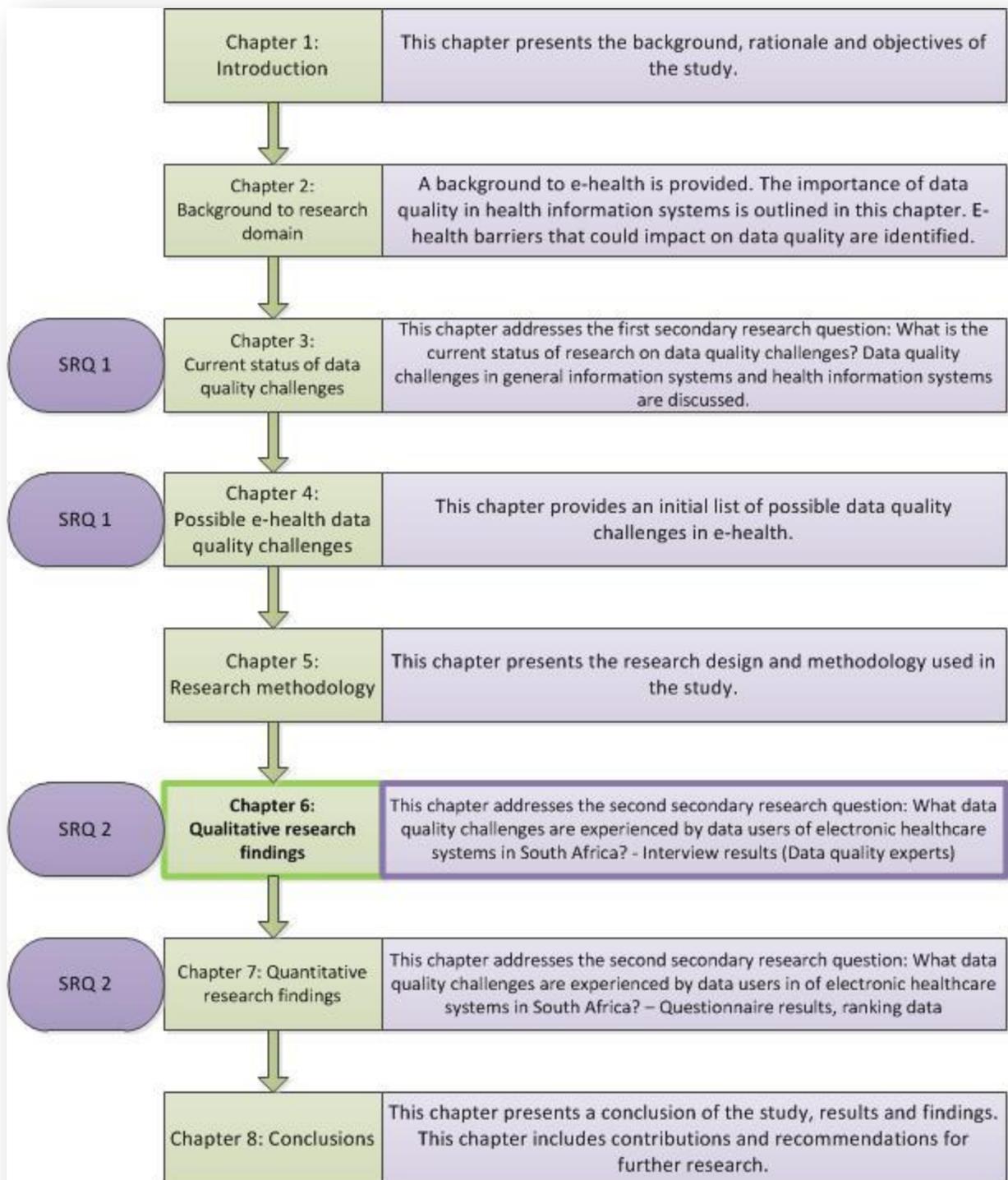


Figure 6.1: Chapter 6 – stage in research

Structure of Chapter 6:

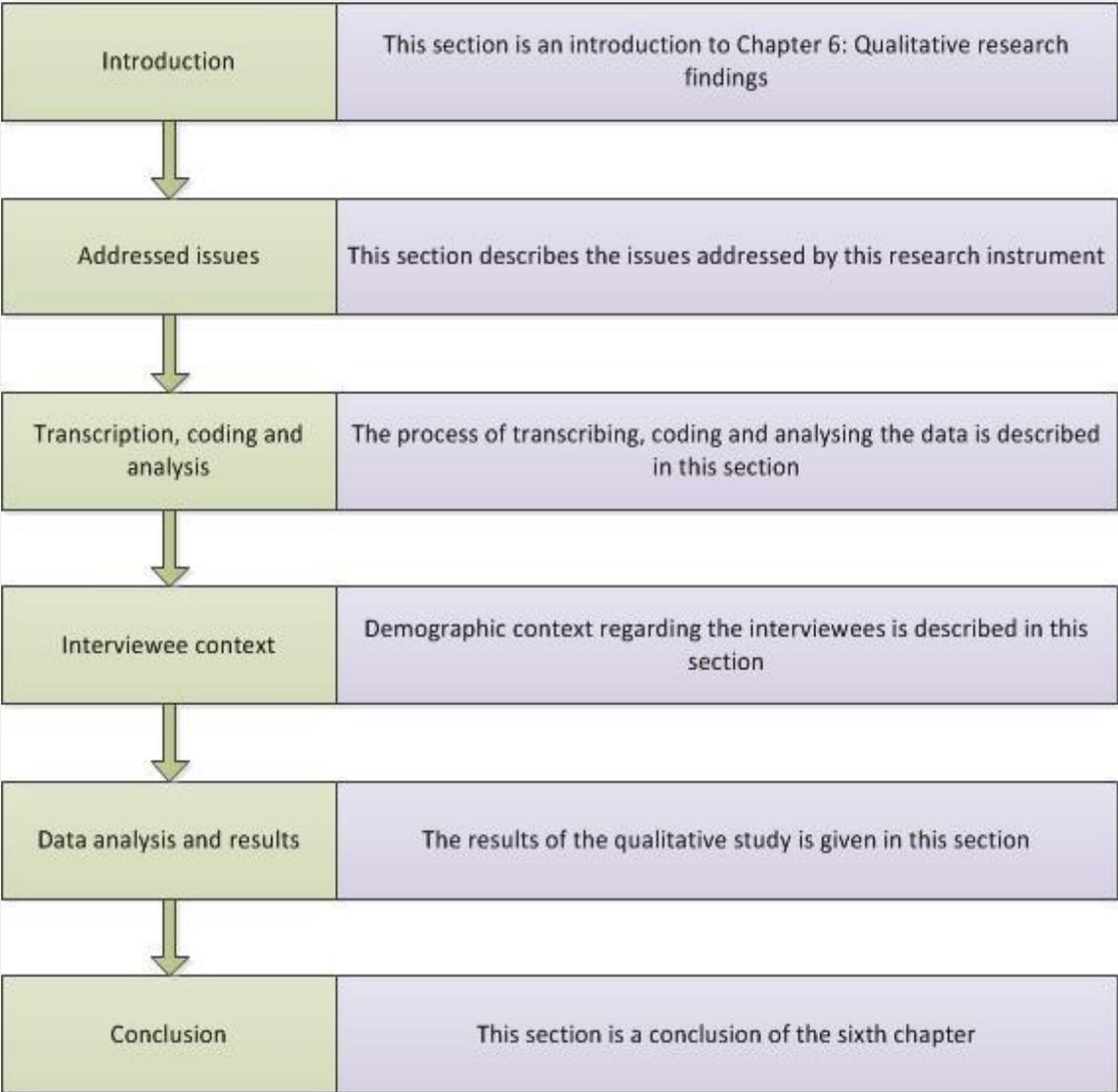


Figure 6.2: Structure of Chapter 6

6.1 INTRODUCTION

In Chapter 5 the research methodology which has been applied in the study was presented. This study adopted a sequential QUAL-quant mixed methods research design to answer the research question. Chapter 6 presents a roadmap to indicate how the study arrived at the qualitative findings presented here. The qualitative study was carried out by interviewing four individual data quality experts, within and outside the e-health domain. The primary purpose of the qualitative phase of the study was to provide an enhanced and practise-specific list of data quality challenges. The product of the qualitative study is an enhanced list of data quality challenges, which emerged from the literature review (see Chapter 4). The purpose was realised by means of a data collection exercise. The instrument used during the qualitative data collection was semi-structured interviews.

In this chapter the results obtained from the qualitative research instrument and the result patterns are presented. These findings are used to partially address the second secondary research question, namely: *What data quality challenges are experienced by data users in electronic healthcare systems in South Africa?* The qualitative research presented in this chapter is used as a basis to finally answer this research question by means of a quantitative study (see Chapter 7). The questions used in the quantitative study are derived from the results of the qualitative study presented in this chapter.

Phase 2 of the research process is presented in Chapter 6. Figure 6.3 illustrates the current phase of the research:

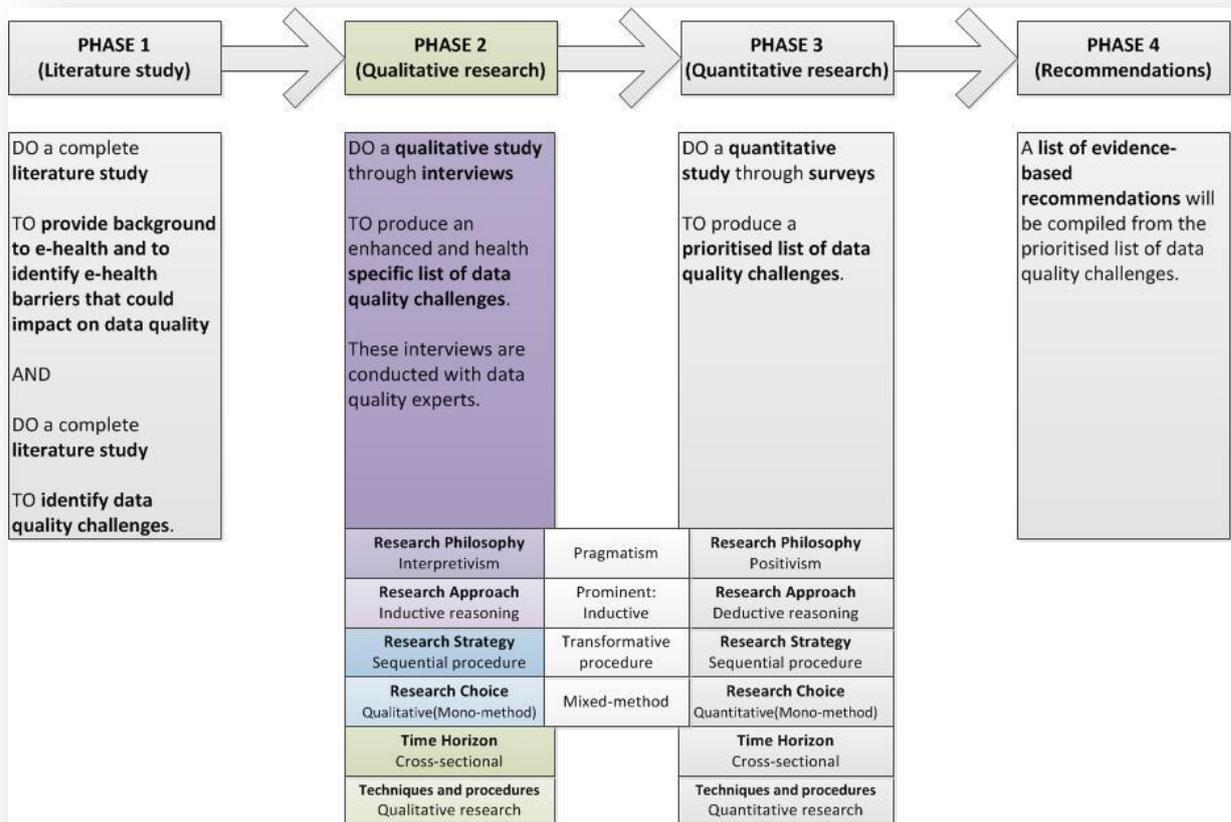


Figure 6.3: Current phase in research

Chapter 6 presents the findings of qualitative interviews to produce an enhanced and health-specific list of data quality challenges.

The remainder of this chapter is structured as follows: after the introduction a description of the issues addressed by this research instrument follows. Consequently a brief description of the process that has been followed to transcribe code and analyse the data is given. This is followed by a brief description of the context of the interviewees' work to provide the reader with a background to the data quality experts who participated in the qualitative research. The data analysis and results are given and finally the chapter conclusion is presented.

This chapter provides the results that emerged from a qualitative study to provide an enhanced and practise-specific list of data quality challenges. The issues addressed by this research instrument follow in Section 6.2.

6.2 ISSUES ADDRESSED BY THE RESEARCH INSTRUMENT

In this study, data was generated from participants' experiences regarding data quality challenges. In order to obtain this information, the researcher made use of semi-structured interviews as data collection method. The participants in this study were purposefully chosen to include only those who have experience with data quality and data quality challenges, in order to compile an enhanced and practise-specific list of data quality challenges experienced by data users and data quality experts.

6.2.1 SEMI-STRUCTURED INTERVIEWS

Individual interviews were conducted in this study. To ensure consistency in the questions asked, the researcher used an interview guide (see Appendix B) which contained the questions posed to the interviewees. A total of four respondents were interviewed at times their schedules allowed. Two of the interviews were done using *Skype*. One interview was done using *GoToMeeting* and one interview was conducted at the respondents' office. The interviews were mainly used to compile an enhanced and practise-specific list of data quality challenges. Table 6.1 provides a summary of the themes and objectives of the interview questions.

Table 6.1: Questionnaire themes

Question number	Themes	Objectives
1-4	Data quality challenges	To identify data quality challenges experienced by the respondent.
5-7	Health data quality challenges	To draw a link between data quality challenges in general and e-health data quality challenges.
8	Importance of study	To verify the importance of the study.

The transcription, coding and analysis processes are discussed in Section 6.3.

6.3 TRANSCRIPTION, CODING AND ANALYSIS PROCESSES

Qualitative data analysis uses themes and classifications in words to analyse the data to provide meaningful results (Oates, 2008). The data that results from the qualitative investigation is studied, manufactured, programmed and grouped. The data is

analysed and grouped by using the open source software system *Hyper Research 3.7.1*.

The qualitative responses from the interviews with four data quality experts were transcribed and grouped by coding responses of the same questions and with regard to the same data quality challenges. Although software has been used as a tool for the qualitative data analysis process, the researcher applied a manual process of coding the qualitative data.

The researcher made use of *Hyper Research 3.7.1* to code data quality challenges mentioned in the interviews together with the interviewees views regarding the specific data quality challenge.

The results that were obtained from coding the interview data were used to exclude data quality challenges from the original literature list (Chapter 4), to add data quality challenges to the list and to combine similar data quality challenges.

In Section 6.4 background information of the interviewees are provided.

6.4 BACKGROUND OF INTERVIEWEES

This section provides background information of the participants in the qualitative study. This demographic information of the participants has been collected by means of a written personal summary requested from the researcher. This information was used by the researcher as a background to consider while coding and analysing the interview data. Contextualised analysis is possible when demographic information about participants is known. The participants in this study were purposefully chosen to include only those who have experience with data quality and data quality challenges. Braun and Clarke (2013) state the importance of background information in qualitative data analysis. Table 6.2 provides a summary with background data of the four participants of the qualitative study.

Table 6.2: Background data of respondents

Participant	Location	Industry	Job title
A	London, England	Data quality	Data quality expert
B	Johannesburg, South Africa	Oil, gas, chemical	Data quality analyst
C	Toronto, Canada	Data quality	Data quality expert

Participant	Location	Industry	Job title
D	Pretoria, South Africa	E-health	E-health and data quality expert

The interviewees were chosen to provide the researcher with experiences from data quality experts. As seen in Table 6.2, the respondents are from different geographical locations and industries, but they are all data quality experts. These experts' experiences are used to enhance and modify the list of data quality challenges identified in the literature (see Chapter 4).

The data analysis and results are given in Section 6.5.

6.5 DATA ANALYSIS AND RESULTS

The study employed hermeneutics as a data analysis technique (as described in Section 5.2.6). The hermeneutical circle, as provided by Gadamer (1998), is applied in this chapter.

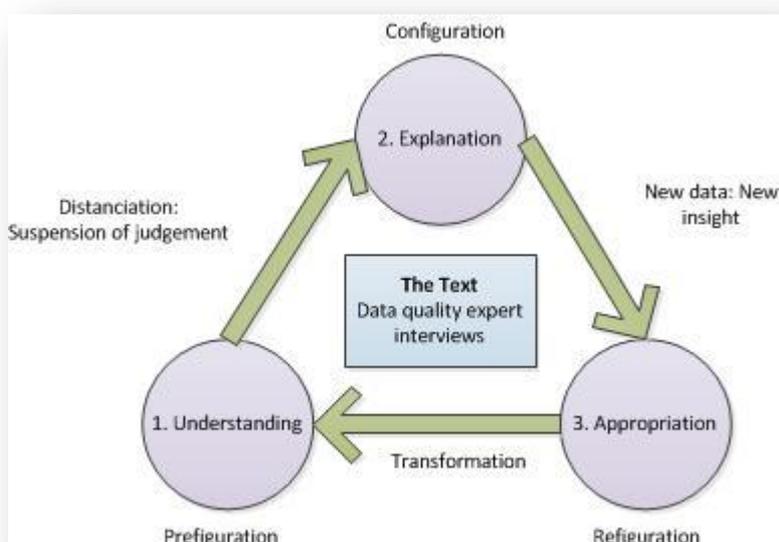


Figure 6.4: The hermeneutical circle (Gadamer, 1998)

The texts analysed are the interviews with four data quality experts. The researcher undergone the stages of i) understanding the text; ii) giving an explanation of the text; and iii) appropriating of the text. The remainder of this section describes the data analysis and results of the qualitative research phase.

The purpose of this chapter is to produce an enhanced and health specific-list of data quality challenges. This section describes the results obtained from the open-ended

interview questions conducted with data quality experts. The process shown in Figure 6.5 has been followed to obtain the final results of this chapter.

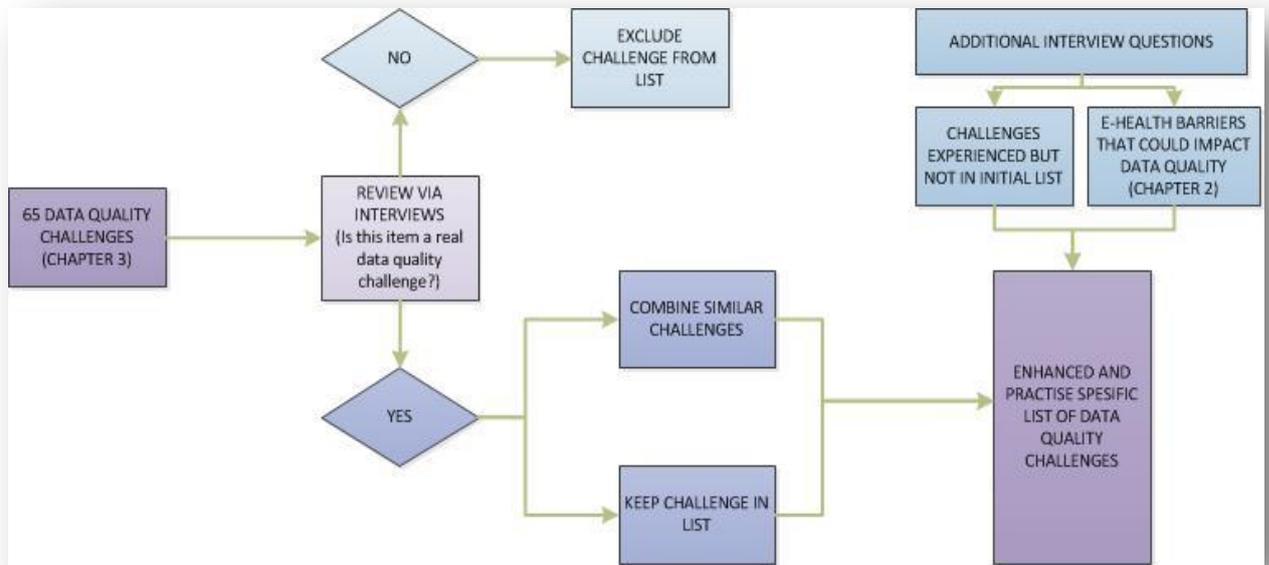


Figure 6.5: Qualitative phase processes

The process of obtaining an enhanced and practise-specific list of data quality challenges through interviews is illustrated in Figure 6.5.

As seen in Figure 6.5 the *enhanced list of data quality challenges* is obtained from two sources, namely: data quality challenges from literature, reviewed by interviewees and additional aspects related to data quality.

1. Additional aspects related to data quality (data quality challenges experienced by the respondents): data quality challenges experienced by one or more of the respondents which were not mentioned in the initial list are included in the enhanced list. These data quality challenges are discussed in more detail in Section 6.5.1.
2. Additional aspects of data quality specific to e-health (e-health barriers): additional aspects regarding e-health were proposed to the interviewees and questions with regard to data quality were asked. E-health barriers which have an effect on data quality are included in the enhanced list of data quality challenges. E-health barriers are discussed in more detail in Section 6.5.2.

3. Reviewed data quality challenges: The researcher asked the interviewees to indicate whether the data quality challenges identified in the literature are challenges that they experience. According to the responses, the researcher either excluded data quality challenges from the list, retained data quality challenges in the list or combined similar data quality challenges. The criteria for this reviewing process are discussed in more detail in Section 6.5.3.

6.5.1 DATA QUALITY CHALLENGES EXPERIENCED BY DATA QUALITY EXPERTS

Figure 6.5 illustrates that data quality challenges experienced by the interviewees are added to the enhanced list of data quality challenges. This section describes the data quality challenges experienced by these data quality experts and concludes by identifying which challenges should be added to the list of data quality challenges. Figure 6.6 illustrates the part of the process analysed in this section:

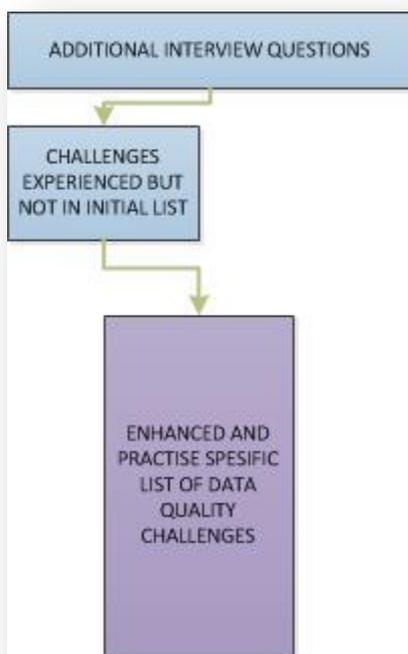


Figure 6.6: Challenges experienced but not in initial list

The researcher firstly asked the respondents to describe in their own words the data quality challenges that they have experienced during their careers. The data quality challenges obtained from these questions will be included in the list of data quality challenges if not included in the list obtained from literature.

The most important challenges experienced by the interviewees, were transcribed by using the process described in Section 6.3. For the purposes of context and readability,

direct quotations from the interviewees as well as a summary of the responses are given in this section.

6.5.1.1 FINDINGS

Each of the interviewees provided the researcher with major data quality challenges they experience every day. The main data quality challenges experienced by these data quality experts are:

- **IT architecture**

Respondent A: “The data quality challenges I used to find, a lot of the problems were created the wrong, I guess IT architecture.... So you always have different systems, right? And what you’ll end up having is just silos of information. And they get out of sync. So synchronization is a really big data quality problem. That’s one of the biggest ones, I would say.”

- **Lack of standards between systems**

Respondent A: “it’s just with lack of standardisation of data quality rules are quite common as well. Other issues are things like standards, yeah, data standards; so one department will have a standard for naming a piece of equipment or location and another department will have a completely different naming convention. So, those are some of the issues I see. Just a lack of synchronization between systems, a lack of standards...”

Respondent D: “I think another big problem is standardisation. The fact that the data sources aren’t standardised really leads to poor data quality.”

- **Outdated technology**

Respondent A: “another big issue is outdated technology as well. Just technology that where, the big problem is where you would have a data model, when it was initially conceived it was mapped to a business model for a business at that time... So, what happens sometimes is you get this drift gap emerging between what the original design of the system was for and how it is servicing the business now. And that’s where you get a lot of data quality issues being created as well.”

- **Data collection methods**

Respondent A: “. It’s a big problem. In terms of this data collection thing you’ve got going on, the e-health thing – some of the issues there around data collection would be things like, uhm, the company isn’t sharing data calls and validation rules, so for example, you’ll have one form has a postal address validation but another form doesn’t or one form looks for data in a certain format but another form doesn’t, it’s just with lack of standardisation of data quality rules are quite common as well.”

Respondent D: “the data collection methods currently in use, lead to data with poor quality. I think that there is a big need to standardise the data collection methods used in the clinics. The different data collection methods really open the door for human error.”

- **Data ownership/governance**

Respondent B: “The next one is, well the challenge that we used to have, was that we didn’t have data owners for the specific data areas or data domains. ...If you don’t have ownership, then you don’t have someone to ensure that data quality is good.”

Respondent C: “The biggest quality measure I think that is missing is the measure of quality. There is no measure of quality...Data should be monitored, audited, controlled.... most organisations don’t actually even give people the trust to bring forward data quality issues.”

Respondent D: “Another issue is data ownership – people don’t want to own the data – they want to keep their names of the data. I mean, it contributes tremendously to poor quality data.” “, so I really think that management is the biggest data quality challenge we have, and it will really take a few years to resolve this issue.”

- **Lack of user skills**

Respondent D: “The data users don’t have sufficient skills and knowledge to produce high quality data.” “I think there should definitely be more training for everyone, from the higher level, like the managers, right down to the data capturers.”

- **Lack of resources**

Respondent D: "I would say that there is a big lack of people. We don't have enough people to capture data, for example. We don't have enough clinical staff to handle data and health related tasks... we have a shortage of people. There are just not enough people." "Another challenge we have at the clinics is that there is a lack of appropriate hardware."

- **Financial challenges**

Respondent D: "I think what I would say is very challenging are things like the financial barrier of implementing this. Money is a big challenge..."

6.5.1.2 SUMMARY

The data quality challenges identified by the data quality experts before seeing the initial list of data quality challenges, included: *IT architecture, lack of standards between systems, outdated technology, lack of user skills, lack of resources and financial challenges.*

From the challenges mentioned above, *a lack of standards between systems, a lack of user skills, IT architecture and a lack of resources* were included in the initial list of data quality challenges obtained from literature. Thus, challenges that should be included in the enhanced data quality challenges list are: *outdated technology and financial challenges.*

The discussions regarding e-health barriers are described and analysed in section 6.5.2.

6.5.2 ADDITIONAL ASPECTS REGARDING E-HEALTH AND DATA QUALITY

Figure 6.5 illustrates that e-health barriers, which could impact on data quality, are added to the enhanced list of data quality challenges. This section describes the importance of a specific domain on data quality challenges, as well as e-health barriers which could have an effect on data quality. Figure 6.7 illustrates the part of the process analysed in this section:

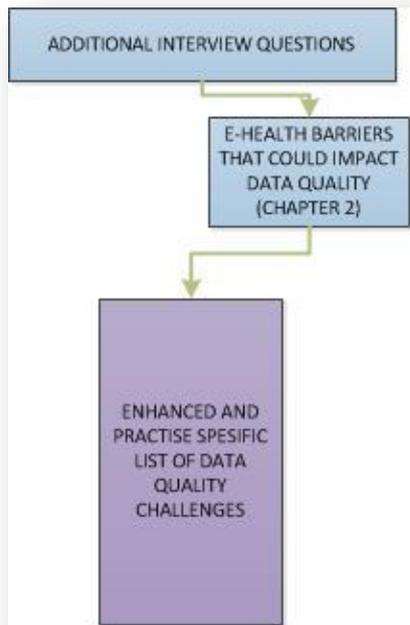


Figure 6.7: E-health barriers that can impact on the data quality

It is clear from the interviews that data quality challenges are similar across domains. Therefore the researcher can conclude that the data quality challenges identified in Chapter 2, are relevant in the e-health domain. Quotations from the respondents are given below to support the statement above:

Respondent A: "It's the same problems. You tackle it in the same way. The only thing that is different is the business context of the data. Some data is more critical than others. But the actual techniques and the problems you face, are the same I would say."

Respondent B: "Yes, it is generic over any domain."

Respondent C: "You find the same challenges, but the difference in culture milieu that they exist in and the capabilities for working with the data, consciousness of data quality can all be different. So it would be lovely if you could say that 'so everybody must do data governance', but it doesn't work because every department is different and the organisation is shifting all the time, the industry is shifting."

Respondent D: "Yes, I think so. I mean, healthcare, specifically in South Africa, has its own set of additional challenges, like we discussed earlier. Things like financial barriers, hardware barriers, people barriers... But in general I do think that data quality is a universal problem, with the same difficulties and challenges across all domains... The

data quality challenges we experience should, generically speaking, be the same as in any other domain.”

Section 6.5.2.1 describes the findings with regard to e-health barriers that may affect data quality.

6.5.2.1 FINDINGS

The respondents were asked to state whether e-health barriers could have an impact on data quality. Two of the respondents (B and D) were asked to provide specific details about the e-health barriers provided in Chapter 2. The reason behind interviewing respondents B and D on specific aspects regarding data quality and e-health is because they are South African citizens, and e-health barriers may differ across countries. Respondent D is also both an e-health and data quality expert. Direct quotations from the respondents are given to describe the findings.

- **Financial barrier to purchase necessary hardware and cost challenges in general**

Respondent B: “Financial barrier - high, because of people don’t realise the importance of data, they won’t be willing to spend high amounts of money to purchase the needed hardware.”

Respondent D: “The financial barrier we experience is tremendous. Clinics don’t have all the necessary equipment and I can say without a doubt that the quality of data is influenced by this.”

- **Implementation of health information systems in a small and rural facility**

Respondent B: “Implementation of health info systems in a small and rural facility...’ – high.”

Respondent D: “This is a major challenge. Internet connectivity is a really big challenge in South Africa. This could definitely have an impact on the quality of data. Along with this, people without computer experience have difficulties conforming to systems.”

- **Sustainability of using health information systems**

Respondent B: “Sustainability’ – high”

Respondent D: “The quality of data could be influenced by this e-health challenge in our opinion”

- **Physicians are hesitant to change existing processes**

Respondent B: “Physicians are hesitant to change existing processes’ – I agree.”

Respondent D: “This is definitely a major challenge we experience. They start off using the system and then just... stop. This has an influence on the quality of data. It mainly causes the data to be incomplete.”

- **It is difficult to form electronic health records as part of your daily routine**

Respondent B: “Forming electronic health records as part of daily routine’ – yes”

Respondent D: “This is almost the same as the physicians not wanting to change their ways. Data users might not have time during the day to complete their data tasks on the computers – especially nurses. Once again, this leads to incomplete data in most cases.”

- **Logistics**

Respondent B: “Logistics’ – I agree”

Respondent D: “Big challenge. Without the necessary logistics in terms of, well everything basically, the data quality could be influenced negatively.”

- **Shortage of necessary infrastructure**

Respondent B: “Infrastructure’ – well if you don’t have infrastructure then it might have a high influence on data quality.”

Respondent D: “This is close to the rural facility one we spoke about earlier. Once again, I would like to stress the challenge we experience with regard to internet connectivity. Without a doubt, this has a great influence on data quality.”

- **Nurses’ notes may go unread by physicians**

Respondent B: “Nurses’ notes may go unread by physicians’ – yes I agree.”

Respondent D: “This is definitely a challenge we experience. I think it is in the same category as the physicians hesitant to change the processes. The data quality could be great, but if it is not executed or used properly, why bother?”

6.5.2.2 SUMMARY

Section 6.5.2.1 described e-health barriers which could have an effect on data quality. E-health barriers, such as a lack of resources, interoperability, data privacy, usability, transformation from paper records, meeting needs at provider level and data access were not discussed, since it was listed as data quality challenges as well as e-health barriers in the literature review.

Barriers that should be included in the enhanced data quality challenges-list are: *financial barrier to purchase necessary hardware and cost challenges in general, implementation of health information systems in a small and rural facility, sustainability of using health information systems, physicians are hesitant to change existing processes, it is difficult to form electronic health records as part of your daily routine, logistics, shortage of necessary infrastructure and nurses’ notes may go unread by physicians.*

The review of data quality challenges is discussed in Section 6.5.3.

6.5.3 REVIEW OF DATA QUALITY CHALLENGES

Figure 6.5 illustrates the process of reviewing data quality challenges previously identified from literature. This section describes the process of reviewing the data quality challenges to eventually exclude and include certain challenges in the enhanced list. Figure 6.8 illustrates the part of the process analysed in this section:

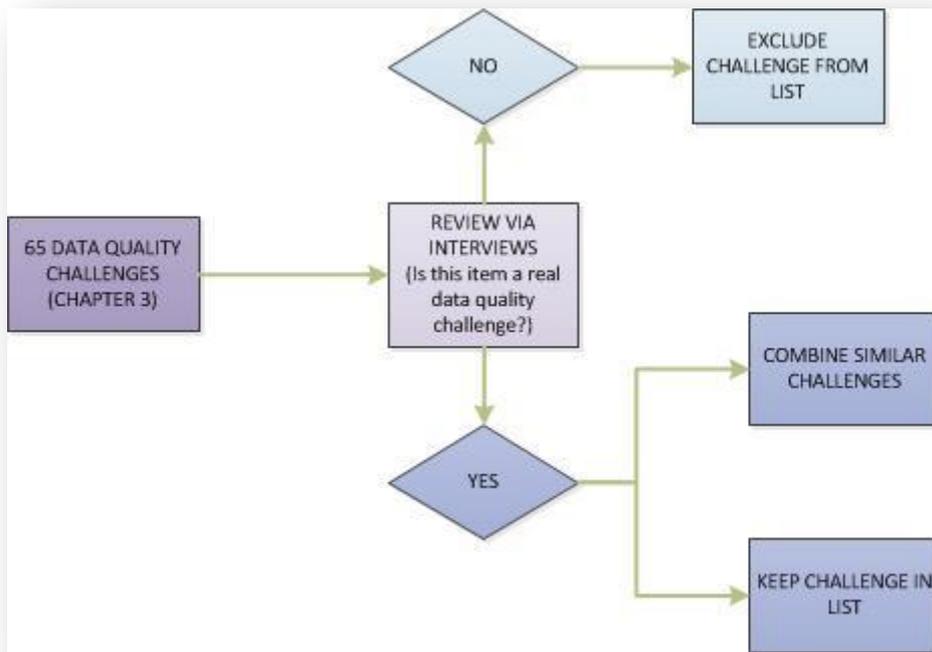


Figure 6.8: Data quality challenges review

The researcher asked the interviewees to validate each of the data quality challenges extracted from the literature (Chapter 3). The inclusion and exclusion criteria are described in Section 6.5.3.1.

6.5.3.1 INCLUSION AND EXCLUSION CRITERIA

This section describes the criteria for including and excluding data quality challenges to the enhanced list of data quality challenges.

Each of the 65 data quality challenges identified in the literature had to be validated by data quality experts. Table 6.3 provides a summary of the inclusion criteria:

Table 6.3: Data quality challenges inclusion criteria

Accepted by number of respondents:	Add to enhanced list
0	No
1	No
2	Yes
3	Yes
4	Yes

The researcher concluded that a data quality challenge may be included in the list if it is accepted by two or more respondents. The reason for this is that each respondent did

not comment on every data quality challenge. In some cases, the data quality experts did not comment on a certain item because it is something that they have not experienced before or are not familiar with. Therefore, two acceptances can be regarded the majority in most cases.

If it was advised by one or more data quality expert to combine or rename data quality challenges, the data quality challenges were revised accordingly and added to the enhanced list of data quality challenges presented in Section 6.6.

The findings of the data quality review are described in Section 6.5.3.2.

6.5.3.2 FINDINGS

The findings of this section are summarised in Table 6.4. The first column in the table is the number associated with the data quality challenge. The data quality challenge is given in the second column in the list. The third column indicates whether the challenge is kept in the list (according to the criteria in Section 6.5.3.1); this column follows a column indicating which of the respondents accepted the relevant item as a true data quality challenge. In the case of combining or renaming data quality challenges, the fifth column provides information indicating the challenge combinations (by number) and the renaming of challenges. The last column indicates which respondent advised the combination or renaming.

Table 6.4 provides the analysis of the review process:

Table 6.4: Data quality challenges review analysis

No.	Data quality challenge	Included in list	Accepted by	Combined with other challenge	Advised by
1	Absent assignment of responsibilities regarding data	Yes	A, B, C, D	2, 3 as data governance	A, C, D
2	Ambiguity of roles with regard to data construction, use and maintenance	Yes	A, B, C, D	1, 3 as data governance	A, C, D
3	Lack of adequate administration focus with regard to data quality	Yes	A, B, C, D	1, 2 as data governance	C, D
4	Lack of adequate rewards in terms of data quality	Yes	A, B, C		
5	Lack of adequate training for data consumers	Yes	A, B, D		
6	Need for written data quality policies and procedures	Yes	A, B, C	1, 2, 3 as data governance	B, C
7	Managers' lack of thorough emphasis on the importance of data quality	Yes	A, B, C, D	1,2,3,6 as data governance	B, C
8	Misspelled data	Yes	A,C, D	23, 12, 13	A, C, D
9	Conflicting use of special characters	No			
10	More than one source of data, each producing different values	Yes	A, B, C, D		
11	Information is produced by subjectivity	Yes	C, D		
12	A certain data element is not	Yes	A, B, D	8, 23, 13	A, C, D

No.	Data quality challenge	Included in list	Accepted by	Combined with other challenge	Advised by
	captured				
13	A certain data element is captured, but not in the required format	Yes	A, B, D	8, 12, 23	A, C, D
14	A certain data element is captured in the correct format for one system or part of a system, but not for another	No			
15	Transformation from paper records	Yes	A, B, C, D		
16	Accidental deletions	No			
17	Accidental edits	No			
18	The default value in the system was never corrected	Yes	A, C		
19	Inaccurate data at collection	Yes	A, C, D		
20	Questionnaire mode	No			
21	Questionnaire length	Yes	C, D		
22	Structure of questions	Yes	A, C, D		
23	Data entry errors	Yes	A, D	8, 12, 13	
24	Ineffective organisational procedures	Yes	A, B	1,2,3,6,7 as data governance	B, C
25	Lack of appropriate software for data management	Yes	A, B		
26	Lack of adequate input options in	No			

No.	Data quality challenge	Included in list	Accepted by	Combined with other challenge	Advised by
	data quality systems				
27	Unfortunate usability of information technology systems	Yes	B, D	Rename: technology not usable	A
28	Absent columns	Yes	A, B	30, 34, 36, 48, 50 as data structure problems	A, B, C, D
29	Insufficient candidate data sources	No			
30	Insufficient familiarity of inter-data dependencies	Yes	A, B	28, 34, 36, 48, 50 as data structure problems	A, B, C, D
31	Changing timeliness of data sources	Yes	A, D		
32	Unforeseen alterations in source systems	Yes	A, B		
33	Numerous sources for the same data	Yes	B, D		
34	Extra columns	Yes	A, B, D	28, 30, 36, 48, 50 as data structure problems	A, B, C, D
35	Data values drift from their field descriptions and business rules	Yes	A, B, C		
36	Unsuitable data relationships	Yes	A, B	28, 30, 34, 48, 50 as data structure problems	A, B, C, D
37	Lost information because of	No			

No.	Data quality challenge	Included in list	Accepted by	Combined with other challenge	Advised by
	systematic errors				
38	Inconsistent definitions and formats because of different systems used	Yes	A, C		
39	Interoperability issues	Yes	A, C, D		
40	Lack of sufficient computer and other resources	Yes	A, B, C, D		
41	Unavailable queries	No			
42	Information not stored	No			
43	Uncertainties in legal aspects of medical records	Yes	B, D	59	D
44	Balancing patient care with electronic health record tasks	No			
45	Information needed for required fields	Yes	A, B, C		
46	Managing large amounts of data	Yes	A, B, D		
47	Managing heterogeneous data	No			
48	Data structure problems	Yes	A, B, C	28, 30, 34, 36, 50 as data structure problems	A, B, C, D
49	Outdated data	Yes	A, B		
50	Data in the wrong field	Yes	A, B	28, 30, 34, 36, 48 as data structure problems	A, B, C, D
51	Data in the wrong sequence	No			

No.	Data quality challenge	Included in list	Accepted by	Combined with other challenge	Advised by
52	Omissions of data	Yes	A, B, D		
53	Data clustered incorrectly	No			
54	Inconsistent data within a single file	Yes	B, C		
55	Authenticity of data elements	No			
56	Lack of data quality assessments	Yes	A, B	1,2,3,6,7, 24 as data governance	B,D
57	Occurrence of outliers	Yes	A, B		
58	Timeframe restrictions	No			
59	Privacy and confidentiality issues	Yes	A, B, D		
60	Data segmentation	No			
61	Tracking completed tasks	No			
62	Interfaced systems may not display data correctly	No			
63	Data not meeting all relevant needs	Yes	A, B		
64	Information is not usable for analytic purposes	Yes	A, B		
65	Data is difficult to access	Yes	B, C		

The conclusion for this section is provided in Section 6.5.3.3.

6.5.3.3 SUMMARY

From Table 6.4 the researcher concluded that 18 of the data quality challenges identified in the literature will be excluded from the list of data quality challenges. This is due to the fact that less than two respondents acknowledged these data quality challenges. Furthermore 20 of the initial data quality challenges could be combined with one another.

The reviewed and accepted or combined challenges will be included in the enhanced and practise-specific list of data quality challenges presented in Section 6.6.

6.6 CONCLUSION

The purpose of Chapter 6, namely to provide an enhanced and practise-specific list of data quality challenges, was realised by conducting interviews with four data quality experts.

As described in Section 6.5.1, 6.5.2 and 6.5.3, the enhanced list of data quality challenges is compiled by including data quality challenges experienced by the interviewees, including e-health barriers that have an effect on data quality and including challenges verified through a review process.

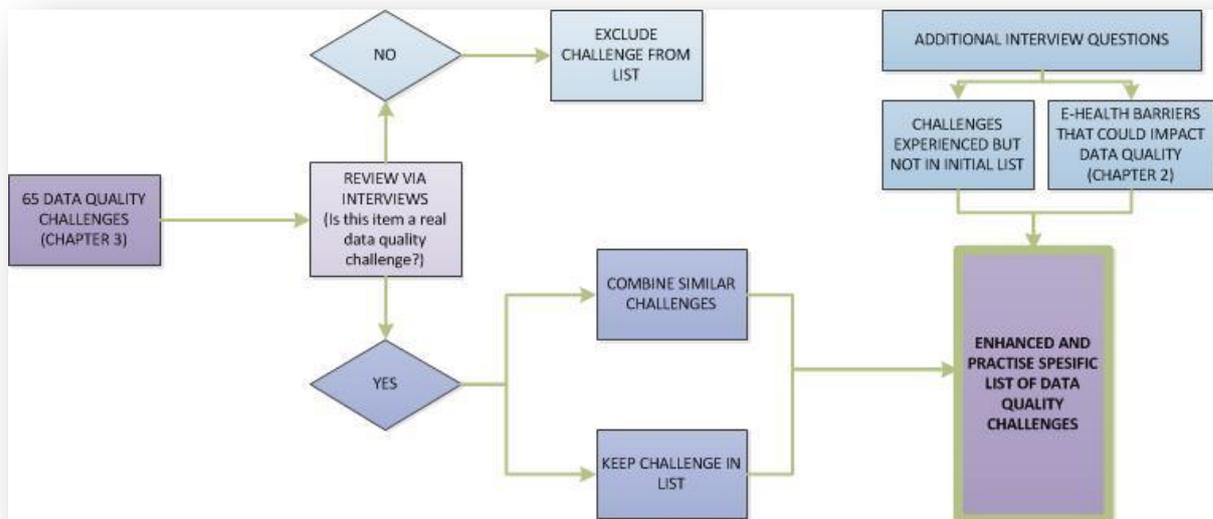


Figure 6.9: Enhanced and practise-specific list of data quality challenges

The enhanced and practise-specific list of data quality challenges is provided in Table 6.5. All data quality challenges, along with the relevant section in Chapter 6 are provided in the Table 6.5:

Table 6.5: Enhanced list of data quality challenges

Data quality challenge	Chapter 6 Reference
Data governance (this includes challenges such as: lack of assignment of responsibilities regarding data; lack of data administration; ambiguity of roles with regard to data tasks; a need for written data quality policies and procedures; managers' lack of thorough emphasis on the importance of data quality; ineffective organisational procedures; lack of data quality assessments)	Table 6.4
There is a lack of adequate rewards in terms of data quality	Table 6.4
There is a need for more training for data users	Table 6.4
Data entry errors (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)	Table 6.4
More than one source of data, each producing different values	Table 6.4
Information is produced by subjectivity of the data capturer	Table 6.4
Transformation from paper records to electronic records	Table 6.4 & Section 6.5.2.2
The default value in the system was never corrected	Table 6.4
Inaccurate data is collected from the patient	Table 6.4
The length of the form is too long	Table 6.4
The structure of the questions on the form produces inaccurate data	Table 6.4
Lack of appropriate software for data management	Table 6.4
Technology is not usable or user friendly	Table 6.4 & Section 6.5.2.2
Data structure problems (this includes problems such as: absent columns in the database; inter-data dependency issues; extra columns in the database; unsuitable data relationships; data in the wrong field of the form or database)	Table 6.4
Changing relevance of data sources	Table 6.4
Unforeseen changes in source systems	Table 6.4

Data quality challenge	Chapter 6 Reference
Numerous sources for the same data	Table 6.4
Data values drift from their field descriptions and business rules	Table 6.4
Inconsistent definitions and formats because of different systems used (standardisation)	Table 6.4 & Section 6.5.2.2
Interoperability issues and standardisation (Interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged.)	Table 6.4 & Section 6.5.2.2
Lack of sufficient computer and other resources.	Table 6.4 & Section 6.5.2.2
Privacy, confidentiality and legal challenges	Table 6.4 & Section 6.5.2.2
Information is needed for required fields (you don't have all the information required by the system)	Table 6.4
Managing large amounts of data	Table 6.4
Outdated data	Table 6.4
Omissions/errors of data	Table 6.4
Inconsistent data within a single file	Table 6.4
Occurrence of outliers	Table 6.4
Data not meeting all relevant needs	Table 6.4 & Section 6.5.2.2
Information is not suitable for analytic purposes	Table 6.4
Data is difficult to access	Table 6.4 & Section 6.5.2.2
Financial barrier to purchase necessary hardware and cost challenges in general	Section 6.5.1.2 & Section 6.5.2.2
Implementation of health information systems in a small and rural facility	Section 6.5.2.2
Sustainability of using health information systems	Section 6.5.2.2
Physicians are hesitant to change existing processes	Section 6.5.2.2
It is difficult to form electronic health records as part of daily routine	Section 6.5.2.2
Logistics with regard to electronic health information systems	Section 6.5.2.2
The shortage or absence of the necessary infrastructure, such as	Section 6.5.2.2

Data quality challenge	Chapter 6 Reference
internet connections	
Nurses' notes may go unread by physicians	Section 6.5.2.2

Table 6.5 provides 39 data quality challenges. These challenges consist of:

- data quality challenges experienced by data quality experts that were not included in the list of data quality challenges from the literature;
- e-health barriers which could have an effect on the quality of data; and
- data quality challenges from literature, reviewed and accepted by data quality experts.

Challenges specific and unique to South Africa, as found during the qualitative interviews are: interoperability, a lack of training for data users, financial barriers, the implementation of health information systems in small and rural facilities and the shortage or absence of the necessary infrastructure, such as internet connections.

The 39 data quality challenges in the enhanced and practise-specific list were used as questions in the questionnaire in the quantitative phase of the research. The quantitative research findings are discussed in Chapter 7.

CHAPTER 7: QUANTITATIVE RESEARCH FINDINGS

Stage in research:

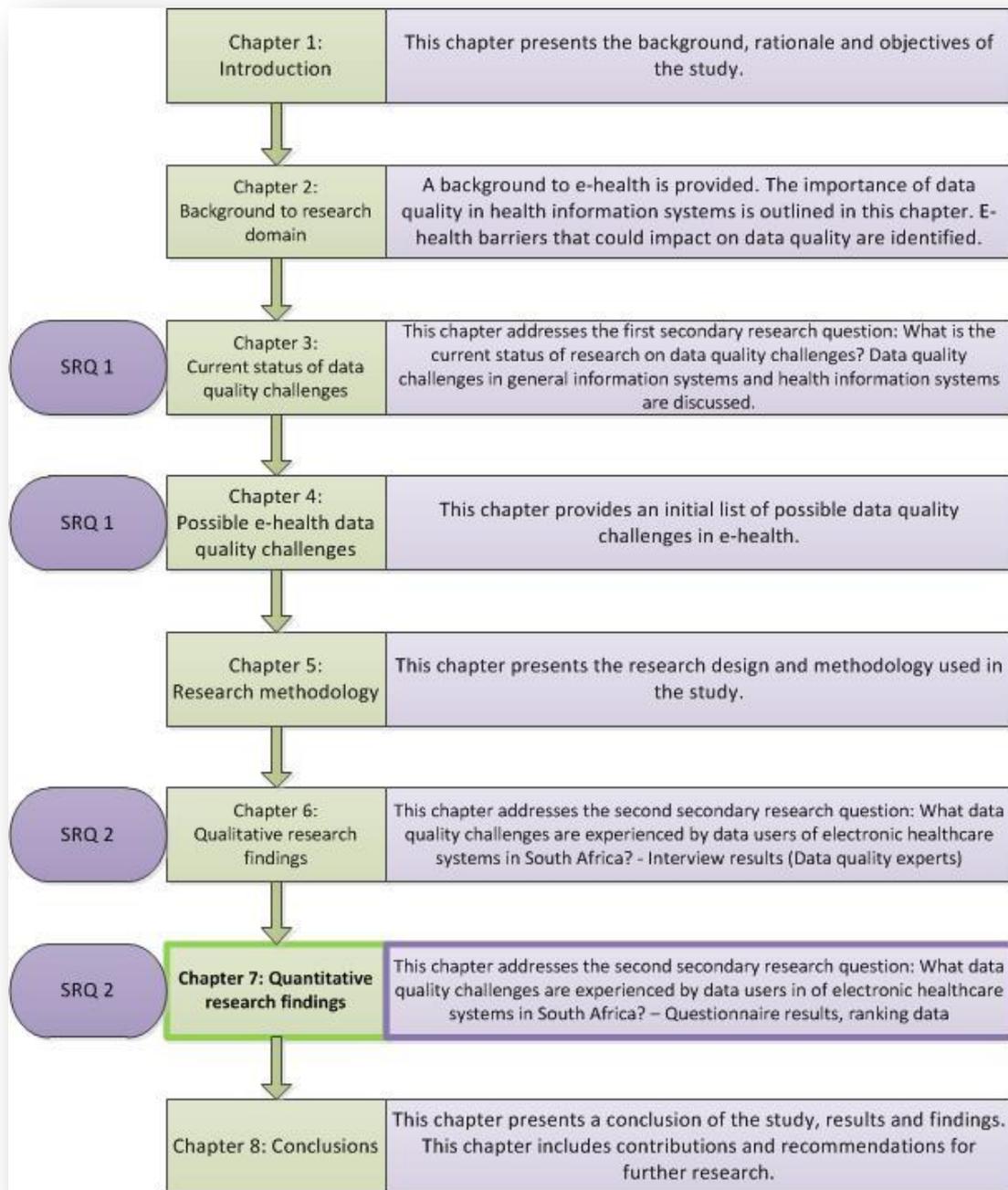


Figure 7.1: Chapter 7 – Stage in research

Structure of Chapter 7:

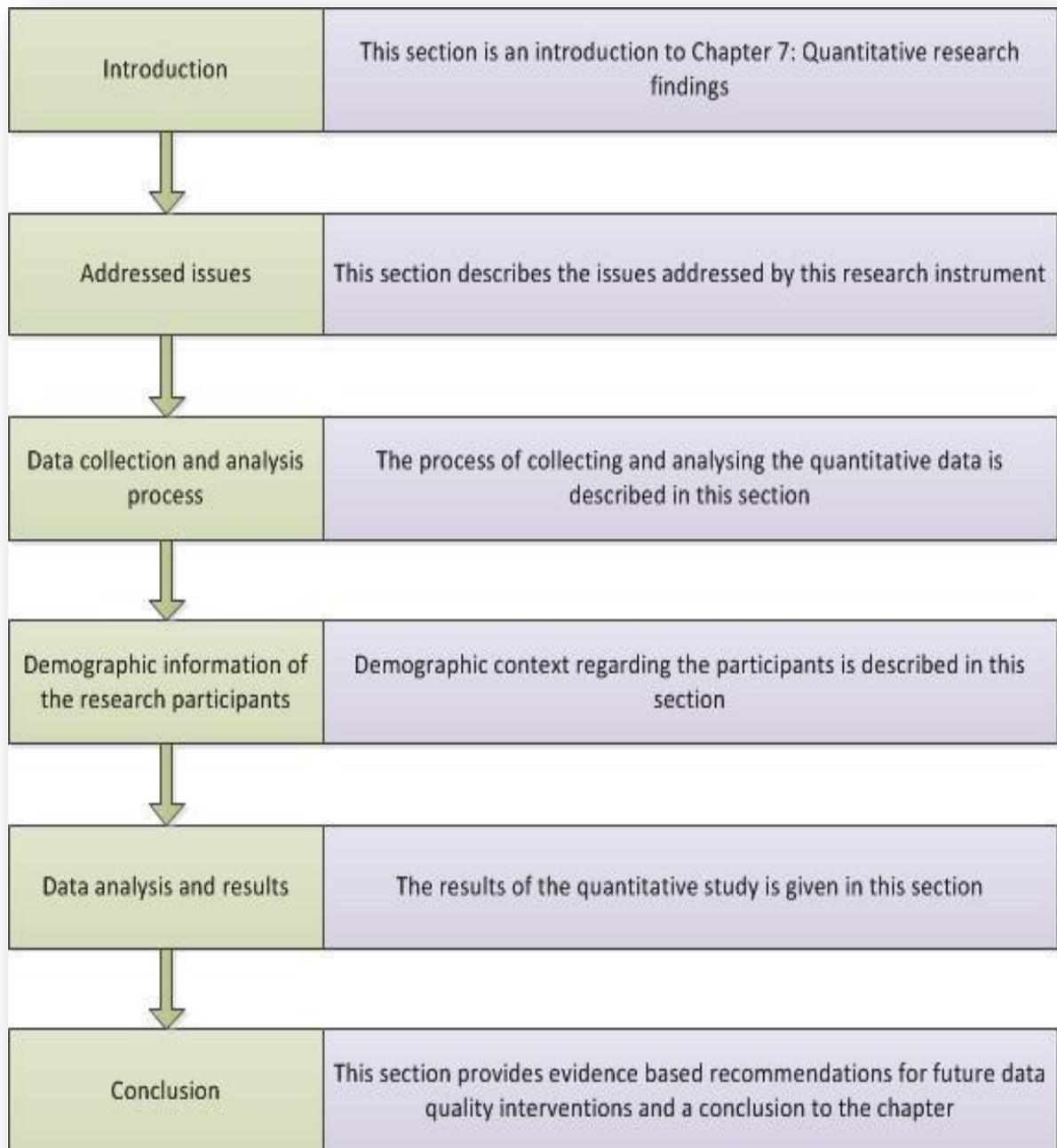


Figure 7.2: Structure of Chapter 7

7.1 INTRODUCTION

Chapter 6 presented the qualitative research findings, obtained from interviews conducted with data quality experts. This study adopted a sequential QUAL-quantitative research design to answer the main research question. Chapter 7 presents a roadmap showing how the study arrived at the quantitative findings presented here. The quantitative study was carried out by collecting questionnaires from data users of electronic healthcare systems in South Africa. The data users included administrative personnel, medical personnel and IT personnel within the health domain. The primary purpose of the quantitative study was to provide a prioritised list of data quality challenges. The results obtained from the qualitative study (see Chapter 6) were used to compile the questionnaire used in the quantitative study. The purpose was realised by means of a data collection exercise. The instrument used during the quantitative data collection was questionnaires.

In this chapter the results obtained from the quantitative research instrument and the patterns of the results are presented. These findings are used to partially address the second secondary research question, namely: *What data quality challenges are experienced by data users in electronic healthcare systems in South Africa?* The quantitative research presented in this chapter is used to prioritise the data quality challenges obtained from the literature (Chapter 4) and verified through the qualitative study (Chapter 6).

Phase 3 of the research is presented in Chapter 7. Figure 7.3 illustrates the current phase of research:

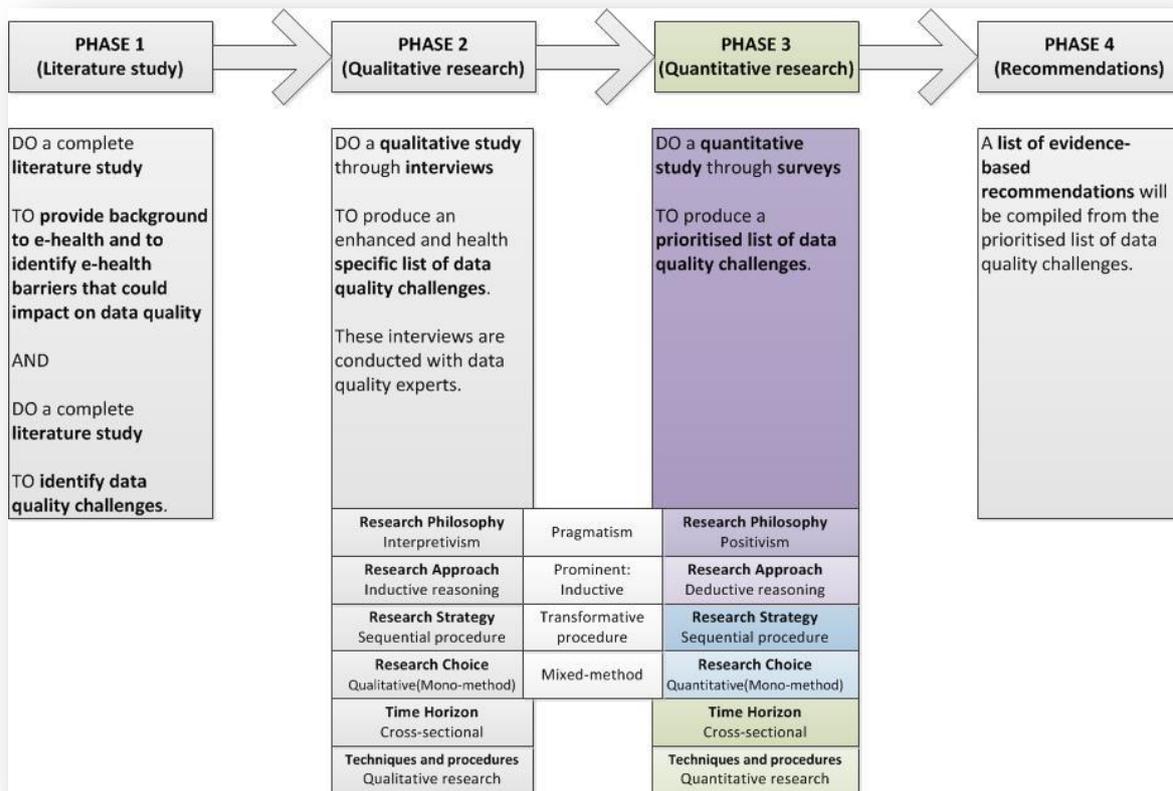


Figure 7.3: Current phase in research

Chapter 7 presents the results of a quantitative survey to produce a prioritised list of data quality challenges. The prioritised list of data quality challenges is followed by a list of evidence-based recommendations.

The remainder of this chapter is structured as follows: after the introduction a description of the issues addressed by this research instrument follows. Consequently a brief description of the process that was followed to collect and analyse the data is given. This is followed by a brief description of the demographic information of the research participants. The data analysis and results are given and finally the chapter conclusion is made.

This chapter provides the results that emerged from a quantitative study to provide a prioritised list of data quality challenges. The issues addressed by this research instrument follow in Section 7.2.

7.2 ISSUES ADDRESSED BY THE RESEARCH INSTRUMENT

In this study, data was generated from participants' experiences with regard to data quality challenges. In order to obtain this information, the researcher made use of questionnaires as a data collection method. The participants in this study were purposefully chosen to include only data users of electronic healthcare systems in South Africa in order to compile a prioritised list of data quality challenges experienced by data users.

7.2.1 QUESTIONNAIRES

Questionnaires were used to gather the data for the quantitative study (see Appendix C). The questionnaire used in this study was compiled by using data quality challenges identified through a literature review (Chapter 4) and verified by data quality experts (Chapter 6). A total of 82 participants completed the questionnaire. The questionnaire was mainly used to compile a prioritised list of data quality challenges. Table 7.1 provides a summary of the themes and objectives of the questionnaire questions.

Table 7.1: Themes and objectives of questionnaire

Question number	Themes	Objectives
1-2	Participant demographic information	To identify the participants' job title and years of experience in the health domain.
3-41	Ranking of data quality challenges	To identify the importance of each data quality challenge.

The questions in the questionnaire were specifically asked to allow the researcher to prioritise the list of data quality challenges in electronic healthcare systems in South Africa.

The data collection and analysis process is discussed in Section 7.3.

7.3 DATA COLLECTION AND ANALYSIS PROCESS

The data collection of the quantitative phase of the study comprised a questionnaire. The purpose of the questionnaire was to prioritise data quality challenges experienced by data users in electronic healthcare systems in South Africa.

A data triangulation exercise took place to produce the final prioritised list of data quality challenges. Figure 7.4 illustrates the data triangulation applied in the study:

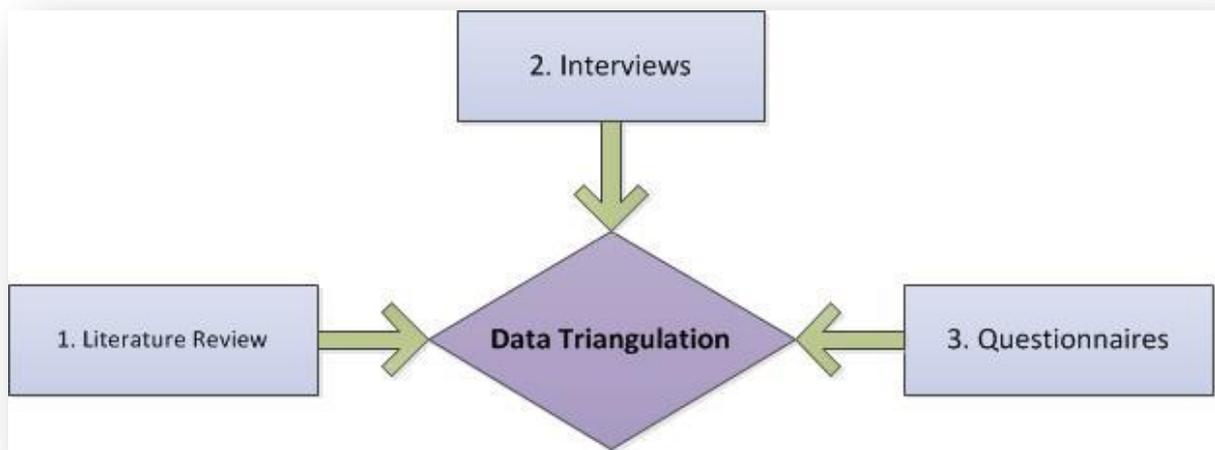


Figure 7.4: Data triangulation

The data triangulation was done in a sequential manner: i) literature review; ii) interviews; and iii) questionnaires. All of the data sources contributed to the final results of the study, as depicted in Figure 7.4. This section describes the data collection and analysis process of the quantitative phase of the research by means of questionnaires.

The data users of electronic healthcare systems (questionnaire respondents) were selected by visiting medical facilities in South Africa. The managers of these facilities were asked whether they make use of electronic healthcare systems. The questionnaires were given to the medical facilities who indicated that they make use of electronic healthcare systems.

A total of 100 questionnaires were handed out for completion, with a total of 82 returned. The response rate for the questionnaires was 82%.

39 data quality challenges were listed in the questionnaire. The respondents had to indicate to which extent they experience each challenges. The questionnaire prompted the respondents to rate each challenge against “*This is not a challenge I experience*”; “*I experience this challenge, but I regard it as a small challenge*”; or “*This is a challenge I experience and I regard it as high priority*”.

The research questionnaire was specifically designed to allow the researcher to prioritise data quality challenges. Statistically insignificant data quality challenges were identified by setting up an appropriate level C confidence interval for each of the data quality challenges. If it could not be statistically proven that more than 50% of the respondents regarded the item as a data quality challenge, the item was regarded as statistically insignificant.

The remaining (statistically significant) data quality challenges were prioritised by calculating the weighted total score for each of the items. The researcher used Microsoft Excel to analyse the data.

The data analysis process is discussed in Section 7.5.

7.4 DEMOGRAPHIC INFORMATION OF THE RESEARCH PARTICIPANTS

This section provides the demographic information of the participants of the quantitative study. The demographic information of the participants was collected in the first two questions in the questionnaire. This information was collected by the researcher to identify the data users of electronic healthcare systems in South Africa. Contextualised analysis is possible when demographic information about participants is known. Braun and Clarke (2013) state the importance of background information in data analysis.

Figure 7.5 provides a summary of the distribution of the data users of electronic healthcare systems in this sample.

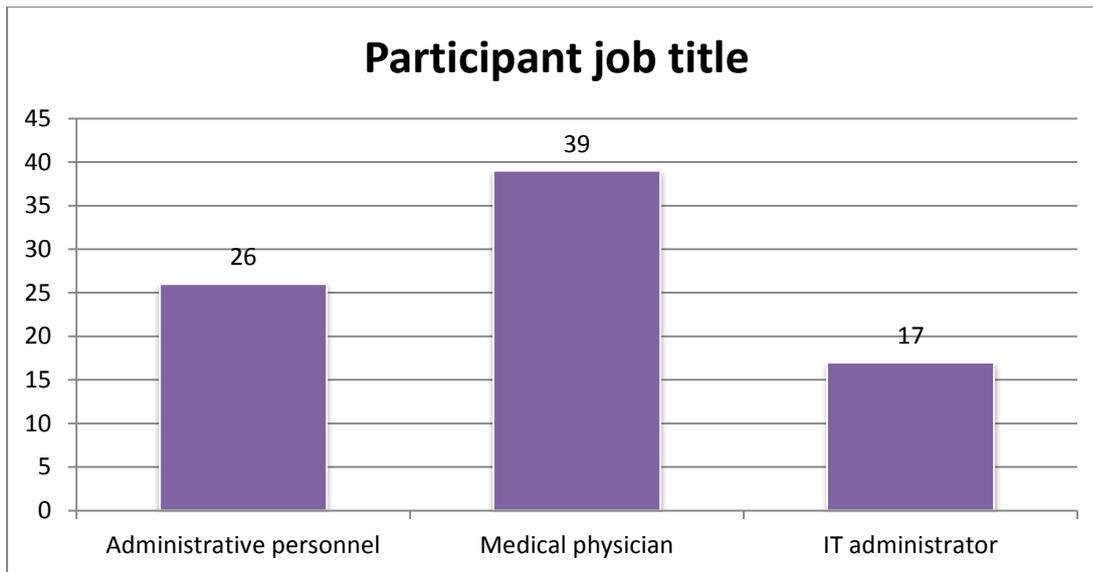


Figure 7.5: Participant job title

The pie chart in Figure 7.5 shows that 47% of the research participants were medical physicians which include doctors, nurses, pathologists, pharmacists and optometrists. 32% of the research participants were administrative personnel and 21% of the participants were IT administrators.

The purpose of the study is to compile a prioritised list of data quality challenges experienced by data users of electronic healthcare systems in South Africa. The data users include medical physicians, administrative personnel and IT administrators. The responses of all the above data users are regarded as equally important for the purpose of the study. Care was taken to ensure that each of the three groups was included in the study and it was anticipated that this broad range of respondents would contribute to the generalisability of the results.

Figure 7.6 provides a summary of the distribution of the experience of the data users in the health domain:

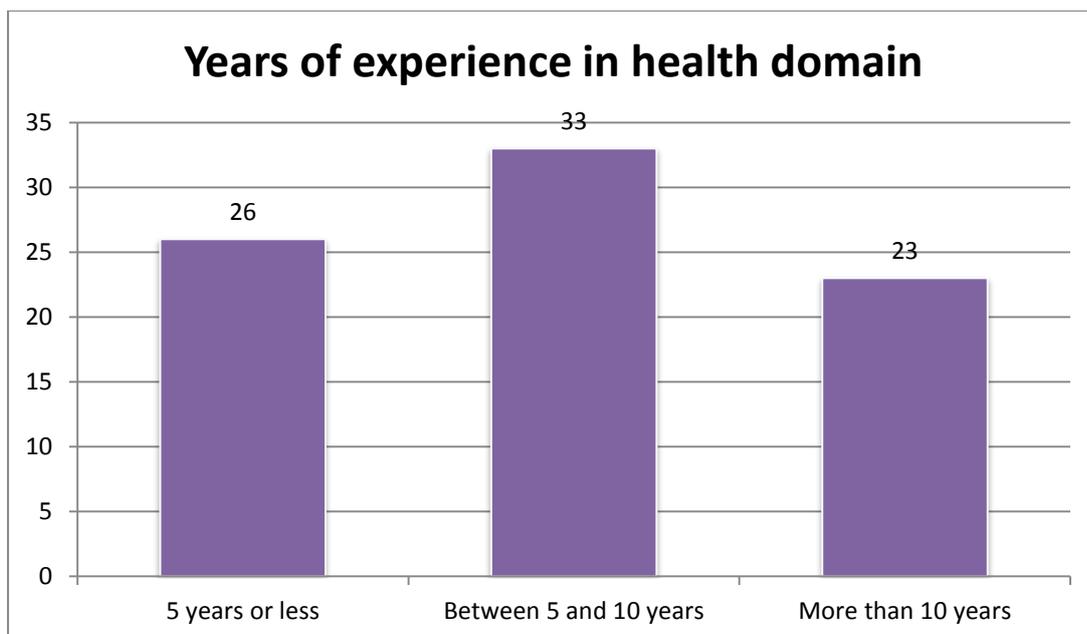


Figure 7.6: Years of experience in health domain

The chart in Figure 7.6 shows that 40% of the participants have been working between 5 and 10 years in the health domain; 32% have been in the health domain for 5 years or less; and 28% for more than 10 years.

The distribution of the years of experience in the health domain provides a sound representation of participants in the study. Data users with a distributed amount of experience are used in this study. Care was taken to ensure that each of the three groups was included in the study and it was anticipated that this broad range of respondents would contribute to the generalisability of the results.

The data analysis and results are discussed in Section 7.5.

7.5 DATA ANALYSIS AND RESULTS

The purpose of this chapter is to produce a prioritised list of data quality challenges. This section describes the results obtained from the questionnaires completed by data users of electronic healthcare systems in South Africa.

The data quality challenges which prove to be regarded as challenges by more than 50% of the participants will be included in the list of prioritised data quality challenges.

To prove that each item is regarded as a data quality challenge by more than 50% of the participants, a 95% confidence interval will be determined for the proportions.

An appropriate level C confidence interval for p is $P \pm z^* \sqrt{\frac{p(1-p)}{n}}$ where z^* is the upper $(1-C)/2$ critical value from the standard normal distribution (Altman *et al.*, 2014). The critical value for a 95% confidence interval is 1.96 (Rice, 2007). If 0.5 (50%) lies within the boundaries of the confidence interval, there is no statistical evidence that more than 50% of the respondents regarded the item as a challenge they experience (Rice, 2007). Table 7.2 displays the confidence intervals for each of the data quality challenges. The items that do not prove to be regarded as significant are displayed in grey.

Table 7.2: Data quality challenges with 95% confidence intervals

Data quality challenge	95% confidence interval
<i>Data governance</i> (this includes challenges such as: lack of assignment of responsibilities regarding data; lack of data administration; ambiguity of roles with regard to data tasks; a need for written data quality policies and procedures; managers' lack of thorough emphasis on the importance of data quality; ineffective organisational procedures; lack of data quality assessments)	(0.5822 ; 0.7836)
There is a lack of adequate rewards in terms of data quality	(0.4536 ; 0.6683)
There is a need for more training for data users	(0.7334 ; 0.9008)
<i>Data entry errors</i> (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)	(0.7049 ; 0.8804)
More than one source of data, each producing different values	(0.4661 ; 0.6802)
Information is produced by subjectivity of the data capturer	(0.5299 ; 0.7384)
Transformation from paper records to electronic records	(0.5169 ; 0.7269)
The default value in the system was never corrected	(0.3555 ; 0.5713)
Inaccurate data is collected from the patient	(0.5299 ; 0.7384)
The length of the form is too long	(0.5042 ; 0.7153)
The structure of the questions on the form produces inaccurate data	(0.3796 ; 0.5959)
Lack of appropriate software for data management	(0.5559 ; 0.7612)
Technology is not usable or user friendly	(0.4914 ; 0.7037)

Data quality challenge	95% confidence interval
<i>Data structure problems</i> (this includes problems such as: absent columns in the database; inter-data dependency issues; extra columns in the database; unsuitable data relationships; data in the wrong field of the form or database)	(0.4914 ; 0.7037)
Changing relevance of data sources	(0.4287 ; 0.6445)
Unforeseen changes in source systems	(0.4914 ; 0.7037)
Numerous sources for the same data	(0.4914 ; 0.7037)
Data values drift from their field descriptions and business rules	(0.4411 ; 0.6565)
Inconsistent definitions and formats because of different systems used (standardisation)	(0.5690 ; 0.7724)
<i>Interoperability issues and standardisation</i> (interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged.)	(0.4914 ; 0.7037)
Lack of sufficient computer and other resources.	(0.4914 ; 0.7037)
Privacy, confidentiality and legal challenges	(0.5690 ; 0.7724)
Information is needed for required fields (you don't have all the information required by the system)	(0.4787 ; 0.6919)
Managing large amounts of data	(0.4787 ; 0.6919)
Outdated data	(0.5042 ; 0.7153)
Omissions/errors of data	(0.5042 ; 0.7153)
Inconsistent data within a single file	(0.4163 ; 0.6325)
Occurrence of outliers	(0.3796 ; 0.5959)
Data not meeting all relevant needs	(0.5559 ; 0.7612)
Information is not suitable for analytic purposes	(0.5042 ; 0.7153)
Data is difficult to access	(0.4163 ; 0.6325)
Financial barrier to purchase necessary hardware and cost challenges in general	(0.5299 ; 0.7384)
Implementation of health information systems in a small and rural facility	(0.4163 ; 0.6325)
Sustainability of using health information systems	(0.4287 ; 0.6445)
Physicians are hesitant to change existing processes	(0.5042 ; 0.7153)
It is difficult to form electronic health records as part of your daily routine	(0.4535 ; 0.6683)
Logistics with regard to electronic health information systems	(0.4287 ; 0.6445)
The shortage or absence of the necessary infrastructure, such as	(0.5559 ; 0.7612)

Data quality challenge	95% confidence interval
internet connections	
Nurses' notes may go unread by physicians	(0.5042 ; 0.7153)

The data quality challenges displayed in grey are not regarded as significant data quality challenges in this study. The confidence intervals show that there is no statistical proof that more than 50% of the respondents regard these items as challenges they experience.

To prioritise the remaining data quality challenges identified in literature and verified qualitatively, the weighted total scores for various data quality challenges were calculated to determine how it is perceived. These rankings were calculated by summing the ordinal values multiplied by the frequency of responses. The ordinal values are given in Table 7.3:

Table 7.3: Ordinal values of questionnaire questions

This is not a challenge I experience	I experience this challenge, but I regard it as a small challenge	This is a challenge I experience and I regard it as high priority
0	1	2

The maximum ranking value a data quality challenge can achieve is $(82 \times 2 = 164)$ (Kruss, Visser, Aphone & Haupt, 2012). This would occur when all 82 respondents marked a certain item as a challenge with high priority. The lowest ranking value a data quality challenge can achieve is 0. This would occur when all 82 respondents marked a certain item as not being a data quality challenge at all $(82 \times 0 = 0)$ (Kruss *et al.*, 2012).

If n is the number of respondents to an item and,

k_1 respondents experience a certain item not as a challenge,

k_2 respondents experience a certain item, but regard it as a small challenge,

k_3 respondents experience a certain item as a challenge with high priority and

$$k_1 + k_2 + k_3 = n$$

the formula for calculating the weighted total score is (Kruss *et al.*, 2012):

$$(0.k1) + (1.k2) + (2.k3)$$

Figure 7.7 displays the weighted total scores of the significant data quality challenges:



Figure 7.7: Weighted total scores of significant data quality challenges

Figure 7.7 displays the weighted total score of the 18 statistically significant data quality challenges. It is evident from the chart that two challenges were considered as high priority challenges. There were no significant differences in the importance of the other data quality challenges.

To give sense to the significant data quality challenges, a summary of the responses for each of the challenges are given below:

- *There is a need for more training for data users*

Figure 7.8 displays the distribution of the responses regarding this data quality challenges.

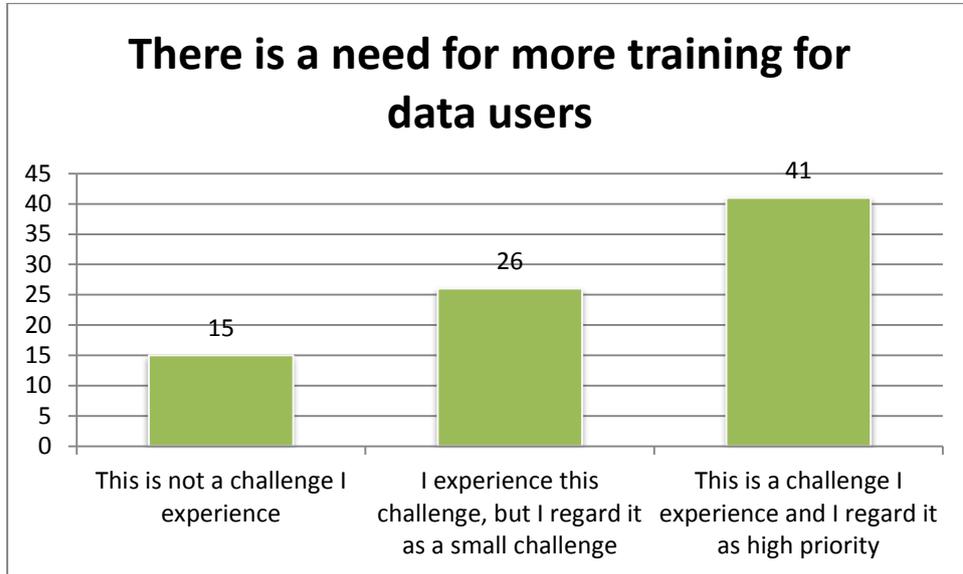


Figure 7.8: Responses – need for more training for data users

Eighteen percent (18%), thus 15 of the respondents did not experience this challenge; 32% (26) of respondents regarded this as a small challenge, while 50% (41) regarded it as a high priority challenge.

The weighted total score for this item is thus $(26 \times 1) + (41 \times 2) = 108$.

- *Data entry errors*

Figure 7.9 displays the distribution of the responses regarding this data quality challenges.

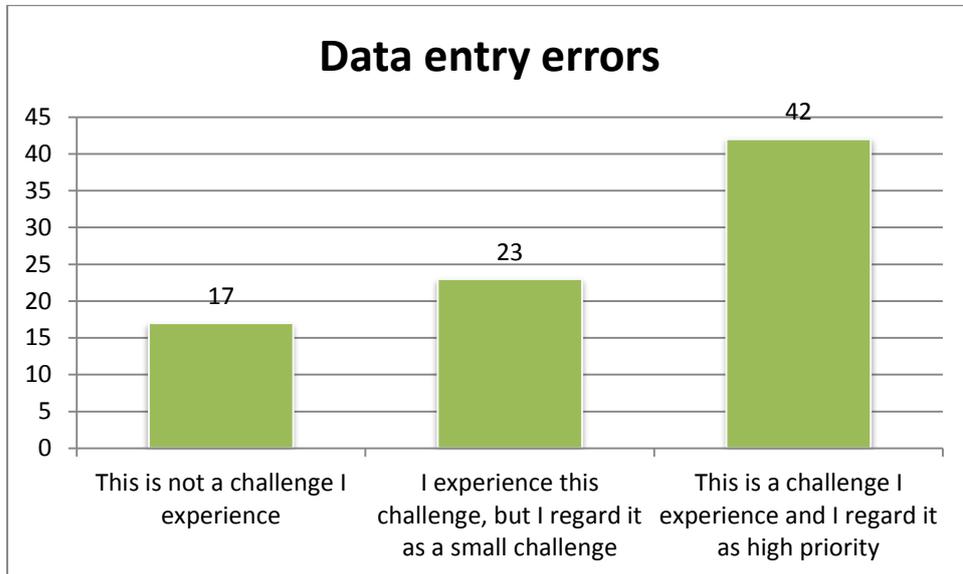


Figure 7.9: Responses – data entry errors

Twenty one percent (21%), thus 17 of the respondents did not experience this challenge; 28% (23) of respondents regarded this as a small challenge; while 51% (42) regarded it as a high priority challenge.

The weighted total score for this item is thus $(23 \times 1) + (42 \times 2) = 107$.

- *Inaccurate data is collected from the patient*

Figure 7.10 displays the distribution of the responses regarding this data quality challenges.

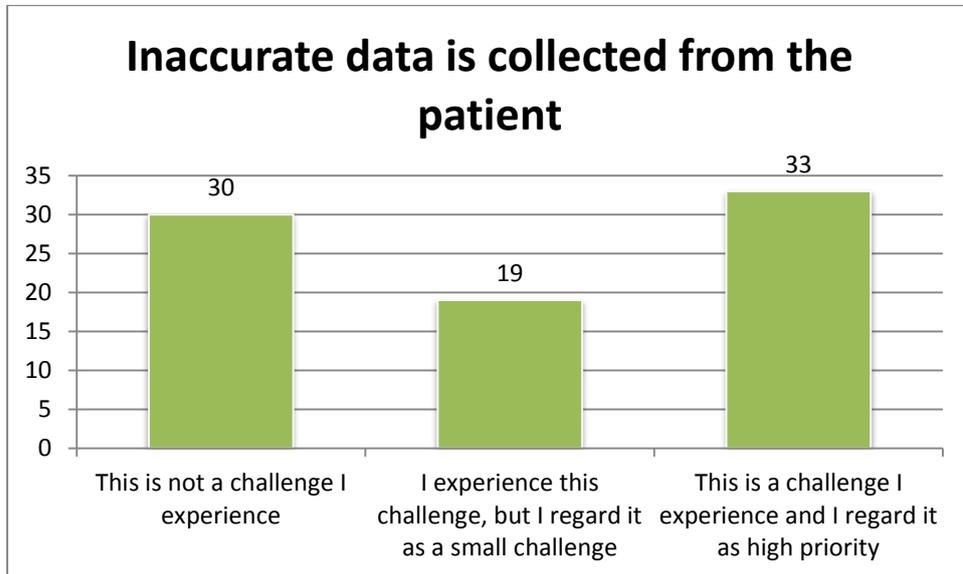


Figure 7.10: Responses – inaccurate data is collected from the patient

Thirty seven percent (37%), thus 30 of the respondents did not experience this challenge; 23% (19) of respondents regarded this as a small challenge; while 40% (33) regarded it as a high priority challenge.

The weighted total score for this item is thus $(19 \times 1) + (33 \times 2) = 85$.

- *The shortage of necessary infrastructure*

Figure 7.11 displays the distribution of the responses regarding this data quality challenges.

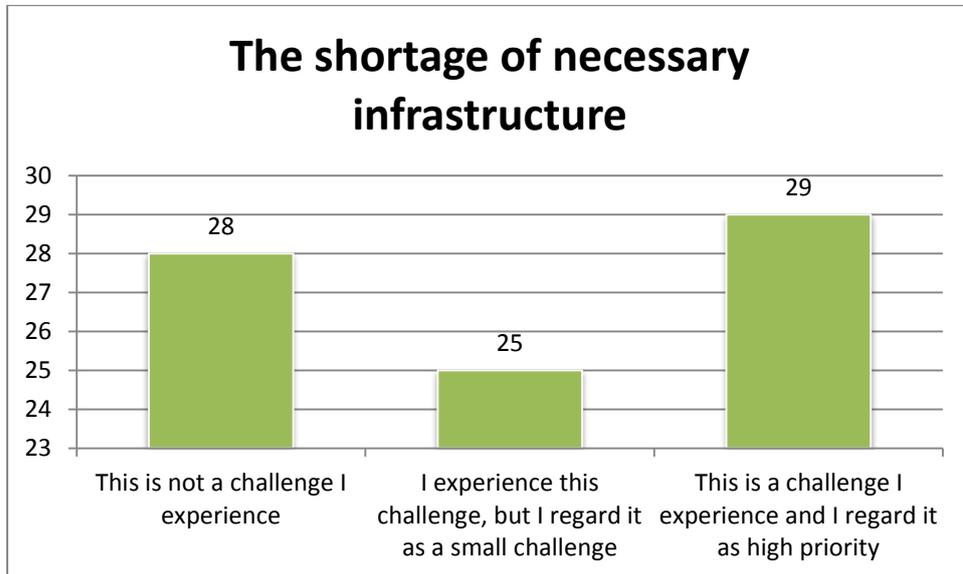


Figure 7.11: Responses – the shortage of necessary infrastructure

Thirty four percent (34%), thus 28 of the respondents did not experience this challenge; 31% (25) of respondents regarded this as a small challenge; while 35% (29) regarded it as a high priority challenge.

The weighted total score for this item is thus $(25 \times 1) + (29 \times 2) = 83$.

- *Inconsistent definitions and formats because of different systems used*

Figure 7.12 displays the distribution of the responses regarding this data quality challenges.

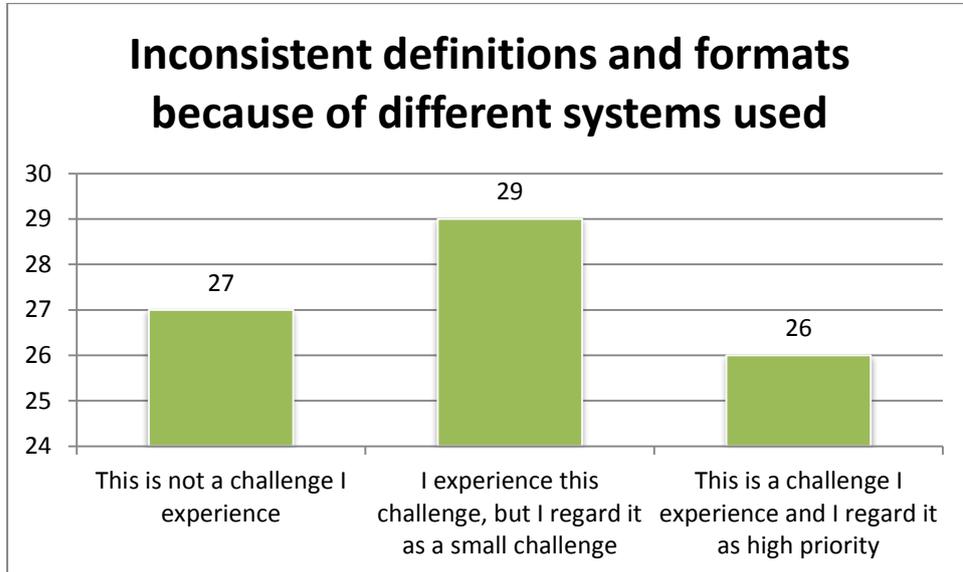


Figure 7.12: Responses – inconsistent definitions and formats because of different systems used

Thirty three percent (33%), thus 27 of the respondents did not experience this challenge; 35% (29) of respondents regarded this as a small challenge; while 32% (26) regarded it as a high priority challenge.

The weighted total score for this item is thus $(29 \times 1) + (26 \times 2) = 81$.

- *Physicians are hesitant to change existing processes*

Figure 7.13 displays the distribution of the responses regarding this data quality challenges.

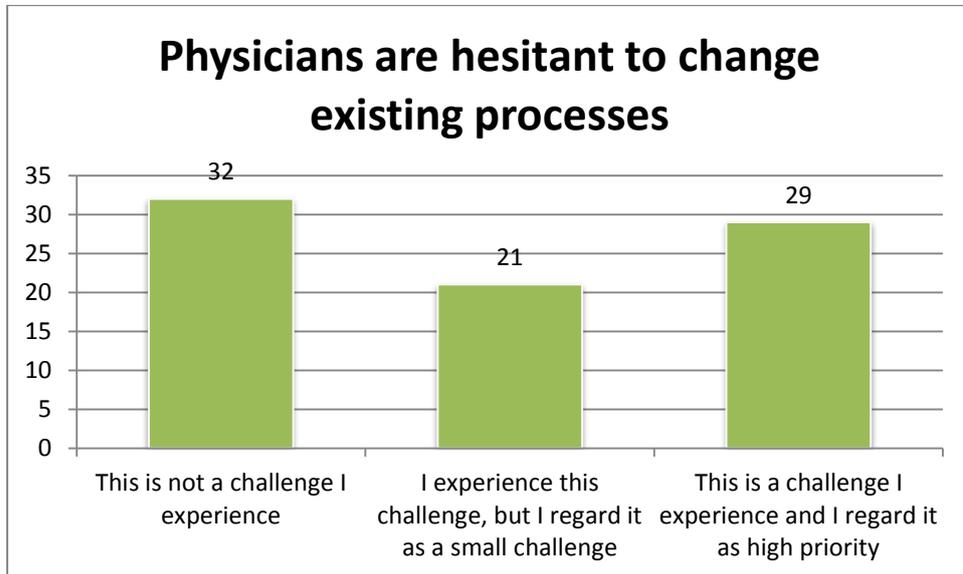


Figure 7.13: Responses – physicians are hesitant to change existing processes

Thirty nine percent (39%), thus 32 of the respondents did not experience this challenge; 26% (21) of respondents regarded this as a small challenge; while 35% (29) regarded it as a high priority challenge.

The weighted total score for this item is thus $(21 \times 1) + (29 \times 2) = 79$.

- *Data governance*

Figure 7.14 displays the distribution of the responses regarding this data quality challenges.

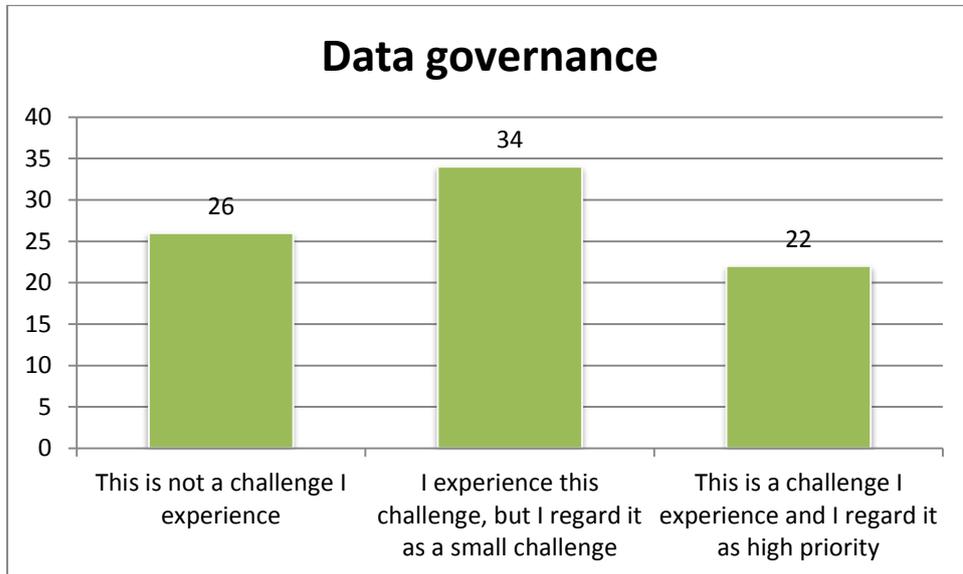


Figure 7.14: Responses – data governance

Thirty two percent (32%), thus 26 of the respondents did not experience this challenge; 41% (34) of respondents regarded this as a small challenge; while 27% (22) regarded it as a high priority challenge.

The weighted total score for this item is thus $(34 \times 1) + (22 \times 2) = 78$.

- *Information is produced by subjectivity of the data capturer*

Figure 7.15 displays the distribution of the responses regarding this data quality challenges.

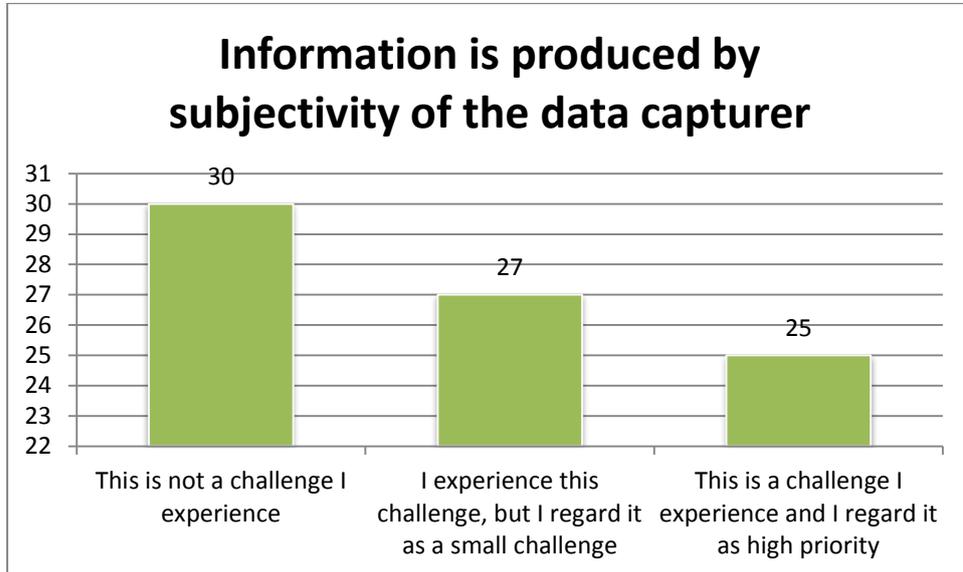


Figure 7.15: Responses – information is produced by subjectivity of the data capturer

Thirty seven percent (37%), thus 30 of the respondents did not experience this challenge; 33% (27) of respondents regarded this as a small challenge; while 30% (25) regarded it as a high priority challenge.

The weighted total score for this item is thus $(27 \times 1) + (25 \times 2) = 77$.

- *Lack of appropriate software for data management*

Figure 7.16 displays the distribution of the responses regarding this data quality challenges.

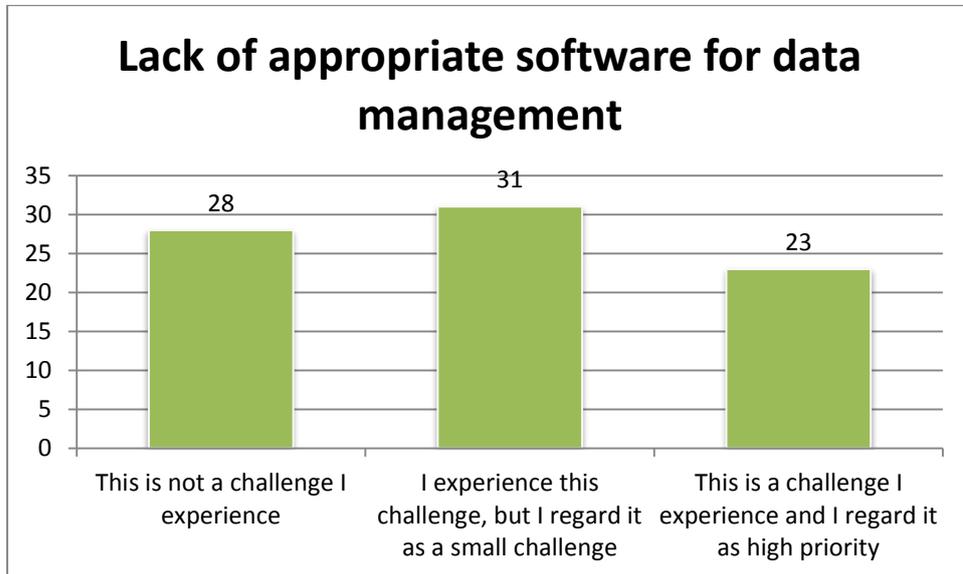


Figure 7.16: Responses – lack of appropriate software for data management

Thirty four percent (34%), thus 28 of the respondents did not experience this challenge; 38% (31) of respondents regarded this as a small challenge; while 28% (23) regarded it as a high priority challenge.

The weighted total score for this item is thus $(31 \times 1) + (23 \times 2) = 77$.

- *The form is too long*

Figure 7.17 displays the distribution of the responses regarding this data quality challenges.

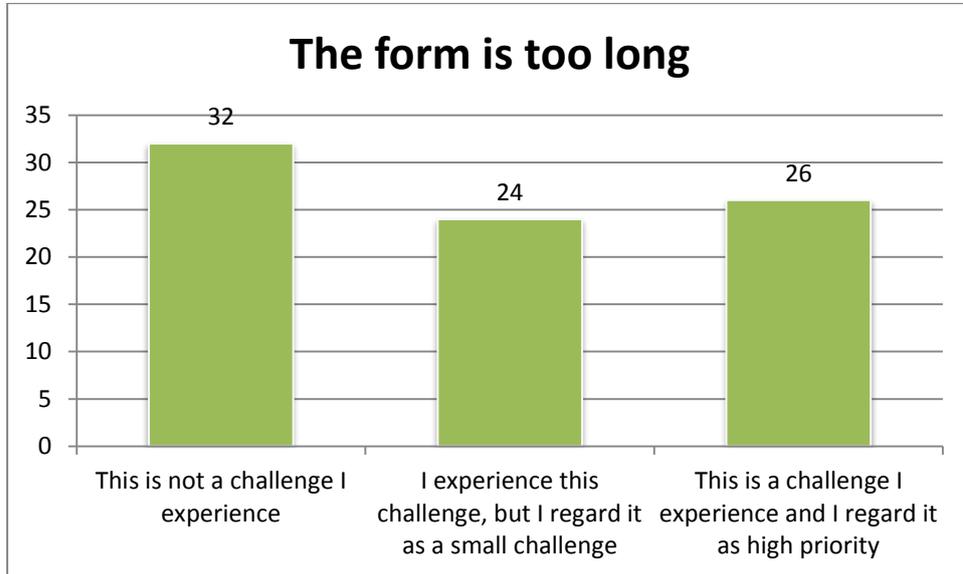


Figure 7.17: Responses – the form is too long

Thirty nine percent (39%), thus 32 of the respondents did not experience this challenge; 29% (24) of respondents regarded this as a small challenge; while 32% (26) regarded it as a high priority challenge.

The weighted total score for this item is thus $(24 \times 1) + (26 \times 2) = 76$.

- *Nurses' notes may go unread by physicians*

Figure 7.18 displays the distribution of the responses regarding this data quality challenges.

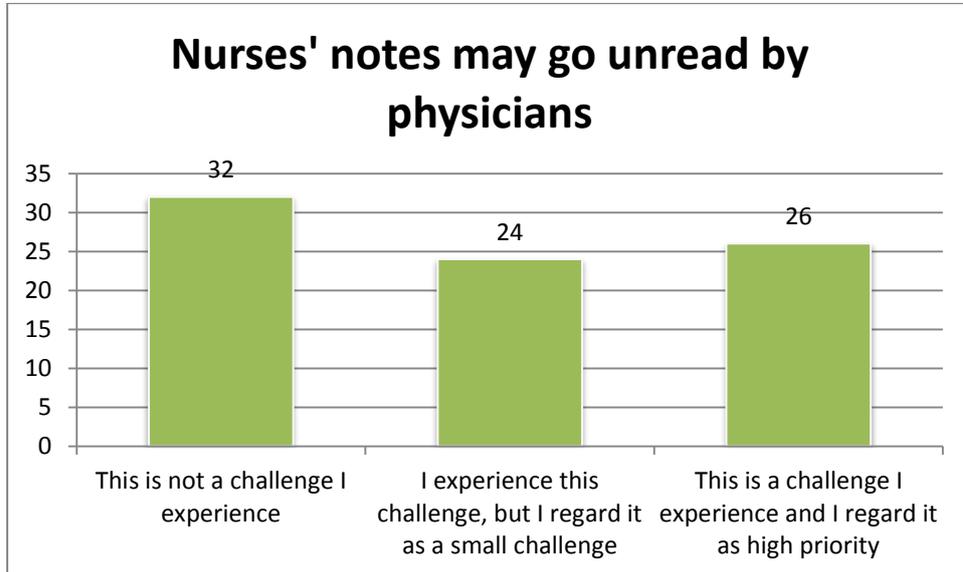


Figure 7.18: Responses – nurses' notes may go unread by physicians

Thirty nine percent (39%), thus 32 of the respondents did not experience this challenge; 29% (24) of respondents regarded this as a small challenge; while 32% (26) regarded it as a high priority challenge.

The weighted total score for this item is thus $(24 \times 1) + (26 \times 2) = 76$.

- *Transformation from paper records to electronic records*

Figure 7.19 displays the distribution of the responses regarding this data quality challenges.

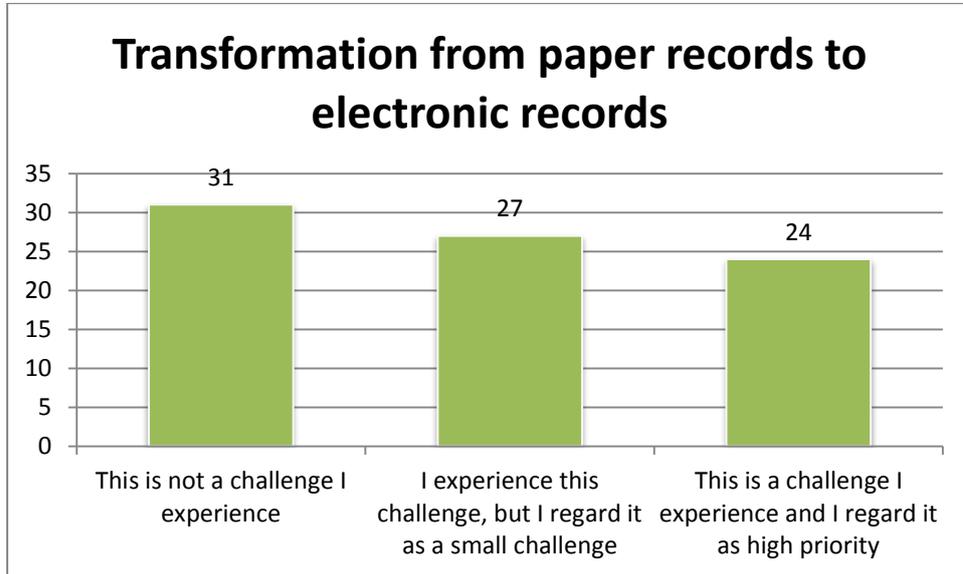


Figure 7.19: Responses – transformation from paper records to electronic records

Thirty eight percent (38%), thus 31 of the respondents did not experience this challenge; 33% (27) of respondents regarded this as a small challenge; while 29% (24) regarded it as a high priority challenge.

The weighted total score for this item is thus $(27 \times 1) + (24 \times 2) = 75$.

- *Privacy, confidentiality and legal challenges*

Figure 7.20 displays the distribution of the responses regarding this data quality challenges.

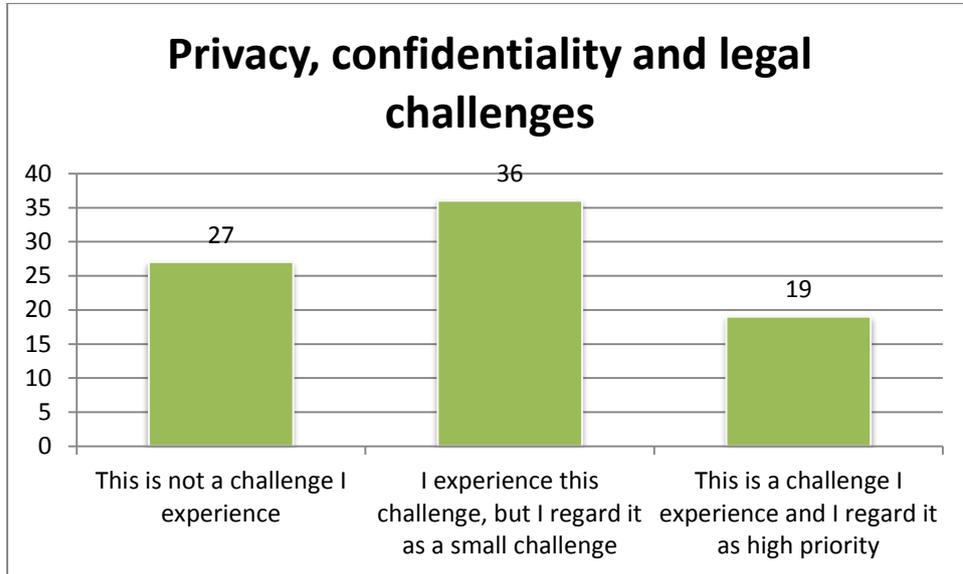


Figure 7.20: Responses – Privacy, confidentiality and legal challenges

Thirty three percent (33%), thus 27 of the respondents did not experience this challenge; 44% (36) of respondents regarded this as a small challenge; while 23% (19) regarded it as a high priority challenge.

The weighted total score for this item is thus $(36 \times 1) + (19 \times 2) = 74$.

- *Omissions of data*

Figure 7.21 displays the distribution of the responses regarding this data quality challenges.

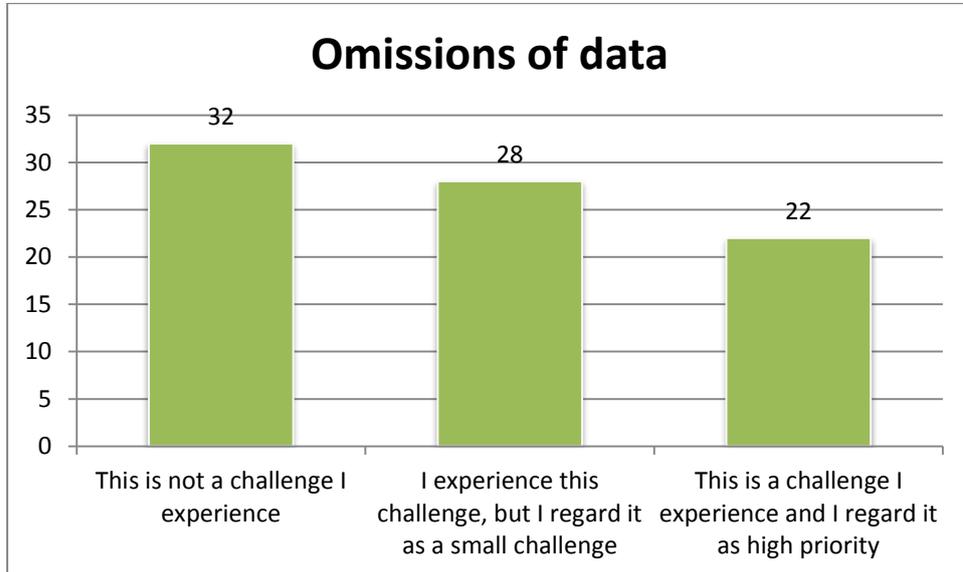


Figure 7.21: Responses – omissions of data

Thirty nine percent (39%), thus 32 of the respondents did not experience this challenge; 34% (28) of respondents regarded this as a small challenge; while 27% (22) regarded it as a high priority challenge.

The weighted total score for this item is thus $(28 \times 1) + (22 \times 2) = 72$.

- *Data not meeting all relevant needs*

Figure 7.22 displays the distribution of the responses regarding this data quality challenges.

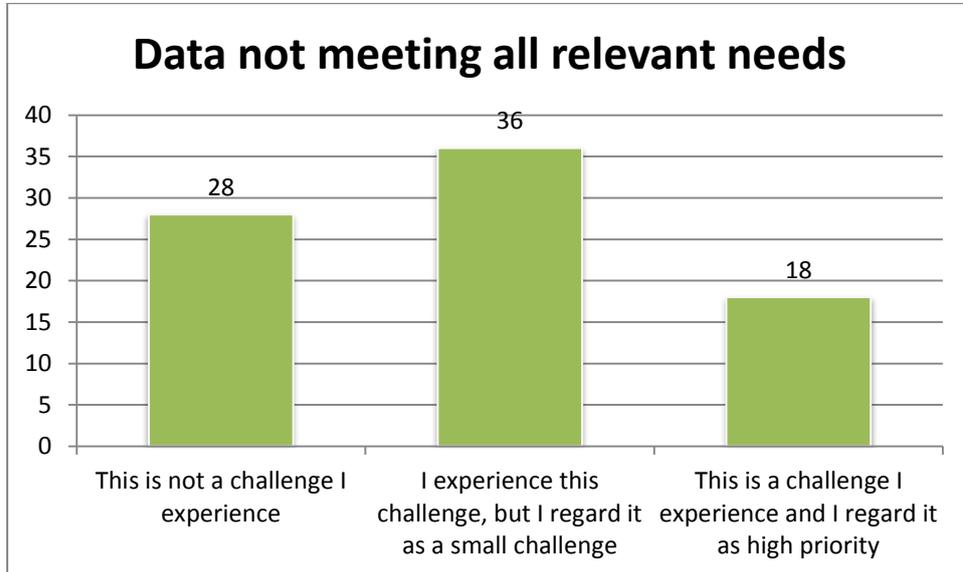


Figure 7.22: Responses – data not meeting all relevant needs

Thirty four percent (34%), thus 28 of the respondents did not experience this challenge; 44% (36) of respondents regarded this as a small challenge; while 22% (18) regarded it as a high priority challenge.

The weighted total score for this item is thus $(36 \times 1) + (18 \times 2) = 72$.

- *Information not suitable for analytic purposes*

Figure 7.23 displays the distribution of the responses regarding this data quality challenges.

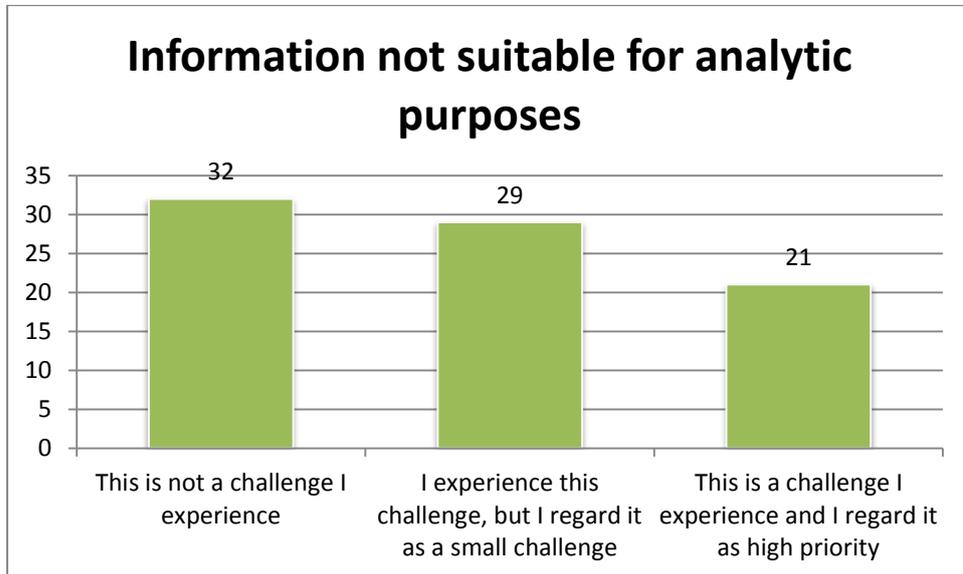


Figure 7.23: Responses – information not suitable for analytic purposes

Thirty eight percent (38%), thus 32 of the respondents did not experience this challenge; 35% (29) of respondents regarded this as a small challenge; while 26% (21) regarded it as a high priority challenge.

The weighted total score for this item is thus $(29 \times 1) + (21 \times 2) = 71$.

- *Financial hardware to purchase necessary hardware*

Figure 7.24 displays the distribution of the responses regarding this data quality challenges.

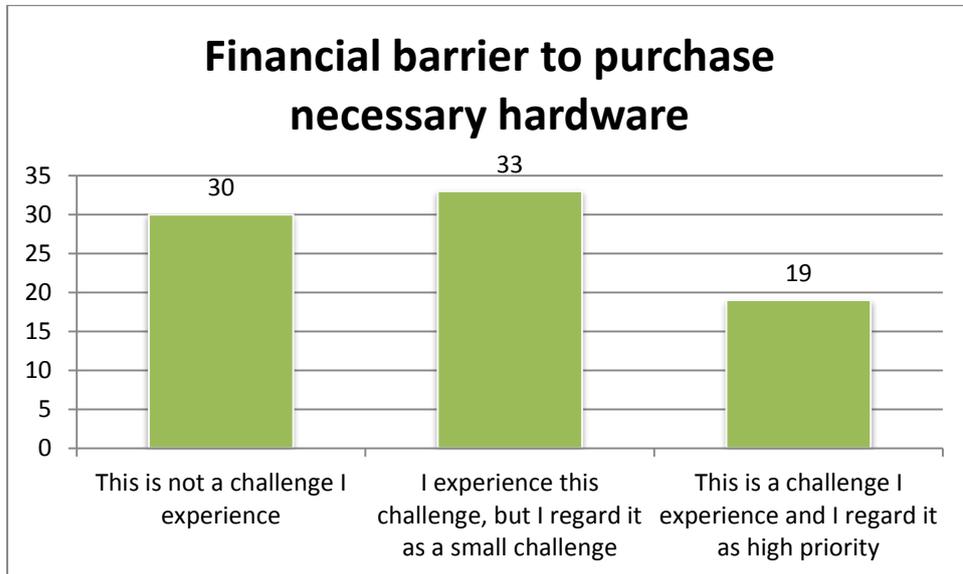


Figure 7.24: Responses – financial barrier to purchase necessary hardware and cost challenges in general

Thirty seven percent (37%), thus 30 of the respondents did not experience this challenge; 40% (33) of respondents regarded this as a small challenge; while 23% (19) regarded it as a high priority challenge.

The weighted total score for this item is thus $(33 \times 1) + (19 \times 2) = 71$.

- *Outdated data*

Figure 7.25 displays the distribution of the responses regarding this data quality challenges.

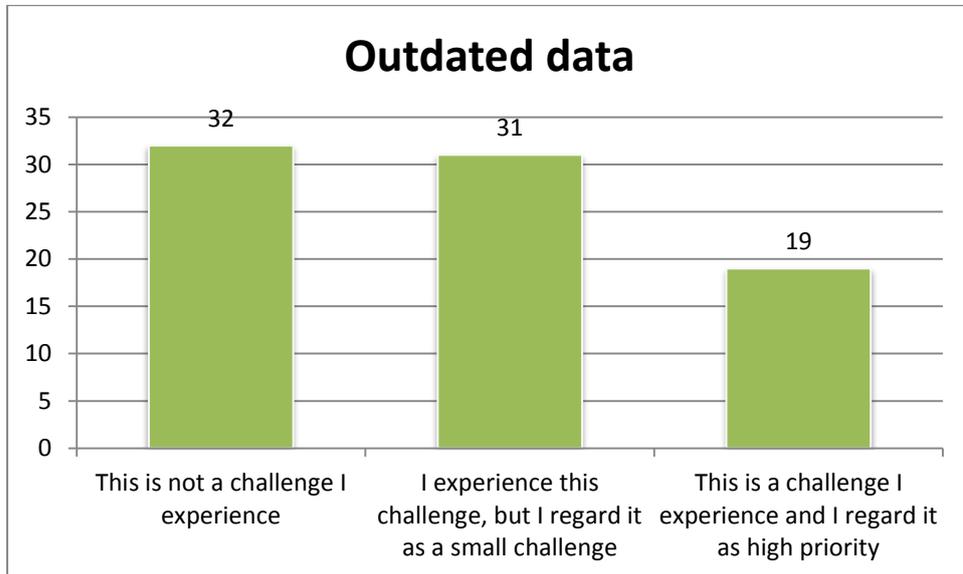


Figure 7.25: Responses – outdated data

Thirty nine percent (39%), thus 32 of the respondents did not experience this challenge; 38% (31) of respondents regarded this as a small challenge; while 23% (19) regarded it as a high priority challenge.

The weighted total score for this item is thus $(31 \times 1) + (19 \times 2) = 69$.

Table 7.4 provides the 18 significant data quality challenges in a ranked order, according to their weighted total scores.

Table 7.4: Ranked data quality challenges according to weighted total scores

Data quality challenge	Weighted total score	Rank
There is a need for more training for data users	108	1
<i>Data entry errors</i> (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)	107	2
Inaccurate data is collected from the patient	85	3
The shortage or absence of the necessary infrastructure, such as internet connections	83	4
Inconsistent definitions and formats because of different systems used (standardisation)	81	5
Physicians are hesitant to change existing processes	79	6
<i>Data governance</i> (this includes challenges such as: lack of assignment of responsibilities regarding data; lack of data administration; ambiguity of roles with regard to data tasks; a need for written data quality policies and procedures; managers' lack of thorough emphasis on the importance of data quality; ineffective organisational procedures; lack of data quality assessments)	78	7
Information is produced by subjectivity of the data capturer	77	8
Lack of appropriate software for data management	77	9
The length of the form is too long	76	10
Nurses' notes may go unread by physicians	76	11
Transformation from paper records to electronic records	75	12
Privacy, confidentiality and legal challenges	74	13
Omissions/errors of data	72	14
Data not meeting all relevant needs	72	15
Information is not suitable for analytic purposes	71	16
Financial barrier to purchase necessary hardware and cost challenges in general	71	17
Outdated data	69	18

The data quality challenges are ranked according to their weighted total scores and according to priority.

A conclusion and evidence based recommendations are given in Section 7.6.

7.6 CONCLUSION

As articulated in Section 1.3 the purpose of the study is to provide a prioritised set of data quality challenges experienced by users of healthcare systems in South Africa, to guide future health data interventions. This prioritised list will allow for recommendations which can assist health institutions in South Africa to ensure future data quality.

Eighteen statistically significant data quality challenges were identified and prioritised in this chapter. The prioritised list of data quality challenges could serve as a foundation for future health data quality interventions.

Chapter 7 presented the issues addressed by this research instrument, the data collection and analysis processes, the demographic information of the research participants and the data analysis and results of the quantitative research phase.

The final conclusion chapter is given in Chapter 8.

CHAPTER 8: CONCLUSION

Stage in research:

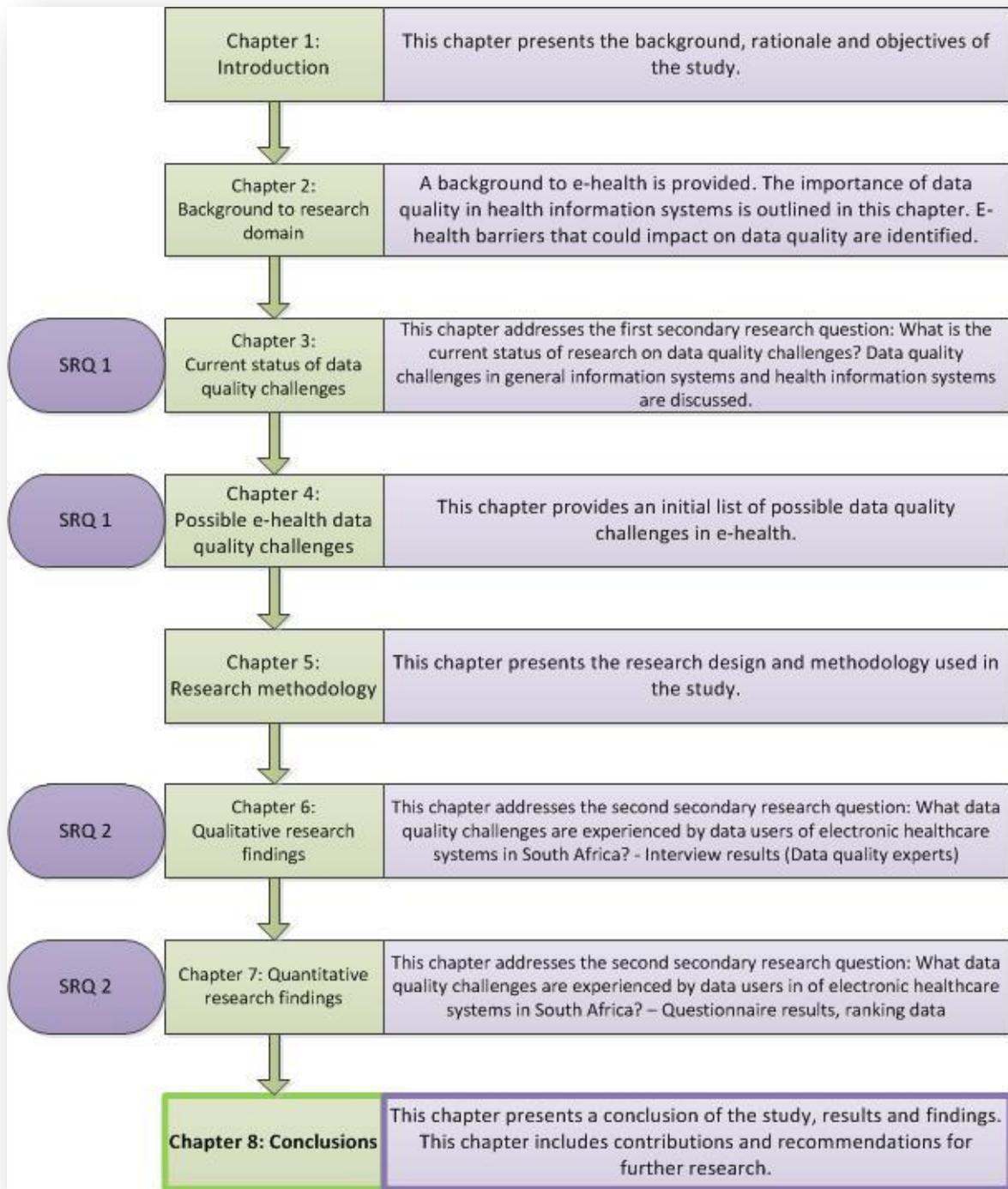


Figure 8.1: Chapter 8 – stage in research

Structure of Chapter 8

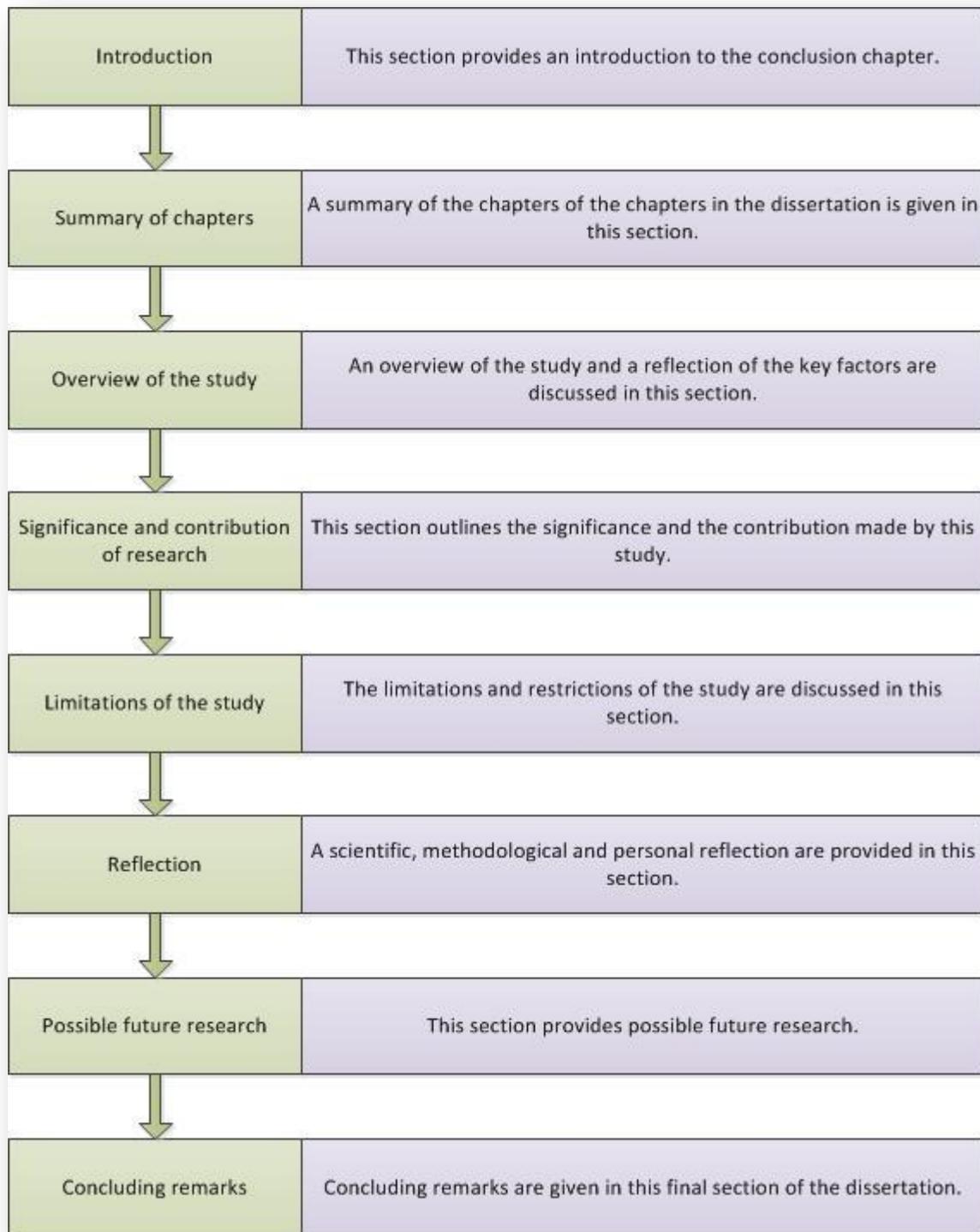


Figure 8.2: Structure of Chapter 8

8.1 INTRODUCTION

Chapter 7 presented the findings of the quantitative research. Data quality challenges experienced in electronic healthcare systems in South Africa were prioritised according to the quantitative data collected from data users of electronic healthcare systems in South Africa.

In this final chapter, the research results are briefly summarised with reference to the research questions. A summary of the chapters of the dissertation is given in Section 8.2. The overview of the study is discussed in Section 8.3 and the success of the research in answering the research question is reviewed. Section 8.4 reviews the contributions made by this study and an overview of limitations and restrictions is given in Section 8.5. Section 8.6 presents a reflection that leads to suggestions for further research in Section 8.7 which is followed by concluding remarks in Section 8.8.

Phase 4 of the research is presented in Chapter 8. Figure 8.1 illustrates the current and final phase of the research:

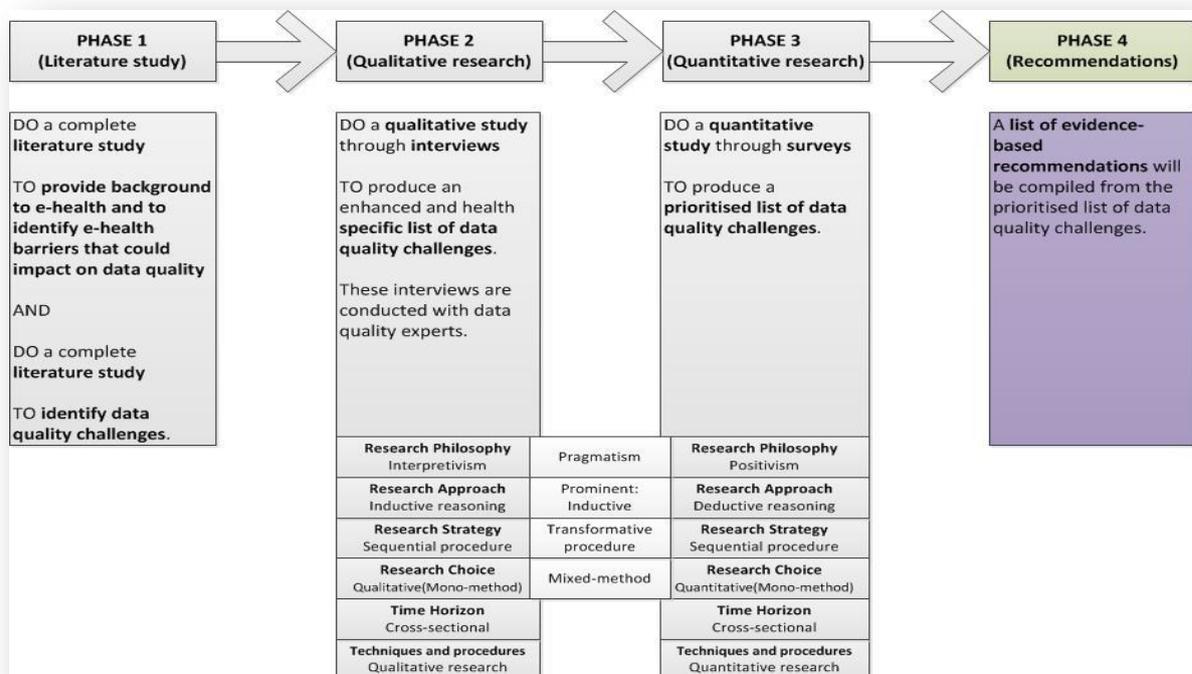


Figure 8.3: Current phase of research

8.2 SUMMARY OF CHAPTERS

Section 1.10 presented the structure for this dissertation and the focus of each of the 8 chapters. In this section, a brief summary highlighting the focal points of Chapter 1-7 is given.

- *Chapter 1: Introduction* defined this study. The research rationale and the context of the study were set by the research questions in Section 1.4.
- *Chapter 2: Background to the research domain – e-health* presented the current status of e-health in South Africa, the purpose and definition of e-health, benefits and challenges regarding e-health and finally outlined the importance of data quality in e-health.
- *Chapter 3: Current status of research on data quality challenges* provided the definition of data quality. Data quality challenges in general and health information systems were explored. The data quality challenges formed the backbone of this study – verified and extended through a qualitative study (Chapter 6) and prioritised by means of a quantitative study (Chapter 7).
- *Chapter 4: Possible e-health data quality challenges.* The possible e-health data quality challenges were identified as a result of the exploration in Chapter 2 and Chapter 3. The identified possible e-health data quality challenges were used to underpin the qualitative and quantitative phases of this research study.
- *Chapter 5: Research methodology* depicted the research methodology in the form of an adapted research process onion (Figure 5.23). The research design specifically catered to the needs of this study so as to ensure the required data would be collected and validated for reliability.
- *Chapter 6: Qualitative research findings* presented the results of the qualitative research study. Interviews with four data quality experts led to an enhanced and verified list of data quality challenges.
- *Chapter 7: Quantitative research findings* presented the results of the quantitative research study. The researcher used questionnaires to collect the

opinions and experiences of data users in electronic healthcare systems in South Africa in order to prioritise the data quality challenges identified in the literature and verified through interviews.

Section 8.3 presents the overview of the study.

8.3 OVERVIEW OF THE STUDY

The overview of the study is discussed in this section. The section provides the process of answering the research question (Section 8.3.1) and a reflection of the key findings of the research (Section 8.3.2).

8.3.1 RESEARCH PROCESS

This dissertation documents the exploration of the research problem articulated in Chapter 1 (Section 1.2.2). The research was contextualised in the e-health domain. Data quality challenges in electronic healthcare systems were explored in literature, verified through qualitative interviews and prioritised according to quantitative data collected from data users of electronic healthcare systems in South Africa. The research question that guided and framed this enquiry was:

What are the prioritised data quality challenges in electronic healthcare systems in South Africa?

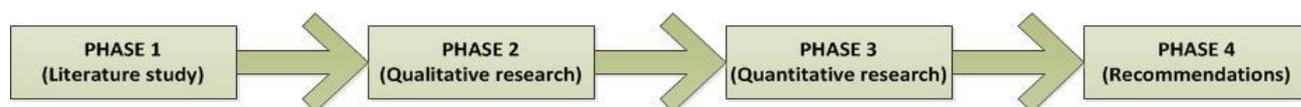
The following secondary research questions were formulated to support the investigation:

What is the current status of research on data quality challenges? and

What data quality challenges are experienced by data users of healthcare systems in South Africa?

The exploration followed a sequential exploratory mixed method approach towards answering the research question. The research process consisted of four phases. The process is briefly summarised in Table 8.1.

Table 8.1: Summary of research process



Chapter 2, 3 & 4	Chapter 6	Chapter 7	Chapter 8
<p>Phase 1 consisted of a literature review that included a background to e-health in South Africa, a list of e-health barriers which could impact on data quality and the importance of data quality in health information systems (Chapter 2) as well as the current status of research on data quality challenges – both in general and health information systems (Chapter 3). Chapter 4 provided an initial list of possible e-health data quality challenges, synthesising the results obtained in Chapter 2 and Chapter 3 of the dissertation.</p>	<p>The objective of this phase was to verify and enhance the initial list of possible data quality challenges that emerged from the literature reviews. Interviews with four data quality experts were conducted to realise the purpose of this research phase. The questions asked in the open-ended interviews were derived from the exploration of literature in Phase 1.</p>	<p>The objective of this phase was to prioritise the enhanced and verified list of data quality challenges that resulted from the second phase of research. Questionnaires were gathered from 82 data users of electronic healthcare systems in South Africa to realise the purpose of this research phase. The participants were asked to rate the data quality challenges that emerged from Phase 2 of the research.</p>	<p>The objective of this research phase was to provide evidence based recommendations regarding data quality challenges in electronic healthcare systems in South Africa.</p>
<p>Output</p> <p>The exploration in Phase 1 provided insights which enabled the identification of existing data quality challenges. The identified data quality challenges served as a foundation for the qualitative research in Phase 2.</p>	<p>Output</p> <p>The interviews resulted in an enhanced and practise-specific list of data quality challenges. This enhanced list included verified data quality challenges from the literature, excluded data quality challenges that could not be verified by the data quality experts, combined similar data quality challenges according to expert opinions and included additional e-health barriers and other data quality challenges recommended by the data quality experts.</p>	<p>Output</p> <p>The questionnaires resulted into a prioritised list of data quality challenges in electronic healthcare systems in South Africa. The prioritised list of data quality challenges excludes data quality challenges that were not statistically proven to be significant. The statistically significant data quality challenges were prioritised according to the weighted total score of the responses.</p>	<p>Output</p> <p>Evidence based recommendations for data quality interventions were provided. The recommendations were developed according to the prioritised list of data quality challenges which emerged from Phase 3 of the research.</p>
<p>Publications: See <i>list of publications</i>, numbers 1, 2 & 3.</p>	<p>Publications: See <i>list of publications</i>, number 4.</p>	<p>Publications: See <i>list of publications</i>, number 4.</p>	<p>Publications: See <i>list of publications</i>, numbers 4 & 5.</p>

Chapter 2, 3 & 4	Chapter 6	Chapter 7	Chapter 8
	Publications: See <i>list of publications</i> , number 4.		

8.3.2 REFLECTION OF KEY FINDINGS

The findings were derived from the literature, from the interviews with data quality experts and from the survey on the data quality challenges experienced by data users of electronic healthcare systems in South Africa.

Secondary research question 1

What is the current status of research on data quality challenges?

Secondary research question 1 was addressed in Chapter 2 and Chapter 3 of the dissertation. Chapter 2 provided a background to the research domain, namely e-health. The importance of data quality in health information systems was outlined in this chapter. Chapter 3 provided literature on the current status of research on data quality challenges. Much research was found on the definition of data quality and the causes of substandard data quality. The prior research mainly consisted of data quality challenges in general. A few researchers have done research on data quality in the e-health domain. Amongst many other challenges, data quality is depicted as a key challenge in the implementation of e-health. Although data quality interventions have been done in e-health, the interventions mainly focussed on one or two causes of substandard data quality. No research has been conducted to prioritise data quality challenges in the e-health domain. The research done in Chapter 2 and Chapter 4 of this dissertation resulted in the identification of possible e-health data quality challenges which served as a foundation for the qualitative and quantitative research phases.

Secondary research question 2

What data quality challenges are experienced by data users of healthcare systems in South Africa?

Secondary research question 2 was addressed in Chapter 6 and Chapter 7 of the dissertation. Interviews were conducted with data quality experts. The results of the

interviews led to the compilation of an enhanced and practise-specific list of data quality challenges. The enhanced list of data quality challenges was compiled by asking the participants to verify the data quality challenges identified in literature and by asking other relevant supporting questions. The data quality challenges that were not verified by the data quality experts were excluded from the list. Additional data quality challenges, identified through open-ended questions, were added to the list. Some related data quality challenges were combined with recommendation of the interviewees. The qualitative interviews showed that data quality challenges are the same across all domains, but that certain e-health barriers (such as the lack of proper infrastructure) could have a direct impact on the quality of data in electronic healthcare systems.

The verified and enhanced list of data quality challenges was used in the compilation of the research questionnaire. The questionnaire results led to a prioritised list of data quality challenges experienced by data users of electronic healthcare systems in South Africa. Data quality challenges were excluded from the list if it was not proven to be statistically significant.

The collection of the data quality challenges experienced by data users of electronic healthcare systems in South Africa (Chapter 4) were guided by a literature review (Chapter 2 and 3).

Main Research Question

What are the prioritised data quality challenges in electronic healthcare systems in South Africa?

The main research question addressed in this study relates to the data quality challenges experienced by data users of electronic healthcare systems in South Africa. The two secondary research questions led to the realisation of the main research question. Data quality challenges were identified through literature, verified and enhanced through interviews and prioritised through questionnaires.

A total of 18 data quality challenges were found to be statistically significant. The prioritised list of data quality challenges in electronic healthcare systems in South Africa is depicted in Table 8.2:

Table 8.2: Prioritised data quality challenges

DATA QUALITY CHALLENGE	PRIORITY
There is a need for more training for data users	1
<i>Data entry errors</i> (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)	2
Inaccurate data is collected from the patient	3
The shortage or absence of the necessary infrastructure, such as internet connections	4
Inconsistent definitions and formats because of different systems used (standardisation)	5
Physicians are hesitant to change existing processes	6
<i>Data governance</i> (this includes challenges such as: lack of assignment of responsibilities regarding data; lack of data administration; ambiguity of roles with regard to data tasks; a need for written data quality policies and procedures; managers' lack of thorough emphasis on the importance of data quality; ineffective organisational procedures; lack of data quality assessments)	7
Information is produced by subjectivity of the data capturer	8
Lack of appropriate software for data management	9
The length of the form is too long	10
Nurses' notes may go unread by physicians	11
Transformation from paper records to electronic records	12
Privacy, confidentiality and legal challenges	13
Omissions/errors of data	14
Data not meeting all relevant needs	15
Information is not suitable for analytic purposes	16
Financial barrier to purchase necessary hardware and cost challenges in general	17
Outdated data	18

Evidence-based recommendations follow to guide future health data interventions. The prioritised data quality challenges could further be divided into related groups. It is recommended that researchers or vendors, who wish to conduct data quality interventions in electronic healthcare systems in South Africa, should consider the prioritised list of data quality challenges.

The data quality challenges are divided into three groups, namely data quality challenges with regard to data users, data quality challenges with regard to IT and general data quality challenges. It is recommended that these groups should be considered when data quality interventions are conducted.

Table 8.3, 8.4 and 8.5 display the data quality challenges prioritised within their groups.

Table 8.3: Data quality challenges with regard to data users

Data quality challenges with regard to data users	Priority
There is a need for more training for data users	1
<i>Data entry errors</i> (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)	2
Physicians are hesitant to change existing processes	3
<i>Data governance</i> (this includes challenges such as: lack of assignment of responsibilities regarding data; lack of data administration; ambiguity of roles with regard to data tasks; a need for written data quality policies and procedures; managers' lack of thorough emphasis on the importance of data quality; ineffective organisational procedures; lack of data quality assessments)	4
Information is produced by subjectivity of the data capturer	5
Nurses' notes may go unread by physicians	6
Omissions/errors of data	7

The data quality challenges, displayed in Table 8.3, are associated with the data users. It is recommended that interventions on these data quality challenges should include training and teaching sessions with data users.

Table 8.4: Data quality challenges with regard to IT

Data quality challenges with regard to IT	Priority
The shortage or absence of the necessary infrastructure, such as internet connections	1
Inconsistent definitions and formats because of different systems used (standardisation)	2
Lack of appropriate software for data management	3
The length of the form is too long	4
Transformation from paper records to electronic records	5
Privacy, confidentiality and legal challenges	6
Data not meeting all relevant needs	7
Information is not suitable for analytic purposes	8
Outdated data	9

The data quality challenges displayed in Table 8.4 are associated with IT and computer hardware. It is recommended that interventions on these data quality challenges should include electronic healthcare systems assessments, the conforming to e-health standards and other technical interventions.

Table 8.5: General data quality challenges

General data quality challenges	Priority
Inaccurate data is collected from the patient	1
Financial barrier to purchase necessary hardware and cost challenges in general	2

The data quality challenges presented in Table 8.5 cannot be associated with the data users or IT. The significance and contribution of the research is discussed in Section 8.4.

8.4 SIGNIFICANCE AND CONTRIBUTION OF RESEARCH

In Section 1.6: *Rationale of research*, the contribution and importance of the study were discussed. The results of the study indicate that the research findings are important and can contribute to the community within information systems research, particularly for researchers whose area of expertise focuses on data quality, e-health and data quality within electronic healthcare systems.

The findings from this study are important to the academic body of knowledge in that:

- the findings offer exploratory insight into data quality and causes of substandard data quality in general and in the e-health domain;
- the findings offer an understanding of the e-health domain and the benefits and challenges of using electronic healthcare systems;
- data quality challenges specific to electronic healthcare systems in South Africa were identified;
- the study serves as a basis for further research initiatives regarding the quality of data in the e-health domain of South Africa; and
- the prioritisation of data quality challenges serves as a guide for future data quality interventions in electronic healthcare systems.

Data users, system developers and vendors of electronic healthcare systems may also benefit from this study, as they seek to enhance and manage the quality of data in electronic healthcare systems. It is recommended that data users (including practice managers), system developers and vendors should consider the prioritised list of data quality challenges in order to manage and improve the quality of data in electronic healthcare systems. The research findings from this study are important to data users, system developers and vendors of electronic healthcare systems, in that:

- the findings offer insight into general causes of substandard data;
- the findings offer an understanding of the e-health domain and the challenges associated with the usage of electronic healthcare systems;
- the study provides key, significant challenges experienced by data users of electronic healthcare systems in South Africa; and
- the study provides a prioritised list of data quality challenges to guide the priority of future health data quality interventions.

The knowledge gained from this study in essence resulted in a summarised, refined, relevant and prioritised list of data quality challenges in electronic healthcare systems in South Africa. This list could guide future research and data quality interventions.

The limitations of the study are discussed in Section 8.5.

8.5 LIMITATIONS OF RESEARCH

Shortcomings in research are not dismissible (Hofstee, 2006). Patton (2004) states that it is important to be clear about a study's limitations in order to address criticisms. The limitations for this study are briefly discussed.

- One of the challenges of the quantitative research design was that the research participants are from different medical settings (doctors, pathologists, pharmacists and nurses). Some of the participants did not understand some of the data quality challenges mentioned in the survey, because it did not apply to their setting. An example is that some of the participants only work on one electronic healthcare system, but one of the data quality challenges mentioned in the questionnaire is interoperability. If the study was conducted in one hospital with many systems, this limitation could have been prevented.
- The purposeful selection of research participants affects generalisability. The respondents were purposefully selected to participate in the study. Only data users of electronic healthcare systems in the health domain were selected. Many possible participants do not make use of electronic healthcare records, and were thus excluded from the study.
- The quantitative phase of research mostly included data users from the private health sector in South Africa. A more representative list of data quality challenges in South Africa could have included input from data users in the public health sector of South Africa as well.

The data quality challenges that were prioritised in this study therefore represent a single perspective in context. Some of the limitations can be regarded as possibilities for future research in the same research area.

8.6 REFLECTIONS

This section provides a scientific, methodological and personal reflection of the study.

8.6.1 SCIENTIFIC REFLECTION

This study outlined data quality challenges within electronic healthcare systems in South Africa. The prioritised data quality challenges can be regarded as a basis for future research and data quality interventions. Although data quality is a global challenge across many domains, the study aimed to summarise data quality challenges in the e-health domain. The interdisciplinary nature of data quality and e-health could lead to specific investigations, interventions and theories. The challenge for researchers working in this domain would be to find ways to enhance the data quality in electronic healthcare records. Another challenge that researchers may have, is to approach the research problem from an information system point of view, instead of the data users' point of view. The data quality challenges identified and prioritised in this study could serve as a foundation to enhance data quality in electronic healthcare systems in South Africa.

8.6.2 METHODOLOGICAL REFLECTION

The methodological reflection is provided by answering the following questions, significant to the study:

- *Was the use of the research methodology chosen the best in answering the research questions?*

The research methodology followed in this study was chosen to accommodate the two domains of research in this study and to support the pragmatic approach of the practical application of theory.

The qualitative and quantitative phases of research were used to support, enhance and summarise the theory of data quality challenges in the e-health domain. Possible e-health data quality challenges were identified and served as a foundation for the qualitative and quantitative phases of the research that followed. The mixed method approach was followed to firstly get in depth insight into data quality challenges in electronic healthcare systems and to finally prioritise the challenges.

Within the opportunities and limitations of the study, I suggest that the most appropriate method was applied to answer the research question.

- *Was the sample that was used in the research justifiable?*

This study included two samples: i) data quality experts used in qualitative interviews; and ii) data users of electronic health systems in South Africa.

A sample of four data quality experts were used for the qualitative research. As discussed in Section 5.2.3 a sample size between three and four experts is the most appropriate in the field of Human Computer Interaction (Nielson & Launder, 2000; Ouma, 2014).

A total of 82 responses were received for the quantitative phase of the study. As discussed in Section 5.2.3, although the sample size may not practically be representative of all data users of electronic healthcare users in South Africa, the sample size is statistically significant according to the following sample size formula (Falk, Marohn & Tewes, 2002): $n = 2 \times m + 1$, where n is the sample size and m is the number of questions on the questionnaire.

We can therefore conclude that the sample size for both the qualitative and quantitative phases of research is justifiable.

- *Were the data analysis techniques justifiable?*

The qualitative research is based on the hermeneutical principles provided by Klein and Meyers (1999). In Section 5.2.6 the researcher explains how the study conforms to these standards. The manner, in which hermeneutics are applied in this study, is explained in Section 6.5.

Furthermore, data triangulation was used to reach the final results of the study. The triangulation exercise in the study is explained in Section 5.2.6 and Section 7.5.

Since the data analysis techniques are grounded on methodological principles, I conclude that the data analysis techniques were justifiable.

8.6.3 PERSONAL REFLECTION

The road to the finish line of this research study taught me to stay in focus despite of all the changes that came my way. The initial idea of the research was to prioritise data quality challenges in public clinics. This changed when I learned that most public clinics do not yet use electronic healthcare systems. The change of research setting did not affect the study, but challenged me to stay focussed.

When I look back on the study, I am satisfied with the theoretical contribution made, the methodology followed and my personal development during this time.

8.7 POSSIBLE FUTURE RESEARCH

This study could serve as the foundation for various research projects in the future. Possible future research may include:

- Data quality interventions could be done based on the prioritised list of data quality challenges.
- Data quality challenges could be identified from an information system viewpoint, instead of the user viewpoint.
- The study could be narrowed to a specific healthcare system used in South Africa.
- The study could be implemented in several other domains, such as business enterprises and the education sector.
- Although e-health has been studied over the past few years and in various countries, literature shows that data quality remains a challenge. Similar studies could be conducted in other countries.
- One limitation of the research (Section 8.6) is the generalisability of the findings. Similar studies with larger and randomised samples could be done in the future.
- A comparative study that is focussed on the other end of the demographical spectrum in South Africa, namely data quality challenges related to public health in rural areas.

The development of e-health across the globe is a major project. Data quality proves to be a challenge in this domain. As shown in literature, data quality plays a significant role in the quality of healthcare. Although this study proposes data quality challenges to improve, a vast amount of research could still be done in this field of research across the world.

8.8 CONCLUDING REMARKS

The purpose of this study was to compile a prioritised list of data quality challenges experienced by data users of electronic healthcare systems in South Africa to guide future data quality interventions in this regard. The researcher followed a pragmatic

research philosophy to realise the purpose of this study. Qualitative and quantitative phases of the research were conducted sequentially to practically apply theories of data quality and e-health.

The study contributed to the academic body of knowledge in the fields of data quality and e-health. The practical contribution of the study could lead to improved data quality in electronic healthcare systems.

The effect of the quality of data on patient care, communication, decision-making and finances cannot be ignored. Data quality is the building block of all organisations, including electronic healthcare systems.

The first step in fixing a problem lies in the acknowledgement thereof. The acknowledgement of data quality challenges in electronic healthcare systems in South Africa could lead to improved quality of data and therefore improved healthcare.

REFERENCES

- Accenture. (2010). Information Governance: The foundation for effective e-Health. Retrieved from http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_100473_InfoGovPoV_Final.pdf
- Adebesin, T. F. (2014). *A method for the selection of e-health standards to support interoperability in healthcare information systems* (PhD). Nelson Mandela Metropolitan University.
- Ahmadnezhad, E. (2009, November). *Research Design Mixed Methods*. Tehran University of Medical Sciences.
- Altman, D., Machin, D., Bryant, T., & Gardner, M. (Eds.). (2013). *Statistics with confidence: confidence intervals and statistical guidelines*. John Wiley & Sons.
- Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health Information systems—problems and challenges. *International journal of medical informatics*, 71(2), 125-135.
- Anderson, J. G. (2007). Social, ethical and legal barriers to e-health. *International journal of medical informatics*, 76(5), 480-483.
- Atreja, A., Gordon, S. M., Pollock, D. A., Olmsted, R. N., & Brennan, P. J. (2008). Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention and control. *Am J Infect Control*, 36, 37–46.
- Babbie, E. (2005). *The Basics of Research*. Belmont, CA: Thomson Wadsworth.
- Beacon Community Program. (2013, July). Capturing High Quality Electronic Health Records Data to support Performance Improvement. Retrieved from <http://www.healthit.gov/sites/default/files/onc-beacon-lg3-ehr-data-quality-and-perform-impvt.pdf>
- Botha, M., Botha, A., & Herselman, M. (2014, December). Data quality challenges: A content analysis in the e-health domain. In *Information and Communication Technologies (WICT), 2014 Fourth World Congress on* (pp. 107-112). IEEE.
- Bowen, M. (2012). eHealth Observatory Data Quality Evaluation Guide Data Quality Evaluation Guide. Retrieved from <https://www.google.co.za/#q=emr+data+quality+evaluation+guide+bowen>

- Bowling, J. M., Rimer, B. K., Lyons, E. J., Golin, C. E., Frydman, G., & Ribisl, K. M. (2006). Methodologic challenges of e-health research. *Evaluation and program planning, 29*(4), 390-396.
- Braa, J., & Hedberg, C. (2002). The struggle for district-based health information systems in South Africa. *The information society, 18*(2), 113-127.
- Braa, J., Monteiro, E., & Sahay, S. (2004). Networks of action: sustainable health information systems across developing countries. *Mis Quarterly, 337-362*.
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Thousand Oaks, CA: Sage.
- Burns, F. (2000). *Information for Health: An information Strategy for The Modern NHS 1998-2005: A national Strategy for Local Implementation*. Department of Health Promotions.
- Calman, N., Hauser, D., Lurio, J., Wu, W. Y., & Pichardo, M. (2012). Strengthening public health and primary care collaboration through electronic health records. *American journal of public health, 102*(11), e13-e18.
- Car, J., Black, A., Anandan, C., Cresswell, K., Pagliari, C., McKinstry, B., ... Sheikh, A. (2008). The Impact of eHealth on the Quality & Safety of Healthcare. Imperial College. Retrieved from <https://www1.imperial.ac.uk/resources/32956FFC-BD76-47B7-94D2-FFAC56979B74/>
- Carter, J. (2012, November 26). Wrestling with EHR Data Quality. *EHR Science: Explorations in the Design and Implementation of Clinical Information Systems*. Retrieved from <http://ehrscience.com/2012/11/26/wrestling-with-ehr-data-quality/>
- Chapman, A. (2011). Positivism & Objective Quantitative Research. April Chapman's Sociology Lessons. Retrieved from <http://www.slideshare.net/aprichelle/a2-lesson-2>
- Clandinin, D. J., & Connelly, F. M. (2000). Narrative inquiry: Experience and story in qualitative research.
- Clark, V. L. P., & Creswell, J. W. (2011). Designing and conducting mixed methods research.
- Cooper, D. R., & Schindler, P. S. (2003). *Business research methods*.
- Creswell, J. W. (2007). Qualitative enquiry and research design: Choosing among five approaches.

- Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Incorporated.
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.
- Darlington, Y., & Scott, D. (2003). *Qualitative research in practice: Stories from the field*. Allen & Unwin.
- Deetz, S. (1996). Crossroads-describing differences in approaches to organization science: Rethinking Burrell and Morgan and their legacy. *Organization science*, 7(2), 191-207.
- De Clercq, E. (2008). From a conceptual problem-oriented electronic patient record model to running systems: A nationwide assessment. *International journal of medical informatics*, 77(5), 346-353.
- De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, 21(2), 233.
- Denscombe, M. (2001). *The Good Research Guide for small-scale social research projects*. Buckingham: Open University Press.
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Transaction publishers.
- Department of Health South Africa. (2012). eHealth Strategy South Africa 2012-2016. Retrieved from http://www.doh.gov.za/docs/stratdocs/2012/eHealth_Strategy_South_Africa_2012-2016.pdf
- De Villiers, M. R. (2005). Three approaches as pillars for interpretive information systems research: development research, action research and grounded theory. *Proceedings of the 2005 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries.*, 142–151.
- Donnelly, J., & Trochim, W. (2007). The research methods knowledge base.
- Eckerson, W. (2014, April). *Data Governance for the Enterprise: Trends in the use of Data Quality, Master Data Management and Metadata Management*. Retrieved from Techtarget website: <http://www.techtarget.com>

- Eysenbach, G. (2001). What is e-health?. *Journal of medical Internet research*, 3(2).
- Falk, M., Marohn, F., & Tewes, B. (2002). Categorical Data Analysis. In *Foundations of Statistical Analyses and Applications with SAS* (pp. 139-183). Birkhäuser Basel.
- Fetterman, D. M. (Ed.). (2010). *Ethnography: Step-by-step* (Vol. 17). Sage.
- Fontenot, S. F. (2013). The Affordable Care Act and Electronic Health Care Records: Does today's technology support the vision of a paperless health care system? *PEJ*, 72–76.
- Francis, T. (2013). Electronic Health Records: Where we are and where we're going. *Physician executive*, 39(4), 82.
- Fry, J. P., & Neff, R. A. (2009). Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. *Journal of medical Internet research*, 11(2).
- Gadamer, H. G. (1998). *Praise of theory: Speeches and essays*. Yale University Press.
- Gerber, T., Olazabal, V., Brown, K., & Pablos-Mendez, A. (2010). An agenda for action on global e-health. *Health Affairs*, 29(2), 233-236.
- Glaser, B. G., & Strauss, A. L. (2009). *The discovery of grounded theory: Strategies for qualitative research*: Transaction Publishers.
- Goldkuhl, G. (2004). Design theories in information systems-a need for multi-grounding. *Journal of Information Technology Theory and Application (JITTA)*, 6(2), 7.
- Greiver, M., Barnsley, J., Aliarzadeh, B., Krueger, P., Moineddin, R., Butt, D. A., ... & Kaplan, D. (2011). Using a data entry clerk to improve data quality in primary care electronic medical records: a pilot study. *Informatics in primary care*, 19(4), 241-250.
- Groves, R. M. (1987). Research on survey data quality. *Public opinion quarterly*, S156-S172.
- Haug, A., Stentoft Arlbjørn, J., Zachariassen, F., & Schlichter, J. (2013). Master data quality barriers: an empirical investigation. *Industrial Management & Data Systems*, 113(2), 234-249.
- Häyrynen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics*, 77(5), 291-304.

- Health Information and Quality Authority. (2011, April 26). International Review of Data Quality. Health Information and Quality Authority. Retrieved from <http://www.hiqa.ie/publications/international-review-data-quality>
- Health System Trust. (2011). South African health review. Health System Trust.
- Health System Trust. (2008). South African health review. Health System Trust.
- Heeks, R. (2006). Health information systems: Failure, success and improvisation. *International journal of medical informatics*, 75(2), 125-137.
- Hess, J., Moore, J., Pascale, J., Rothgeb, J., & Keeley, C. (2001). The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality. *Public Opinion Quarterly*, 65(4), 574-584.
- Hofstee, E. (2006). Constructing a good dissertation: a practical guide to finishing a Master's, MBA or PhD on schedule. EPE.
- Horbatuk, E. (2011, July). *Electronic Health Records (EHR): Benefits and Challenges for Data Quality*. Presented at the The Fifth MIT Information Quality Industry Symposium.
- Iakovidis, I. (1998). Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe. *International journal of medical informatics*, 52(1), 105-115.
- Jarosławski, S., & Saberwal, G. (2014). In eHealth in India today, the nature of work, the challenges and the finances: an interview-based study. *BMC medical informatics and decision making*, 14(1), 1.
- Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., ... & Blumenthal, D. (2009). Use of electronic health records in US hospitals. *New England Journal of Medicine*, 360(16), 1628-1638.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.
- Kaplan, B., & Maxwell, J. A. (1994). Qualitative Research Methods for Evaluating Computer Information Systems. *Evaluating Health Care Information Systems: Methods and Applications*, 45-68.

- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Prentice-Hall, Inc.
- Kern, L. M., Edwards, A., & Kaushal, R. (2014). The patient-centered medical home, electronic health records, and quality of care. *Annals of internal medicine*, 160(11), 741-749.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1), 7-15.
- Khalifehsoltani, S. N., & Gerami, M. R. (2010, January). E-health challenges, opportunities and experiences of developing countries. In *e-Education, e-Business, e-Management, and e-Learning, 2010. IC4E'10. International Conference on* (pp. 264-268). IEEE.
- Khan, M. I., Singh, R. K., & Dey, P. K. (2010). Simplified Approach for Data Warehouse Quality Measurement. *International Journal of Computer Applications*, 12(7), 32-35.
- Kimchi, J., Polivka, B., & Stevenson, J. S. (1991). Triangulation: operational definitions. *Nursing research*, 40(6), 364-366.
- Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS quarterly*, 67-93.
- Kruss, G., Visser, M., Haupt, G., & Aphane, M. (2012). *Academic Interaction with External Social Partners: Investigating the Contribution of Universities to Economic and Social Development Paperback*. HSRC Press.
- Kwankam, S. Y. (2004). What e-Health can offer. *Bulletin of the World Health Organization*, 82(10), 800-802.
- Lee, A. S. (1989). A scientific methodology for MIS case studies. *MIS quarterly*, 33-50.
- Leedy, P. D., & Omrod, J. E. (2005). Practice Research. *Planning and Design*, 8.
- Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, 9(1), 181-212.
- Lobach, D. F., & Detmer, D. E. (2007). Research challenges for electronic health records. *American Journal of Preventive Medicine*, 32(5), S104-S111.

- Lucas, A. (2010, June). Corporate data quality management: From theory to practice. In *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on* (pp. 1-7). IEEE.
- Mahmood, K. (2013). *Mixed Methods Research*. University of the Punjab, Lahore, Pakistan.
- Maree, K., & Van der Westhuizen, C. (2007). Planning a research proposal. *First steps in research. Pretoria: Van Schaik*, 23-45.
- McNiff, J. with Whitehead, J.(2002). *Action research: Principles and practice*, 2.
- Measor, L. (1985). Interviewing: A strategy in qualitative research. *Strategies of educational research: Qualitative methods*, 55-77.
- Minear, M. Hospital EHR: An eHealth bootcamp for rural providers. In *Annual Rural Health Conference, California*.
- Motsoaledi, A. (2014). National Health Act (61/2003): National Health Normative Standards Framework for Interoperability in eHealth. Government Gazette 37583 (Government Notice 314).
- Moustakas, C. (1994). *Phenomenological research methods*. Sage Publications.
- Mphatswe, W., Mate, K. S., Bennett, B., Ngidi, H., Reddy, J., Barker, P. M., & Rollins, N. (2012). Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa. *Bulletin of the World Health Organization*, 90(3), 176-182.
- Myers, M. D. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*, 21, 241-242.
- Myers, M. D. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*, 21, 241-242.
- Nielsen, J., & Landauer, T. K. (2003). Why You Only Need to Test with 5 Users. Jakob Nielsen's Alertbox. 2000.
- Nix, M., & Rosenau, P. (2011, July). *Anticipating Data Quality Challenges in EHR Implementations*. Presented at the The Fifth MIT Information Quality Industry Symposium.

- Nuq, P. A., & Aubert, B. (2013). Towards a better understanding of the intention to use eHealth services by medical professionals: The case of developing countries. *International Journal of Healthcare Management*, 6(4), 217-236.
- Oates, B. J. (2008). *Researching Information Systems and Computing* (2d Ed.). London: Sage Publications Ltd.
- Olivier, M. (2009). *Information Technology Research: A Practical guide for Computer Science and Informatics*. Van Schaik Publishers.
- Ouma, S. (2013). *M-health user experience framework for the public healthcare sector* (PhD). Nelson Mandela Metropolitan University.
- Ouma, S. (2008). *Assessing technology for rural hospitals* (Masters). Tshwane University of Technology.
- Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1), 130-168.
- Penoyer, D., Kendall, H., Noblin, A., Bullard, T., Talbert, S., Wilson, J., ... Briscoe, J. (2014). Use of Electronic Health Record Documentation by Healthcare Workers in an Acute Care Hospital System. *Journal of Healthcare Management*, 59(2), 130–144.
- Recommended Requirements for Enhancing Data Quality in Electronic Health Records. (2001). RTI International. Retrieved from https://www.rti.org/pubs/enhancing_data_quality_in_ehrs.pdf
- Redman, T. C. (1995). Improve data quality for competitive advantage. *MIT Sloan Management Review*, 36(2), 99.
- Reynolds, R. B. and Bowman, E. D. (2010). Paper-Based and Hybrid Health Records. In K. M. LaTour and S. E. Maki (Eds.), *Health Information Management: Concepts, Principles, and Practice*. Chicago: AHIMA Press.
- Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning.
- Rodrigues, J. R. (2008). Compelling issues for adoption of e-health. *The Commonwealth Ministers Preference book*.
- Rodríguez, C. C. G., & Riveill, M. (2010). e-Health monitoring applications: What about Data Quality?.
- Rosaldo, R. (1989). Imperialist nostalgia. *Representations*, 107-122.

- Salati, M., Brunelli, A., Dahan, M., Rocco, G., Van Raemdonck, D. E. M., & Varela, G. (2011). Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) Database. *European Journal of Cardio-Thoracic Surgery*, *40*, 91–98.
- Samyshkin, Y., & Timoshkin, A. (2004). An Information Strategy for the modern NHS. *Centre for health Management, Imperial College London*.
- Saunders, M., Lewis, P., & Thornhill, A. (2015). *Research methods for business students* (7th Ed.). Pearson.
- Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires* (Vol. 2). Rowman Altamira.
- Schiavo, R. (2013). *Health communication: From theory to practice*. John Wiley & Sons.
- Singh, R., & Singh, K. (2010). A descriptive classification of causes of data quality problems in data warehousing. *International Journal of Computer Science Issues*, *7*(3), 41-50.
- Smithson, S., & Cornford, T. (1996). *Project Research in Information Systems*.
- Soares, S. (2015). *The Chief Data Officer Handbook for Data Governance* (1st ed.). USA: MC Press.
- Soellner, R., Huber, S., & Reder, M. (2014). The concept of eHealth literacy and its measurement: German translation of the eHEALS. *Journal of Media Psychology*, *26*(1), 29–38.
- Strode, A., Slack, C., Grant, C., Bamjee, S., & Mushariwa, M. (2009). Directory of the Legal Rights of Child and Adolescent Research Participation in South Africa. *HIV/AIDS Vaccines Ethics Group, University of KwaZulu-Natal*.
- Strong, D. M., Yang, W. L., & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM*, *40*(5), 103–110.
- Strong, D. M., Yang, W. L., & Wang, R. Y. (1997). 10 Potholes in the Road to Information Quality. *Cybersquare Computer*, 38–46.
- Subar, A. F., Ziegler, R. G., Thompson, F. E., Johnson, C. C., Weissfeld, J. L., Reding, D., ... & Hayes, R. B. (2001). Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *American journal of epidemiology*, *153*(4), 404-409.

- The Markle Foundation. (2006). Background Issues on Data Quality. Connecting for health common framework. Retrieved from http://library.ahima.org/xpedio/groups/public/documents/external/bok1_031232.pdf
- Theobald, D. (2014). The. Road to Health Data Equity. *Harvard International Review*, 35(4), 48.
- Thurmond, V. A. (2001). The point of triangulation. *Journal of nursing scholarship*, 33(3), 253-258.
- Tomhave, B. L. (2005). Alphabet soup: Making sense of models, frameworks, and methodologies. *George Washington University*.
- Trauth, E. M., & Jessup, L. M. (2000). Understanding computer-mediated discussions: positivist and interpretive analyses of group support system use. *MIS Quarterly*, 43-79.
- Ure, J., Procter, R., Martone, M., Porteous, D., Lloyd, S., Job, D., ... Anderson, S. (2006). Data Integration in eHealth: A Domain/Disease Specific Roadmap. UK National e-Science Centre.
- Verma, R. (2012). Data quality and clinical audit. *Anaesthesia and Intensive Care Medicine*, 13(8), 397–399.
- Vaziri, R., & Mohsenzadeh, M. (2012). A Questionnaire-Based Data Quality Methodology. *International Journal of Database Management Systems*, 4(2), 55.
- Waegemann, C. P. (2002). Status report 2002: electronic health records. *Boston: Medical Records Institute*.
- Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, 41(2), 58-65.
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 5-33.
- Wentzel, J., van Velsen, L., van Limburg, M., de Jong, N., Karreman, J., Hendrix, R., & van Gemert, J. E. W. C. (2014). Participatory eHealth development to support nurses in antimicrobial stewardship. *BMC medical informatics and decision making*, 14(1), 45.
- Westra, B. L., Delaney, C. W., Konicek, D., & Keenan, G. (2008). Nursing standards to support the electronic health record. *Nursing outlook*, 56(5), 258-266.

- Wills, J. W. (2007). *Foundations of Qualitative Research: Interpretive and Critical Approaches*. Thousand Oaks, CA: Sage.
- World Health Organisation. (2003). Improving data quality: a guide for developing countries. Retrieved from http://www.wpro.who.int/publications/docs/Improving_Data_Quality.pdf
- Wright, G., & Odama, A. (2012). Health data ownership and data quality: clinics in the Nyandeni District, Eastern Cape, South Africa. *Engineering Management Research*, 1(2), p146.
- Yakout, M., Elmagarmid, A. K., & Neville, J. (2010, March). Ranking for data repairs. In *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on* (pp. 23-28). IEEE.
- Yin, R. K. (2003). Case study research design and methods third edition. *Applied social research methods series*, 5.

APPENDIX A: INTERVIEW CONSENT FORM

Research Information and Consent Form

Data quality challenges in electronic healthcare systems in South Africa

Introduction

This is to get consent for your participation in the research conducted by **Marna Botha**. The research is for a MSc in Computing that I am currently undertaking with the University of South Africa (UNISA), under the supervision of **Dr. Adele Botha** (Email: abotha@csir.co.za, Telephone: 012 841 3265) and **Prof Marlien Herselman** (Email: mherselman@csir.co.za, Telephone: 012 841 3081)

Purpose of research

The study is aimed at gathering your views and perceptions regarding the data quality challenges that you experience. The purpose of the study is to compile a prioritised or ranked list of data quality challenges as experienced by data users.

Procedure

The interview will require approximately 40 minutes of your time. The researcher will ask you a number of open ended questions regarding data and data quality challenges. The interview will be recorded for analysis purposes.

Confidentiality

The input you provide will be treated confidentially and only used towards the completion of the afore-mentioned qualification. All data will be used in summary form without reference to any individual.

Participation

Participation in this research study is voluntary, and you have the right to, at any time, withdraw or refuse to participate. The purpose of the study will be explained in the questionnaire.

Benefits and compensation

There aren't any direct benefits for your participation. All findings will be used for the completion of the academic qualification mentioned. No compensation will be provided to anyone partaking in this research.

Risks and discomforts

There are no risks or discomfort associated with your participation. All answers from you and other participants will be analysed collectively. Individual answers will therefore not be linked to any names, positions and companies of participants.

Participant consent

I have read and understood all the above. I willingly choose to participate in this study.
Full name (optional) _____
Date: _____
Signature: _____

APPENDIX B: OPEN INTERVIEW QUESTIONS

Interview Process and Open Ended Questions

1. Interview Process

1.1 Opening

A. Establish Rapport

- Thank you [Name of participant] for taking time out of your schedule and participating in this research.

B. State purpose

- I am conducting this research for my MSc in Computing at the University of South Africa (UNISA). It is aimed at gathering your input regarding data quality challenges experienced by data users.

C. State time and procedure

- This session will take approximately 40 minutes. I will be asking you a number of questions related data, data quality and data quality challenges, and I need you to provide me with the answer you deem fit.

1.2 Body

A. Prompt to read and sign research consent form

- Before we proceed, I need you to please go through the research consent form. This in addition to what we have discussed is to make you understand what the research is about and all the surrounding conditions for your participation. Once you have read and understood everything, please provide your signature at the bottom of the page. We will then begin.

B. Start with open ended interview questions

1. What data quality challenges/difficulties/issues do you experience every day?

2. Go through list [give the interviewee a list with data quality challenges identified from literature], state the validity/ importance of each of the identified data quality challenges.
3. Which data quality challenges are in your experience the greatest challenges? Please explain.
4. Which data quality challenges are the easiest to resolve? Please explain.
5. Do you believe that data quality challenges are the same across all domains? Please explain.
6. What data quality challenges do you think are the greatest challenges in the case of developing countries health information systems?
7. Which of the e-health barriers [give interviewee a list of e-health barriers identified in literature] could, in your opinion have a direct impact on data quality in health information systems?
8. Why do you think it is important to identify data quality challenges?

1.3 Closing

- We are at the end of the interview. I appreciate your time and input. As previously stated, all input gathered from you will be treated confidentially. Thank you, and enjoy the rest of your day.

APPENDIX C: QUESTIONNAIRE CONSENT FORM

Research Information and Consent Form

Data quality challenges in electronic healthcare systems in South Africa

Introduction

This is to get consent for your participation in the research conducted by **Marna Botha** (Email: mbotha@csir.co.za, Telephone: 0828383526). The research is for a M.Sc in Computing that I am currently undertaking with the University of South Africa (UNISA), under the supervision of **Dr. Adele Botha** (Email: abotha@csir.co.za, Telephone: 012 841 3265) and **Prof Marlien Herselman** (Email: mherselman@csir.co.za, Telephone: 012 841 3081)

Purpose of research

The study is aimed at gathering your views and perceptions regarding the data quality challenges that you experience. The purpose of the study is to compile a prioritised or ranked list of data quality challenges as experienced by data users. *Data quality challenges refer to the challenges or difficulties you experience with regards to the quality of data in the electronic health system you make use of.*

Procedure

The questionnaire will require approximately 15 minutes of your time. It is divided into 2 sections (Sections A and B), both aimed at gathering specific information about the data quality challenges that you experience. You will be required to complete sections A and B of the questionnaire.

Confidentiality

The input you provide will be treated confidentially and only used towards the completion of the afore-mentioned qualification. All data will be used in summary form without reference to any individual.

Participation

Participation in this research study is voluntary, and you have the right to, at any time, withdraw or refuse to participate. The purpose of the study will be explained in the questionnaire.

Benefits and compensation

There aren't any direct benefits for your participation. All findings will be used for the completion of the academic qualification mentioned. No compensation will be provided to anyone partaking in this research.

Risks and discomforts

There are no risks or discomfort associated with your participation. All answers from you and other participants will be analysed collectively. Individual answers will therefore not be linked to any names, positions and companies of participants.

Participant consent

I have read and understood all the above. I willingly choose to participate in this study.
Full name (optional) _____
Date: _____
Signature: _____

APPENDIX D: QUESTIONNAIRE

QUESTIONNAIRE: BACKGROUND

My name is Marna Botha, and I am conducting this research for my M.Sc in Computing at the University of South Africa (UNISA). It is aimed at gathering your input regarding the data quality challenges that you experience. In order to collect representative data, I would like you to please complete this questionnaire that should take you approximately 15 minutes.

QUESTIONNAIRE: INSTRUCTION

The questionnaire is divided into the following three sections:

	WHAT IS COVERED
SECTION A	Data regarding the participant's relationship in terms of data and data quality within the e-health domain.
SECTION B	Questionnaire prompting you to rate different data quality challenges

- Please go through the sections and where relevant:
 - Answer the question in the box provided
 - Use the rating system provided in the section to indicate your preference in the box provided
- The input you provide will be treated confidentially and only used towards the completion of the aforementioned qualification.
- Please read the consent form, and once you understand everything that is described and you agree, please sign it and return it back to the researcher along with this questionnaire.

Your co-operation is highly appreciated.

SECTION A: USER INFORMATION (DEMOGRAPHIC INFORMATION)

- Please indicate your job title with an "X":

Administrative personnel	<input type="checkbox"/>	Medical physician	<input type="checkbox"/>	IT administrator	<input type="checkbox"/>
--------------------------	--------------------------	-------------------	--------------------------	------------------	--------------------------

- Please indicate the years of experience you have in the medical domain with an "X":

5 years or less	<input type="checkbox"/>	Between 5 and 10 years	<input type="checkbox"/>	More than 10 years	<input type="checkbox"/>
-----------------	--------------------------	------------------------	--------------------------	--------------------	--------------------------

SECTION B: RATING OF DATA QUALITY CHALLENGES

This section lists data quality challenges identified by the researcher in a literature study. Please read the following challenges or difficulties that can influence the quality of data in your electronic healthcare system. Indicate with an "X" to which extent you experience the following data quality challenges:

DATA QUALITY CHALLENGE	THIS IS NOT A CHALLENGE THAT I EXPERIENCE	I EXPERIENCE THIS CHALLENGE, BUT I REGARD IT AS A SMALL CHALLENGE	THIS IS A CHALLENGE THAT I EXPERIENCE AND I REGARD IT AS HIGH

			PRIORITY
Data governance (this includes challenges such as: lack of assignment of responsibilities regarding data; in short of data administration; ambiguity of roles with regards to data tasks; a need for written data quality policies and procedures; the importance of data quality is not emphasized enough by managers; ineffective organisational procedures; lack of data quality assessments)			
There is a lack of rewards in terms of data quality			
There is a need for more training for data users			
Data entry errors (this includes errors such as: misspelled data; a certain data element is not captured; a certain data element is captured in the wrong format)			
More than one source of data, each producing different values			
Information is produced by subjectivity of the data capturer			
Transformation from paper records to electronic records			
The default value in the system was never corrected			
Inaccurate data is collected from the patient			
The length of the form is too long			
The structure of the questions on the form produces inaccurate data			
In short of appropriate software for data management			
Technology is not usable or user friendly			
Data structure problems (this includes problems such as: absent columns in the database; inter-data dependency issues; extra columns in the database; unsuitable data relationships; data in the wrong field of the form or database)			
Changing relevance of data sources			
Unforeseen changes in source systems			
Numerous sources for the same			

data			
Data values drift from their field descriptions and business rules			
Inconsistent definitions and formats because of different systems used (standardisation)			
Interoperability issues and standardisation (Interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged.)			
Lack of sufficient computer and other resources.			
Privacy, confidentiality and legal challenges			
Information is needed for required fields (you don't have all the information required by the system)			
Managing big amounts of data			
Data is out of date			
Omissions/errors of data			
Inconsistent data within a single file			
Occurrence of outliers			
Data not meeting all relevant needs			
Information is not suitable for analysis purposes			
Data is hard to access			
Financial barrier to purchase necessary hardware and cost challenges in general			
Implementation of health information systems in a small and rural facility			
Sustainability of using health information systems			
Physicians are hesitant to change existing processes			
It is difficult to form electronic health records as part of your daily routine			
Logistics with regards to electronic health information systems			
The shortage or absence of the			

necessary infrastructure, such as internet connections			
Nurses' notes may go unread by physicians			

Thank you very much for your co-operation. Your input is greatly appreciated.

APPENDIX E: ETHICAL CLEARANCE



Dear Ms M Botha (55768458)

Application number:
019/MRB/2015

REQUEST FOR ETHICAL CLEARANCE: (Prioritizing Data Quality Challenges in public clinics in the City of Tshwane)

The College of Science, Engineering and Technology's (CSET) Research and Ethics Committee has considered the relevant parts of the studies relating to the abovementioned research project and research methodology and is pleased to inform you that ethical clearance is granted for your research study as set out in your proposal and application for ethical clearance.

Therefore, involved parties may also consider ethics approval as granted. However, the permission granted must not be misconstrued as constituting an instruction from the CSET Executive or the CSET CRIC that sampled interviewees (if applicable) are compelled to take part in the research project. All interviewees retain their individual right to decide whether to participate or not.

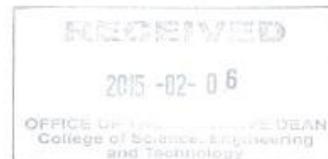
We trust that the research will be undertaken in a manner that is respectful of the rights and integrity of those who volunteer to participate, as stipulated in the UNISA Research Ethics policy. The policy can be found at the following URL:
http://cm.unisa.ac.za/contents/departments/res_policies/docs/ResearchEthicsPolicy_apprvCounc_21Sept07.pdf

Please note that the ethical clearance is granted for the duration of this project and if you subsequently do a follow-up study that requires the use of a different research instrument, you will have to submit an addendum to this application, explaining the purpose of the follow-up study and attach the new instrument along with a comprehensive information document and consent form.

Yours sincerely

Prof Ernest Mnkandla
Chair: College of Science, Engineering and Technology Ethics Sub-Committee

Prof IOG Moche
Executive Dean, College of Science, Engineering and Technology



University of South Africa
College of Science, Engineering and Technology
The Science Campus
C/o Christiaan de Wet Road and Pioneer Avenue,
Florida Park, Roodepoort
Private Bag X6, Florida, 1710
www.unisa.ac.za/cset

