A QUALITY ASSURANCE FRAMEWORK FOR DIGITAL HOUSEHOLD SURVEY PROCESSES IN SOUTH AFRICA

MOGAMAT MAHIER HATTAS

A thesis submitted in fulfilment of the requirements for the Doctoral degree in

INFORMATION SYSTEMS

in the

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES OF THE UNIVERSITY OF THE WESTERN CAPE

SUPERVISOR: DR JOHAN BREYTENBACH

December 2019

Student Number: 2103745

DECLARATION

I, Mogamat Mahier Hattas, declare that this thesis entitled: "A QUALITY ASSURANCE FRAMEWORK FOR DIGITAL HOUSEHOLD SURVEY PROCESSES IN SOUTH AFRICA" is my own work, and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

Mogamat Mahier Hattas

……………………………………

SIGNATURE

……………………………………

DATE

ACKNOWLEDGEMENTS

All praise is due to Almighty GOD, ALLAH, the LORD of the worlds, who has provided me with the ability, strength, patience and guidance needed to complete this study.

I wish, further, to express my utmost gratitude to the following pesons:

Dr Johan Breytenbach, my supervisor, for your ever-present ability to make time for me, as well as all your other students and peers. This includes your endearing wisdom for encouraging me to register for this research topic, trusting and believing in me, as well as your selfless sacrifices and astute guidance throughout this study; to Prof. Theunis Kotze of the Division for Postgraduate Studies (DPGS), for his guidance and support; Prof. Olushola Rotimi Adeniyi, Dr Stephen Kyakulumbye, DPGS Statistical Coach; the former Statistician-General, Dr Pali Lehohla from Statistics South Africa, who ignited a fire within me to follow my passion approximately 9 years ago, i.e. the need to bring the realms of digital innovation within the operational confines of the public sector, in order to add value to the broader public and similar organisations or institutions at large.

I also need to thank the current South African Statistician-General, Mr Risenga Maluleke, who approved the right to analyse the required data, and for his support, be it budgetary, academic, technical or strategic; to the Western Cape Provincial Chief Director, Mr Marius Cronje, for continuously providing me with a platform to use research to explore alternative methods, processes and techniques – with the aim of enhancing business or staff processes.

Furthermore, I would like to thank the staff at Statistics South Africa, based in the various national, provincial and district offices, who participated in this study, either as respondents, or to permit access to the required information; the Corporate Data Processing Chief Director, Mr Luqmaan Omar, who invited me to participate nationally, as a technical professional on the Computer Assisted Personal Interviewing (CAPI) team, ensuring the largest digital implementation roll-out of official household surveys to date; my CAPI colleagues for their technical support; and the Human Capacity Development Division, which sponsored the study; and to Dr PJS Goldstone as editor of this thesis.

Last, but not least, I would like to thank my supportive and beautiful wife (Mrs Wadeedah Hattas); my late father (Yusuf Hattas), my mother (Ms Badroe-nesa), my children (Zafirah, Ubayd, Mikaeel, Ebrahim) and all other family and friends, for your enduring love and encouragement to complete my studies.

PUBLICATIONS

Refereed/Peer-reviewed conference:

Hattas, M. & Breytenbach, J. (2017). Improving the quality of official survey data collection in South Africa through digital data collection. In the proceedings of the XXVIII International Population Conference. Cape Town. Retrieved from: https://iussp.confex.com/iussp/ipc2017/meetingapp.cgi/Paper/4002

ABSTRACT

Official household-based survey statistics is predominantly collected using the paper-and-pen data collection (PAPDC) method. In recent times, the world has seen a global rise in the use of digital technology, especially the use of mobile handheld devices for the collection of survey data in various fields of statistical collection. Various sectors in the population require data for a multitude of purposes, from planning, monitoring and during the evaluation of projects and programmes. The pressure of attaining the data often requires data or information producers to gather more data or information more frequently with improved quality, efficiency, and accuracy.

The quality of data or information collected remains uncertain as more surveys enter the global arena. The overall survey quality needs to improve continuously. The data used may not be trustworthy and users should be aware of this. There should be a continuous holistic assessment of the validity and reliability of data before these are used (T. Chen, Raeside, & Khan, 2014). Digital data collection (DDC) offers national statistical organisations (NSOs) in Africa possible, albeit partial, solutions to several current quality, performance, and cost-efficiency concerns. Potential benefits found in the literature for DDC methods over PAPDC methods include, inter alia: increased speed of data collection, increased data accuracy, timeous availability of data, higher data quality, effective data security and lower costs for data-collection processes.

Most NSOs in Africa, including South Africa, currently rely on manual, paper-based data collection methods for continuous official household survey collection. Paper-based methods tend to be slower, to rely on manual reporting and involve more survey-intensive resources. With the rise of handheld mobile Global Positioning Systems (GPS) enabled devices, official household surveys are able to monitor surveys spatially, and in real-time. The information could be securely synchronised to a central secure database, to allow for immediate post-processing and data analysis.

The feasibility of DDC for official household survey statistics depends on meeting the requisite quality standards required for official statistics. The quality criteria used for PAPDC was not designed for a DDC approach. The South African Statistical Quality Assessment Framework (SASQAF) for PAPDC was assessed to determine the compatibility with a DDC process. The primary focus of the study identified gaps and improvements in the quality framework that need to be considered for DDC through the proposed development of an updated Quality Assurance framework.

This study applied the methodological approach for continuous problem solving and improvement, known as the Shewhart Cycle for Learning and Improvement, based on statistical methods from the viewpoint of quality control. The Plan-Do-Check-Act (PDCA) cycle was developed by Shewart and modified by Deming as the Plan-Do-Study-Act (PDSA) cycle (Shewhart and Deming, 1986). Processes and products through quality productivity management using the PDSA cycle are continuously improved (Weinstein & Vasovski, 2004).

The PDSA was an appropriate approach because it lends itself to iteratively seeking to address quality processes, products or service improvements in organisations, such as NSOs. Stage 1 of this study consisted of a review of the literature and deliberation with organisational experts in the fields of official data collection from across the domains of statistical methodology, geography, information technologists and senior management. The review in the first stage resulted in the design of a conceptual framework. The framework integrated the qualitative and quantitative data over three iterations in Stage 2. During 2014/2015 and 2016/2017, three iterative cycles were conducted using mixed-methods research. The sample size was incrementally increased and lessons learned from previous iterations were incorporated as inputs towards the latter iterations.

To iteratively develop and improve on the design aspects of the proposed DDC quality framework; semi-structured interviews, focus-group results, data collected from a survey questionnaire, and the lessons learned after each iteration, were combined and analysed after every iteration. Stage 3 consisted of a non-parametric statistical analysis test (Kruskal-Wallis) to identify any significant relationships across the three iterations, given the common, key, quality-related measurements across each iteration. As a result, this research proposed a DDC developmental quality assurance (QA) framework for NSOs seeking to collect official household surveys when using digital handheld devices for the data collection.

Keywords: Digital data collection, Official household surveys, Mobile handheld devices, Quality Assurance framework, Paper-and-pen-data collection, Digital devices

Preface (Note on writing style)

In this dissertation, when I refer to "the researcher", it relates to the "author" of this study.

I used "study" and "this research" interchangeably and used the following words interchangeably: "data collector", "enumerator", "interviewer", "survey officer" or "fieldworker".

Writing of numbers as numerals versus words

- In some cases, certain numbers less than ten are written in words, example 8 and 10 are written as "eight" and "ten" respectively.
- All numbers above ten are written in the numerical form, for example, 12 and 30 are written as "12" and "30" or as-is.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS/ABBREVIATIONS

CAPI	Computer Assisted Personal Interviewing
CS	Community Survey 2016
DDC	Digital Data Collection
DU	Dwelling Unit (See Definition: DU)
EA	Enumeration Area (See Definition: EA)
HO	Statistics South Africa Head Office, Pretoria, South Africa
NSO	National Statistics Organisation
NSS	National Statistics System
PAPDC	Paper and Pen Data Collection
QA	Quality Assurance
QD	Quality Dimension
QI	Quality Indicator
SASQAF	South African Statistical Quality Assessment Framework
SASQAF LITE	South African Statistical Quality Assessment Framework - Lite version
SDDC	Secure Digital Data Collection
SO	Survey Officer (Collects household survey data)
SRQ	Sub-Research Question
Stats SA	Statistics South Africa
SVC	Statistical Value Chain
WC	Western Cape Province (South Africa)

DEFINITIONS AND GLOSSARY:

Dwelling unit (DU): “It is any structure or part of a structure or group of structures occupied or meant to be occupied by one or more than one household. It can also be described as any structure, or part of a structure, which is vacant and/or under construction, but can be lived in. It is also important to note that the DU, as defined here, is the major listing unit and the selection unit for the sample at Statistics South Africa” (Statistics South Africa, 2010a).

Enumerator(s): Are “fieldworker(s)” or “survey officer(s)” or “data collector(s)” collecting official household survey data.

Enumeration Area (EA): “denotes the smallest geographical unit (piece of land) into which the country is divided for census and survey purposes. For survey purposes, EA’s can be combined to form Primary Sampling Units (PSUs). Each EA has been allocated a unique eight-digit number, for the purpose of record-keeping. Also referred to as the EAQN in this study” (Statistics South Africa, 2010a).

Household: “Is a group of persons who live together and provide themselves jointly with food and/or other essentials for living, or a single person who lives alone” (Statistics South Africa, 2010a).

National statistics: “Statistics produced by an organ of state that is within the public domain, but has not been designated as official” (Statistics South Africa, 2011).

Non-Response: A form of non-observation present in most surveys. Non-response means failure to obtain a measurement on one or more study variables for one or more elements selected for the survey. In Statistics South Africa, the non-response rate has been defined as the proportion of eligible households for which a questionnaire could not be completed, in contrast to the total number of eligible households.

There are two types of non-response (Sarndal C.E., Swensson B., 2002):

> “Firstly, a sampled unit that is contacted may fail to respond. This represents "**unit non-response**". Secondly, the unit may respond to the questionnaire incompletely. This is referred to as "**item non-response".** Missing data occur primarily because of item non-responses i.e. no information was provided for one or more data items. Missing values can reduce the representativeness of the sample; and they can falsify any inference about the population”

Official statistics: “In South Africa, the head of department (also referred to as the "Statistician-General (SG) of Statistics South Africa") declares statistics produced by any organ of the state as "official statistics" in terms of the Statistics Act 1999 of South Africa (Section 14 (7))" (Statistics Act, 1999).

Out-of-scope rate: The out-of-scope rate is defined as the proportion of dwelling units in which no eligible household was found in contrast to the total number of sampled dwelling units. There are several reasons why dwelling units may not contain eligible households. At the time of enumeration, the dwelling unit could have been vacant or unoccupied; the dwelling unit could have been demolished or converted into a shop; or the structure could have been erroneously classified as a dwelling unit on the frame.

Let d_g be the total number of dwelling units, sampled from the geographic area g and $d_g^{(os)}$ the corresponding number of dwelling units with no eligible household. The out-of-scope rate is then given by:

$$Out\ of\ Scope\ Rate_g = \frac{d_g^{(os)}}{d_g} \times 100$$

Response rate: The response rate is defined as the proportion of eligible households, which completed a questionnaire with usable information to the total number of eligible households.

Let n_g be the number of eligible households in the dwelling sample from the geographic area; and n_g^r the corresponding number of respondent households, where eligible households include both respondent and non-respondent households, but exclude out-of-scope households. The response rate was calculated as:

$$Response\ Rate_g = \frac{n_g^r}{n_g} \times 100$$

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Information is defined as the communication, reception of knowledge or intelligence (Merriam-Webster Dictionary, 2018). The standard definition of data includes information. The definition of information is "data that is (1) accurate and timeously, (2) specific and organised for a purpose, (3) presented within a context that gives it meaning and relevance, and (4) can lead to an increase in understanding and a decrease in uncertainty" (BusinessDictionary, 2017). It is also valuable because it can affect behaviour, a decision, or an outcome. Information provides insight into opportunities for communities and nations to prosper and grow. Investing in acquiring high-quality information for developmental purposes is fully justified by balancing the essential resource requirements to attain good, reliable information (Üstun, Chatterji, Mechbal, & Murray, 2005).

A statistical information system is an information system oriented towards the collection, storage, transformation and distribution of statistical information (OECD, 2001). Statistical information is essential for measuring a country's progress in reaching its respective local and global developmental goals. Many developing countries have weak statistical systems and mechanisms for measuring results and providing essential information about the effectiveness of policies and programmes (EUROSTAT, 2009). High quality data assists governments to improve their policies, to be transparent, and accountable for the delivery of developmental results. As a data user, governments can benefit by applying advanced analytics, such as predictive analytics, machine learning, deep learning algorithms, etc. This could improve internal decision-making, promote the creation of new offerings and increase the accountability of the results (Baig, Cadena, Chui, Dua, Farrell, Law, Mendonca, Remes, Suder, & Woetzel, 2014).

Reliable statistics is a key element towards better measurement, monitoring, and management of results for development (EUROSTAT, 2009). Population statistics are critical in setting and monitoring government policy. Significant levels of public funding are allocated to targeted populations; and the correct allocation of these funds often depends on accurate population statistics (Frank, 2014). With the growing demand for accurate and timely socio-economic information, decision-makers at all levels of any organisation, including government departments and business enterprises, have become increasingly dependent on statistics as fundamental data in the decision-making process.

Thus, the quality of statistics used is essential when using them as information for decision-making purposes. The National Statistical Organisations (NSOs) are primarily responsible for the collection of "official statistics" (see Definitions and Glossary) in any given country. Over time, NSO institutions have become increasingly more aware of the importance of good quality statistics (United Nations Economic Comission for Europe, 2012) . Digital Data Collection (DDC) offers NSOs possible, albeit partial, solutions to several current performance and profitability concerns, when conducting household-based surveys and censuses. The benefits of DDC methods versus Paper-And-Pen Data Collection (PAPDC) methods included: (1) the increasing speed of collecting data, (2) increased data accuracy, (3) the timeous availability of data, (4) higher data quality, (5) improved data security, and (6) lower costs of data collection (M. P. Couper, 2005; Hattas, 2013; Pace & Staton, 2005; Rai et al., 2017; Zimmermann, 2008).

Furthermore, attaining high-quality data more frequently at a cheaper cost directly increases the probability of improving planning, policies and programme outcomes for national, country or local governmental departments or agencies, reliant on NSOs for these data or this information. Most NSOs in Africa especially sub-Saharan Africa, however, still rely heavily on manual, paper-based data-collection methods for collecting household-based surveys (Ahmed et al., 2018; Teganga, Emukule, Wambugu, Kabore, & Mwarogo, 2009). In Africa, the processes of alignment of the “organisational structure” of NSOs to interface with newer technologies is often found to be slower than in developed countries, primarily because of the political nature of the public sector and related departments that house NSOs. Processes are not improved by simply adding or applying the latest technological enhancements, but by improving the related subjective, contextual decision-making processes within these organisations (Tokosi, 2016).

Paper-based data collection methods often involve tedious and expensive data collection processes. These include the printing of paper questionnaires, as well as the forward and reverse logistics of paper and boxes amongst far outlying regions of a country to its respective headquarters. It includes the capturing of data manually or scanning data from the printed questionnaires. Advancing or adapting towards a full-scale DDC process may prove challenging, especially given the varying spatial distribution of connectivity across rural, remote and urban-dense landscapes. In the health sector, for example, the cost of technological infrastructure at a rural area level were still very high. The internet and network connections were still slow when compared to their urban counterparts (Adesina, Agbele, Februarie, Abidoye, & Nyongesa, 2011; Bukachi & Pakenham-Walsh, 2007). Besides, the availability of a computer, maintenance, and network experts remains largely inadequate in rural or remoter regions in Africa and South Africa (Adesina et al., 2011).

1.2 DISCUSSING THE CONCEPTS

It is important to discuss some of the common concepts used during this study in order to avoid possible misinterpretation and confusion of terms such as concepts used to express ideas, which may be in contradiction of their everyday use. PAPDC refers to the process of collecting official household based surveys by using a printed questionnaire administered to a household respondent by an interviewer, survey officer or fieldworker. PAPDC is also referred to as a “Pencil-and-Paper Interview” (PAPI), which is one of the techniques used for collecting data in social research. The interviewer proceeds by asking question after question, according to the questionnaire template; and a respondent answers these questions. The interviewer records the answers to the questionnaire.

As with Lauren (2003), this research draws some parallels with the conversion from PAPI to Computer-Assisted Personal Interviewing (CAPI); and it examines the implications of the move to CAPI concerning data quality. This research does not aim to compare PAPI with CAPI, however it seeks to identify the quality and process improvements in iterative cycles of CAPI, in order to maintain or achieve higher or improved levels of overall survey performance. DDC and CAPI is similar to each other during this study, given that both utilises a handheld tablet or electronic device to capture a respondent’s information or responses to survey questions.

A personal-digital-assistant (PDA) based survey interview is similar to a DDC survey. Fieldworkers visit selected households using a Global Positioning System (GPS) tool to conduct a household survey interview (Eng et al., 2007). Traditionally when comparing the quality of data collected using a DDC process versus a PAPDC, survey researches tended to compare traditional PAPDC using various digital device options (Antoun, Couper, & Conrad, 2017). In general, these studies have generally compared CAPI with PAPI (D. L. Wright, Aquilino, & Supple, 1998). The research reported here intends to compare the responses from handheld mobile devices over time through three iterations, as a way of evaluating the impact of using mobile handheld devices on the quality of the household surveys conducted.

1.3 PURPOSE AND MOTIVATION FOR THE STUDY

Countries depend on their respective leaders, high-level authorities, traditional monarchies, governments, or other ruling authorities to act responsibly when developing effective policies for their institutions. When measuring the global and local impact of the ruling constituencies in respective countries; timely, accurate and reliable statistics are required. It is increasingly important to have high quality statistics on the population and projections of the population for policy development, planning and providing public services (Kelegama, 2016). Good quality data promotes growth within a country. The quality of statistical data often facilitates the respective democracies of countries positively if the data produced from these institutions are of “good quality”; and adversely if the opposite is true. Many NSOs, including Statistics South Africa (Stats SA) and Statistics Australia, define quality as “fitness for use”, which is primarily dependent on data uses and users (Australian Bureau of Statistics, 2015; Pistorius, 2010; Statistics South Africa, 2008).

Whether the product, i.e. data and statistical information, are ‘fit for use’, depends on the intended use and the characteristics of the data, or the statistical information. Quality includes the conformity to specifications, it indicates how well a product or service meets the design specifications. The "fitness for use" evaluates the performance of a product against its intended use. The value for money, which outlines the utility level of a product or service against cost; and support services, refer to quality in terms of after-sales service (Reid & Sanders, 2013). The primary motivation for this study was to add value to human development, growth, and progress by advocating the importance of quality collection of data using DDC. Furthermore, the ability to provide regular and reliable (quality) data to a country’s economy, and the well-being of its population, is an indicator of the existence of good policies and institutions (EUROSTAT, 2009).

1.4 PROBLEM STATEMENT

It is vital that official statistics produced by NSOs are used for global monitoring of government programmes and planning purposes. Furthermore, both public and private assistance to NSOs to produce these key development statistics and indicators should be prevalent. To ensure people have access to basic services, such as water, electricity, health, etc., local governments require good quality statistics measuring the progress made in this area. Good quality data, concerning infrastructure, like roads, rail, etc. also enable local businesses to interact more competitively in economic activities (Kniivilä, 2007). Developing countries continue to require assistance in improving their statistical systems to measure their development progress. The statistical systems assist government departments in monitoring programmes and help guide policy (UN system task team, 2013).

African NSOs rely heavily on paper-based data collection methods for household survey data collection. Recent technology based innovations such as the internet, mobile, and geographical coding technologies are rapidly changing how national and international statistical services collect, process and disseminate statistics. The use of handheld computers to phase out paper questionnaires altogether is very appealing to most NSOs, since the potential exists for improving the quality of data. For example, by using DDC skipping instructions and enforcing them consistently and correctly throughout the questionnaire, brings about immediate quality benefits in the data, as well as a reduction in data cleaning and editing post fieldwork collection (Lauren, 2003).

Developing countries have not fully adopted DDC as an approach to collect official surveys, even though the technology has been around for more than two decades. Even in the most advanced countries', paperless questioning or DDC has so far been restricted to relatively simple exercises, such as employment surveys and the collection of prices for the consumer price index (Munoz, 2003). Working within the limitations of financing, human-resource skills and legal responsibilities for data quality and confidentiality, statistical services are responding to these opportunities to work more efficiently and productively. It is important to note that mobile handheld technology has evolved rapidly during the last decade. New technologies emerge daily, which could potentially improve the processing time and reduce the cost of household based survey data (Hattas & Eloff, 2011). In the context of poor research infrastructure and the increasing demand for large-scale surveys, the affordability and availability of mobile phones and wireless networks makes DDC a viable alternative to traditional PAPDC methods (Tomlinson, Solomon, Singh, Doherty, Chopra, Ijumba, Alexander, & Jackson, 2009).

To keep abreast of these ever-changing technologies, current methodologies, processes, systems, and quality standards, such as the "South African Statistical Quality Assessment Framework (SASQAF)" (Statistics South Africa, 2010b) should be revised for DDC. NSOs should seek potential gains in social efficiencies, increasing levels of quality, increasing productivity, as well as enhancing the economic potential of a country's developmental objectives.

1.5 AIM AND OBJECTIVES OF THE STUDY

This research aims to identify the potential benefits of using handheld DDC for official household survey data collection by comparing it against itself over numerous iterations. As stated previously, this research does not aim to compare PAPI with CAPI, however it seeks to identify the quality and process improvements in iterative cycles of CAPI, in order to maintain or achieve higher or improved levels of overall survey performance. The process gaps and improvements required when collecting data using DDC as opposed to PAPDC are discussed. The advantages of using handheld devices or PDAs are numerous, for both respondents and researchers (Graham, Smith, Caldwell, & Mathews, 2008; Lauren, 2003). Respondents find PDAs to be user-friendly, comfortable, and enjoyable (Seebregts, Zwarenstein, Mathews, Fairall, Flisher,Seebregts, Mukoma, & Klepp, 2009). When asked directly which format they prefer, there is evidence that respondents prefer PDAs to paper interviewing (Seebregts et al., 2009).

As an added benefit to respondents, PDA-based assessments or surveys require less time to complete than paper assessments (Forster et al., 1992; Yu, de Courten, Pan, Galea, & Pryor, 2009). Various African NSOs and partners within their respective National Statistics System's (NSSs) have shown a tremendous amount of interest in the use of mobile handheld technology, given its potential to improve the frequency and the quality of data collected. If DDC contain a significant benefit for NSO's internal data collection processes, the Statistical Value Chain (SVC) and quality frameworks, such as SASQAF should be revisited. A re-engineering of the PAPDC, SVC and SASQAF processes may be required, to capitalise fully on these benefits. This study seeks to capacitate NSOs' data collection practitioners and management on how to improve the quality of their household survey data collection processes.

Deriving from the above problem statement, the **primary objectives** of the study were:

- to identify the shortfalls of SASQAF (if any) related to DDC processes, and to make suggestions for the enhancement of the existing PAPDC framework.
- to investigate the current "best practices" applied in official household survey collection within the global and the South African context.
- to design and describe a quality framework for adopting an effective digitised household-survey process in South Africa.

1.6 RESEARCH QUESTIONS AND SUB-QUESTIONS

This study sought to answer the research question(s) (RQ(s)) by analysing the quality of the household survey data collected over three iterations of a case study. This was an iterative design process; since each iteration formed part of the same case study. After each iteration, improvements or gaps in data quality was identified. Before Iteration 1, collecting the data for household surveys was almost exclusively paper-based. The process followed was to look objectively for improvements or gaps in the quality of data after each iteration of DDC and propose a DDC developmental Quality Assurance (QA) framework.

Indications of process improvements were determined from the data collected during the iterations. In terms of process improvements; these were identified, adjusted or included in the newly developed QA framework for DDC. In recent times, NSOs found themselves in positions where available processes and technologies are constantly changing. The SASQAF and SVC does not cater for the ever-changing processes and technologies that are required for DDC. It is within this context, the research question of this study was as follows:

How can the South African Statistical Quality Assurance Framework (SASQAF) be enhanced to support digitised household survey data collection in South Africa?

The main research question is subdivided into three sub-research questions (SRQs):

SRQ 1: What are the shortfalls of the SASQAF related to digital data collection?

SRQ 2: Which components of the household survey process require changes for ensuring the creation of an effective developmental QA framework for DDC?

SRQ 3: What would an effective framework entail to ensure high quality DDC?

1.7 THE RESEARCH METHODOLOGY

Research in Information Systems (IS) has had a long positivist tradition (Cardoso & Ramos, 2012). IS assumes a single reality and truth and is often presented objectively and can be described by measurable properties that are independent of the observer and his instruments (Cardoso & Ramos, 2012). This research is be defined as a: “positivist”; “longitudinal”, “single-case study” conducted over three iterations (see Figure 1.2). The NSO, Stats SA is the case in this study. Stats SA’s organogram (see Figure 1.1) displays the various divisions within the organisation (Statistics South Africa, 2016f). The research philosophy, in this case, is ‘positivism’, as is indicated by (Saunders, Lewis, & Thornhill, 2008), which contains important assumptions about the way the researcher views the world, and thereby inherently utilised the appropriate research strategy, methods or (choices) over the appropriate time horizons.

The research design utilised both qualitative and quantitative research methods, or the mixed-method approach, which is an accepted practice within IS (Johnson & Onwuegbuzie, 2004)(Creswell, 2003; Johnson & Onwuegbuzie, 2004; Wisdom & Creswell, 2013; Zachariadis, Scott, & Barrett, 2013). Data collection consisted of both primary and secondary data sources; some of the latter included government publications, technical documents, organisational reports, etc. Data generation comprised of the triangulation method (Oates, 2006); and it incorporated the observations and documents over a period of time. Secondary data covered different sources; and it provided the essential preparation for the selected interviews, as well as supporting the exploration of particular responses during these interviews.

The primary data techniques used in this research were semi-structured interviews or surveys, participant observation, and group discussions through debriefing reports. This case study method necessitated the collection of ‘rich’ qualitative information from a variety of sources, to address the complexity of the organisational processes within the context in which it is studied. The case study research approach is widely used in several different ways within the IS community. Positivist, deductive case studies are an important research approach used within IS. They provide a sound and systematic approach for conducting research and they may be used in conjunction with other approaches, to provide richer and more reliable research results (Shanks, 2002).

"Positivism" also known as "logical" positivism, is based on the view that science is the only foundation for true knowledge. It implies that the methods, techniques, and procedures used in the natural sciences offer the best framework for investigating the social world. Positivist research is value-free; and it is based on precise observations and verifiable measurements (Kawulich, 2015). In IS, the use of both quantitative and qualitative data in a mixed-method selected choice is an accepted practice. It is consistent with researchers who suggest that a peaceful co-existence of multiple paradigms is feasible in a research inquiry (Venkatesh, Brown, & Bala, 2013).

Furthermore, mixed methods within case studies conform to positivist research in IS. The discussions in Chapter 3 focused on the research methodology used for this study. The Plan-Do-Study-Act model improved the outputs and outcomes amongst the three iterations iteratively.

1.7.1 Qualitative data

Using an iterative process, the researcher addressed key functional areas concerning the quality of the survey collection processes. To derive a set of guidelines, methods, and processes to address the potential or actual evolution from PAPDC to DDC. Working with key personnel over an extended period provided the researcher with an opportunity to interact with various role-players and act in various roles himself. The researcher gained invaluable experience in understanding the humanitarian, managerial and technical intricacies of DDC. Furthermore, the researcher solicited the self- and informant-perceptions of various members, who piloted the various DDC initiatives or related projects. The researcher targeted the following user population within the NSO for this study:

- Directors
- Deputy Directors
- Assistant Directors: District Survey Co-ordinators
- Geographic Information System (GIS) specialists
- Information Technology (IT) and
- Survey Officers (SOs)

A sectional themed-structured questionnaire was administered to a small sample of the user population (i.e. staff mentioned above), which provided useful information on DDC.

These themes were as follows:

A. Background information
B. Expectations of digital handheld device technology (DHDT) and experience thereof
C. Equipment access
D. Data quality
E. Best practices in official household survey-data collection
F. Pilot experiences for DDC

Furthermore, qualitative data included organisational statistical reports, statistical releases, participant's feedback, debriefing reports and test-report outcomes.

1.7.2 Quantitative data

Data collected from three DDC iterations were analysed. These were termed "output" datasets in each of the three iterations of increasing sample size and complexity. For Iteration 2 and Iteration 3 iteratively, a statistically random sample selection of records were selected to ensure a 95% Confidence Level and a 5% confidence interval or precision. In Iteration 1, 100% of the data were analysed, due to the original small sample size covering the population of interest (i.e. households within the Western Cape Province, South Africa).

The data collected contained some common household measurements; however, the households or dwellings where the data were collected across the three iterations were not the same (i.e. different samples spatially). In addition, the primary quality measurements included the cost and time elapsed across three iterations to determine the consistency of the collected data. The level of responses, non-responses, and out-of-scope was analysed to determine the response rate of the data collected.

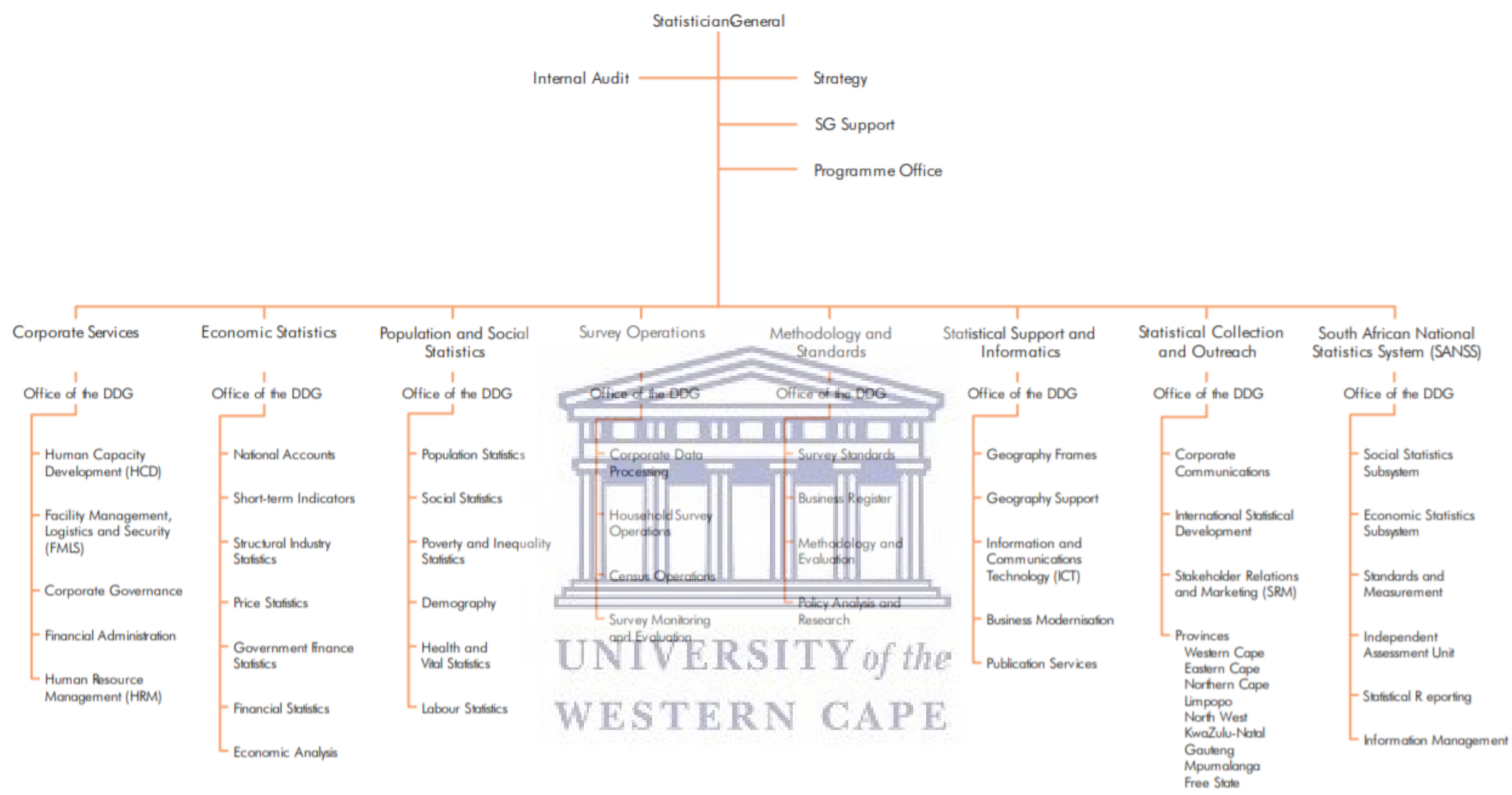


Figure 1.1: Organogram for Statistics South Africa 2016/2017

Source: (Statistics South Africa, 2017c)

1.8 IMPORTANCE AND APPLICATION OF THE STUDY

The study proposed a DDC developmental QA framework, which is required for collecting official household based surveys using a DDC mode of collection. All stakeholders using information from an NSO should be confident in the final estimates produced by any of the household survey results emanating from an NSO using a PAPDC or DDC approach. Research completed recently discussed the viability of conducting secure DDC by NSOs (Hattas, 2013).

The study makes both theoretical and practical contributions to the field of DDC. The current study contributes theoretically to the emerging literature on DDC, by investigating the quality factors when developing an updated quality framework for DDC. This study proposes a conceptual framework of key quality criteria that NSOs should consider, and it provides a view of the issues of organisational adoption towards DDC. The contributions identified throughout the study is for senior management and lower-level employees to consider when using DDC. The conceptual developmental QA framework for DDC highlights the need for NSOs to consider the factors required when developing and implementing DDC for official household surveys.

1.9 SCOPE AND LIMITATIONS OF THE STUDY

There are limitations associated with this study, which may narrow the scope of the findings and point to potential directions for future studies. The proposed developmental QA framework for DDC for NSOs needs further validation and testing. This should lead to the development of best practices, standards, and guidelines for a fully-fledged QA framework for DDC. This will support senior management in NSOs, internal staff, researchers and external stakeholders when considering using DDC. It would seem that the next step(s) in further exploring the phenomenon of quality DDC should be quantitative testing that would cover a much greater number of NSOs / similar organisations. In conclusion, this study notes that the organisational transition to DDC and the use of information systems and ICT for its support are extremely important. It will require more complex research problems that may require further studies and analysis, to capitalise fully on the DDC potential for future collections.

The primary objective of the study focused on official household survey statistics collected by an NSO, which imposed some limitations to this study's generalisation. The case study and provided insights concerning the statistical quality framework required for DDC. The results cannot be generalised to all NSOs globally (especially across developed countries), without further research being conducted. Brix's empirical evidence in his research indicated that "more team knowledge creation occurred in the "case" organisation as compared to the alternative amount of knowledge presented to the decision-makers to enable organisational learning" (Brix, 2017).

In conducting this research, the researcher envisaged the following limitations:

- The study is not a comparative analysis between PAPDC and DDC for household surveys. It is an analysis of three iterations of DDC surveys to identify the process gaps or improvements required to collect quality official household surveys. The study monitors the overall performance of data quality over the three iterations in terms of survey responses, non-responses and out of scopes.
- The availability of respondents was not always guaranteed, and there was the likelihood that some respondents might not make themselves available for interviews, and they could not be replaced at the last minute. This resulted in a reduced response rate that could potentially limit the outcome statements of the research.
- The study is limited in its ability to prove the quality concept of DDC for official household surveys within the South African context.

Key international/national frameworks and standards considered in the study included the:

- "South African Statistical Quality Assessment Framework (SASQAF)" (Statistics South Africa, 2010b)
- "South African National Standard: Information Technology – Security techniques – Code of practice for information security management" (SANS 27002, 2008)

The laws/legislation relevant to the study included the:

- Statistics Act No. 6 of 1999 (Statistics Act, 1999)
- United Nations Statistics Division, "Fundamental Principles of Official Statistics" (United Nations Statistics Division, 2007)

1.10 ETHICAL CONSIDERATIONS

This study adhered to ethical norms in IS research. Ethical guidelines specified by the University of the Western Cape were evident and applied throughout the research process. Ethical clearance was obtained prior to the study (see Appendix A). Research participants were encouraged to partake in the study through an invitational information sheet or letter which contained details about the study (see Appendices D, E, and F). Staff members (or participants) who participated in the study, were informed about their involvement in the research before the start of each iteration. Participation was voluntary, and staff were provided the option to withdraw from the study at any given time. All participants signed consent forms (see Appendix G). This indicated that staff members received sufficient information to enable them to make decisions about whether they wanted to participate as part of the research or not. The confidentiality of the data supplied to the researcher was honoured and respected.

There are several reasons why it is important to adhere to ethical norms in research. Norms promote the aims of the research, such as knowledge, truth, and the avoidance of error. For example, prohibitions against fabricating, falsifying, or misrepresenting research data promote the truth and minimise error (David B. Resnik, 2015). Ethics are norms or standards of behaviour that guide moral choices about personal behaviour and interpersonal relationships (Tokosi, 2016). The researcher foresaw few ethical problems in this study, for example, the concern for confidentiality by all the interviewees. The researcher intended to ensure compliance to all prescribed ethical norms during the research process.

Concerning "researcher bias", a researcher's personal beliefs can be reflected in the choice of methodology and the interpretation of the findings, as well as in the choice of a research topic (Mehra, 2001). The researcher's bias in this study was that digital technology provides value to the public sector; and, thus, this belief determined the study. The researcher's bias was minimised by first piloting a questionnaire, and then directly adapting some questionnaire items to form the interview questions for the main study. This minimised the influence of the researcher's views and beliefs in developing interview questions for the study.

To avert potential negative connotations, the researcher:

- ensured that internal reports sourced from the NSO maintained the required levels of confidentiality.
- notified the staff participating in the study of their right to freedom of choice; and ensured that consent was obtained from them when they participated, as respondents.
- sought permission from the necessary authorities or gatekeepers to undertake this study, and the respondents were advised that the information released would be treated confidentially.
- ensured the findings and results presented were those of the facts stated in the interviews.

1.11 CONTRIBUTION OF THE STUDY

From a practical viewpoint, this research identified the potential use of modern handheld digital technologies to collect quality official household surveys in South Africa by:

- introducing an updated quality assurance framework to cater for DDC

- indicating that household surveys should be able to fully migrate to DDC using handheld devices by:
  - identifying some of the processes involved when migrating from PAPDC to DDC and ensuring statistics remained a “conduit of trust” for all the users. This was done through the development of a proposed Quality Assurance framework for DDC.
  - reducing the overall data-collection costs significantly in the long term for an NSO, and in turn, government reduce expenditure to run existing surveys on the current platform (PAPDC), by highlighting key cost savings between PAPDC and DDC.

Training using digital handheld devices for data collection is a relatively new approach, as opposed to PAPDC training. This research identified the need for an adjustment to a more standardised practical approach to training data collectors for household-based surveys (Hattas & Eloff, 2011). According to Census 2011, the digital divide is closing very slowly in terms of internet usage. More recently, in 2016, under 50% of the total population in South Africa was considered to be active internet users (Shezi, 2016). Only 21% of South Africans households had access to a computer; while 65% of all South Africans did not have access to the internet in 2011 (Statistics South Africa, 2012).

Furthermore, in terms of mobile coverage, 89% of all South African households had mobile phones; while only 16% of South Africans could access the internet through this medium. DDC is also used in a variety of other applications and industries, besides official social household surveys (Hattas, 2013). DDC potentially provides NSOs with an opportunity to collect more surveys more frequently. Finally, an updated QA framework for DDC would assist NSOs and their partners in the National Statistics System (NSS) ensure that quality statistical products for household-based surveys are produced for all types of users.

Lastly, the study yielded a theoretical and methodological contribution by using a Plan-Do-Study-Act (PDSA) or Deming cycle (Shewhart & Deming, 1986; Weinstein & Vasovski, 2004) as a theoretical underpinning in IS, to execute a research project.

1.12 THESIS OUTLINE

1.12.1 Chapter One: Introduction

This chapter provides an overview and outline of the research. It introduces the research; and it includes the purpose of the study, its aims, objectives, the research questions, the methodology, the ethical considerations, and the contributions thereof.

1.12.2 Chapter Two: Literature review - Existing quality assurance frameworks for official household survey data collection

Understanding the literature concerning both PAPDC and DDC within national statistical organisations including South Africa is a key ingredient when unpacking related theories underpinning DDC processes. Unique challenges of DDC within SA were unpacked, ranging from organisational transitioning to DDC, security concerns and the current SASQAF that does not cater fully for DDC, etc.

1.12.3 Chapter Three: Research design and methodology

Chapter 3 discusses the research design and the methods employed to answer the research questions in this study. The various data collection instruments used, processes, standards, and the procedures are all detailed.

1.12.4 Chapter Four: Mapping of the Statistical Value Chain (SVC) and the South African Statistical Quality Assessment Framework (SASQAF)

This chapter presents the official quality framework used in South Africa, namely the SASQAF. It also highlights the Statistical Value Chain (SVC) used by the South African NSO for the collection of official household-based statistics, using PAPDC techniques (Statistics South Africa, 2010). For DDC, the requisite quality dimensions, including the indicators and standards required for the development of a usable developmental quality assessment framework for South Africa are discussed. Each iteration discussed in the following Chapters (5-7) focuses on the various thematic dimensions to consider within the scope of the proposed DDC developmental QA framework, e.g. training, user interface, technology acceptance, security, organisational (fit, processes, transition), etc.

1.12.5 Chapter Five: Iteration 1 - Western Cape Test 2015

Chapter 5 presents the results from the first iterative cycle of DDC conducted in the Western Cape, South Africa during 2015. Chapter 5 comprises quantitative and qualitative data as primary and secondary data sources respectively. The data used for analysis included: captured data, semi-structured interviews, test data, focus groups and survey questionnaires.

1.12.6 Chapter Six: Iteration 2 - KwaZulu-Natal Citizen Satisfaction Survey 2015

Chapter 6 presents the results from the second iterative cycle of DDC conducted in the KwaZulu-Natal Province, South Africa, during 2015. It includes both quantitative and qualitative data analysis, based on both primary and secondary data sources.

1.12.7 Chapter Seven: Iteration 3 - Community Survey 2016

Chapter 7 presents the results from the third iterative cycle of DDC conducted in South Africa during 2016. It includes quantitative and qualitative data analysis, based on both primary and secondary data sources.

1.12.8 Chapter Eight: A Proposed Digital Data Collection Quality Framework

This chapter discusses the findings in general and interprets the results of the previous three iteration Chapters (5-7). The Kruskal-Wallis statistical technique compared the distributions of the three iterations. Chapter 8 presents a proposed DDC developmental QA framework for NSOs.

1.12.9 Chapter Nine: Conclusion and recommendations

The final chapter concludes, summarises and finalises the study by discussing the extent to whether the research objectives were satisfied. The contributions to the research domain are indicated and possible further research suggested. Figure 1.2 provides a graphical representation of the layout of the dissertation chapters.

1.13 CHAPTER SUMMARY

A proposed DDC developmental QA framework is a systematic process that offers NSOs a platform to assess the quality of data if DDC is the selected method of collection. The SASQAF forms the basis of the developmental QA framework developed in this study. Research studies on quality assurance frameworks for DDC are largely non-evident. A survey administered to NSO staff that participated in one of the tests or pilot studies throughout this research may have had similar experiences with DDC elsewhere within the NSO. The initial adoption of the SASQAF explained the need for the development of a new quality assessment framework for DDC for official household survey collection.

Chapter 2 is a review of the literature on quality assurance frameworks and systems in place for PAPDC, including the history, the concepts, and terminologies. The first chapter laid the foundation for an understanding of the research title, its aims, objectives and the basic terminologies used throughout the study. It also positions the applied research method that provides an outline of the content to follow. Furthermore, Chapter 2 addresses the related literature on data collection within the context of household based surveys for PAPDC and DDC.

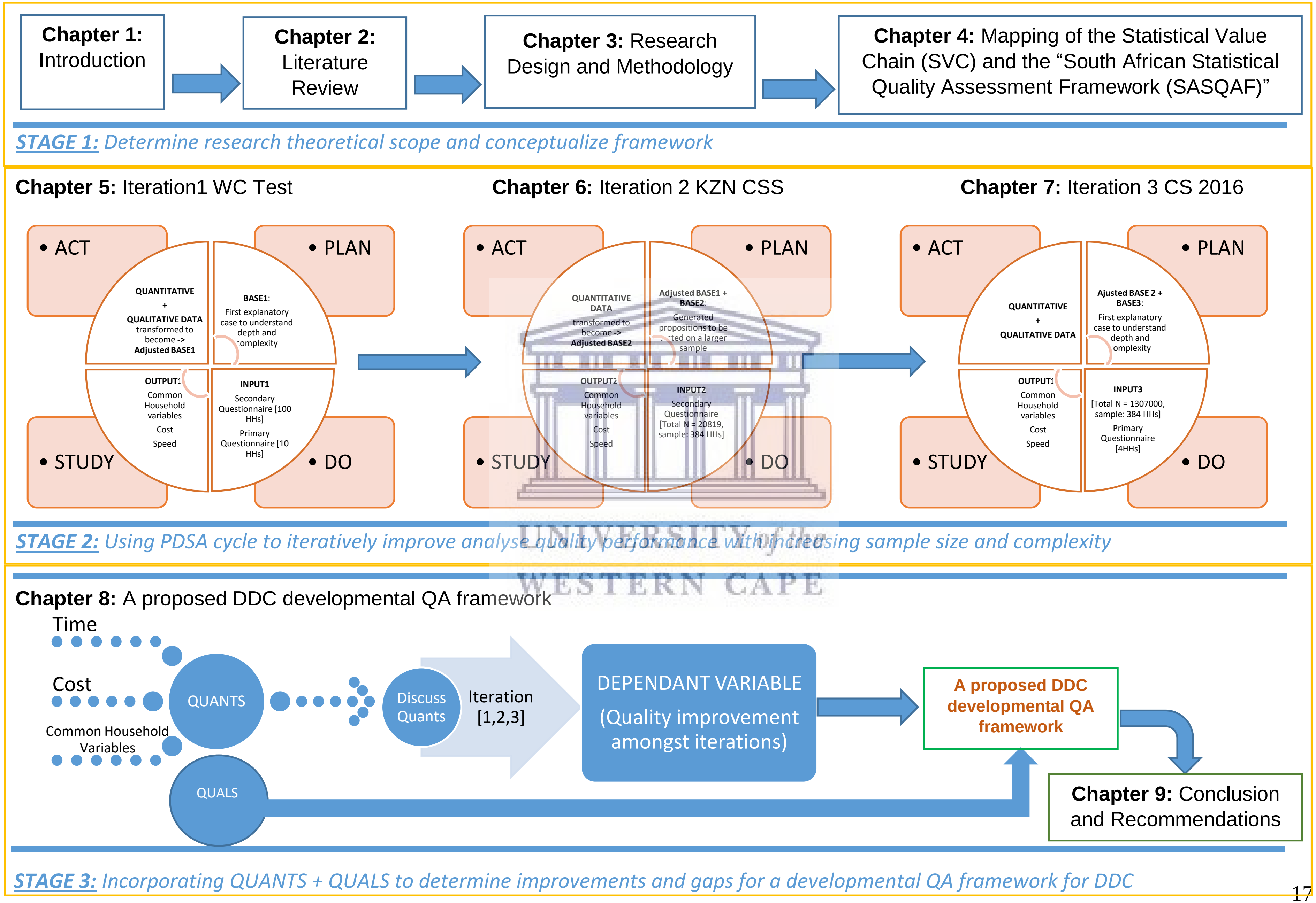


Figure 1.2: Chapter Outline - Conceptual Framework

CHAPTER 2: LITERATURE REVIEW - EXISTING QUALITY ASSURANCE FRAMEWORKS FOR OFFICIAL HOUSEHOLD SURVEY DATA COLLECTION

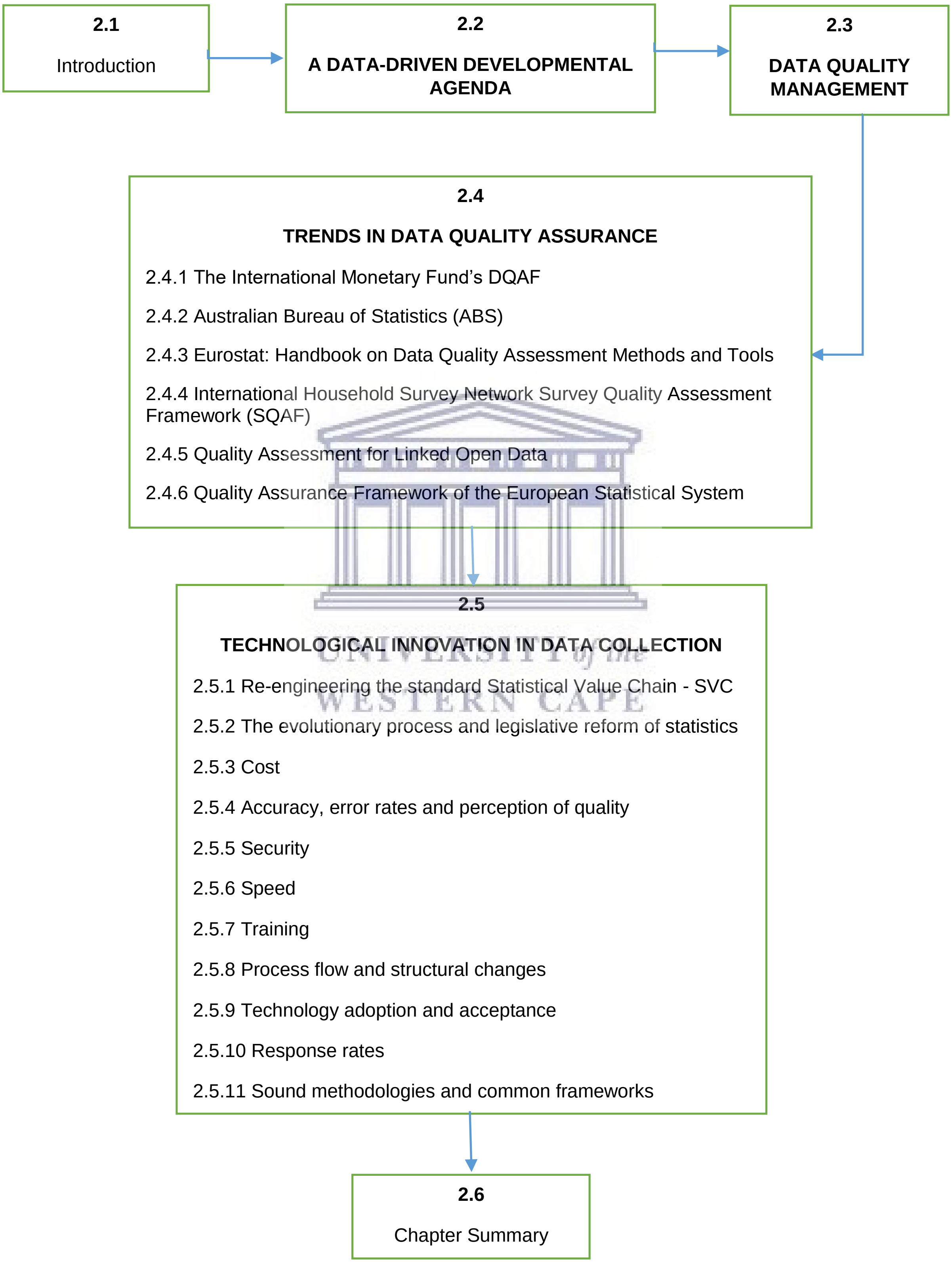


Figure 2.1: Chapter 2 Layout

CHAPTER 2: LITERATURE REVIEW - EXISTING QUALITY ASSURANCE FRAMEWORKS FOR OFFICIAL HOUSEHOLD SURVEY DATA COLLECTION

This chapter comprises of two core, related components. The first component highlights the need for good quality data to drive the developmental policy agendas of countries (Section 2.2). The corresponding, second component follows on from the previous one and the discussions spans over three sections. Firstly, the SVC for a generic quality management system of a country, as well as its corresponding quality assurance framework provides the reader with an opportunity to identify the core quality dimensions prevalent throughout this study (Section 2.3). Secondly, a global summary of other international quality assurance frameworks dealing with country specific data quality issues provides an indication of the similarities and cohesive nature of the quality dimensions (Section 2.4). Thirdly, technological innovations found in DDC and related themes used by NSOs to improve on their core mandate of collecting good quality statistics are addressed (Section 2.5).

2.1 INTRODUCTION

Changes to countries' policy agendas at all levels require revolutionary and innovative thinking around the measurement systems – in terms of data, information, and statistics required as evidence to inform their respective sustainable development agendas. Developing countries continually require assistance in improving their statistical systems to measure their development progress. This also includes the need to participate effectively in their national monitoring programmes to guide policy formulation. A target for achieving sound statistical systems is required to further support capacity development for basic statistics and the analysis of data in countries (UN system task team, 2013).

Historical public policy agendas are important for understanding the social fabric or make-up of a country. For example, the Policy Agendas Project has coded the content of the US policy process since 1900, according to a common scheme. Although there is basic agreement on the methodology and measures, scholars would like to use the data (informing policies) to solve a range of substantive problems. A natural extension was to examine the content of policies outside the US, to generate hypotheses about the nature of policy-making in the different institutional, country and cultural contexts.

A major difference for policy was the different set of states or even country activities. Reflecting the fact that less well-developed welfare states or country codes are incomprehensible to a non-US audience (John, 2016).

While goals themselves may be aspirational, numerical targets should balance ambition with realism. They should challenge preconceptions of what is possible to achieve, and inspire concerted public efforts to meet them within a reasonable timeframes. They should be results-oriented in terms of wellbeing and sustainability, and sufficiently specific to relate clearly to public and policy concerns. Indicators should be mainly “outcome-based”, to keep the focus on long-term results. Furthermore, indicators should target-orientated, be measurable over time by using the data collected in countries cost-effectively and practically. Indicators should be helpful in informing policy, clear and easy to communicate to the public and civil society alike (see Figure 2.2).

These indicators should provide a clear basis for understanding the problem statement in the development and transformation agenda of a country.

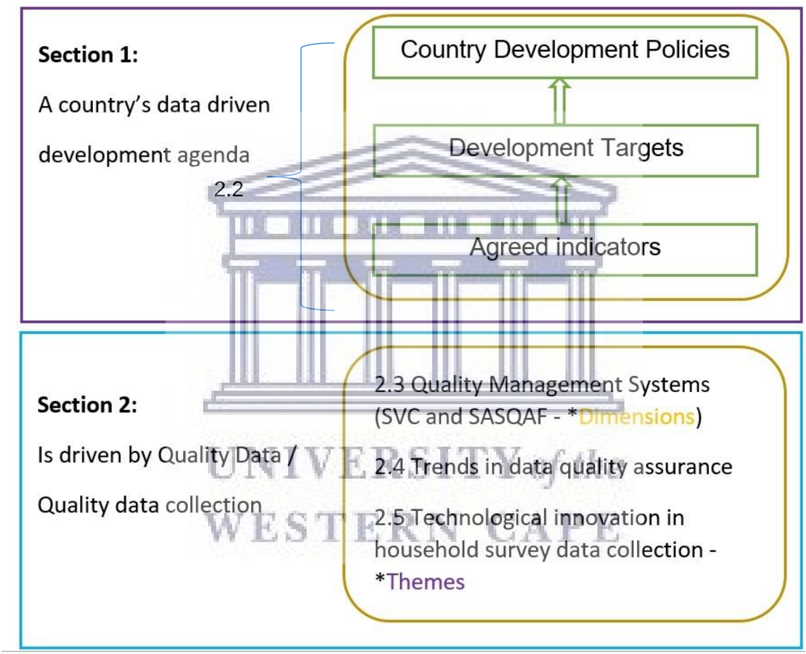


Figure 2.2: A Data Driven Developmental Agenda

2.2 A DATA DRIVEN DEVELOPMENTAL AGENDA

The framework for statistics and indicators in the post-2015 agenda should maintain data quality control mechanisms, such as those already developed in the Millennium Development Goals (MDG) indicators programme. These mechanisms include strategic, technical and policy reviews in the statistical commission and the technical responsibility for the compilation of data (UN system task team, 2013). Countries are measured against their developmental performance indicators which were agreed upon in global and / or continental agreements, such as: (1) the UN Millennium Project (United Nations, 2015); (2) the Sustainable Development Goals (SDGs) (UNDP, 2015) and (3) the African Union’s Agenda 2063 (African

Union Commission, 2015; Brand South Africa, 2015). It is important for countries to provide regular feedback on these developmental indicators, using accurate and high quality agreed upon data sources. Monitoring progress and development, as part of Agenda 2030, will require a concerted effort on the part of both countries and their international development partners to improve the availability, quality, timeliness and relevance of data required to track the SDGs (Kilic, Serajuddin, Uematsu, & Yoshida, 2017).

High quality data enables countries to measure themselves more effectively against these targets. “Official” and “national” statistics on a national and sub-national level are often used as the primary source of demographic and other key population data, as required by countries. The policy agenda at a global level contains SDGs. A new universal set of goals, targets, and indicators that the United Nations member states have committed to frame their development agendas and political policies over the next 10 to 15 years. The SDGs follow on and expand the MDGs (“UN Millennium Project | About the MDGs,” 2006) which were agreed on by governments across the globe in 2000, and which expired at the end of 2015. The post-2015 Global Development Agenda, called SDGs, outlined 17 goals, 169 targets and 231 indicators to measure progress towards sustainable development. It was aptly themed: "Leave no one behind" (UNDP, 2015).

The South African government has embarked on a process to develop an integrated indicator framework (Figure 2.3) that aligns policy agendas at all the different spatial levels:

- Global agendas (SDGs) - (UNDP, 2015)
- Continental agendas - Agenda 2063 (African Union Commission, 2015),
- Regional agendas - the Southern African Development Community (SADC) common agendas refer to a set of key principles and values that guide the regional integration agenda with member states (including South Africa) (Southern African Development Community (SADC), 2017) and
- South African domestic agendas and programmes, such as the:
  - National Development Plan 2030 (NDP) (South African Ministry of Planning, 2012),
  - Medium-Term Strategic Framework (MTSF) sector plans (Republic of South Africa, 2014),
  - Provincial Growth and Development Plans (PGDPs) (Republic of South Africa: The Presidency, 2005) and
  - Municipal Integrated Development Plans (IDPs) (Republic of South Africa: Department of Provincial and Local Government, 2000)

All of the above agendas and indicator frameworks require good quality data, which are to be measured within the national statistics system (NSS) of a country.

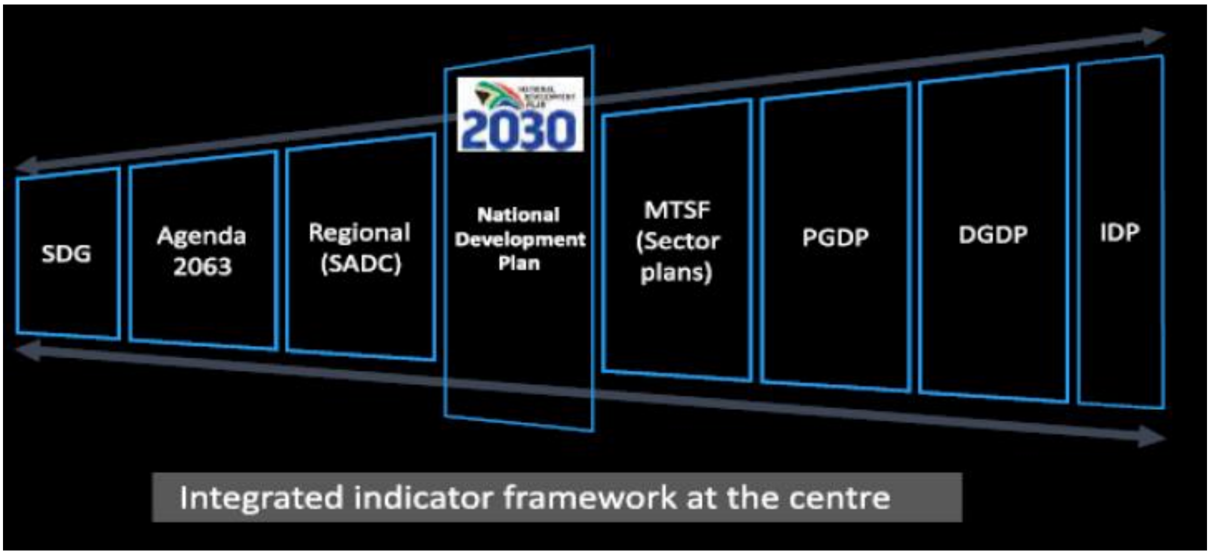


Figure 2.3: South African Integrated Indicator Framework

Concerning the policy agenda at a continental level, most African countries have committed to this developmental Agenda 2063 over the next 2 decades. The African Union launched a 50-year vision called Agenda 2063 (African Union Commission, 2015) for Africa in 2013. Agenda 2063 aims to harness the continent's comparative advantages, such as its people, history and cultures. Furthermore, Agenda 2063 also addresses the natural resources, the position and repositioning in the world to effect equitable and people-centered social, economic and technological transformation, including the eradication of poverty. It seeks to develop Africa's human capital to build social assets, infrastructure, and public goods. This is to empower women and youth, promote lasting peace, security and to build effective developmental states through participatory and accountable institutions and governance.

The African Agenda complements the UN Sustainable Development Agenda by reinforcing the urgent need for accountability concerning institutions. The eradication of poverty and ensuring that capital in all its forms be it human and natural resources is shared equally in a manner that ensures positive development and growth. With the policy agenda on the domestic front, the 2030 National Development Plan (South African Ministry of Planning, 2012) provided the framework to achieve a radical socio-economic agenda for South Africa. It recognised the need for a capable and developmental state, a thriving business sector, as well as strong civil society institutions with shared and complementary responsibilities. It identified decent work, education and the capacity of the state as being particularly important priorities. It also highlighted the need to improve the quality of administration of many government activities. Government-wide monitoring and evaluation (M&E) in South Africa requires detailed knowledge across and within all the government spheres, including interactions among planning, budgeting, and implementation functions in the public sector. The situation is complicated even further when the machinery of government is decentralised, with powers and functions distributed across three spheres of government (Engela, Ajam, Banque mondiale, & Banque mondiale, 2010).

On a larger scale, the research problem lies within and across the domains of data quality and quality assurance, in particular.

2.3 DATA QUALITY MANAGEMENT

Quality management for NSOs is necessary; to ensure confidence in the quality of the information it produces. Throughout the phases of a survey life cycle, or SVC, quality should be embedded and monitored during the planning and implementation phases (Figure 2.4) (Statistics South Africa, 2010b). Quality management systems have evolved across the globe to ensure that household data collection in surveys is accurate and trusted by all users of the data (United Nations, 2017). With the current technological innovations in handheld digital devices, NSOs, particularly in Africa, have identified an opportunity to utilise the opportunities at hand to partially and eventually fully integrate digital handheld devices within their strategies for household-based survey collection.

The SASQAF developed a tool to assess the quality of data produced by mainly NSS partner agencies for certification of the data as official statistics. In the South African context, this framework has been widely adopted by government state departments and across all three spheres of government, i.e. by the local, provincial and national government (National Treasury, 2010). The Government-Wide Monitoring and Evaluation (GWM&E) system was institutionalised in the South African public service and uses the SASQAF as the primary QA reference tool when evaluating departmental data quality (Marais, 2017; United Nations Department of Economic and Social Affairs, 2005). A collaborative effort was required to develop this framework, primarily coordinated by the NSO of the country. This framework has had a major impact on society in South Africa, even seeking to address spatial transformation through sub-city spatial indicators (Maritz et al., 2016). Relating to the goals of this research, where data quality is the overarching theme, both Chapters 2 and 3 respectively, will seek to respond to the primary research objectives, which read as follows:

- to investigate the current "best practices" applied in official household survey collection within a global and South African context.
- to design and describe a quality method or framework for adopting an effective digitised household survey process in South Africa.

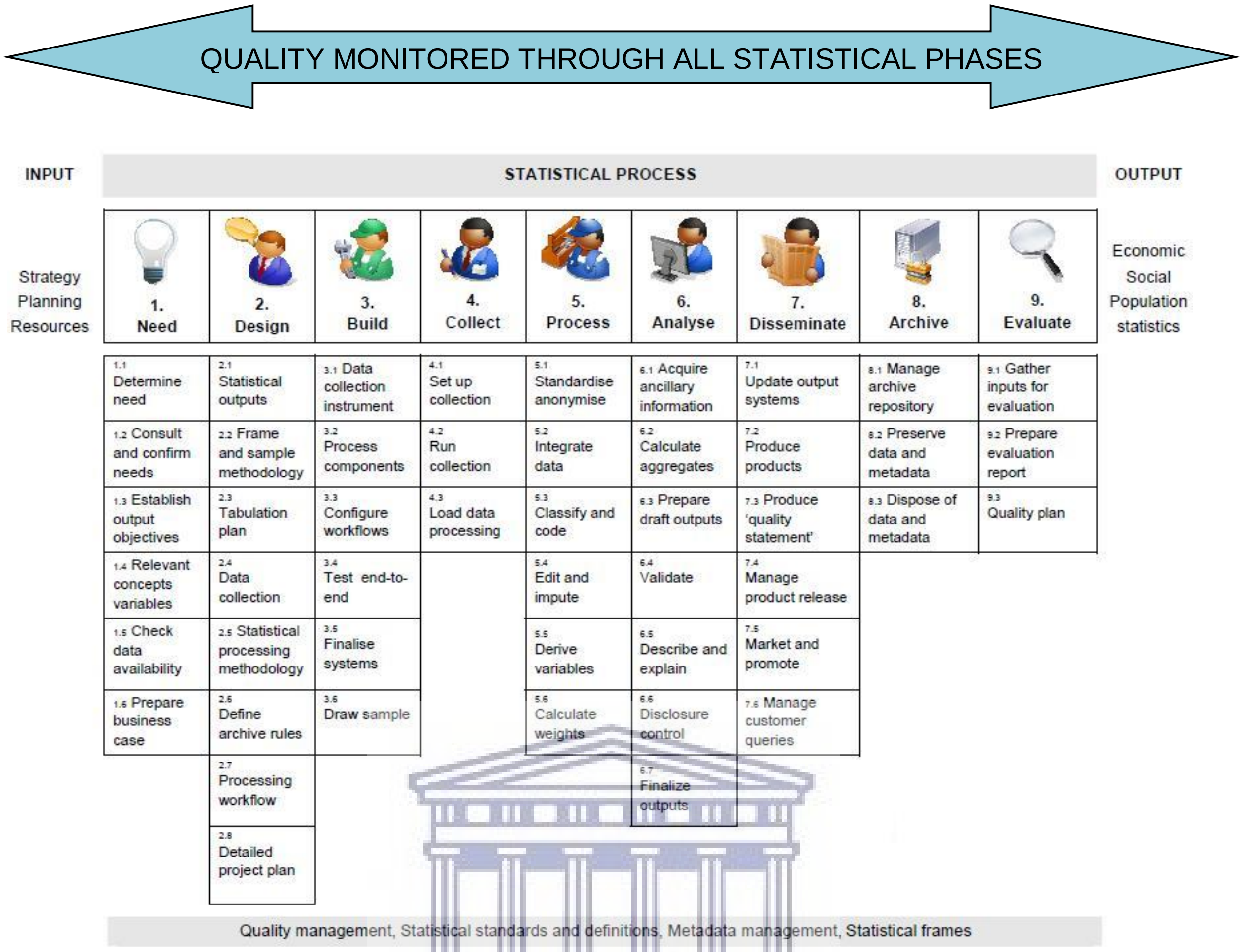


Figure 2.4: Statistical Value Chain (SVC) High Level Process

Source: (National Statistical System, 2010)

The primary research question is:

How can the South African Statistical Quality Assurance Framework (SASQAF) be enhanced to support digitised household survey-data collection in South Africa?

The SASQAF (first edition) provided producers of statistics with transparent procedures and criteria for the evaluation of statistical data. Institutions and organisations across both the public and private sectors, including research bodies or other non-governmental organisations (NGO's) also produce statistics. "If these statistics produced are in the public domain, they will most likely influence the development of government policies or the measurement and monitoring of government programmes. Statistics produced by the NSS partners can use the SASQAF to evaluate the quality of the data collected" (Statistics South Africa, 2011).

When assessing the quality of statistics, it is important to ensure that it is done using a methodolically sound, sustainable and transparent quality framework. Statistical producing bodies, listed previously, can use the SASQAF as a quality framework for evaluating and certifying the statistics they produce. Within the NSS framework, the SASQAF distinguishes between two types of statistics, namely "official" or "national' statistics". "National statistics" refer to those statistics used in the public domain; but which the Statistician-General (SG) has not certified it as being “official statistics” given that the statistics have not been evaluated formally through the SASQAF. The formal process of certification of statistics produced by organs of state involves a standard assessment procedure undertaken by a Data Quality Assessment Team (DQAT), which was established by the SG.

For the assessment of data for quality to begin (Figure 2.5), the submitting organ of state and the statistics under review need to comply with three initial criteria, namely the:

- producing agency should be a member of the NSS;
- statistics need to meet the user needs beyond those needs only being specific and internal need of the producing agency; and
- statistics produced should be part of a sustainable series and not a once-off collection.

On meeting these initial criteria, the assessment of the data can then begin. The DQAT is required to report on the statistics, classifying the quality levels of statistics as either of of the following:

- quality statistics;
- acceptable statistics;
- questionable statistics; or
- poor statistics.

If the outcome of an evaluation of statistics are deemed “quality statistics”, the DQAT is required to make certain recommendations indicating the areas for improvement for statistics to reach a higher level of quality, specifically to the level of “quality statistics”. In addition, if the outcome of an evaluation of statistics are “quality statistics”; in line with the quality dimensions set out in SASQAF, the SG will then formally designate the data as being "official statistics". These statistics are then subjected to periodic reviews determined by the SG in consultation with the head of the producing agency or department within the NSS.

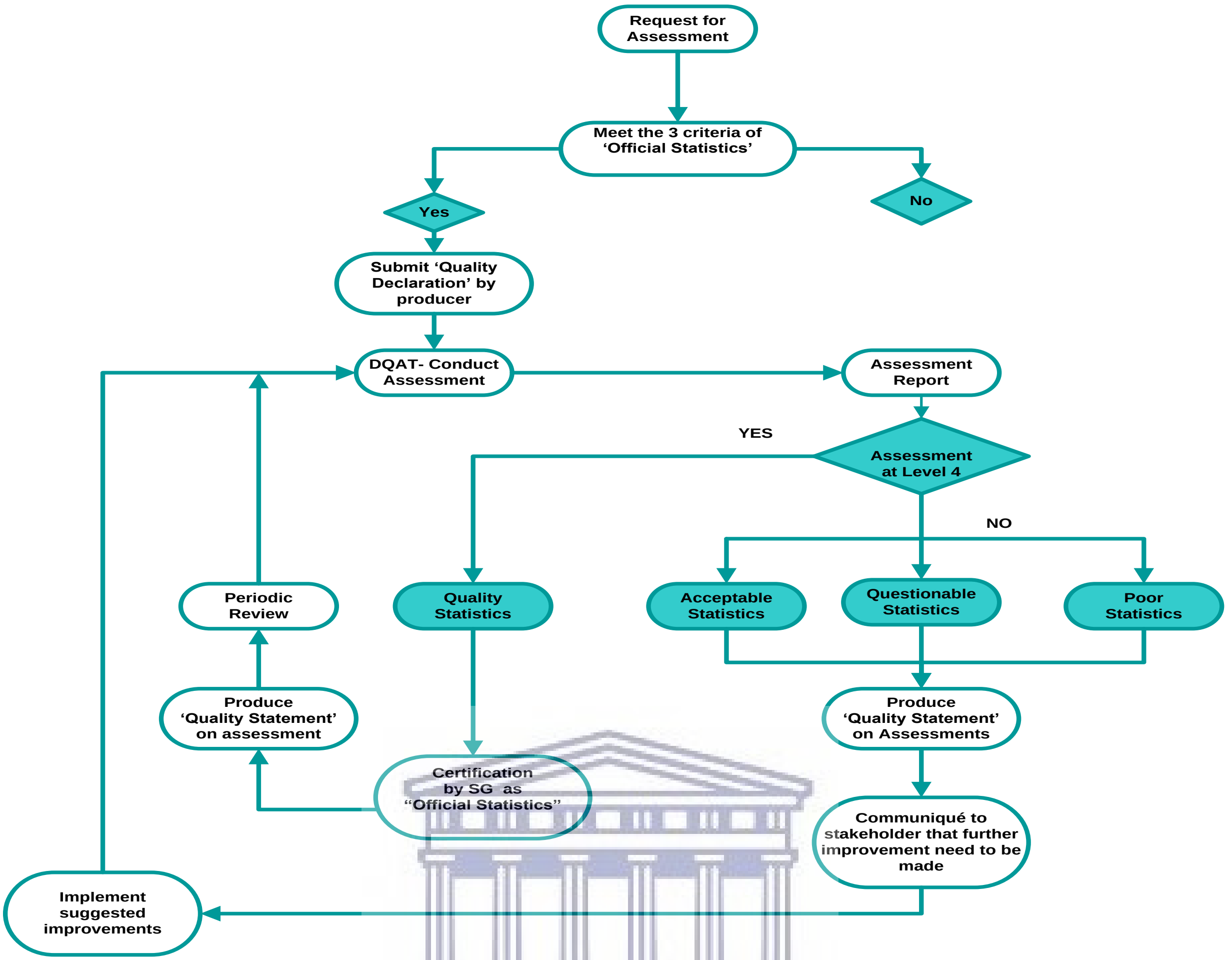


Figure 2.5: How to assess the quality of statistics and designate it as official?

Source: (National Statistical System, 2010)

The first edition of the SASQAF was published in 2008 (Statistics South Africa, 2008). The SASQAF (second edition) appeared in the government gazette on 23 September 2009 and published during 2010. At this stage, there was an increased demand and interest from a variety of stakeholders, both internal to the government, as well as external, to utilise SASQAF as a tool for measuring data quality (Statistics South Africa, 2010b). The increase in demand for quality data came primarily from data producers aiming to produce statistics that the public can trust and use. A SASQAF "Lite" version was designed as a self-assessment QA tool used by data producers (see Appendix I). The SASQAF Lite version determined the initial overall data quality level before applying the full version of SASQAF (second edition).

The Statistics Act of 1999 mandates the SG to develop standards for all organs of state and other agencies that produce statistics. To designate statistics as official statistics, or to specify the class of statistics produced by any other organ of the state. SASQAF provides clear criteria and transparent procedures for the evaluation of official statistics and other data (Statistics South Africa, 2010b). SASQAF Lite contains all the quality dimensions of the SASQAF (second edition). The difference is the reduced selection of quality indicators for self-assessment. Table 2.1 below demonstrates the use of the SASQAF Lite Self-Assessment tool by data producers and the selected Key Survey Performance Areas (KSPA's).

Table 2.1: SASQAF Lite Self-Assessment tool

SASQAF Quality Dimension	Score per Quality Dimension	Key Survey Performance area (KSPA)	Score per (KSPA)
	Self-assessed		Self-assessed
Prerequisites of quality	3,25	Responsibility	4,0
		Data Sharing	4,0
		Confidentiality	4,0
		Resources	3,3
		Quality Management	1,0
Relevance	4,00	User Database	4,0
		User Satisfaction	4,0
Accuracy	1,18	Non-Sampling Errors	1,4
		Flexibility	1,0
Timeliness	1,00	First Release Lag	1,0
		Final Release Lag	1,0
		Planned Timelines	1,0
Accessibility	1,83	Accessibility to public	2,0
		Information Restrictions	2,0
		Pre-announced release schedule	1,0
		Statistical Products	1,0
		Data dissemination policy	1,0
		Metadata	4,0
Interpretability	4,00	Statistical releases	4,0
Comparability and Coherence	3,00	Consistency	4,0
		Consistency checks	2,0
Methodological soundness	1,00	Concepts, Definitions, and Classifications	1,0
		Questionnaire	1,0
Integrity	1,33	Conditions for policy-makers	2,0
		Methodology changes	1,0
		Code of conduct	1,0

The SASQAF (second edition) provides a structure for the assessment of statistical products based on:

- **Dimensions of quality**
- **Indicators**, and
- **Benchmarks**

Each of the nine quality dimensions consists of several indicators. Within the indicators, several benchmarks are identified relating to a 4-point scale:

- Level 4 - Quality statistics
- Level 3 - Acceptable statistics
- Level 2 - Questionable statistics
- Level 1 - Poor statistics

Each dimension has associated quality indicators, standards, and benchmarks.

2.3.1 Dimensions of quality

The 9 dimensions of quality for the full SASQAF (Second edition) are (Statistics South Africa, 2010b):

0. Pre-requisites of quality:

The "pre-requisites of quality" refers to the basic conditions prelisted by organisations or institutions that may have an impact on data quality. The minimum set of the necessary or required conditions need to be in place in order to produce good "quality statistics". These minimum conditions for quality statistics forms the basis or foundations of which all other dimensions of quality is based on. The first quality dimension is the "pre-requisites of quality". Table 2.2 describes the SASQAF Lite KSPA in relation to the SASQAF (Second edition) Quality Indicators. The SASQAF Lite KSPAs for the prerequisites of quality are the:

- responsibility for producing statistics;
- data-sharing arrangements;
- confidentiality;
- availability of resources to collect statistics; and
- quality management system plans.

Table 2.2: SASQAF (Second edition) vs. SASQAF Lite (Statistics South Africa, 2010b)

SASQAF Indicator Number	Full SASQAF v2.0 Indicators	SASQAF Lite Key Survey Performance Area (Self-Assessment tool for data producers) – see Appendix I
1.1	The responsibility for producing statistics is clearly specified.	Responsibility
1.2	Standards and policies are in place to promote the consistency of methods and results.	
1.3	Data sharing and coordination among data-producing agencies are clearly specified.	Data Sharing
1.4	Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.	Confidentiality

1.5	Measures to oblige response are ensured through law.	
1.6	Resources are commensurate with the needs of statistical programmes. • Staff • Facilities • Computing resources • Financing	Resources
1.7	Measures to ensure efficient use of resources in 1.6 are implemented.	
1.8	Processes are in place to focus on, monitor and check quality.	Quality Management

1. Relevance:

This is the degree to which the data meet the real needs of clients (Minister of Industry, 2017; Statistics South Africa, 2008). “Relevance” refers to whether the statistics are required when it was produced. Relevance sheds light on the issues of most importance to users. Relevance also covers methodological soundness, particularly the extent to which the concepts, definitions, and classifications correspond to users’ needs. Assessing relevance is subjective, and it depends on the varying needs of the users. The statistical agency’s challenge is to weigh and balance the conflicting needs of current and potential users. Statistics should satisfy the user’s priority needs within the given resource constraints. Relevance can be seen as having the following three components: completeness; user needs; and user satisfaction (NATIONAL INSTITUTE OF STATISTICS OF RWANDA, 2014; Statistics South Africa, 2008). European Statistics defines relevance as meeting the needs of users. The three indicators of relevance are (EUROSTAT, 2015):

- processes in place to consult users, to monitor the relevance and the utility of the existing statistics in meeting their needs, and to consider their emerging needs and priorities.
- priority needs being met and reflected in the work programme.
- user satisfaction monitored regularly, and it is systematically followed up.

The SASQAF Lite KSPAs for relevance are:

- user database
- user satisfaction

2. Accuracy:

“Accuracy” is the degree to which the statistical output correctly describes the phenomena it was designed to measure. In the general statistical sense, accuracy denotes the closeness of computations or the estimates of the exact or true values (Organisation for Economic Co-operation and Development, 2007). NSOs need to develop, produce and disseminate statistics that are accurate and that reliably portray reality. It is usually characterised in terms of the estimation of sampling and non-sampling errors. These types of errors are delineated into bias (systematic error) and variance (random error) components. They reflect the major sources of error (e.g. errors linked to sampling, coverage, measurement, non-response, and processing).

Reliability concerns whether the statistics consistently measure the reality that they are designed to represent over time (NATIONAL INSTITUTE OF STATISTICS OF RWANDA, 2014; Statistics South Africa, 2008). If the data quality were limited to the accuracy dimension only, then different aspects of the survey accuracy could be related to various error sources in surveys. Five of these error sources can be identified: sampling error, non-response error, coverage error, measurement error, and processing error (Alwin, 2007; Osborne, 1942).

The SASQAF Lite KSPAs for accuracy:

- non-sampling errors
- flexibility

3. Timeliness:

"Timeliness" denotes the delay between the reference point to which the information pertains and the date on which the information becomes available. Timeliness also addresses those aspects of periodicity and punctuality of the production activities within the statistical value chain. According to Anderson et al., (1997), timeliness is the length of time between its availability and the event or phenomenon it describes. Timeliness can refer to several concepts, including the length of the data collection's production time, the time from data collection until the first availability of a product, and the frequency of the data collection. Timeliness can, however, be difficult to characterise; since the characteristics of the data collection can often reflect the availability of the data (Andersson, C., Lindstrom, H. L. and Lyberg, 1997). Some of the key indicators for timeliness include:

- the statistical production time
- timely receipt of administrative records
- the periodicity of statistical release
- the punctuality of statistical release

The SASQAF Lite KSPAs for timeliness:

- first release lag

- final release lag
- planned timelines

4. Accessibility:

The “accessibility” of statistical information and metadata refers to the ease with which data are retrieved from the NSO or statistical producing agency. In addition to whether information exists, "Accessibility" from a statistical user's perspective includes the format or medium through which it can be attained from the NSO or statistical producing agency (Statistics South Africa, 2010b). The cost of the information may also be an aspect of accessibility for some users. Accessibility also refers to the physical conditions under which users can obtain the data: where to go, who to contact, the availability on a website, the format of the data (degree of aggregation), the costing policy, the marketing conditions (copyright, etc.), the availability of micro- or macro- data in various formats (paper, files, DVD, Internet etc.) (Statistics Canada, 1998).

Accessibility also implies that the data products include adequate documentation and discussion to allow proper interpretation of the survey results. This can be enhanced by the provision of user workshops and training. Data products tend to have a higher value if they are easily accessible.

The SASQAF Lite KSPAs for accessibility include:

- accessibility to public
- information restrictions
- pre-announced release schedule
- statistical products
- data-dissemination policy
- metadata

5. Interpretability

“Interpretability” reflects the ease with which the user may understand, properly use and analyse the data. This requires an adequate definition of the concepts, target populations, variables and terminology underlying the data; and the information describing the limitations of the data, if any, largely determines the degree of interpretability (T. Chen et al., 2014). The information needed to understand or interpret statistical data falls under four broad headings (United Nations Department of Economic and Social Affairs, 2005):

- Concepts, definitions, and classifications that underlie the data;
- Metadata on the scientific methodology used to collect and compile the data;

- Key findings, giving a summary of the results;
- Presentation of the statistics in a meaningful and useful format.

The SASQAF Lite KSPA for interpretability:

- statistical releases contain a summary of the key findings for ease of interpretation

6. Coherence

Ideally, we would use the same definitions from all sources. This is not always possible, but there should be a good reason for any differences. "If the data are the responsibility of different agencies, it is likely that the requirements would be different. There may be neither opportunity nor incentive for harmonisation, in which case coherence then suffers. Frameworks, such as the System of National Accounts, are extremely valuable for achieving coherence. Even within one country, regional differences may affect coherence and consistency. If the definitions are the same, collecting the same information about events from different sources would, with all probability, provide different results" (Eurostat, European Commission, & Statistika Centralbyran, 1999).

"Coherence" is the degree to which statistical datasets are connected to each other within a broad analytical framework over time. "It is the extent to which the differences between two sets of statistics are attributable to differences between the estimates and the true value of the statistics; and this is commonly characterised by statistics"(Statistics South Africa, 2010b):

- the use of common concepts and definitions within and between series;
- the use of common variables and classifications within and between statistical series;
- the use of common methodology and systems for data collection and processing within series;
- the use of common methodology for various processing steps of a survey, such as editing and imputations within a series.

The SASQAF Lite KSPAs for coherence:

- consistency
- consistency checks

7. Methodological Soundness

Methodological soundness is the application of international standards and agreed-on practices to produce statistical outputs. The application of such standards fosters national and international comparability. The key indicators for methodological soundness are (Statistics South Africa, 2010b):

- international norms and standards on methods;
- data compilation methods employ acceptable procedures;
- other statistical procedures employ sound statistical techniques;
- transparent revision policy and studies of revisions are done and made public.

The SASQAF Lite KSPAs for methodological soundness:

- concepts, definitions, and classifications
- questionnaire

8. Integrity

"Integrity refers to the values and the related practices that maintain users' confidence in the agency producing the statistics, and ultimately, in the statistical product (Organisation for Economic Co-operation and Development, 2007). This includes, among others, the need for the statistical system to be based on the United Nations (UN) Principles of Official Statistics. It includes the principles of objectivity in collection, compilation, and dissemination of the data, to ensure unbiased statistics, which are not subject to confidentiality breaches, or premature releases" (Statistics South Africa, 2010b). The key indicators for integrity are (Statistics South Africa, 2010b):

- professionalism and ethical standards, which guide policies and practices;
- assurances that statistics have been produced on an impartial basis;
- ethical standards are guided by policies and procedures.

The SASQAF Lite KSPAs for integrity includes the:

- conditions for policy-makers
- methodology changes
- code of conduct

The SASQAF includes the institutionalisation of quality management to improve the comparability and the accuracy of statistical information by institutionalising an end-to-end quality management system in line with the Fundamental Principles of Official Statistics (United Nations Department of Economic and Social Affairs, 2005) and the African Charter on Statistics (United Nations Statistics Division, 2007). In summary, SASQAF contains 9 dimensions of quality, which are the following (see Table 2.3):

Table 2.3: SASQAF quality dimensions references

SASQAF QUALITY DIMENSIONS	References
0. Prerequisites of quality	(Statistics South Africa, 2010b)
1. Relevance	(EUROSTAT, 2015; NATIONAL INSTITUTE OF STATISTICS OF RWANDA, 2014; Statistics South Africa, 2008)
2. Accuracy	(Bergdahl, Ehling, Elvers, & Földesi, 2007; NATIONAL INSTITUTE OF STATISTICS OF RWANDA, 2014; Organisation for Economic Co-operation and Development, 2007; Osborne, 1942; Statistics South Africa, 2008)
3. Timeliness	(Andersson, C., Lindstrom, H. L. and Lyberg, 1997; Statistics South Africa, 2008)
4. Accessibility	(Minister of Industry, 2017; Statistics Canada, 1998; Statistics South Africa, 2008)
5. Coherence & Comparability	(Eurostat et al., 1999; Statistics South Africa, 2008)
6. Interpretability	(T. Chen et al., 2014; Statistics South Africa, 2008)
7. Methodological Soundness	(Statistics South Africa, 2010b)
8. Integrity	(Organisation for Economic Co-operation and Development, 2007; Statistics South Africa, 2008)

Table 2.4 displays the comparison of quality dimensions of the SASQAF between countries and other well-known statistical agencies. This entails the development of an end-to-end Quality Management System (QMS), which is a set of interrelated or interacting elements of an organisation to establish policies, objectives, processes to achieve those objectives and to address quality management as a single discipline.

The management system's elements establish the organisation's structure, roles, and responsibilities, planning, operation, policies, practices, rules, beliefs, objectives, and processes to achieve those objectives. Examples of a quality management system include Total Quality Management (TQM) – see Section 8.1.1 , Lean Manufacturing, Six Sigma and Process Management (Andersson, Eriksson, & Torstensson, 2006).

The design and implementation exercise for a QMS takes into account the key elements of a QMS that should exist and as outlined below, the (Statistics South Africa, 2017b) :

- quality framework and quality-assessment framework, e.g. SASQAF;
- quality policy;
- production or manufacturing process flow, e.g. SVC;
- quality manual, standard operating procedures, and guidelines;
- system for continuous quality improvement of surveys (e.g. TQM, six sigma)
- internal auditing methods for monitoring, evaluation, and follow-up on surveys conducted (e.g. control surveys);

- external auditing methods for monitoring, evaluation, and follow-up on surveys conducted;
- defined organisational structure/establishment, which must be aligned with quality management activities;
- identifiable information-management system for data quality, including capturing systems and data quality assessment tools. paper-based and computerized tools including a quality scorecard;
- document management and archiving facilities: all processes should be documented, and archiving facilities should be provided.

Table 2.4: Dimensions of quality used by various national statistical organisations, agencies or institutions

World Bank	Eurostat	South Africa SASQAF	Canada	Finland
Prerequisites of quality		Prerequisites of quality		
	Relevance	Relevance	Relevance	Relevance
Accuracy & ***Reliability**	Accuracy	Accuracy	Accuracy	Accuracy and reliability
	Timeliness & Punctuality	Timeliness	Timeliness	Timeliness & Punctuality
Accessibility	Accessibility & ***clarity**	Accessibility	Accessibility	Accessibility and ***clarity**
	Coherence	Coherence & Comparability	Coherence	***Consistency** and comparability
		Interpretability	Interpretability	
Methodological Soundness		Methodological Soundness		
Integrity		Integrity		
	Comparability			
***Serviceability**				

2.4 TRENDS IN DATA QUALITY ASSURANCE

To address the high costs and time requirements associated with household-based surveys, higher-income countries (HICs) developed and employed predominantly telephone surveys to collect health and demographic estimates. Mobile phone ownership and access became more common with 94 subscriptions per 100 inhabitants in developing countries (ITU, 2014). Opportunities arose to revolutionise the current methods of data collection in lower and middle-income countries (LMICs) mainly found in Africa (Gibson et al., 2017). There are several QA frameworks used across the globe by HICs and LMICs for quality improvement (Zeng, Gheorghe, & Nair, 2016), these include :

- The International Monetary Fund's DQAF
- Eurostat: Handbook on Data Quality Assessment Methods and Tools
- The International Household Survey Network - Survey Quality Assessment Framework (SQAF)
- The Quality Assessment for Linked Open Data
- The Quality Assurance Framework of the European Statistical System
- Data Quality Online: Australian Bureau of Statistics (ABS)

2.4.1 The International Monetary Fund's DQAF

The IMF's Data-Quality Assessment Framework (DQAF) identifies the quality related governance features of statistical systems, statistical processes, and statistical products. Rooted in the UN Fundamental Principles of Official Statistics, the DQAF was derived from the Special Data Dissemination Standard (SDDS); the General Data Dissemination System (GDDS) as well as the IMF's initiatives on data dissemination. The IMF's DQAF is based on internationally accepted core standards, principles, and practices for official statistics. It covers various quality aspects ranging from data collection and processing to dissemination. The DQAF was used within and across industrial and developing countries to measure compliance with standards and codes (Worldbank Group, 2017).

The DQAF allows countries to perform their data-quality assessments, and it applies to eight different datasets. These include the national accounts statistics, consumer price index, producer price index, government finance statistics, monetary statistics, balance-of-payment statistics, external debt statistics, and household income statistics. This framework uses five dimensions of quality, which are broken down into key quality elements and indicators. The five dimensions of quality include the assurances of integrity, methodological soundness, accuracy and reliability, serviceability, accessibility of data quality as well as a set of prerequisites for data quality founded by the center of the IMF Data Quality Assessment Framework (DQAF) (International Monetary Fund, 2012).

2.4.2 Data Quality Online: Australian Bureau of Statistics (ABS)

The Australian Bureau of Statistics (ABS) allows its users to use their website[1] to develop a sense of data quality. The 'General' data quality statements contain a broad range of

[1] Data Quality Online Tool: https://www.nss.gov.au/dataquality/index.jsp

applications, including various data sources, such as administrative, survey and multiple sources. The Data Quality Online Tool (See Figure 2.6) was developed by the ABS as a national statistical service initiative to assist users to find information and receive guidance on the quality management of data (Australian Bureau of Statistics, 2015).

The purpose of the ABS data-quality statement tool is to encourage informed decision-making by:

- Assisting data users, data producers, data custodians, and data owners to communicate key data quality issues in the form of a data-quality statement
- Providing information about data quality to enable information use and re-use by the wider community
- Assisting people using statistical data to understand how the data can be used, and assessing whether they can be used for the purpose they have in mind

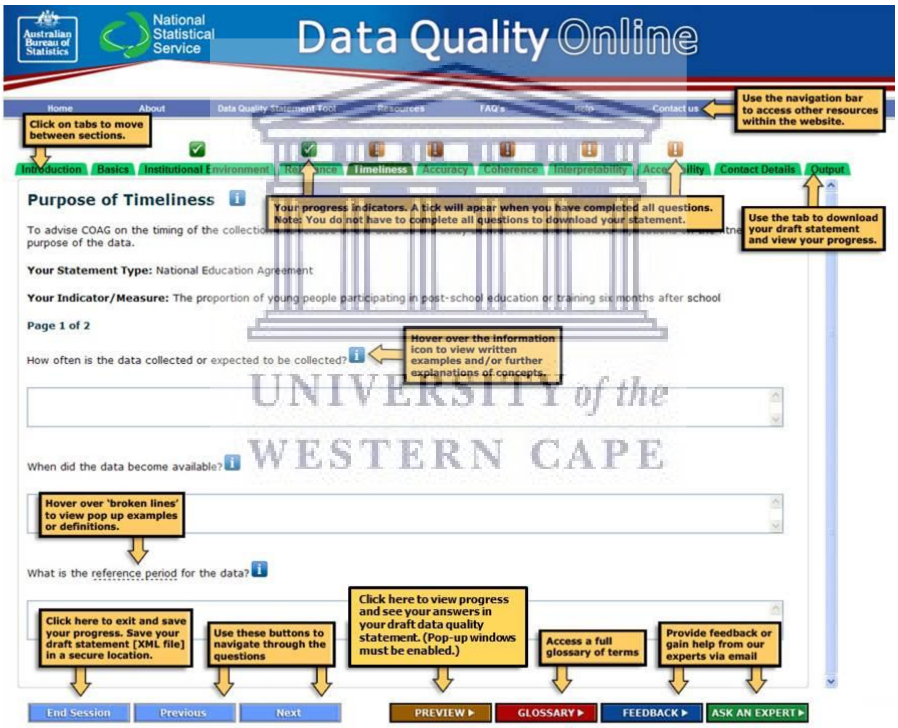


Figure 2.6: Australian Bureau of Statistics Quality Assessment Online Framework

The benefits of the “data quality” statement tool include:

- a clear, simple and methodical process for developing a data quality statement
- the ABS data-quality framework, which was an endorsed framework, to assess the data quality
- information expressed in plain English that could be readily understood

- a method for compiling a data quality statement. This enabled data users, data producers, data custodians, and data owners to apply a broad and multi-faceted approach to assessing data quality. In addition to communicating the key issues in a data-quality statement

2.4.3 Eurostat: Handbook on Data Quality Assessment Methods and Tools

The Eurostat handbook on data-quality assessment methods and tools provides a concise description of the data-quality assessment methods in use. The handbook provides recommendations on how these methods and tools are used. When combined, it was found to be the most efficient and cost-effective use of the methods (Bergdahl et al., 2007). The handbook focused on the use of methods in the field of surveys based on primary data collection. The methods may, however, also be used in other areas of statistical activity, such as administrative registers and national accounts.

The Handbook on Data-Quality Assessment Methods and Tools (DatQAM) aimed at facilitating a systematic implementation of the data quality assessment process. It presented the most important assessment methods, namely: quality reports, quality indicators, measurement of process variables, user surveys, self-assessment, and auditing, as well as the approaches of labelling and certification. This framework, however, does not cater for DDC.

2.4.4 The International Household Survey Network - Survey Quality Assessment Framework (SQAF)

The international household survey network - Survey Quality Assessment Framework (SQAF) is a series of questions written to assist survey managers to improve the quality of their surveys. Questions on multiple aspects of the survey process are required from survey managers in relation to quality. The framework encourages checking, documentation, implementation of systems, minimising the introduction of errors and ensure the completeness of the information (Statistical Services Centre of the University of Reading, 2009). The questionnaire included questions around data entry (scanning or CAPI). This tool consists of a checklist, which does not assess data either quantitatively or qualitatively.

2.4.5 The Quality Assessment for Linked Open Data

"The development and standardisation of semantic web technologies have resulted in an unprecedented volume of data published on the web, as Linking Open Data (LOD). However, these contained widely varying data quality ranging from extensively curated datasets to crowd-sourced and extracted data of relatively low quality. Data quality is commonly conceived as fitness for use" (Zaveri et al., 2013). The aim was to provide researchers and data curators with a comprehensive understanding of the existing work. Thereby encouraging further experimentation and the development of new approaches focusing on data quality, specifically for LOD. Zaveri et. al (2013) focused on research related to data quality and guidelines, as well as recommendations on how to publish "good" data.

However, there is less focus on how to use these "good" data, although the focus is mainly web-based; handheld DDC are not discussed.

2.4.6 The Quality Assurance Framework of the European Statistical System

The Quality Assurance Framework of the European Statistical System (ESS QAF) is a supporting document aimed at assisting the implementation of the European Statistics Code of Practice (CoP). This framework dealt primarily with the "Institutional environment" and "commitment to quality" aspects. It identified possible activities, methods, and tools that could provide guidance and evidence for the implementation of the indicators of the CoP. In general, the ESS QAF is used as a reference framework by all the actors participating in the production and dissemination of European Statistics, such as national statistical institutes or other statistical authorities (EUROSTAT, 2015).

In this section, we have highlighted the quality dimensions from SASQAF, IMF's DQAF, ABS, Eurostat, SQAF, and Data Assessment for linked open data, ESS QAF. Other notable international quality assurance frameworks include:

- the generic National Quality Assurance Framework (NQAF). United Nations Statistical Division – UNSD (UNCTAD, 2017).
- a proposal for the structure of a regional code of good statistical practice for Latin America, and the Caribbean Institutional strengthening working group (EUROSTAT, 2015).
- Statistics Canada's Quality Assurance Framework (Minister of Industry, 2017).
- Statistics Netherland's standard for statistical processes (Nederpelt, 2011).
- The CARTA framework specifies universal core data quality dimensions (Marais, 2017).

2.5 TECHNOLOGICAL INNOVATION IN DATA COLLECTION

NSOs globally started using more efficient methods and tools for conducting their surveys. In the area of data collection, for example, several newer methods such as Computer Assisted Personal Interviewing (CAPI), Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Web Interviewing (CAWI) are used. Besides, several specialised software systems allow for real-time capturing, editing and monitoring of survey data from the field. Advancements in digital Geographic Information Systems (GIS) have allowed for the integration of maps and satellite images more seamlessly into the survey process. In South Africa, as found in most countries in the developing world, the financial costs and the time associated with conducting surveys traditionally have become costly and unsustainable.

2.5.1 Re-engineering the standard Statistical Value Chain - SVC

Statistics South Africa recently embarked on a strategy to innovate and modernise the current Statistical Value Chain (SVC). The aim was to achieve efficiencies and improve survey product quality (see APPENDIX H). The business model generally used by the South African NSO to conduct its surveys used the traditional PAPI method. Most organisational and management structures of NSOs survey areas are aligned for PAPI survey processes.

However, this business model is unsustainable for several reasons. Firstly, there have been several innovations in the area of survey methodology and design over the last decade with the advancement of technology. To that effect, the initial steps taken towards the implementation strategy were to collect the data digitally by using a handheld tablet or device. Thereby introducing the CAPI technique of household field-data collection. During the research period, the following large sample size DDC projects conducted included: the KwaZulu-Natal (KZN) Citizen Satisfaction Survey in 2015, the South African Demographic and Health Survey (SADHS) 2016, and the Community Survey 2016. This study excludes the SADHS 2016.

The current economic environment and budget cuts imposed by the South African National Treasury compelled the organisation to use its limited financial resources more efficiently. The demand for timely statistics and data at lower levels of dissemination and thematic spread (i.e. covering more areas/themes) is ever increasing. Technological innovation is a key element an NSO requires in order to explore these technologies and remain relevant. This is to respond to all user needs and challenges. Operationally, in South Africa, the NSO has already achieved success with previous DDC projects that utilised the CAPI model in the survey process, as opposed to PAPDC (see Table 2.5).

To address the challenge of becoming more relevant to the policy and data needs, the organisation adopted a long-term strategic direction of "Integration by Design" for the work it does. This approach allows the organisation to become more efficient, responsive and agile to the data needs in the country. For DDC however, it would require a change in the business model of the organisation for the collection processes.

For example, technological advancements in GIS and survey methodology e.g. DDC, monitoring and fieldwork management processes would need to play a key role in surveys conducted in the future.

Table 2.5: CAPI results vs. previous PAPDC household surveys conducted at Stats SA

CAPI Benefits	CAPI CS2016 versus PAPDC CS2007	CAPI KZN CSS 2015	Comment
Survey results are available sooner	CS 2007 collected data from 284 000 households with 6146 field staff. It took 3 weeks in the field and the time between processing and release was 7 months. In contrast, CS2016 involved 1.4 million households, 11 828 field staff collected data over 6 weeks, while 2 months were allocated for processing and the release of results through publication.	Data was collected from an initial 20 819 households; employing over 200 field staff; Fieldwork took 6 weeks and the data was released two months later for publication.	CAPI resulted in no transportation and storage of questionnaires to the central warehouse for processing. Instead, questionnaires were electronically submitted or synchronised to the central warehouse (server). Resulting in the elimination of data processing and reduction on post-processing and editing processes.
Improved data quality and response rates	Response rates: CS 2007 was 93.9% and CS2016 was 90.5% A statistically good result when seen in the context of the sample size.	83% response rate and 15% out of scope	Reduction of “out of scope” cases as well as real-time analysis of live data, real-time field management (remote monitoring) and support. Elimination for the need of coding and in-built error trapping eliminated the possibilities of incorrect, missing or excess entries. Filters, controls, skips and jumps also contributed to more effective and accurate data gathering. These guided the fieldworker and reduced the risk for human error and the temptation to take shortcuts. Ensured that the enumerated dwelling unit is the actual sampled dwelling unit.
Cheaper to conduct	Approximate total survey cost reduction between CS 2016 and CS 2007 was 53%.	Only 2% of the total cost of CS 2016 was used to conduct the KZN CSS 2015.	Cheaper based on the sample size covered and thus useful for larger sampled surveys than previously possible with PAPI. Digital handheld devices could accommodate more questions than PAPI to meet user needs including the availability of physical point data, which was used with GIS applications.

2.5.2 The evolutionary process and the legislative reform of statistics

The Statistics Act of 1999 (Statistics Act, 1999) is the legal framework that governs statistical production and use in South Africa. South Africa's official supplier of data has embarked on the revision of the statistics legislation to drive statistical reform, particularly in statistical coordination. Co-ordination between statistical agencies is essential for consistency and efficiency. The department's proposed legislative reform will focus on: strengthening coordination between the organs of State; improving the capabilities of information collection, analysis, use, retrieval, storage and archiving; creating a State-wide statistics service, including the establishment a professional and sustainable NSS; the data revolution; statistical geography and institutional arrangements. The revised statistics legislation will form the bedrock of statistical development and practice in the country to inform planning, policy development, monitoring, and evaluation, as well as decision-making.

Stats SA, the South African NSO used as a case subject for this study, is a well-recognised state or national government institution in South Africa. It upholds its independence from any political interference. Independence is important and is prevalent throughout all phases of the SVC in terms of statistical production, i.e. conducting business and large household surveys, managing registers and the dissemination of statistics. As an organisation, it has demonstrated over the past decade that it has evolved with the times and increasingly delivered on its broad mandate, namely to "collect official statistics". The external environment has not remained static as the needs of stakeholders or users have changed and grown over time. Technologies have evolved providing new opportunities; and easier access to data represents both an opportunity and a threat to the organisation. Staff and skills profiles have also changed during this period.

Data collected will become more real-time, services and product delivery will be digital by default.. Decision-makers across the spectrum – individuals, businesses and government – will value these services as being vital to them. Confusion about statistics will be much rarer; as the information base delivered as "national" and "official" statistics will be accepted and used with confidence. Statistics should remain relevant and improve the delivery of an efficient statistical service to the nation. Statistics should also be meaningful by continuously innovating and being at the forefront of statistical production and integrated statistical products and services (Statistics South Africa, 2014). This led to the review of the SVC to deliver a different output and outcome.

The institution has been conducting household surveys primarily using the manual PAPDC method. This is more costly and inefficient. The use of pen and paper to create a listing frame for sampling DUs, manual monitoring and progress reporting for survey operations using paper has become costly (Fitzgerald & Fitzgibbon, 2014). The current business model of conducting surveys is not sustainable, given the current budgetary constraints in South Africa.

2.5.3 Cost

In Africa, the cost of a survey is a key factor when deciding to conduct one. Information costs may also be a deciding factor hampering the "accessibility" of data for some statistical users. In the SASQAF, the Pre-Requisites of quality, quality indicator: *"1.6 Resources are commensurate with the needs of statistical programmes"* focuses on the cost or budget availability for all survey-related items (Statistics South Africa, 2010b). The survey implementation cost per household is significantly higher than in other regions of the world. Household surveys are therefore conducted infrequently, typically due to the high costs in personnel and transportation associated with household survey implementation and the face-to-face nature of data collection, especially in LMICs (Gibson et al., 2017; Kilic et al., 2017; Lukwago et al., 2013; Somda et al., 2009, 2010; Toscano et al., 2013).

If technological improvements during training of SOs are incorporated, the survey implementation costs in Africa should decline. Survey costs could increase in real terms due to the increased complexity of questionnaire instruments and fieldwork protocols. To meet the demands for higher-quality data collected at the individual level on priority topics (Kilic et al., 2017). In the study conducted by Kilic et al, it was important to undertake further analyses to understand why the survey implementation cost in Africa is so high. Their findings suggested that household survey implementation in Africa costs at least USD 300 per household (or R4440.27[2]). It is approximately five times higher in comparison to other developed countries such as Europe, Central Asia, Latin America, and the Caribbean (Kilic et al., 2017).

2.5.4 Accuracy, error rates and perception of quality

In the SASQAF, the integrity of data is to be upheld by the NSO producer through accurate collection of data. The quality indicator standard: "*9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations."* reaffirms the need to use transparent, scientific statistical methodologies to support rational decisions when conducting surveys. This refers to the perception of quality from the users of data, as well as the data collectors themselves. From an accuracy perspective "*3.2 Measures of non-sampling errors are calculated, viz.: i.e. Misclassification errors",* ensures that actual data are collected at the correct sampled point based on methodological and scientific probability sampling techniques from the data producer. GIS systems and other quality-control activities support

[2] using the 31 March 2016 (USD to ZAR exchange rate) - https://www.poundsterlinglive.com/best-exchange-rates/us-dollar-to-south-african-rand-exchange-rate-on-2016-03-31

these activities via quantitative and qualitative verification in the field by quality monitors where available.

For the error-rate quality indicator the *“3.2 Measures of non-sampling errors are calculated”* or more specifically, the standard of "*3.2.16 Data entry must average an acceptable accuracy rate”* should be upheld.

2.5.5 Security

Improved data security and respondent confidentiality are important factors to ensure when conducting official household-survey data collection. In terms of security, the mode of data collection is insignificant, i.e. whether PAPDC or DDC. The security of staff collecting the data, the security of the respondent while participating in the surveys, as well as the respondent's information, should be “planned” for before the data collection phase. The SVC indicates during the "Collection phase" or "4.1 Set up collection" phase that security should be ensured (Statistics South Africa, 2010b). For DDC, the synchronisation and uploading of encrypted surveys to a central secure location and attaining access to the web-based interface contributes to improved data security and respondent confidentiality (Hattas & Eloff, 2011; Tomlinson et al., 2009).

Security and confidentiality are prevalent in both the SASQAF and SVC (see Table 2.6):

Table 2.6: Security referenced to SASQAF and SVC

SASQAF Dimensions	SVC phases
Pre-requisites of quality (1.4)	Collect (4.1)
Accessibility (5.3, 5.4)	Process (5.1)
Integrity (9.1)	Analyse (6.6)

2.5.6 Speed

The standard time for the releasing of statistics by an NSO is to be made public (Statistical Services Centre of the University of Reading, 2009; Statistics Canada, 1998; Statistics South Africa, 2008). If deviations of tight time-bound schedules are to be avoided, all processes on the SVC need to adhere to specific internal time arrangements and agreements. Reducing the overall survey time refers to the speed at which survey processes occur. The "timeliness" dimension within SASQAF will influence its relevance or importance to the users of data (Statistics South Africa, 2010).

The impact of speed can affect an NSO both internally and externally. Internally, speed determines whether NSOs meet the SVC processes deadlines. Externally, speed dictates whether users of the information receive timely, relevant data that would satisfy their needs.

Speed is correlated to the accuracy at which all downstream processes within the SVC are required to occur. In an article by Frank, he asks: "what do users want?" Amongst others, the user will also expect a faster turnaround frequency and speed. Results need to be instantaneous, i.e. data today about today. Moreover, they need to be frequent, i.e. data needs to be available for every year or every quarter, certainly not just for every decade! (Frank, 2014).

2.5.7 Training

The questionnaire design process generally takes long, i.e. 2 to 3 months. It involves many different people (subject matter specialists, survey practitioners, etc.). With a PAPDC process, this is done by preparing, distributing, discussing and piloting numerous "versions" or "generations" of the questionnaire until the final version is agreed upon. For DDC, the equivalent steps for questionnaire design and training still need to be defined. Interviewer training will need to be re-designed around the new digital handheld technology. PAPDC training usually involves theoretical sessions, mock interviews, training manuals, etc. However, if DDC is employed, relevant training approaches and outcomes need to be developed (Hattas & Breytenbach, 2017; Hattas & Eloff, 2011; Munoz, 2003).

Training interviewers to collect data accurately is important to maintain the credibility and reputation of the collecting agency or the NSO. Negative organisational reputation often leads to poor performance, distrust amongst the users, and often, among the sponsors of data collection agencies or NSOs. The effectiveness of training for either paper-based surveys or paperless surveys needs to be determined. Training techniques, methodologies and resultant performance during these aspects are to be transparent, whether DDC or PAPDC processes (Statistics Act, 1999; Statistics South Africa, 2008).

On the SVC, training is discussed under the "collect" phase; however, more emphasis should be provided within one of the quality dimensions within SASQAF.

2.5.8 Process flow and structural changes

2.5.8.1 Generic end-to-end DDC flow:

The PAPDC process flow defined by the standard SVC (Need, Design, Build, ***Collect***, Process, Analyse, Disseminate, Archive and Evaluate) may not fully cater for a DDC process. Therefore, the need to build and develop a standard generic end-to-end process flow for DDC, and to determine which phases of the standard SVC will be affected, is quite evident. DDC affects all processes on the standard SVC, however, the "Collect" phase(Hattas, 2013) (Hattas, 2013) was identified as a significant component. It is affected due to the impact it has on the (external) respondent and the (internal) interviewer (Hattas, 2013).

2.5.8.2 Access to data (internal):

The NSO collecting official data primarily utilises staff employed by the NSO, either on a permanent or contract basis. Each official survey is conducted using project management principles, consisting of human and non-human resources. A hierarchical staff structure, based on the office and staff responsible for the data collection in the field forms part of the project team responsible for collecting official surveys. With a PAPDC process, internal staff only receive access to the data collected from the field, once they have been collected, packaged and sent to a central location for processing. Often, staff with the appropriate access rights would only then be able to determine the actual quantity and quality of the data collected upon commencing the SVC downstream processes (Process → Analyse → Disseminate → Archive →Evaluate).

For DDC, data can be accessed in real-time to monitor the data collection processes. Relevant field managers can interact with fieldworkers to assist with any queries or challenges faced in the field (M. M. Hattas, 2013; Munoz, 2003; United Nations Department of Economic and Social Affairs, 2005).

2.5.8.3 Access to data (external):

The SASQAF indicator (5.1) requires statistical products (e.g. data, metadata) to be made available to the public (Statistics South Africa, 2008; United Nations Department of Economic and Social Affairs, 2005). Regardless of the mode, PAPDC or DDC, access to data should be made available on a platform, which is accessible to all the users of data. At the other extreme of the survey project timeline, the notion is no longer that the end product expected from the survey is a printed publication. With a collection of statistical tables, it has been replaced by that of a database. This can be used by the statistical agency to prepare the initial tables; but it would also be accessible to researchers, policy-makers and the public in general.

The descriptive summary report of survey results are no longer seen as the final step. It is seen as the starting point of a variety of analytical endeavours that may last for many years after the project is officially closed and the survey project team is disbanded (Munoz, 2003).

2.5.8.4 Organisational structure:

For a PAPDC, many household surveys still considered data entry and editing as activities to be conducted in a central location. Centralised data entry was the only known option before the emergence of microcomputers, and it is still used today in many surveys. It considers data entry to be an industrial process, to be conducted in centralised data-entry processing centers after the interviews have been completed. The objective of the operation is to convert PAPDC questionnaires into an intermediate product or machine-readable files. This needs to be further refined using editing programs and clerical processes to obtain a so-called "clean" database as a final product (Munoz, 2003).

Depending on the NSOs, the organisational structure, staff resources, and teams should be created and determined beforehand, when collecting household surveys. The roles, functions, and responsibilities of the various project teams within an NSO need to be outlined with the clear directives of responsibilities. All various internal organisational divisions or components need to be defined. The question is, for DDC, what will the organisational structure comprise, when conducting official household surveys?

2.5.8.5 Organisational process changes:

In a study conducted by Meyer (2008), it was found that amongst firms that had adopted new technologies the share of firms whose workplace organisation had changed in the last three years is higher than amongst firms that did not adopt these new technologies (J. Meyer, 2008). On the one hand, this can be a signal for the generally higher propensity to change and innovate in certain firms. On the other hand, it reflects the complementary relationship between ICT and workplace organisation. Process changes are inevitable in a rapidly changing corporate or business environment embracing new technologies.

Technologies are enablers of methodological principles and standards. What would the changes of processes concerning DDC as opposed to PAPDC be?

2.5.9 Technology adoption and acceptance

In the study conducted by Moussiegt et al. (2017), digitisation is occurring more rapidly with the adoption rates of ICT occurring much faster than in the past (see Figure 2.7). It took 35 years since the telephone was commercially available for 25% of the US population to adopt the technology. However, the diffusion of the internet and smartphones reached 25% in the US population after roughly seven years. Faster rates of ICT diffusion; increased diversity;

convergence of digital products and services means that countries should ensure they have the right policies in place to increase the adoption of technology. If not, countries might quickly fall behind. Policy-makers need to better understand the degree to which firms are investing in emerging (digital) technologies; and the need to identify the likely drivers of this digital use by firms (Moussiegt, DeStefano, & De Backer, 2017).

2.5.9.1 User interface:

Handheld digital devices, such as PDAs are small pocket-sized computers containing a visual display screen and input and/or output interface, like an external or touch-screen keyboard. Handheld computers, particularly smartphones and PDAs can handle surveys through applications available for survey collection. The use of handheld computers to replace paper questionnaires altogether is very appealing, because of the advantages of automating certain parts of the interviews, such as skipping instructions (Munoz, 2003).

In terms of the user interface or aesthetics, Hattas (2013) recommended that tabs on top of the screen for each person in a household should be incorporated in the software design. In this way, a question can be completed for all persons; and the next question can then follow. This would eliminate the repetition of the questions for all the persons, as well as that of the entire questionnaire per person (Hattas, 2013).

The design of user interfaces for mobile devices imposes several challenges regarding the interactive nature of surveys on mobile devices. Screen real estate is very small due to the limited physical size of handheld devices. Thereby stressing the design of graphical interfaces (Bergman, 2000); and forcing designers to explore the use of new means of output (Kjeldskov & Stage, 2002). Companies offer user-designer capabilities for DDC; however, on the managerial side, concerns about standardisation, tool selection, and cost prevail in many of the software options/packages available. Further issues are the availability of tool expertise, knowledge about tools, and learning the use of tools (Keller, 1995). How the interviewer (user) and respondent adopt to DDC is based on the experiences of the user interface and their belief in sound methodologies existing when using DDC.

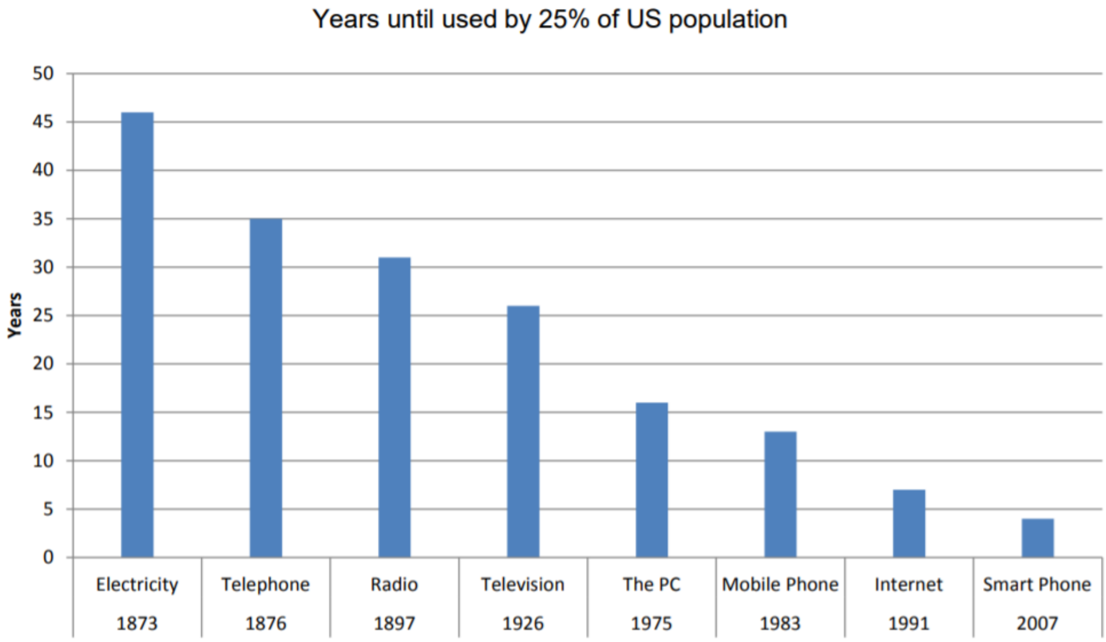


Figure 2.7: The Acceleration of Technology Adoption Over time

2.5.9.2 Technology acceptance:

The existing literature contains a rich source of technology acceptance theories; for example: Davis' (1989) Technology Acceptance Model (TAM) (Davis, 1989); Rogers' (2003) Diffusion of Innovation Model (Rogers, 2003); Venkatesh et. al (2003) Unified Theory Of Acceptance And Use Of Technology Model (UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003) and Venkatesh et. al (2012) extension of the Unified Theory Of Acceptance And The Use Of Technology (UTAUT2) (Venkatesh, Thong, & Xu, 2012). However, only a few of these studies have tested how their model fits among different demographic profiles of consumers, as in the case of Venkatesh, Morris, Davis, and Davis (2003). Venkatesh et al. (2003) integrated four main moderators (including age, gender, experience, and voluntariness) into the UTAUT.

It is therefore important to determine the level of acceptance by NSO staff to DDC, analysing their understanding of digital technology, the approval thereof, and expectations. Furthermore, the role of gender cannot be neglected, when studying technology adoption (Viswanath Venkatesh, Sykes, & Zhang, 2018). As an example it can be argued that studying gender differences in mobile adoption and use is crucial for the case of developing countries in general, and Arab countries in particular. There is great potential for mobile phone (smartphone) adoption in the Arab region. Studying technology adoption in the Arab region is important due to its large population of young people. This population group forms the highest segment; thus making the Arab region a significant market with great potential (Ameen, Willis, & Hussain Shah, 2018).

2.5.9.3 Access to digital devices

For DDC, access and the use of mobile devices are some of the key components in understanding the competencies of staff in NSOs in relation to embracing digital technologies. A higher access to digital devices increases the probability of users being more comfortable and accessible to digital devices.

2.5.9.4 Digital data-collection communications:

Communicating to staff and users of NSOs effectively on the integration or transition from PAPDC to DDC is important. DDC needs to inform all roleplayers on the effect it has on staff, the processes and methodologies.

2.5.9.5 Experience and ability to use digital handheld device technology:

The study conducted by Riddell and Song (2011) concluded that education increases the probability of using computers on the job. Employees with a higher education level have longer work experiences in using computers than those with a lower education level. However, education does not influence the use of computer-controlled devices, computer-assisted devices, or other technological devices, such as cash registers and sales terminals. Their study estimates were consistent with the view that formal education increases the use of technologies that require workers to carry out higher-order tasks. However formal education does not increase the use of technologies that necessitates routine workplace tasks (Riddell & Song, 2011).

The experience and ability to use DDC technologies hinges on the educational level, the staff level of experiences with DDC technologies, and their respective ability to use DDC.

2.5.10 Response rates

“Response rates” are often used as a measure of product quality; although they are strictly a measure of process, which is used as a proxy for the quality of the data. The higher the response rate, the less likely there is to be non-response bias in the survey results. Non-response is appreciable in most general-population surveys; however, it has been increasing over time (Smith, 2011). Often, non-sampling errors are not measured, but indicators of their likely impact on the bias or variances of the estimates are calculated. The most commonly used ones are the unit or item response rates. These indicators will indicate that bias exists and give some indication of its likely extent; but they do not measure the magnitude of the bias directly (Dobbs, Gibbins, Martin, Davies, & Dodd, 1991).

It is therefore critical for interviewers to ensure high response rates, regardless of the mode of collection. Modes of questionnaire administration are likely to affect the quality of the data collected. The effects appeared to be more marked between interview and self-administration modes, rather than within collection modes (Bowling, 2005). In terms of respondent preferences, while the biasing effects of the mode of preference are unknown, studies examining respondents' preferences report that people prefer face-to-face interviews to telephone interviews (Nicholaas G, Thomson K, 2000). On the otherhand, electronic self-completion questionnaires are preferred to paper self-completion questionnaires (Ryan JM, Corry JR, Attewell R, 2002).

The effects of the mode of collection on the response rates are to be determined and reported. This includes the overall actual response codes, for each sampled assignment within a survey. On the SASQAF, the response rates refer to the quality dimension of "Accuracy", where the assessment of the response rates and the estimates of the impact of imputation are to be calculated (Statistics South Africa, 2008; United Nations Department of Economic and Social Affairs, 2005):

"3.2 Measures of non-sampling errors are calculated, viz.: non-response errors:

- overall response rate
- item response rates
- unit non-response (e.g. weighted and un-weighted response rates)."

Furthermore, on the SVC, the response rate is located in the "collect" phase and reported during the "analyse" phase.

"6.4 Validate

- This sub-process is where statisticians verify the quality of the outputs produced, in accordance with a general quality framework. Verification activities can include:
  - checking that the population coverage and the response rates are as required".

2.5.11 Sound methodologies and common frameworks

"Methodological soundness" refers to the application of international, national, or peer-agreed standards, guidelines, and practices to produce good statistical outputs (Statistics South Africa, 2008; United Nations Department of Economic and Social Affairs, 2005). Adhering to a set of well-developed methodologies and common frameworks fosters local, regional, national and international comparability. NSO staff, in general, need to adhere to these standards in terms of methodology and common frameworks. If the data collection mode changes; NSO staff are required to adapt to these changes swiftly. For example a traditional census is a survey of households; and it collects information on several dimensions, namely: *Individual*, *Family*, *Household*, and *Dwelling*.

The first phase of the census is known as the enumeration or listing of dwellings. This is conducted to determine the number of households within a dwelling. Methodologically, this initial work is developing the population frame of households, with the secondary process of completing the census questionnaires, and determining the population of individuals. Thereafter a process for determining families is derived through a relationship to the head of the household. Official surveys, on the other hand, have often used this same methodological approach, built around a household. The household (which can be wider than the family) is seen as an economic unit, in which people live who can provide something in the way of shared income and consumption. The household is now commonly defined around shared living, shared meals, or suchlike. The household is important, in that members of the household are no longer independent. Collectively, they can rely on a shared or common income (Frank, 2014).

2.6 CHAPTER SUMMARY

This chapter identified that the global developmental goals of a country require data that are of good quality. Respective NSOs are usually responsible for country reporting obligations at local, national and international level (African Union Commission, 2015; UNDP, 2015). The quality of data is dependent on the quality management systems and frameworks that underpin and regulate how the data are collected through a full SVC. PAPDC is still a common method used by many NSOs, especially in Africa, for the collection of official household-based surveys. Household-based surveys are facing unprecedented challenges from both societal and technological changes (Mick P. Couper, 2017).

The increasing availability of digital handheld devices at lower costs with higher processing abilities assists NSOs to utilise the available opportunities to improve on their core mandate. The mandate is to collect good quality official statistics for their respective countries (M. P. Couper, 2005; Fitzgerald & Fitzgibbon, 2014; We Are Social, 2016).

Furthermore, this chapter dealt with questions and issues such as:

- who is responsible for collecting primary good quality data to measure social and developmental indicators for countries?
- what framework determines good “data quality” for official household surveys?
- some unique challenges are associated with data quality issues

- organisational permissions, security concerns (Hattas & Eloff, 2011)
- the current SASQAF for PAPDC not fully catering for DDC complexities, which alludes to issues of quality as pointed out and earlier on in this literature to review.
- linkages between DDC themes and SASQAF quality dimensions (see Table 2.7)

The following chapter describes the research methodology used to answer the research question.

Table 2.7: DDC Themes in relation to SASQAF Dimensions

Technological innovation in data collection - (DDC THEMES)	SASQAF QUALITY DIMENSIONS	Literature References
2.5.3 Cost	0. Prerequisites of quality 4. Accessibility	(Gibson et al., 2017; Kilic et al., 2017; Lukwago et al., 2013; Somda et al., 2009, 2010; Toscano et al., 2013), (Statistics South Africa, 2010b)
2.5.4 Accuracy, error rates and perception of quality	2. Accuracy 8. Integrity	(Statistics South Africa, 2010b)
2.5.5 Security	**	(Hattas & Eloff, 2011; Statistics South Africa, 2008)
2.5.6 Speed	3. Timeliness	(Frank, 2014; Statistics South Africa, 2008)
2.5.7 Training	**	(Hattas & Breytenbach, 2017; Munoz, 2003; Statistics Act, 1999)
2.5.8 Process flow and structural changes • Generic end-to-end DDC flow: • Access to data (internal): • Access to data (external): • Organisational structure: • Organisational process changes:	0. Prerequisites of quality 3. Timeliness (Generic end-to-end DDC flow: Speed) 4. Accessibility 6. Interpretability	(Hattas, 2013; Munoz, 2003)
2.5.9 Technology adoption and acceptance • User interface • Technology acceptance • Access to digital devices • Digital data collection communications • Experience and ability to use digital handheld device technology	2. Accuracy	(Ameen et al., 2018; Bergman, 2000; Davis, 1989; Kjeldskov & Stage, 2002; Munoz, 2003; Rogers, 2003; Venkatesh et al., 2003, 2012; Viswanath Venkatesh et al., 2018)
2.5.10 Response rates	2. Accuracy	(Dobbs et al., 1991; Smith, 2011; Statistics South Africa, 2008)
2.5.11 Sound methodologies and common frameworks	1. Relevance 5. Coherence & Comparability 7. Methodological Soundness	(Frank, 2014; Statistics South Africa, 2008)

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

3.1

Introduction

3.2

RESEARCH METHODOLOGY

3.2.1 The Research Process: Plan-Do-Study-Act Framework

3.2.2 Research philosophy

3.2.3 Research approach

3.2.4 Research strategy

3.2.5 Research choices

3.2.6 Research time horizons

3.2.7 Data-collection techniques and procedures

3.3

DATA ANALYSIS

3.4

Chapter Summary

Figure 3.1: Chapter 3 layout

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

3.1 INTRODUCTION

The previous chapter highlighted the quality assurance frameworks related to the collection of household survey and census data. The primary focus was on how NSOs developed best practices related to the art of official survey data collection nationally and globally. Building on the constructs of the previous chapter, this chapter discusses the research design and the methodology used to collect and analyse data for this study. Research design and methodology comprises the description of the paradigm, approach, design and the rationale for data collection that would enable the researcher to discover new knowledge.

"The research methodology is an important element of the research study required to clarify all the required steps, to achieve the objectives of the research" (Kilani & Kobziev, 2016). The basic purpose of describing the research methodology is that it "identifies the problems which need to be investigated and answered, to reach the set objectives in a research study" (Wedawatta, G., Ingirige, B. and Amaratunga, 2011). "It can be described as a method for collecting data and doing analysis; however, these methods have to be compatible with the research problems to get accurate and realistic results" (Harling, 2002). The research problem is located within the field of “data quality” for official household survey data being collected by NSOs. Changing the mode of collection from a paper-based to a digitally based method may (or may not) have an impact on the accuracy of the survey estimates. However, it is important to ensure that the household data collection for official surveys is accurate and can be trusted by all the users of the data.

The study aimed to update the existing Quality Assurance framework for official household survey data collection in South Africa, based on the objective findings from three iterations of data collection and is not solely based on the interpretations of various stakeholders. If DDC is being considered, what are the key quality considerations? Iteration 1 was, therefore, the first learning set from the Plan-Do-Study-Act (PDSA) cycle (see Section 3.2.1). After that, follow-up DDC iterations of increasing size or complexities were conducted namely: Iteration 2 and Iteration 3.

Where the interpretations did support the findings, they are presented in Chapters 5 – 7, as supporting data. Chapter 3 consists of three core sections. The first section (3.2) discusses the research process, including the philosophy, the approach used, the strategy, the choices, time horizons, and data-collection techniques, as well as the procedures used. The second section (3.3) elaborates on the data collection processes employed in the study. A brief discussion is done on each of the iterations undertaken as well as on the primary questionnaires used. The final section (3.4) explains the analysis component of the study. The research 'onion' developed by (Saunders, Lewis, & Thornhill, 2008) is used to guide the reader in terms of the selections employed in this study, the requisite research philosophies and approaches undertaken (see Figure 3.2).

The outer ring on the research 'onion' provides an additional overarching research process step employed during the study, and this is defined by the Plan-Do-Study-Framework (to be discussed in 3.2.1).

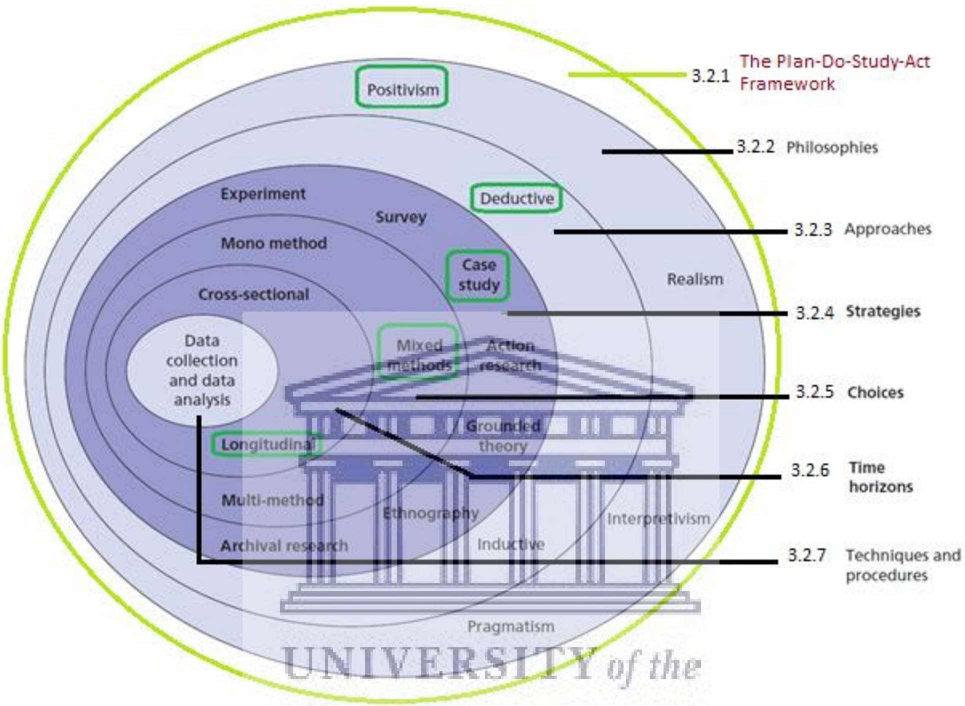


Figure 3.2: The Research Onion (Saunders, Lewis and Thornhill, 2008)

The research design, on the other hand, aims to provide a graphical presentation of the methods employed and the research structure. Data collection aspects, research questions and the sources of the data enables the researcher to outline all the required methods and tools used for the research in terms of selecting the most appropriate theoretical constructs for the research methodology. "The research design is a method of inquiry within different approaches to research, such as 'qualitative', 'quantitative', and 'mixed methods'. Mixed methods refer to a methodology of research that advances the systematic integration of quantitative and qualitative data within a single investigation or sustained program of inquiry" (Wisdom & Creswell, 2013).

Traversing between stages, i.e. from the one stage to the next stage within the methodology requires an action plan, which is also synonymous with the research design. Drawing the units of research clearly; it is very helpful to arrange the collected data in the right format. This provides the reader a map to understand the research; and it is particularly useful for other researchers to follow (R. Yin, 2003).

Exploratory vs Explanatory

To develop a proposed DDC developmental QA framework, this study was not found to be primarily exploratory. The study does not aim to design a framework from scratch. The study is therefore primarily explanatory. Since it is an enhancement from an existing PAPDC framework already in use and institutionalised by the South African NSO.

3.2 THE RESEARCH METHODOLOGY

3.2.1 Theoretical underpinning: Plan-Do-Study-Act Framework

The Plan-Do-Check-Act (PDCA) framework provides a methodical approach to problem-solving and continuous improvement. Dr. Walter Shewhart discussed the concept of the continuous improvement cycle (Plan-Do-Check-Act) in his original 1939 book entitled: *"Statistical Methods From the Viewpoint of Quality Control"* (Shewhart & Deming, 1986). Dr. W. Edwards Deming modified and popularised the Shewhart cycle (PDCA) to what is now referred to as the Deming Cycle (Plan-Do-Study-Act) (Weinstein & Vasovski, 2004). Dr. Walter A. Shewhart and Dr. W. Edward Deming advocated the PDCA concept for productivity management and continuous quality improvement of process and products.

PDCA is the “golden cycle for improvement”. It is a methodical approach for problem-solving and continuous improvement. The PDCA process is considered a never-ending cycle for improvement towards an ideal condition (see Figure 3.3). The “Plan” step is to establish the objectives, processes or countermeasures with expected outcome based on past performances or future forecasting of work. The “Do” step is to implement the processes or countermeasures planned. The “Check” step is to measure the effectiveness of the processes or countermeasures planned between the actual results and the expected results, to ascertain any differences. The “Act” step is to analyze the differences to identify the causes of the “Gap(s)”, and to take the necessary action to improve the changes. Deming again modified the Shewhart cycle in 1993 and called it the Shewhart Cycle for Learning and Improvement – the Plan-Do-Study-Act (PDSA) Cycle. He described it as a flow diagram for learning and improvement of a product or a process.

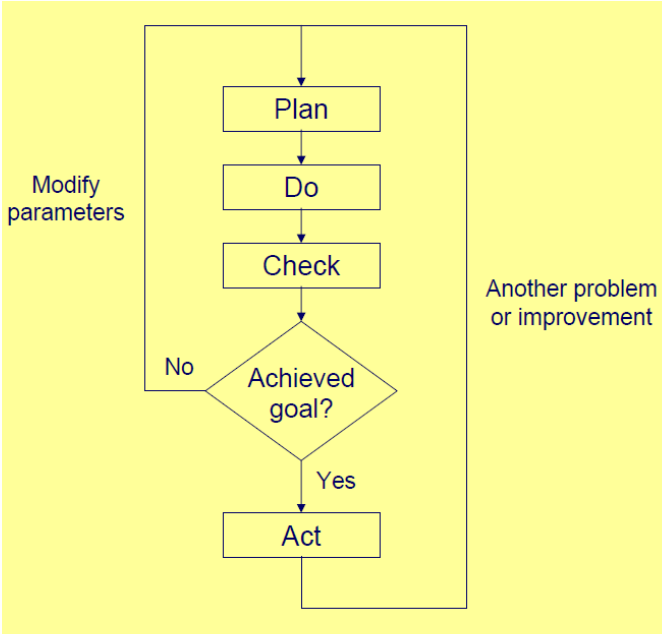


Figure 3.3: Plan-Do-Check-Act cycle (Weinstein & Vasovski, 2004)

Langley, Nolan, and Nolan added three basic questions to supplement the PDSA cycle, see Figure 3.4 (Ronald Moen, 2009). To aid the "Is", "Is Not" and "Therefore" processes, the PDCA/PDSA cycles were used as a guideline for developing a scientific approach towards quality or process improvement to be implemented during this research.

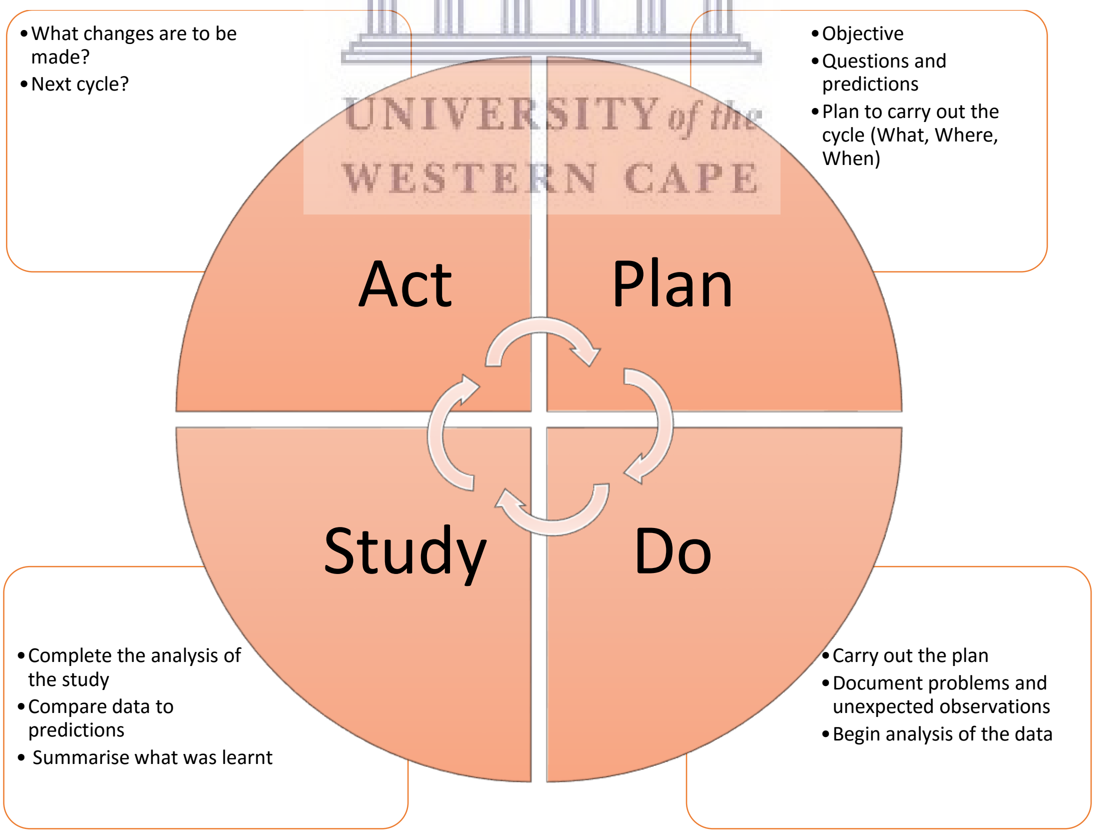


Figure 3.4: Plan-Do-Study-Act cycle (Ronald Moen, 2009)

3.2.2 Research philosophy

“The literal meaning of philosophy is the “love of wisdom or knowledge” (Greek philo- “loving” + sophia- “wisdom”). In its current incarnation, philosophy is typically seen as an academic discipline, part of the broader canon of the humanities. The student of philosophy will typically teach many of the sub-disciplines of philosophy including metaphysics, logic, ethics, aesthetics, epistemology, philosophy of language, political philosophy and others” (John Mingers & Stahl, 2018). All research is philosophy in action, and a lack of attention to an understanding of philosophy can render research and its outcomes misleading. Understanding philosophical questions, on the other hand, can help information systems (IS) researchers ensure that their work is rigorous and insightful (John Mingers & Stahl, 2018). It can also improve the quality of the work itself (Lee, 2004).

"*“Positivism”* is the view that social science procedures should mirror, as nearly closely as possible to those of the natural sciences. The researcher should be objective and be detached from the objects of the research. It is possible to capture ‘reality’ through the use of research instruments, such as experiments and questionnaires. Positivism has been a very predominant way of knowing the social world, which is referred to as the ‘received view’" (Guba & Lincoln, 2005). This can be seen by how many still perceive positivist approaches to be simply a sensible way of conducting research. While there are many varieties of positivism (Crotty, 1998), quantitative approaches that use statistics and experiments are seen as classic examples.

“Positivism emphasises the role of science, as the only method conducive to the truth” (Gonzalez, 2014). It claims that the social world can be described by law-like generalisations stemming from a collection of value-free facts (W. Chen & Hirschheim, 2004; Nandhakumar, J. & Jones, 1997). Positivist research can be identified by the presence of hypotheses, propositions, models, quantitative variables and statistical inferences of the “objective” data (Klein, H. & Myers, 1999).

Positivism is based on five pillars: the unity of the scientific method, the search for causal relationships (through reductionism), empiricism, value-free science, and the logical and mathematical foundations of science (Hirschheim, 1992). The assumption is that the truth is out there; and that it can be reached through the methods of science (Wynn, 2001). Extreme positivism in IS research sees technology as neutral, believes in rational management, ignores power relations and conflict, sees organisations as individual closed entities, and focuses on the business environment (Mitev, 2000).

“Chen and Hirschheim (2004) conducted an empirical study analysing eight major IS publications between 1991 and 2001; and found that the overwhelming majority of articles are based on a positivistic philosophy” (Carlsson, 2006). They examined 1 893 articles published in American and European journals and they found that, on a methodological level, while quantitative methods dominate the USA research culture (71%), 49% of the articles published in European journals apply qualitative methods (W. Chen & Hirschheim, 2004).

On the paradigmatic level, the vast majority of US publications (89%) are characterised by a positivist paradigm (Karthikeyan, P. & K.A. 2010). Facts, as opposed to impressions or subjectivities by the researcher, were given preference during this study. Given that IS are social in nature, it seems clear why many researchers consider positivism to be inadequate for IS research (Hirschheim, 1992). However, despite this criticism, it is still the dominant epistemology (W. Chen & Hirschheim, 2004).

"*"Post-positivism”* was derived from differences of opinions closely related to positivism. Post-positivism although similar to positivism with respect to beliefs, post-positivists argue that we can only know social reality imperfectly and probabilistically. While objectivity remains an ideal, there is increased use of qualitative techniques, to ‘check’ the validity of the findings. Post-positivism holds that only partially objective accounts of the world can be produced; since all methods for examining such accounts are flawed" (Denzin & Lincoln, 2005).

“Interpretivism” approaches to social research see interpretations of the social world as culturally derived and historically situated. Interpretivism is often linked to the work of Weber (Cook, 2012), who suggested that the social sciences are concerned with “understanding”. This is compared to “explaining”, which forms the basis of seeking causal explanations and is the hallmark of the natural sciences. Interpretivism has many variants. These include hermeneutics, phenomenology, and symbolic interactionism. “Critical”, as expected, is when critical social paradigms criticise both positivism and interpretivism, as ways of understanding the social world. The critical inquiry is not research that seeks merely to understand; it is research that challenges, that takes up a view of conflict and oppression; and that seeks to bring about the necessary change (Crotty, 1998).

“Post-modernism” is contradictory to other paradigms that offer grand theories for understanding the social world. It is an advocate of post-modernism. There was an argument that the era of big narratives and theories is over. Locally, temporally and situationally limited narratives are now required (Flick, 1998). Post-modernist approaches seek to overcome the boundaries that are placed between art and the social sciences. Post-modern approaches do not offer a view of rational progression to a better world. All we might expect is that social life

would, in some ways, be different. As with the other paradigms, there is a variety of positions within the broad label of post-modernism.

For this study, the researcher concurred that adopting the natural scientist approach entailed the process of “working with the observable social reality”. The hope is that the end product (or framework) in this research should be able to be generalised, similar to those made by physical and (or) that of the natural scientists (Remenyi, 2002). With this research approach of working with the observable reality, “positivism” was the research philosophy that best described the researcher’s way of conducting the study. Given that observed phenomena led to the production of credible data used for this study.

3.2.3 The research approach

Deductive reasoning is associated with quantitative research; and it uses a top-down process that tests the general premises through a series of steps to reach specific conclusions. Researchers seek to be objective through the research process, and they strive for generalisable findings by testing hypotheses through a deliberate series of steps. In contrast, inductive reasoning is associated with qualitative research; and it develops general conclusions, based on the exploration of how individuals experience and perceive the world around them (Wheeldon & Ahlberg, 2012).

Figure 3.5 provides some of the differences between deductive and inductive reasoning.

The deductive approach for this study was a dominant feature, given that the PAPDC QA framework (SASQAF) was already in existence.

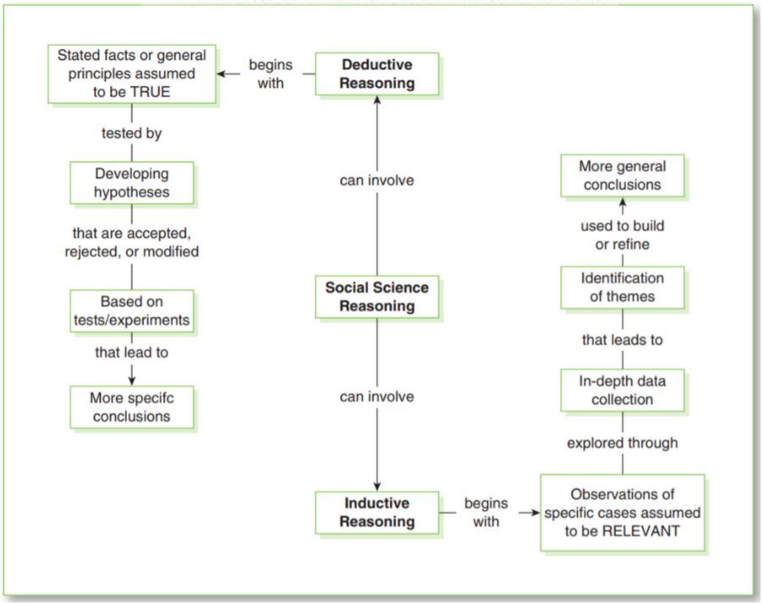


Figure 3.5: Deductive vs. Inductive reasoning (Wheeldon & Ahlberg, 2012)

3.2.4 The research strategy

Saunder et al. defined the research strategy as “the general plan of how the researcher should go about answering the research questions” (Zaidah, 2007). Research strategy, according to Wedawatta (Wedawatta, G., Ingirige, B. and Amaratunga, 2011), provides the overall direction of the research, including the process by which the research is conducted. Rowley (Rowley, 2002) argues that a chosen research strategy depends on the research questions, as well as the goals of the research. Wedawatta et al. (2011) mentioned that the research strategy selected is based on three conditions, which are (Wedawatta, G., Ingirige, B. and Amaratunga, 2011):

- The type of question,
- The control the researcher has over the behavioural events, and
- The focus is on contemporary issues, as opposed to, historical events.

The research strategy can be defined as the development of a research plan aiming to clarify the aims of the research. It enables the researcher to conduct research systematically, rather than haphazardly.

3.2.4.1. The Case Study

To generate a research strategy to collect the data by using a “positivistic” philosophy, the researcher is most likely to use existing theory to develop a hypothesis (Saunders et al., 2008). A “single case study” was employed as the research strategy. A case study can also run over multiple “design” iterations. It is preferable to cross-check the research aims and objectives by using various relevant data sources. Triangulation was used in this study to answer the formulated research question and to meet the objectives of this study. It should be noted that triangulation may possess a risk of taking on too many unfocused questions all at once unless there is sequencing involved and a sense of which technique is primary (Olsen, 2004).

The existing quality assurance framework for PAPDC, the SASQAF (Statistics South Africa, 2010b) will be updated to develop a proposed DDC developmental QA framework. APPENDIX I displays SASQAF LITE (a subset of the full SASQAF for PAPDC). Sections 3.2.5.2 and 3.2.5.3 describe the tools used for both quantitative and qualitative methods to be employed.

Yin (R. Yin, 2009), defined a case study “as an empirical inquiry that investigates a contemporary phenomenon within its real-life context. A case study is conducted when the boundaries between phenomenon and context are not clearly defined.” Furthermore, case studies are good for using mixed methods to validate or update models (or a quality framework) (R. K. Yin, 2013). Zakaria (2014), noted that when “how” and “why” questions are frequently posed, case studies are the preferred strategy (Zakaria, 2004). “In a research strategy, the case study is most commonly used, when the phenomenon cannot be divorced from its context. The case study depends on a qualitative approach as a method to be used in the example of an information system” (Iacono, J., Brown, A. and Holtham, 2009; Wedawatta, G., Ingirige, B. and Amaratunga, 2011).

Furthermore, Onatu (2013), states that “there are three reasons to choose a case study when research is conducted in the field of IS” (Onatu, 2013):

- The case study enables the researcher to study an information system in its natural settings and to generate theories from practice.
- The case study enables the researcher to answer “how” and “why” questions, and to consequently gain more explicit information.
- The case study enables the researcher to "better understand" the nature and the complexity of the process taking place.

Case studies can be classified into a “single” case or “multiple” cases. A single case provides all the required information about the research question from one organisation. In a single case, information and the data from one unit is enough to achieve the aims of the research. Multiple case studies require data from more than one unit to achieve the research objectives. Saunders et al. (2008) defined that a single case is often used where it represents a critical case, or an extreme or unique case (Saunders et al., 2008). They also noted that multiple cases used more than one case to investigate whether the findings of the first case occurred in the other cases also.

Yin (R. Yin, 2009), claims that a single case is more relevant if:

- It is a revelatory case.
- It represents a critical case for testing a formulated theory.
- It is a unique case.

On the other hand, Baxter and Susan (2008) state that multiple cases enable the researchers to study and compare the findings between different cases, and thus to explore differences within and between cases. This enables the researchers to forecast the results (Baxter, P. and Jack, 2008).

3.2.5 Research choices

In terms of possible research choices, the “mixed-method” (mixed-method research using quantitative and qualitative data) in combination with the use of both primary and secondary data sources, was used (Surbhi S, 2016) (See Figure 3.3).

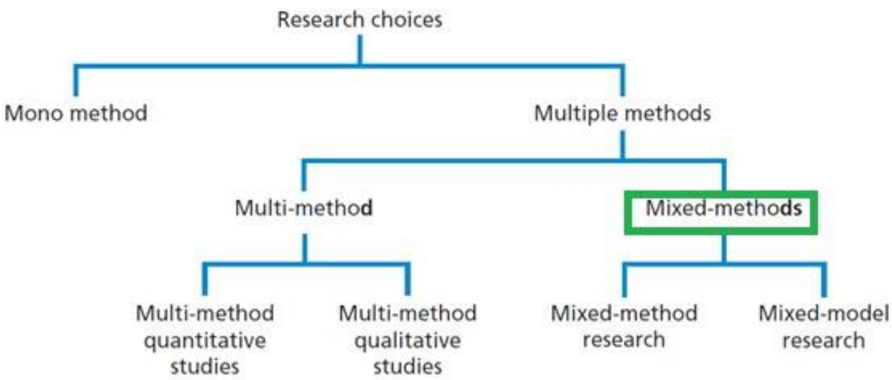


Figure 3.6: Research Choices (Saunders et al., 2008)

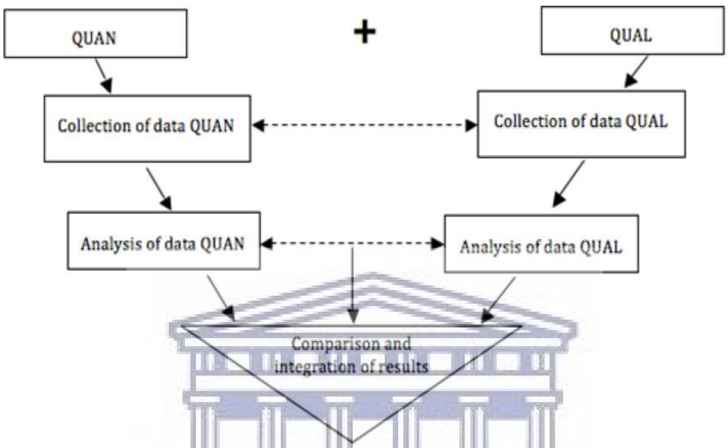
Babbie (2012) identified the research approach as the “systematic and orderly approach taken towards the collection and analysis of data, such that information can be obtained from those data”. It is very important to understand the aims of the research clearly, to determine and choose the appropriate technique(s) to achieve the research's aim(s) (N. Wright & Losekoot, 2012). During the selection of the research approach there are two factors which need to be considered: firstly, what were the characteristics of the topic and the time taken to conduct the research; and secondly, there are three dominant research approaches to be considered, either being a “quantitative” approach, a “qualitative” approach or a “mixed-method” approach (Creswell, 2003).

Both qualitative and quantitative data are used in this case study; and a mixed-method approach was selected, which aided the sampling strategy employed. Qualitative research is often found to be primarily exploratory research, given its intended use. In this study, qualitative research is used to gain an understanding of the underlying factors, reasons, staff opinions, or motivations; and it provides the initial insights into the problem. It helped to develop ideas or the hypotheses for potential quantitative research. Quantitative data were also used to analyse the data through the variables collected when using the appropriate statistical tools and/or techniques.

3.2.5.1. Mixed Methods

Mixed-methods research is a research approach that is widely accepted in IS; and it combines quantitative and qualitative research methods in the same research inquiry (Venkatesh et al., 2013). Considering the strength of the mixed-methods research for understanding and explaining complex organisational and social phenomena, there is a need for IS researchers to conduct and publish research that employs mixed methods (Cao, Crews, Lin, Deokar, Burgoon, & Nunamaker, 2006). The decision to use mixed-method research hinges on the research question, purpose, and context. Although the terms mixed methods and multi-method have been used interchangeably in the social and behavioural sciences, there are significant conceptual differences between the two. In multi-method research, the researchers employ two or more research methods; but this may (or may not) restrict the research to a single worldview (Cao et al., 2006; John; Mingers, 2001)c.

Mixed-method research, in contrast, uses quantitative and qualitative research methods, either concurrently (i.e. independent of each other) or sequentially (e.g. findings from one approach inform the other) to understand a phenomenon of interest (Zachariadis et al., 2013). Figure 3.7 displays the mixed-method approach, indicating the sequential/concurrent quantitative and qualitative integration of the data (Creswell, 2003).



Source: according to Creswell W. J., 2003, ***Research Design: Qualitative and Quantitative, and Mixed Approaches***, Thousand Oaks, CA: Sage.

Figure 3.7: Concurrent strategy of triangulation

In this study, the following questions were asked, and, as described below, some questions were answered by using the quantitative approach; while others lent themselves to be analysed by using a qualitative approach:

Step 1: From Chapter 2, if an NSO required household survey data to be collected digitally by using handheld mobile devices, are there Quality Assurance frameworks available to do so?

- Conduct research or review the literature to determine this
- Result: Chapter 2 revealed that a DDC QA framework does not exist

Step 2: Review the literature on various PAPDC quality assurance frameworks against the processes involved for DDC. In Chapter 2, a systematic search of the literature relating to official household survey quality dimensions was completed.

- Result: Inconclusive, therefore it necessitated the need to test the PAPDC QA frameworks against the processes involved for DDC

- Select the most appropriate PAPDC QA framework, namely: the SASQAF and map the dimensions of quality to the PAPDC statistical value chain (SVC)
- The mapping of the PAPDC processes indicated which areas of the SVC are directly affected by which quality dimension(s) within the SASQAF

Step 3: Would the standard SVC for PAPDC be affected by changing the mode of collection from PAPDC to DDC? If it is affected, in what way is it affected?

- Result: Research which core processes for DDC are required by iteratively identifying them. Grouping them into themes or subthemes and linking them to the relevant SASQAF dimensions. After that determine the gaps and/or identify improvement areas in the SASQAF to cater for DDC)
  - This was done by using key quantitative variables and measuring them against known benchmarks for PAPDC within the respective themes or dimensions for DDC
- Qualitative data provided further meaning to create a better understanding of the gaps and improvements identified by using the views and perceptions of experts or relevant field staff

3.2.5.2. The Qualitative Approach

The qualitative research approach is defined as those data that are formed from words, phrases, sentences, and narratives, instead of numbers (non-numerical data). Examples include: explanation, conversation, interviews, and discussion, which make the collected data qualitatively rich and holistic, with a strong potential for revealing the complexity, by focusing on problems in their social and cultural environments (Jebreen, 2012; N. Wright & Losekoot, 2012; Zakaria, 2004). The qualitative approach method provided explanations to explore a particular phenomenon, theory-building and capturing everyday life through data collection and analysis. This shifts the philosophical assumptions to the appropriate research design and technical data (Andrade, 2009; Zachariadis, M., Scott, S. and Barrett, 2013).

3.2.5.3. The Quantitative Approach

The quantitative approach includes the interpretation of the numerical data, such as percentages, interval or ratio, and using items of analysis such as graphs or diagrams to attain the results (Choy, 2014; Thorne, S., Kirkham, S.R. and O'Flynn-Magee, 2004). The quantitative approach can also be described as an extreme of empiricism, which depends on the control and explanation of the phenomenon. Quantitative researchers are more interested in measuring "how many?", "how often?", or "to what extent?" (Zakaria, 2004). Quantitative research in the analysis of data depends on statistical principles, contrary to qualitative research. Qualitative research is preferred when there is little research into the phenomenon that needs to be investigated, for it to be better understood.

In qualitative and quantitative approaches, there are various things to consider. Table 3.1 outlines the comparison between the two methods in terms of concepts, processes, and analysis.

Table 3.1: Comparison of quantitative and qualitative methodologies

Qualitative	**Quantitative**
1. It is often an inductive process and the language is informal.	1. It is a deductive process and the language is formal
2. Can be faster and cheaper compared with quantitative.	2. Can be relatively slow and more costly compared with qualitative.
3. Concepts are in the form of themes, motifs, and taxonomies.	3. Concepts are in the form of distinct variables.
4. The analysis proceeds by extracting themes or generalisations from evidence and organising data to present a coherent picture.	4. The analysis proceeds by using statistics, tables, or charts.
5. Procedures are particular and replication is difficult.	5. Procedures are standard and replication is assumed.

Source: (Bernard, 2000)

As noted earlier, deduction owes much to what we would think of as scientific research. It involves the development of a theory that is subjected to a rigorous test. As such, it is the dominant research approach in the natural sciences, where laws present the basis of explanation. It allows for the anticipation of phenomena, predicts their occurrence and therefore, it permits them to be controlled (Collis, J., Hussey, 2003). Examining and determining the factors affecting the development of a proposed DDC quality framework is deduced from the iterative cycles that will be mentioned in the next chapter.

The non-parametric Kruskal-Wallis Test (KWT) was used to analyse the three iterations. The KWT compares the mean ranks between the groups. Variables were ranked in terms of the order; after which, the values were converted into ranks and then computed. The assumptions on the data were the following:

- the data were not normally distributed,
- there were some outliers, or they
- failed the homogeneity test of variance; however, the data needed to pass the assumptions for KWT:
    - The distribution has to be similar for each variable across the group.

3.2.6 Research time horizons

Referring to the time dimension, snapshot vs. diary or (cross-sectional vs longitudinal), the case study is best suited to a “longitudinal” time scale. A longitudinal study is a study done over time; where the data on the dependent variable are collected more than once, to answer the research question(s). The primary strength of longitudinal research is the capacity that it has to study change and development over time (Adams, G. and Schvaneveldt, 1991). Saunders et al. (2008) point out that in observing people or events over time, the researcher can exercise a measure of control over the variables being studied, provided that they are not affected by the research process itself (Saunders et al., 2008).

3.2.7 Data-collection techniques and procedures

Data collection in the case study is one of the key processes when conducting research. It requires the most effective method to collect the data and to uncover all the details relevant to the research. Rowley indicates that the data collection and the analysis phases contribute significantly in supporting the research propositions; and he points out that there are three protocols, including the following (Rowley, 2002):

- an overview of the case study project;
- field-procedures, such as the use of different sources of information, and the access arrangements to these sources; and the
- case-study questions, or the questions that the case study researcher needs to keep in mind when collecting the data.

Researchers point out that there are different sources to collecting data, which are: interviews; documents and direct observations (Iacono, J., Brown, A. and Holtham, 2009; C. Meyer, 2001; Njie, B. and Asimiran, 2014; Sharma, G., Bao, X. and Peng, 2014). While Yin identified six methods of data collection for the case study; these **six** sources were (R. Yin, 2009):

- documentation
- archival records
- focus-group interviews
- direct observation
- participant observation
- physical artefacts

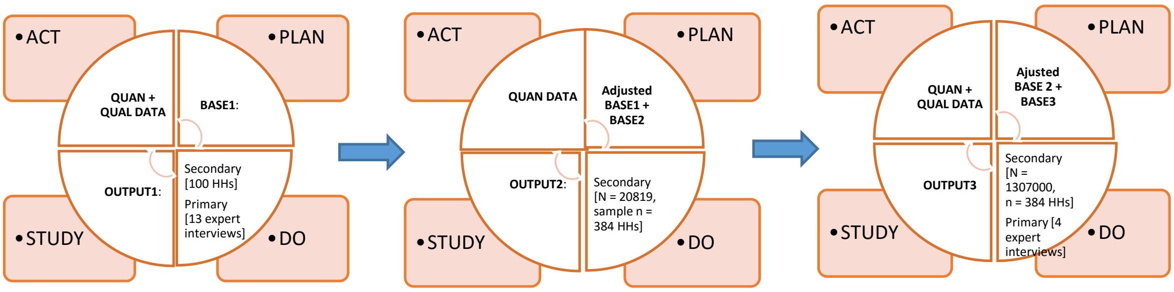
In this study, the following general data-collection methods were used:

- Interviews through a survey questionnaire (semi-structured) questionnaire to collect the primary data for analysis;

- Documentation;
- Participant observation through a survey questionnaire;
- Focus-Group Discussions / Debriefing;
- A secondary online questionnaire (to gather the secondary data for analysis)

The data were collected over three iterations:

Iteration 1- WC Test **Iteration 2**- KZN CSS **Iteration 3 -** CS 2016



The classification of the data utilised across three iterations was as follows:

Iteration 1 - Western Cape Test:

- DOCUMENTATION - Test Cases and Reports
- ARCHIVAL INTERNAL RECORDS - Quantitative Secondary Raw Dataset (100 records)
- QUESTIONNAIRE – Quantitative Primary Completed Questionnaire Responses
- FOCUS GROUP INTERVIEWS – Qualitative Primary Focus Group Debriefing Reports from 13 experts

Iteration 2 – KZN Citizen Satisfaction Survey (CSS):

- DOCUMENTATION – Qualitative Release Notes
- FOCUS GROUP INTERVIEWS – Qualitative Group Presentations on Challenges (What worked, what didn't work)
- ARCHIVAL INTERNAL RECORDS - Quantitative Secondary Raw Dataset (384 records)

Iteration 3 – Community Survey (CS) 2016:

- DOCUMENTATION – Qualitative User Views and communications
- QUESTIONNAIRE - Quantitative Primary Completed Questionnaire Responses
- ARCHIVAL INTERNAL RECORDS - Quantitative Secondary Raw Dataset (384 records)

- FOCUS GROUP INTERVIEWS – Qualitative Primary Focus Group Debriefing Reports from 4 experts

Other:

- ARCHIVAL EXTERNAL RECORDS – SASQAF Framework, SASQAF LITE Framework, Stats Act 1999
- ARCHIVAL INTERNAL RECORDS - Internal related documents (QA Procedures, Procedures/Policies (APPENDIX L), SANS 27002:2008 REFERENCES

Both qualitative and quantitative data were collected in a mixed-method design, which allows different but complementary data to be collected and analysed. By combining both qualitative and quantitative methods, one can address both exploratory and confirmatory questions in the same research inquiry (Venkatesh et al., 2013). The application of mixed-method research can result in rich insights into various phenomena to develop novel theoretical perspectives (Venkatesh et al., 2013; Zachariadis, M., Scott, S. and Barrett, 2013).

The researcher collected and analysed quantitative and qualitative data separately and then triangulated both sets of data to determine or explain proposals for the DDC developmental QA framework. Quantitative data were collected through survey questionnaires and analysed after the collection of the data. The qualitative data emerged from debriefings, observations, semi-structured interviews, test reports and documentary analyses throughout the three iterations.

3.2.7.1. Interviews

The interview can clarify all incomplete or unclear issues, by asking additional questions. This method enables the researcher to understand the interviewee, as well their point-of-view clearly; and the interviewee has flexibility and adapts the situation to each subject. The interviewer has to ensure that the interviewees were informed about the purpose of the interviews (Al-Azri, 2010). Yin states that one of the most important sources of case-study information is the interview (R. Yin, 2003). He suggested other names for interview forms, namely: open-ended, focused, and structured (survey) interviews. Open-ended interviews are two-way open communications, which enable the researchers to ask respondents for the facts, as well as their opinions. Focused (semi-structured) interviews aim to present the research questions purposefully.

The semi-structured interview has more flexibility than other types of interviews, and it can extract more detailed data from the participants and get fresh commentaries and significant insights. Besides, it is a useful method to access individuals to get their opinions, which cannot be sourced by formal questionnaires or open-ended interviews. Flexible questions have a better response than closed questions (Zakaria, 2004). The semi-structured interview is a discussion between two or more individuals, in which a set of questions are posed (Saunders et al., 2008).

In terms of selection for the forms of interview based on Saunders and Thornhill's (Saunders et al., 2008) data-collection techniques and procedures for this study included (Figure 3.8):

(1) Interviewer-administered questionnaires,

(2) Face-to-face interviews, and a

(3) Focus group

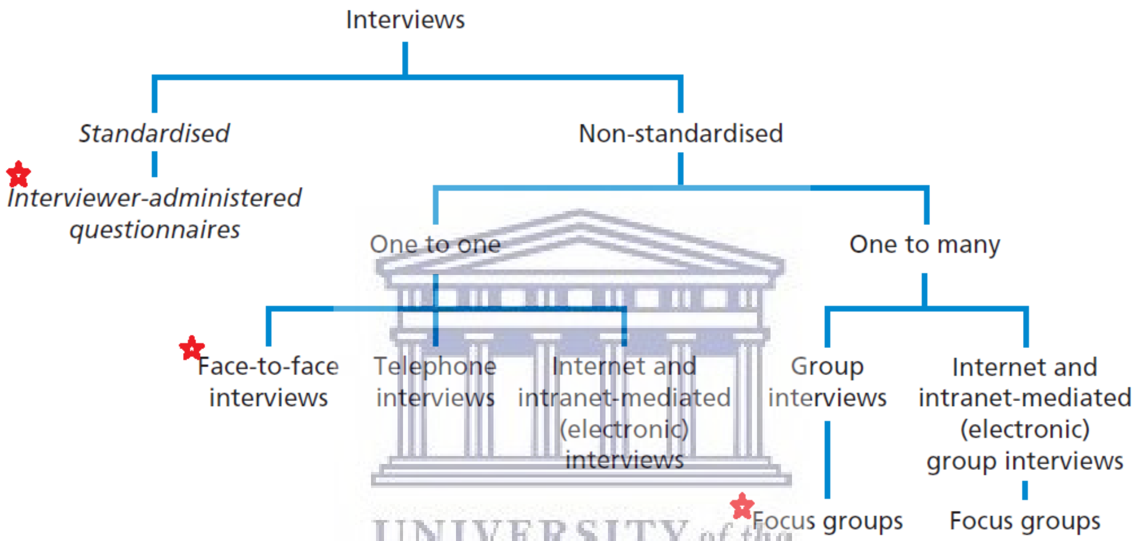


Figure 3.8: Forms of interviews (Saunders et al., 2008)

3.2.7.2. Documentation

Documentation is an important tool to most schemes in social research, because of its location in historical circumstances; also, it has the possibility of repeated reviewing. The document is a way of collecting data and can be an essential source of any investigation or an alternative method of collecting information. Instead of only relying on a questionnaire survey, interview and/or observations; documentation should be used to supplement the research undertaken (Zakaria, 2004). Three reasons make documents one of the most useful methods of data collection (C. Meyer, 2001):

1. They are used as inputs to the interview guide and they save time.
2. They are useful for tracing the history of the organisations and the statements.
3. They help counteract any bias in the interviews.

Documents provide the possibility to supplement any needed information from other resources and to highlight new ideas and insights that need further investigation by a researcher (R. Yin, 2003).

3.2.7.3. Questionnaire

The questionnaire is a research instrument containing a set of questions, which prompts and allows the researcher to gather information from the respondents. The questionnaire is one of the most widely used tools of data collection. Initially, it was designed for statistical analysis and the primary advantage of the questionnaire is the flexibility of collecting these data. However, more recently it is being used in other types of analysis. Nicholas points out that a questionnaire is a flexible tool that enables researchers to organise the questions and to receive replies from respondents, without needing to talk with each one separately (Walliman, 2001). The questionnaire is helpful, but it must take into account the following points:

- Careful design of questions,
- The questionnaire must be easy to understand and uncomplicated,
- The purpose of the questionnaire must be clear,
- Language and terminology must be easy to use and understand.

ITERATION 1: Western Cape Test (Small sample mini-test: 100 households)

Primary data was collected from 13 experts, they were permanent national and provincial Stats SA staff members (viz. methodologists, monitoring and evaluation, provincial directors, provincial senior staff and other professionals) through the use of a survey questionnaire (see APPENDIX B) for the Western Cape province's CAPI WC pilot held in August 2015. The purpose of this survey questionnaire was to examine the use of digital devices for official household survey-data collection by Statistics South Africa staff members. It contained the following themes:

- Background information
- Expectations about digital handheld device technology (DHDT) and experience thereof
- Equipment access
- Data quality
- Best practices in official household survey-data collection
- Pilot experiences for DDC

Secondary data (100 sampled households) were collected, and 100% of the data were analysed to determine the responses and the non-responses (item and unit non-responses).

Iteration 1 is not a representative sample of the Western Cape. The sample size is relatively small (100). It contains a 95% Confidence Level, the Confidence Interval (precision) = 9.8%, where 50% unknown population proportion. For it to be more precise, the researcher is required to collect at least 384 dwellings for a 95% Confidence Level, 5% Confidence Interval, 50% unknown population proportion.

ITERATION 2: KZN Citizen Satisfaction Survey (Large sample survey: approximately 20 819 households).

The primary data was collected from articles, publication data, and debriefing information post Iteration 2. The reason for not collecting similar data from NSO experts was the fact that they were incorporated in Iteration 2. Repeating the same primary questionnaire would result in unduly duplication and risk respondent fatigue or non-response.

Secondary data (20 819 sampled households) were collected; and a sample of 384 for a 95% Confidence Level, 5% Confidence Interval, 50% unknown population proportion was analysed. The data was further analysed to determine the responses and the non-responses (item and unit non-responses). Permission to use and analyse the data for iteration 2 was received from the head of department (Stats SA) – see APPENDIX S. Iteration 2 was a representative sample of KZN Province and the sub-sample (n) selection process was as follows:

- Sort primary data file by: Main Place → Sub-place → EAQN (see Definitions and Glossary)
- Create variable “Random numbers”, use the formula RAND() in Excel and fill all rows (1 – 20 189)
- Create a variable “Random fixed”. Fill the column by pasting special “Random numbers”.
- Create variable “Record numbers” and fill the column with integers [1-20 819]
- Sort Random number (random fixed) from the smallest to the largest.
- Sort variable “Random fixed” from the smallest to the largest.
- Select an appropriate sample (384).
- Select the first 384 records.

ITERATION 3: Community Survey 2016 (Large sample survey: approximately 1.37 million households)

The primary data was collected from 4 expert district staff members based in the Eastern Cape Province using the survey questionnaire (see APPENDIX B). The purpose of this survey questionnaire was to examine the use of digital devices for official household survey-data collections by Statistics South Africa staff members. It contained the following themes:

- Background information,
- Expectations about digital handheld device technology (DHDT) and experience thereof,
- Equipment access,
- Data quality,
- Best practices in official household survey-data collection,
- CS 2016 experiences for DDC.

Secondary data (1 370 809 sampled households) were collected, which comprised a representative sample of South Africa. A sample of 384 for a 95% Confidence Level, 5% Confidence Interval; a 50% unknown population proportion was analysed to determine the responses and the non-responses (item and unit non-responses). Permission to use and analyse the data for iteration 3 was received from the head of department (Stats SA) – see APPENDIX S. Other secondary data sources included Lessons Learnt during CS 2016.

The following appendices related to the primary data-collection tools are:

- APPENDIX C: FOCUS-GROUP DISCUSSION GUIDE
- APPENDIX D - F: INFORMATION SHEET FOR PARTICIPANTS IN (FOCUS GROUPS, RESEARCH QUESTIONNAIRE / SURVEY, SEMI-STRUCTURED INTERVIEWS)
- APPENDIX G: LETTER OF CONSENT FORM (FOCUS GROUPS OR RESEARCH QUESTIONNAIRE / SURVEY OR SEMI-STRUCTURED INTERVIEWS)

In summary, Table 3.2 depicts the questionnaires and the other data tools used to elicit the responses from the relevant staff members participating in the iterations.

Table 3.2: The collection tools utilised in the study with appendix references

Data Source	Title	Appendix	Origin	Sample size	Method
Iteration 1 - Primary data	PHD RESEARCH SURVEY QUESTIONNAIRE	B	Compiled by author	n = 13	Paper-based
	FOCUS GROUP DISCUSSION GUIDE	C	Compiled by author	N/A	Paper-based
	INFORMATION SHEET FOR PARTICIPANTS IN FOCUS GROUPS – (DDC QUALITY FRAMEWORK DEVELOPMENT)	D	Compiled by author	n = 13	Paper-based
	INFORMATION SHEET FOR PARTICIPANTS IN RESEARCH QUESTIONNAIRE/SURVEY	E	Compiled by author	N/A	Paper-based
	INFORMATION SHEET FOR PARTICIPANTS IN SEMI-STRUCTURED INTERVIEWS	F	Compiled by author	N/A	Paper-based
	LETTER OF CONSENT FORM [FOCUS GROUP OR QUESTIONNAIRE OR SEMI-STRUCTURED INTERVIEWS]	G	Compiled by author	N/A	Paper-based
Iteration 1 - Secondary data	WC Test Survey Solutions Questionnaire	N/A	Questionnaire edited by the author and captured by staff	n = 100	Digitally collected using a handheld device
Iteration 2 - Primary data	Articles Publication data Debriefing information	N/A	NSO archives	N/A	Post iteration 2 reports
Iteration 2 - Secondary data	KZN Citizen Satisfaction Survey 2015 Survey Solutions Questionnaire	N/A	Captured by 217 Contract fieldworkers	Initial Dwelling Sample N = 20 819 A sample of n= 384 dwellings was analysed	Digitally collected using a handheld device
Iteration 3- Primary data	CS 2016 RESEARCH SURVEY QUESTIONNAIRE	B (excluding variable: Are you currently working (full-time/part-time)	Compiled by author	n = 4	Paper-based
Iteration 3 - Secondary data	CS 2016Household_QN Survey Solutions Questionnaire	N/A	Captured by 11 245 Contract fieldworkers	Initial Dwelling Sample N = 1 370 810 A sample of n = 384 dwellings was analysed	Digitally collected using a handheld device

3.3 DATA ANALYSIS

Data quality through data collection processes plays an instrumental role when analysing data. The data from all the questionnaires received were recaptured and verified. The data collected were checked for correctness, including ranges checks, valid and completed responses. The use of existing questionnaires or the compilation of new research questionnaires was done through literature surveys of the current and past research conducted in the specific fields of the study discussed in Chapter 2. The validity and reliability of the research instruments in quantitative and qualitative research need to be established (Saunders et al., 2008). The basic goals of the design of the questionnaires were, firstly, to obtain information relevant to a research study; and, secondly, to collect the data with maximum reliability and validity.

The researchers indicated that according to the field of IS, validity and reliability must be achieved in all research studies.

3.3.1 Quantitative Sample Size Calculation

The sample size is a statistical concept that involves determining the number of observations or replicates selected; (i.e. the repetition of an experimental condition used to estimate the variability of a phenomenon) that should be included in a statistical sample. It is an important aspect of any empirical study, requiring that inferences be made about a population, based on a sample. Essentially, sample sizes are used to represent those parts of a population chosen for any given survey or experiment. To carry out this calculation, set the margin of error, ε, or the maximum distance desired for the sample estimate to deviate from the true value. To do this, use the confidence interval equation below; but set the term to the right of the ± sign equal to the margin of error, and solve for the resulting equation for sample size, n.

The equation for calculating the sample size is shown below.

$$n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2}$$

$$n = \frac{1.96^2 \times 0.5(1-0.5)}{0.05^2} = 384.16$$

The variable or measurement distributions found between the full dataset and a selected sample remained consistent for both Iteration 2 and Iteration 3 (see Table 3.3 and Table 3.4 respectively). This implies that a sample of 384 records, randomly selected from the full datasets for Iteration 2 and Iteration 3 would yield similar distributions on all key measurements. It should be noted that in Table 3.3 and Table 3.4, "N" denoted the count of responses for the measurement being analysed, and not the sample size "n" of 384.

Table 3.3: Iteration 2 full dataset (21 106) vs. Sample (384)

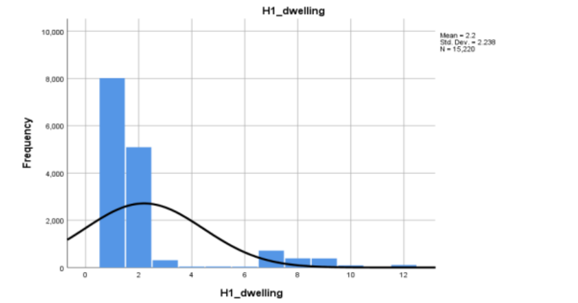
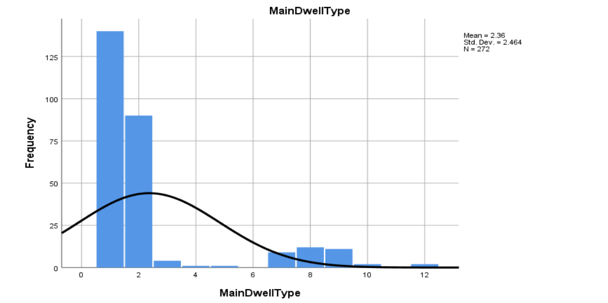
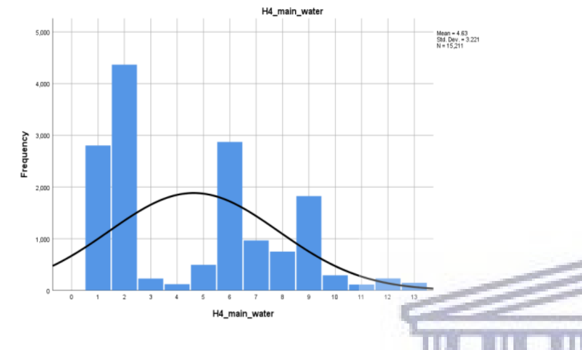
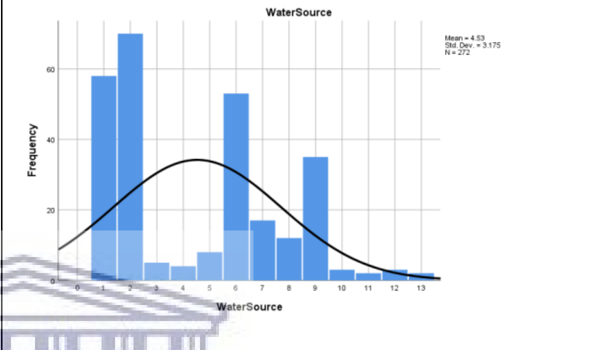
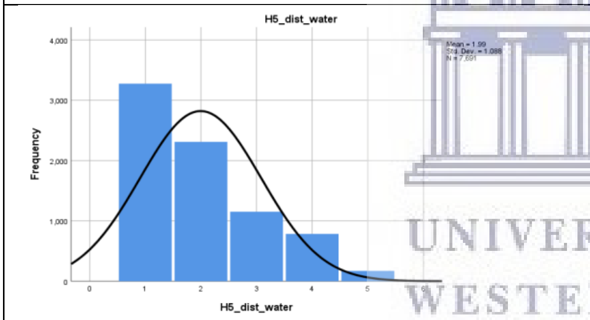
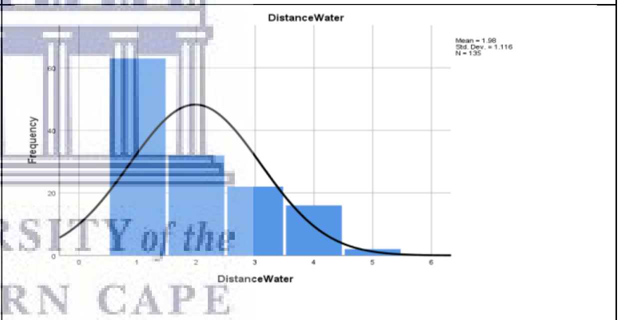
<table>
<tr><th>Iteration 2 - Full Data Set (21 106) including OOS
N (20819) + Out of Scopes (287) = 21106</th><th>Iteration 2 - Sample (384) including OOS
(NB: variables were renamed in the sample to Iteration 3 variable names for later comparative analysis across iterations)</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Table 3.5 below displays the various sample sizes selected for the three iterations of the study. If the confidence interval were to be adjusted to 1%, then the distributions would have remained the same.

In qualitative studies different methods can be used to address both validity and reliability, which includes the triangulation of information among different sources of data, receiving feedback from the informants (i.e. member checking), and expert reviews (Simon, 2011). On the other hand, Meyer (2001) argues that, in his studies, most qualitative researchers focus on the findings, rather than explaining or describing how these results were collected (C. Meyer, 2001).

Table 3.4: Iteration 3 full dataset (approx. 1.3 million) vs. Sample (384)

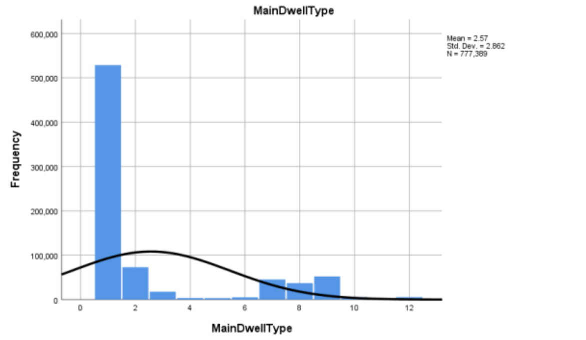
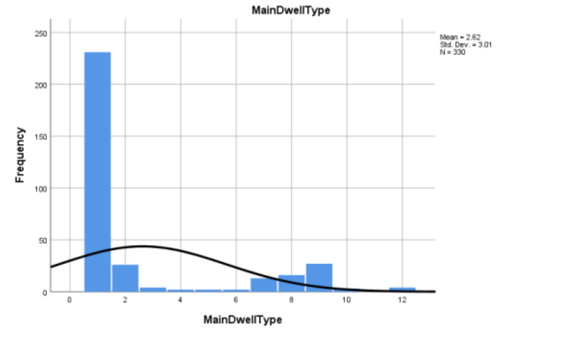
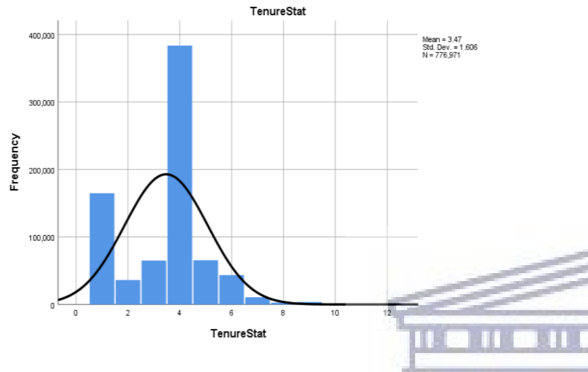
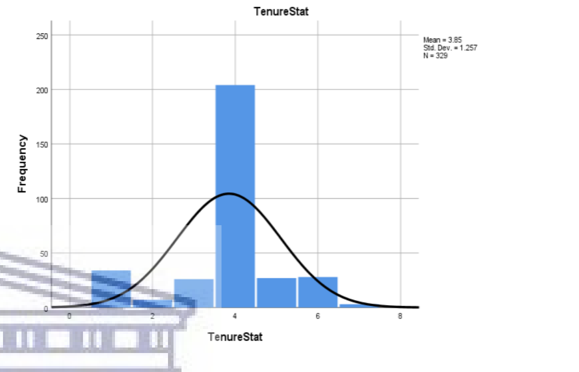
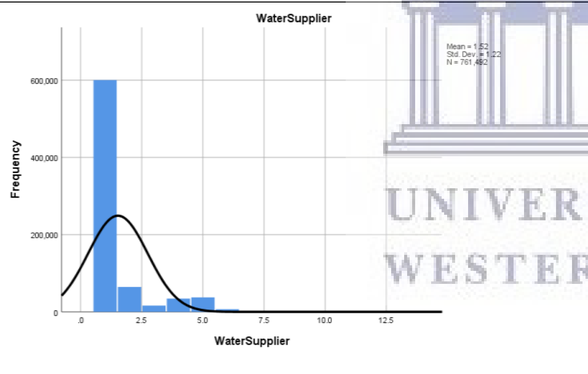
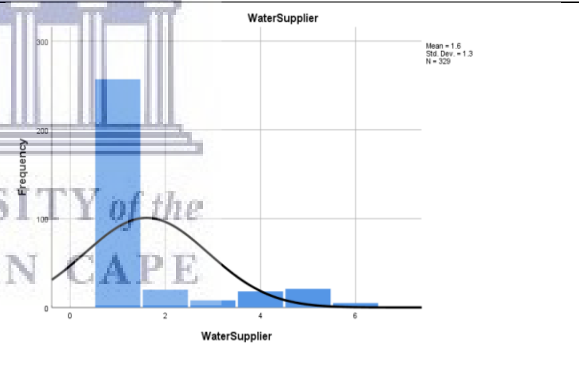
Iteration 3 - Full Data Set (1.4 mill / 887116) including OOS	Iteration 3 - Sample (384) including OOS
	
	
	

Table 3.5: Quantitative Iteration Sample Sizes [1,2,3]

Confidence Level (Proportion within range)	Confidence Interval (Precision)	Population Proportion (unknown)	Sample Size	Iteration 1	Iteration 2	Iteration 3
95%	1%	50%	9604			
95%	5%	50%	384		x	x
95%	9.8%	50%	100	X		
99%	1%	50%	16641			
99%	5%	50%	666			

The validity and reliability can be achieved when there is a clear explanation of the techniques used to collect the required data. Enabling the interviewee to assemble documentation that can support the interview increases the validity and reliability of data collected. The validity and reliability can also be increased depending on the careful design of the individual questions, the clear and pleasing layout of the questionnaire, the lucid explanation of the purpose of the questionnaire, pilot testing as well as carefully planning and execution of administration (Saunders et al., 2008; Simon, 2011).

3.4 CHAPTER SUMMARY

This chapter discussed the research design and the methodology used for the study. Firstly, by identifying and motivating the appropriate research philosophy; after which, the corresponding approaches, strategies, analytical choices, time-horizons, and techniques employed was discussed. The study is a positivist, single case study whereby the time horizons spanned over 3 iterations. The iterations contained both qualitative and quantitative data, thereby indicating that the most appropriate choice deployed was the mixed-method approach. There are various methods of research methodologies, which have been explored by researchers in the field of IS. Each method has its assumptions and approach to data collection, and it has been processed to achieve the objectives of the research. The research methodology centers on the ability to choose the appropriate method for data collection and analysis. On the other hand, research methodology can be defined as a strategy to focus on the research problems and the development of appropriate solutions. In the next chapter, the iterative cycles of DDC and its relation to aspects of the current SASQAF and SVC will be identified and discussed.

CHAPTER 4: THE STATISTICAL VALUE CHAIN (SVC) MAPPED TO THE SOUTH AFRICAN STATISTICAL QUALITY-ASSESSMENT FRAMEWORK (SASQAF)

4.1

Introduction

↓

4.2

A CONCEPTUAL FRAMEWORK

↓

4.3

THE GENERIC STATISTICAL VALUE CHAIN – SVC PHASES

4.3.1 Need

4.3.2 Design

4.3.3 Build

4.3.4 Collect

4.3.5 Process

4.3.6 Analyse

4.3.7 Disseminate

4.3.8 Archive

4.3.9 Evaluate

↓

4.4

DIGITAL DATA-COLLECTION THEMATIC AREAS

↓

4.5

Chapter Summary

Figure 4.1: Chapter 4 layout

CHAPTER 4: MAPPING OF THE STATISTICAL VALUE CHAIN (SVC) AND THE SOUTH AFRICAN STATISTICAL QUALITY ASSESSMENT FRAMEWORK (SASQAF)

4.1 INTRODUCTION

This chapter defines the draft design guidelines and the conceptual framework following the deductive research approach. A mapping of quality dimensions and processes between the SVC and the SASQAF for PAPDC was completed during this chapter. This was to gain insight or a basic understanding of the interrelationships and importance of the broader quality processes involved for official survey collection. The literature discussed in Chapter 2, concerning quality themes, is analysed in conjunction with the SVC to identify linkages through a mapping process between SASQAF and the SVC.

4.2 A CONCEPTUAL FRAMEWORK

The design of a conceptual framework assisted in guiding the study. The framework incorporates a number of existing frameworks, such as the iterative PDSA cycle (Shewhart & Deming, 1986) for process improvements across iterations and the mixed-method approach (Creswell, 2003), indicating the sequential/concurrent quantitative and qualitative integration of data through the process of triangulation (Bentahar & Cameron, 2014). The conceptual framework was constructed to guide the design of a proposed DDC developmental QA framework for management in NSOs to consider when conducting a DDC survey. The set of design themes emerged from a review of the literature in Chapter 2, and a process was followed to systematically review the findings and extract the information.

Figure 4.2 displays the overarching conceptual framework; the first stage addresses the problem statements, the purpose of the research, the literature of themes related to quality. The SASQAF and the corresponding SVC describe the bounds by which NSOs operate to collect official data for household surveys. After the collection and the analysis of iterative data with an increasing sample size, a clear, coherent and integrated presentation of the results is required. This is accomplished by combining the qualitative and quantitative data by using triangulation.

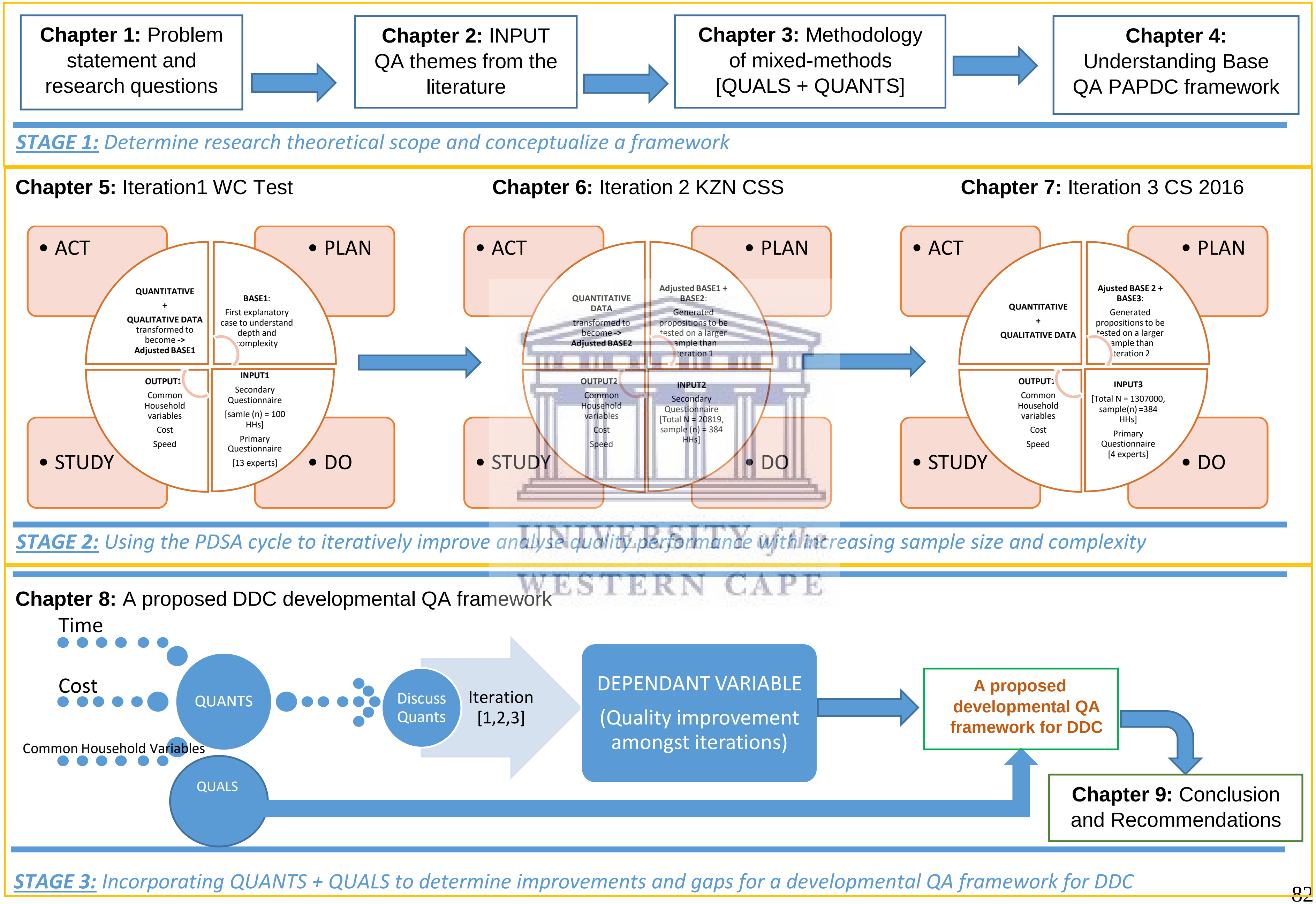


Figure 4.2: Conceptual Framework

4.3 THE GENERIC STATISTICAL VALUE CHAIN – SVC PHASES

A generic statistical value chain (SVC) was adapted by Statistics South Africa in 2009 (Figure 4.3). The joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS) developed the SVC based on a range of statistical operations enabling various support functions within the development of statistical surveys (Statistics South Africa, 2010b). It should be noted that DDC is briefly referred to in the PAPDC SASQAF during the "build" phase. It indicates that DDC is one of the options as a collection instrument, which can also include amongst other collection options such as Computer-Assisted Personal Interviewing (CAPI), paper and pen data collection (PAPDC) questionnaires, administrative data and other data integration techniques (Statistics South Africa, 2010b). However, the current SASQAF framework does not fully address all procedural and/or methodological aspects, when considering collecting the data by using the DDC process. It is therefore important to identify the key components of the SVC; and how they are mapped to the SASQAF.

The 9 phases (level 1) of the SVC contain various sub-processes within (level 2); see Figure 4.3. The SVC can be divided into four levels (Statistics South Africa, 2010b):

- Level 0, the statistical business process;
- Level 1, the nine phases of the statistical business process;
- Level 2, the sub-processes within each phase;
- Level 3, a description of those sub-processes.

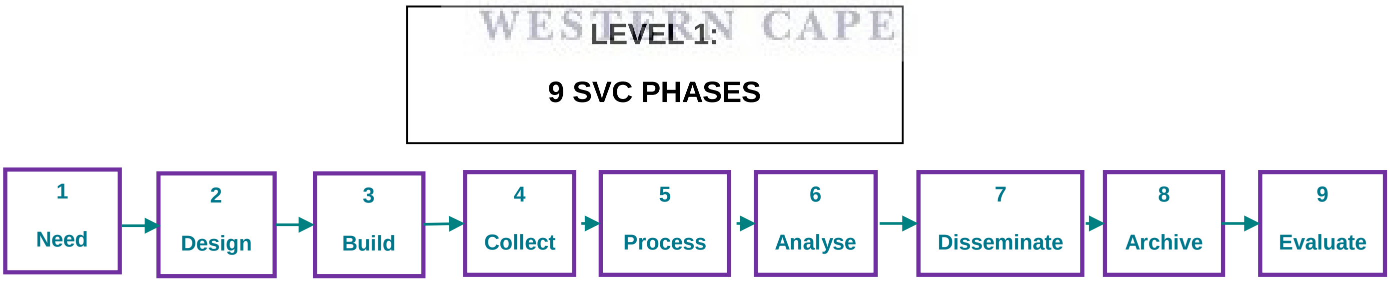


Figure 4.3: Generic Statistical Value Chain – Level 1

Figure 4.4 below displays the various SVC phases. The first level (Level 1) includes the high-level SVC phases and the second level (Level 2) lists the various sub-processes which is included in part one (Part 1). The sub-processes are described in detail in part two (Part 2) . The SVC also contains two overarching processes that apply throughout the nine phases and across the statistical business process, which are the "quality" management and "metadata" management. Each phase and corresponding sub-processes is discussed below (Statistics South Africa, 2010b):

Part 1: Levels 1 and 2 of the Statistical Value Chain

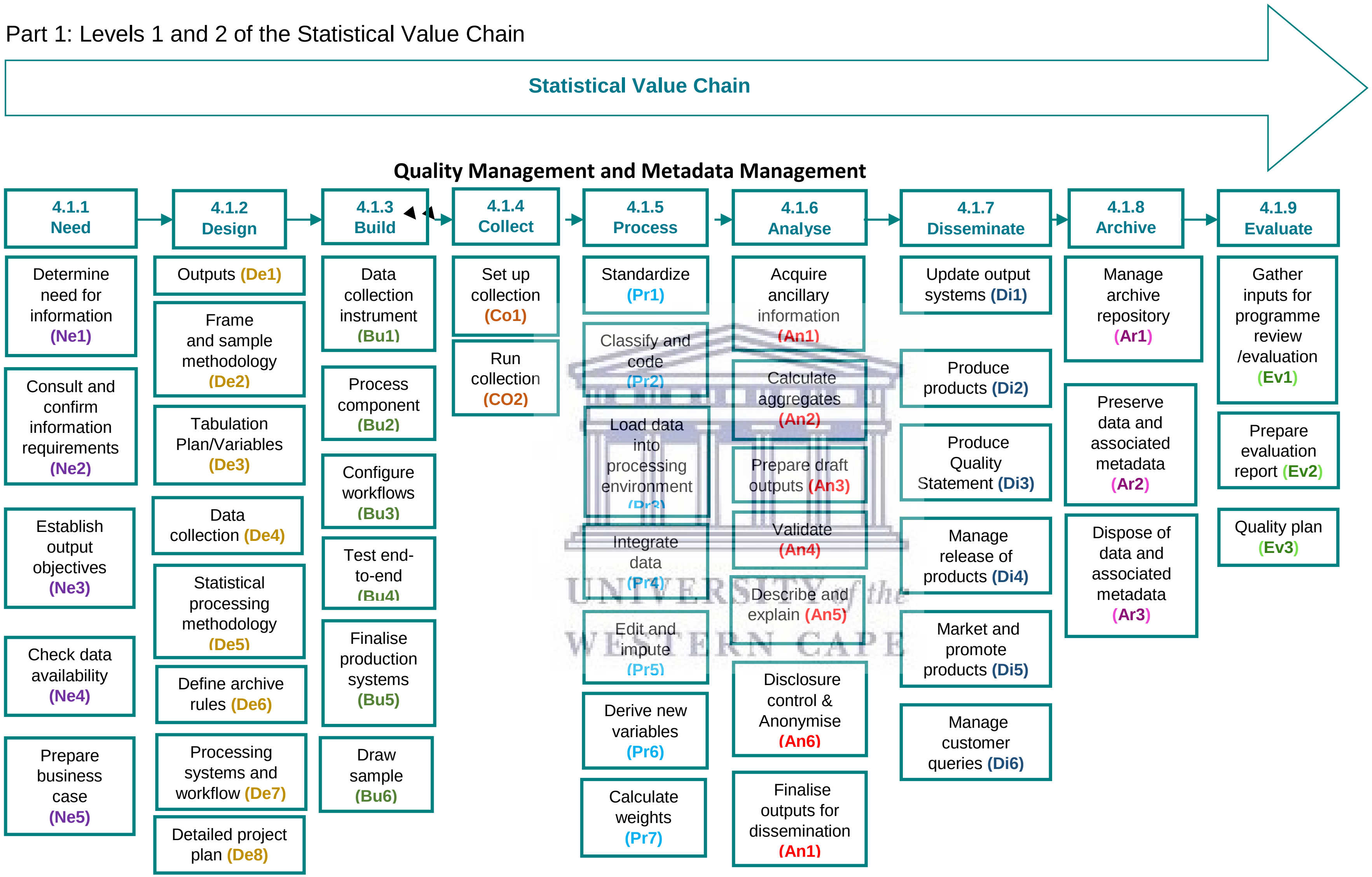


Figure 4.4: Generic Statistical Value Chain – Level 1,2,3 Source: (Statistics South Africa, 2010)

4.3.1 Need

The "**Need**" phase describes the development, design activities and any associated practical research work required to define the statistical outputs, concepts, methodologies, collection instruments, and operational processes. It consists of the following activities, or sub-processes (see Table 4.1):

Table 4.1: Phase 1 of SVC (Need) (Statistics South Africa, 2010b)

PHASE (Level 1)	SUBPROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
1. NEED	1.1 **Determine** the requirements of **need** for information	Ne1	Conduct an initial investigation and identification of what statistics are required and what is needed of the statistics from the users.
	1.2 Consult users and **confirm the need** for information	Ne2	Consulting with relevant stakeholders (users and interested parties) and confirming in detail the need for the statistics.
	1.3 Establish statistical **output objectives**	Ne3	Identify the statistical outputs that are required to meet the user needs to be identified in sub-process 1.2 (i.e. consult and confirm need).
	1.4 Check the **data availability**	Ne4	Check if current data sources available could meet user requirements and the conditions under which they would be available, including any restrictions on their use.
	1.5 Prepare a **business case** of information required	Ne5	Document the findings of the other sub-processes in this phase in the form of a business case. Include an assessment of costs and benefits, as well as any external constraints.

In terms of the mapping of the SVC to SASQAF, the "Need" phase spans 5 dimensions in SASQAF; and it is related to 13 indicators, of which 21 sub-processes of the SVC are affected for PAPDC (see Table 4.2). The 21 co-joint mapping pairs in the NEED phase are denoted as in the form:

- (PRER1; Ne1);
- (RELE1; Ne1);
- (INTE5; Ne3);
- …. etc., as well as all the other co-joint mapping pairs in the corresponding Level 1 phase below.

Table 4.2: SASQAF vs. SVC (Need)

SASQAF Quality Dimensions			Statistical Value Chain				
			Need				
Number	Dimension	SASQAF Code	Ne1	Ne2	Ne3	Ne4	Ne5
0	Pre-requisites of quality	PRER1	X				
		PRER2	X				
		PRER3			X		
		PRER4			X		
1	Relevance	RELE1	X				
		RELE2	X				
		RELE3			X		
		RELE4			X		
		RELE5				X	
3	Timeliness	TIME3	X	X	X	X	X
		TIME4				X	
4	Accessibility	ACCE1				X	
8	Integrity	INTE5	X	X	X	X	X

4.3.2 Design

The "Design" phase describes the development and design activities of survey development. The collection instruments, concepts, methodologies and operational processes related to statistical outputs are defined. It consists of the following activities or sub-processes (see Table 4.3) (Statistics South Africa, 2010b):

Table 4.3: Phase 2 of SVC (Design) (Statistics South Africa, 2010b)

PHASE (Level 1)	SUB-PROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
2. DESIGN	**2.1 Outputs**	De1	Includes the related development work and preparation of the systems and tools used in the "Disseminate" phase.
	2.2 Frame and sample **methodology**	De2	A sampling plan should be made. The actual sample is created during sub-process 3.6 (Draw sample), using the methodology specified in this sub-process.
	2.3 Tabulation plan/Variables	De3	Defines the variables to be collected via the data collection instrument, as well as any other variables that will be derived from them in sub-process 5.6 (Derive new variables), and any classifications that will be used.
	2.4 Data collection	De4	Determines the most appropriate data collection method(s) and instrument(s) to use. These may include computer-assisted

			interviewing, paper and pen data collection or administrative data collection techniques.
	2.5 Statistical **processing methodology**	De5	Designing the processing methodology to be applied during processing and analyses phases. The various testing routines for coding, editing, imputing, estimating, integrating, verifying and finalising data sets are designed.
	2.6 Define **archive rules**	De6	Archiving rules for statistical data and metadata resulting from a statistical business process are determined in this sub-process..
	2.7 Processing **systems and workflow**	De7	Systems and workflow throughout SVC from need, data collection to archiving is listed to ensure it is integrated together efficiently with no gaps or duplication.
	2.8 Detailed **project plan**	De8	Design a project plan giving on tasks to be carried out. The duration and responsibilities for each task should be highlighted.

The "Design" phase spans 6 dimensions in the SASQAF; and it is related to 14 indicators, of which 40 sub-processes of the SVC are affected for PAPDC. See Table 4.4 below:

Table 4.4: SASQAF vs. SVC (Design Phase)

SASQAF Quality Dimensions			Statistical Value Chain							
			Design							
Number	Dimension	SASQAF Code	De1	De2	De3	De4	De5	De6	De7	De8
0	Pre-requisites of quality	PRER6								X
		PRER7							X	
2	Accuracy	ACCU5		X						
		ACCU6				X				
		ACCU7				X				
3	Timeliness	TIME3	X	X	X	X	X	X	X	X
6	Comparability and Coherence	COMP2						X	X	
		COMP3	X	X	X	X	X	X		
7	Methodological soundness	METH1		X	X					
		METH2					X			
		METH3					X			
		METH4							X	
8	Integrity	INTE4	X	X	X	X	X	X		
		INTE5	X	X	X	X	X	X	X	X

4.3.3 Build

All systems need to have been designed, built and tested in a "testing" environment before going "live" into a production environment. Five processes defines this phase and they are conducted either in parallel or in sequence. It consists of the following activities or sub-processes (see Table 4.5):

Table 4.5: Phase 3 of SVC (Build) (Statistics South Africa, 2010b)

PHASE (Level 1)	SUB-PROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
3. BUILD	3.1 Data collection **instrument**	Bu1	Activities to build the collection instruments to be used during Phase 4 (Collect). The collection instrument is generated or built based on the design specifications created during phase 2 (Design).
	3.2 Process **components**	Bu2	Activities to build new components and enhancing the existing software components needed for the business process, as designed in Phase 2 (Design).
	3.3 Configure **workflows**	Bu3	Configures the workflow, systems, and transformations used within the statistical business processes, through all stages of SVC
	3.4 **Test** end-to-end	Bu4	Activities to manage a test or pilot of the statistical business process from the field to office.
	3.5 Finalise **production systems**	Bu5	Activities to put the process, including workflow systems, modified and newly built components into production ready for use by business areas.
	3.6 Draw the **sample**	Bu6	Establishes the frame and selects the sample for this iteration of the collection, as specified in sub-process 2.2 (Frame and sample methodology).

The "Build" phase spans 2 dimensions in the SASQAF; and is related to 2 indicators, of which 12 sub-processes of the SVC are affected for PAPDC. See Table 4.6 below:

Table 4.6: SASQAF vs. SVC (Build Phase)

SASQAF Quality Dimensions			Statistical Value Chain					
			Build					
Number	Dimension	SASQAF Code	Bu1	Bu2	Bu3	Bu4	Bu5	Bu6
3	Timeliness	TIME3	X	X	X	X	X	X
8	Integrity	INTE5	X	X	X	X	X	X

4.3.4 Collect

This phase collects all the necessary data, by using different data-collection modes (including extractions from administrative and statistical registers and databases); and it loads them into the appropriate data environment. It consists of the following activities or sub-processes (see Table 4.7):

Table 4.7: Phase 4 of SVC (Collect) (Statistics South Africa, 2010b)

PHASE (Level 1)	SUB-PROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
4. COLLECT	4.1 **Set up** of collection	Co1	Ensures that the people, processes, and technology are ready to collect data, using various modes of collection as planned (designed). During this period of setting up the collection, the preparation activities are outlined including strategy, training and the statistical business process.
	4.2 **Run** collection	Co2	The implementation of various data collection instruments as designed/planned for.
	4.3 Load **data into the processing** environment	Co3	Initial data validation, as well as loading the collected data and metadata into a suitable electronic environment for further processing in the following phase 5 (Process).

The "Collect" phase spans 2 dimensions in the SASQAF; and is related to 2 indicators, of which 5 sub-processes of the SVC are affected for the PAPDC. See Table 4.8 below:

Table 4.8: SASQAF vs. SVC (Collect Phase)

SASQAF Quality Dimensions			Statistical Value Chain		
			Collect		
Number	Dimension	SASQAF Code	Co1	Co2	Co3
3	Timeliness	TIME3	X	X	X
8	Integrity	INTE5	X		X

4.3.5 Process

This phase describes processing of data firstly through the preparation phase (i.e. cleaning of data) to the analysis phase. It includes the sub-processes of checking, cleaning and transforming the collected data. It consists of the following activities or sub-processes (see Table 4.9):

Table 4.9: Phase 5 of SVC (Process) (Statistics South Africa, 2010b)

PHASE (Level 1)	SUB-PROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
5. PROCESS	5.1 Standardise and **anonymise**	Pr1	To protect the confidentiality of respondent information, processes and procedures need to be defined and executed throughout the SVC. Standardisation and anonymisation may take place before or after sub-process 5.2 (Integrate data).
	5.2 Integrate **data**	Pr2	Integration of data sources can include data from a variety of sources either internally or externally. The idea of integration is to ensure compatability and harmonisation of the data sets.
	5.3 Classify and **code**	Pr3	The classification and coding of input data forms part of this phase. For example, automatic (or clerical) coding routines may

			assign numeric codes to text responses according to a pre-determined classification scheme.
	5.4 Edit and **impute**	Pr4	The editing and imputation applies to collected micro-data. Each record is scanned to identify any missing information or errors on the lo to try to identify (and where necessary correct) missing data, errors and discrepancies.
	5.5 Derive **new variables**	Pr5	Deriving additional variables to aid the statistical outputs and interpretation as required in the design of the survey may be required. Applying mathematical or statistical formulae clearly defined for users of the information is applied using existing variables to derive the new ones.
	5.6 Calculate **weights**	Pr6	This sub-process creates weights for unit data records according to the methodology created in sub-process 2.5: Statistical processing methodology.

The “Process” phase spans 3 dimensions in the SASQAF; and is related to 3 indicators, of which the 13 sub-processes of the SVC are affected for the PAPDC. See Table 4.10 below:

Table 4.10: SASQAF vs. SVC (Process Phase)

SASQAF Quality Dimensions			Statistical Value Chain					
			Process					
Number	Dimension	SASQAF Code	Pr1	Pr2	Pr3	Pr4	Pr5	Pr6
3	Timeliness	TIME3	X	X	X	X	X	X
6	Comparability and Coherence	COMP4				X		
8	Integrity	INTE5	X	X	X	X	X	X

4.3.6 Analyse

The statistics produced are analysed i.e. examined in detail, interpreted, and made ready for dissemination and incudes the sub-processes and activities that enable statistical analysts to understand the statistics produced. The “analyse” phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced. They consist of the following activities or sub-processes (see Table 4.11) (Statistics South Africa, 2010b):

Table 4.11: Phase 6 of SVC (Analyse)

PHASE (Level 1)	SUB-PROCESSES (Level 2)	VARIABLE	DESCRIPTION (Level 3)
6. ANALYSE	6.1 Acquire **ancillary information**	An1	Acquiring a high level of domain intelligence will allow a statistical analyst to understand the data better, and to identify where results might differ from expected values. This allows better explanations of these results in sub-process 6.5 (Describe and explain).

	6.2 Calculate **aggregates**	An2	This sub-process creates aggregate data and population totals from micro-data. It includes summing data for records sharing certain characteristics, determining measures of average and dispersion, and applying weights from sub-process 5.6 to sample survey data to derive population totals.
	6.3 Prepare **draft outputs**	An3	Applying the domain intelligence to the data collected to produce statistical outputs.
	6.4 **Validate**	An4	Validating statistics against pre-defined expectations
	6.5 Describe and **explain**	An5	Statisticians describe and explain statistical data after attaining an in-depth understanding of the outputs.
	6.6 Disclosure control and **anonymise**	An6	Data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality thereby ensuring confidentiality.
	6.7 **Finalize outputs** for dissemination	An7	The statistics and associated information are designed for its purpose and reach the required quality level for dissemination.

The "Analysis" phase spans 4 dimensions in the SASQAF; and is related to 5 indicators, of which 17 sub-processes of the SVC are affected for the PAPDC. See Table 4.12 below:

Table 4.12: SASQAF vs. SVC (Analyse Phase)

SASQAF Quality Dimensions			**Statistical Value Chain**						
			Analyse						
Number	**Dimension**	**SASQAF Code**	**An1**	**An2**	**An3**	**An4**	**An5**	**An6**	**An7**
2	**Accuracy**	**ACCU1**				**X**			
		ACCU2				**X**			
3	**Timeliness**	**TIME3**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
6	**Comparability and Coherence**	**COMP3**				**X**			
8	**Integrity**	**INTE5**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

4.3.7 Disseminate

This phase manages the release of statistical products to the users of the information. For statistical outputs produced regularly, this phase occurs in each iteration. It consists of the following activities or sub-processes (Statistics South Africa, 2010b)(see Table 4.13):

Table 4.13: Phase 7 of SVC (Disseminate) (Statistics South Africa, 2010b)

PHASE (Level 1)	**SUB-PROCESSES (Level 2)**	**VARIABLE**	**DESCRIPTION (Level 3)**
7. DISSEMINATE	7.1 Update **output systems**	Di1	Manages the update of systems where data and metadata are stored for dissemination purposes.
	7.2 **Produce** products	Di2	Produces the products, as previously designed, to meet user needs.

	7.3 Produce **'quality statement'**	Di3	A quality report is produced including a declaration of the level of quality within the metadata.
	7.4 Manage the **release of products**	Di4	All the elements for the release are in place, including managing the timing of the release.
	7.5 Market and **promote products**	Di5	Actively promoting and marketing of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience.
	7.6 Manage **customer queries**	Di6	Customer issues and questions are recorded. Official response tmes and the management of quesries are indicated.

The "Disseminate" phase spans 6 dimensions in the SASQAF; and is related to 25 indicators, of which 35 sub-processes of the SVC are affected for the PAPDC. See Table 4.14 below:

Table 4.14: SASQAF vs. SVC (Disseminate Phase)

SASQAF Quality Dimensions			**Statistical Value Chain**					
			Disseminate					
Number	**Dimension**	**SASQAF Code**	**Di1**	**Di2**	**Di3**	**Di4**	**Di5**	**Di6**
2	**Accuracy**	**ACCU3**			X			
3	**Timeliness**	**TIME1**				X		
		TIME2				X		
		TIME3	X	X	X	X	X	X
4	**Accessibility**	**ACCE1**				X		
		ACCE2				X		
		ACCE3		X				
		ACCE4		X				
		ACCE5				X		
		ACCE6				X		
		ACCE7						X
		ACCE8					X	
		ACCE9				X		
		ACCE10				X		
		ACCE11					X	
		ACCE12				X		
5	**Interpretability**	**INTR1**		X				
		INTR2		X				
		INTR3		X				

7	Methodological soundness	METH5				X		
		METH6				X		
8	Integrity	INTE1			X			
		INTE2			X			
		INTE3			X			
		INTE5	X	X	X	X	X	X

By far, the majority of the SVC and the SASQAF are found to be inter-connected during the "Disseminate" phase. This is largely to reporting obligations for publication to users and metadata completion.

4.3.8 Archive

The archiving and disposal of statistical data and metadata is discussed. Archiving both electronically and physically is defined during this phase. The disposal may include intermediate files from previous iterations, as well as that of disseminated data. It consists of the following activities or sub-processes (see Table 4.15) (Statistics South Africa, 2010b):

Table 4.15: Phase 8 of SVC (Archive) (Statistics South Africa, 2010b)

PHASE (Level 1)	**SUB-PROCESSES (Level 2)**	**VARIABLE**	**DESCRIPTION (Level 3)**
8. ARCHIVE	8.1 **Manage** archive repository	Ar1	The management of one or more archive repositories. These may be electronic databases or physical locations where copies of data or metadata are stored.
	8.2 **Preserve** data and associated metadata	Ar2	Is where the data and metadata from a specific statistical business process are archived.
	8.3 **Dispose** of data and associated metadata	Ar3	Is where the data and metadata from a specific statistical business process are disposed of.

The "Archive" phase spans 2 dimensions in the SASQAF; and it is related to 2 indicators, of which 6 sub-processes of the SVC are affected for the PAPDC. See Table 4.16 below:

Table 4.16: SASQAF vs. SVC (Archive Phase)

SASQAF Quality Dimensions			**Statistical Value Chain**		
			Archive		
Number	**Dimension**	**SASQAF Code**	**Ar1**	**Ar2**	**Ar3**
3	**Timeliness**	**TIME3**	**X**	**X**	**X**
8	**Integrity**	**INTE5**	**X**	**X**	**X**

4.3.9 Evaluate

“This phase manages the evaluation of a specific instance of a statistical business process. It logically takes place at the end of the instance of the process, but it relies on the inputs gathered throughout the preceding phases. For statistical outputs to be produced regularly, evaluation should at least, in theory occur for each iteration, thereby determining whether future iterations should take place or not. If it does then whether any improvements should be implemented. However, in some cases, particularly for regular and well-established statistical business processes, an evaluation may not be formally carried out for each iteration.” (Statistics South Africa, 2010b). It consists of the following activities or sub-processes (see Table 4.17):

Table 4.17: Phase 9 of SVC (Evaluate) (Statistics South Africa, 2010b)

PHASE (Level 1)	**SUB-PROCESSES (Level 2)**	**VARIABLE**	**DESCRIPTION (Level 3)**
9. EVALUATE	9.1 Gather inputs for **programme review/ evaluation**	Ev1	Evaluation material may take many forms, including feedback from users, process metadata and system metrics, and staff suggestions.
	9.2 Prepare **an evaluation report**	Ev2	Analyses of the evaluation inputs and synthesises them into an evaluation report noting any quality issues specific to this iteration of the statistical business process and should make recommendations for changes if required.
	9.3 **Quality plan**	Ev3	This sub-process brings together the necessary decision-making power to form and agree on an action plan based on the preceding evaluation report.

The “Evaluate” phase spans 4 dimensions in the SASQAF; and is related to 4 indicators, of which 8 sub-processes of the SVC are affected for the PAPDC. See Table 4.18 below:

Table 4.18: SASQAF vs. SVC (Evaluate Phase)

SASQAF Quality Dimensions			**Statistical Value Chain**		
			Evaluate		
Number	**Dimension**	**SASQAF Code**	**Ev1**	**Ev2**	**Ev3**
0	**Pre-requisites of quality**	**PRER5**	**X**		
1	**Relevance**	**RELE6**	**X**		
3	**Timeliness**	**TIME3**	**X**	**X**	**X**
8	**Integrity**	**INTE5**	**X**	**X**	**X**

Two cross-cutting processes occur throughout the SVC, namely:

“Quality management: This process is present throughout the SVC. It is closely linked to Phase 9 (Evaluate), which has the specific role of evaluating the individual instances of a statistical business process. The overarching quality-management process, however, has both a deeper and broader scope. As well as evaluating the iterations of a process, it is also necessary to evaluate the separate phases and the corresponding sub-processes. Ideally, this should be done each time they are applied, but at least according to an agreed schedule. These evaluations can apply within a specific process, or across several processes that use common components” (Statistics South Africa, 2010b).

“Metadata management: Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase, either created or carried forward from a previous phase. The key challenge is to ensure that they are captured, as early as possible and stored, and then transferred from phase-to-phase alongside the data to which they refer to. A metadata-management strategy and system(s) are therefore vital to the operation of the SVC” (Statistics South Africa, 2010b).

4.4 DIGITAL DATA COLLECTION THEMATIC AREAS

For each of the three iterations to follow (Chapters 5 to 7); key themes, as discussed in the literature review in Chapter 2, were measured and addressed for DDC. The following table demonstrates the themes that were analysed in more than one iteration (see Table 4.19 and refer to Table 2.5):

Table 4.19: Iterational cross-cutting themes

DDC DIMENSION / THEME	ITERATION 1 **WC TEST 2015** • QUALITATIVE Primary Data Sample = 13 • QUANTITATIVE Sample n = 100 Dwellings,	ITERATION 2 **KZN CSS 2015** • Total N = 20 819 Dwellings • QUANTITATIVE Secondary Data Sample n = 384	ITERATION 3 **CS 2016** • Total N = 1 370 810 Dwellings • QUALITATIVE Primary Data Sample = 4 • QUANTITATIVE Secondary Data Sample n = 384
A. **COST (2.5.3)**	**X**		**X**
B. **ACCURACY**, **ERROR RATE** and **PERCEPTION OF QUALITY (2.5.4)**	**X**		**X**
C. **SECURITY (2.5.5)**	**X**		**X**
D. **SPEED (2.5.6)**		**X**	**X**
E. **TRAINING (2.5.7)**		**X**	**X**

4.5 CHAPTER SUMMARY

This chapter described the critical and complementary roles that the SASQAF and the SVC play when collecting official household survey data. During the process of the household-survey collection, the inter-relationship between the SVC processes and the SASQAF quality dimensions are instrumental, to ensuring that a good statistical product is produced. The mapping of the phases of the SVC to the SASQAF provides components within the NSO with a compartmental, conceptual view of their respective responsibilities and timeframes related to any given official project or household survey. The proposed DDC developmental QA framework, once formulated, should concurrently indicate possible, albeit (yet to be defined), work phases to be considered for further development of a DDC SVC.

The following three chapters (Chapters 5 to 7) describe the themes/phases that were analysed in their respective iterations.

CHAPTER 5: ITERATION 1: WESTERN CAPE TEST 2015

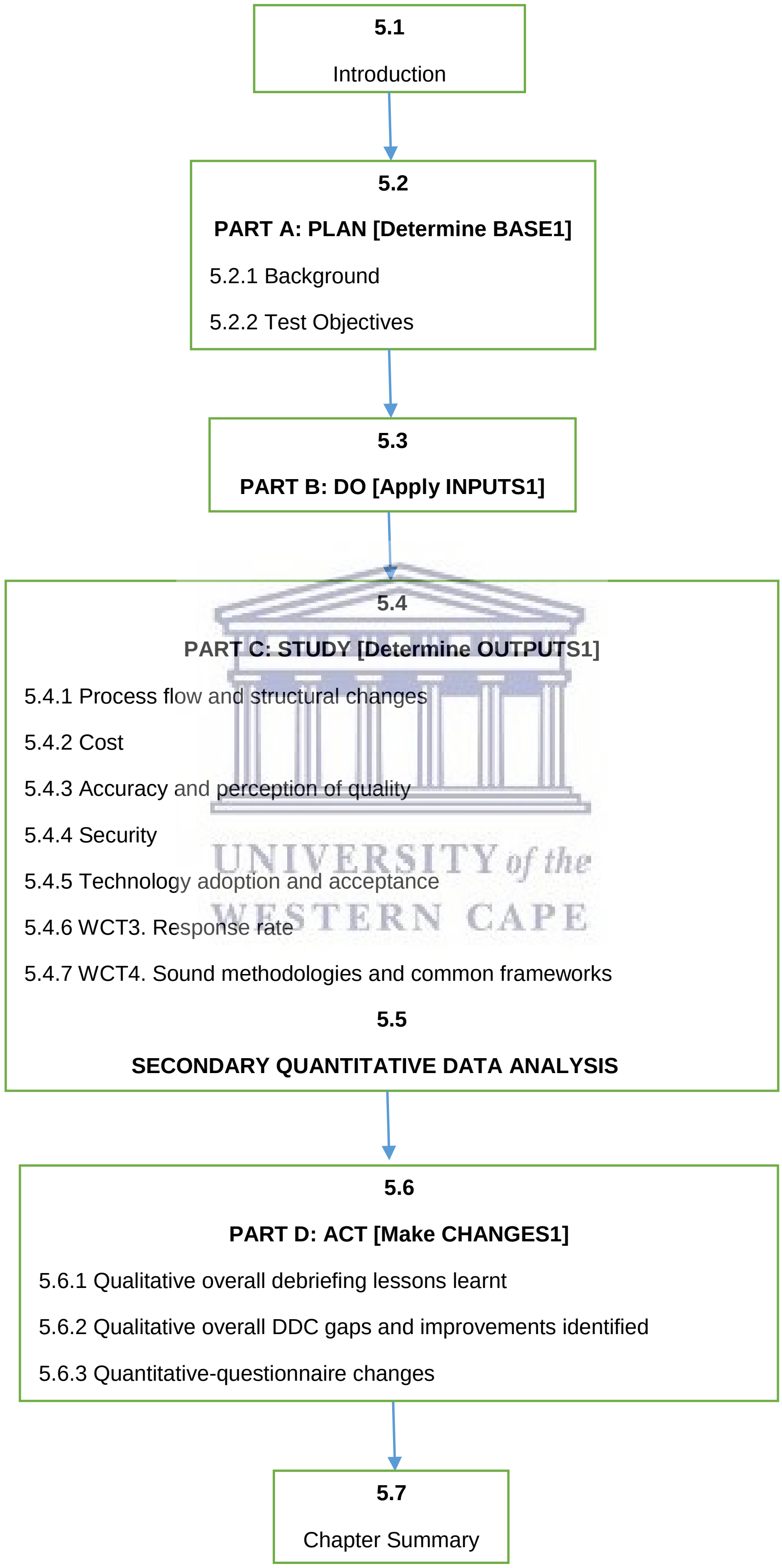


Figure 5.1: Chapter 5 layout

CHAPTER 5: ITERATION 1 - WESTERN CAPE TEST 2015

5.1 INTRODUCTION

This chapter is the first of three "iteration" chapters called Iteration 1. It presents the data from a specific DDC iteration to pinpoint areas within the SASQAF framework that requires changes to accommodate DDC. The chapter resulted in two improvements and two gaps being identified in the SASQAF framework, as presented in the final section of this chapter.

Using the PDSA cycle (Figure 5.2), Iteration 1’s "PLAN" phase ensures a "BASE1" DDC dataset is developed (Section 5.2). In the next phase, namely the "DO" phase, the inputs to "BASE1" are actioned based on the themes from literature (Chapter 2) related to the required quality aspects. These should be considered when conducting official household survey collection (Section 5.3). The net results of applying the inputs from literature are the "OUTPUTS1", or the outcomes of the captured data discussed in the "STUDY" phase (Section 5.4). The secondary data was presented in section 5.4 according to DDC quality themes. Quality measurements were identified during (Chapter 2, sections 2.5.3 to 2.5.11), which included the types of "Responses" and "Non-Responses", while secondary data was obtained from 100 sampled dwellings (Chapter 3, Table 3.2).

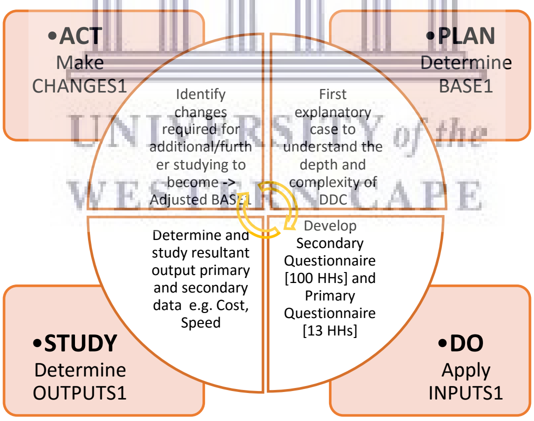


Figure 5.2: Iteration 1 - Plan-Do-Study-Act (PDSA) cycle

Once a “BASE1” was established, the “INPUTS1” applied and the “OUTPUTS1” had been studied, based on the resultant data quality, the “CHANGES1” were identified in the “ACT” phase of the PDSA (see Section 5.5) and proposed as inputs for the “PLAN” phase for the next iteration discussed in Chapter 6.

In addition to analysing the secondary data (household survey data), the primary data (or interviews) were collected from 13 NSO experts, who participated in this iteration. The data were analysed to attain an indication of the procedural experiences of the NSO experts, based on the newly introduced DDC process.

5.2 PART A: PLAN [Determine BASE1]

This section describes the details of the "PLAN" phase for Iteration 1. This is the first explanatory case to understand the depth and complexity of a DDC process for an official household survey collection. The end-to-end DDC process flow consists of the development of a base DDC process flow. The key feature of this iteration is to setup and design a full end-to-end process flow for DDC.

5.2.1 Iteration 1 Background

During 2015, the World Bank developed a DDC software solution called "Survey Solutions (SS)" for handheld mobile devices. SS was created and geared towards standardising the household survey data collection process using digital handheld devices for DDC. The SS for digital handheld device collection process was used for the test in the Western Cape Province, South Africa. The primary purpose of the test was to conduct an end-to-end process validation of the digital survey life cycle by using this solution with a primary focus on data quality throughout the process. In addition to the end-to-end process, experiences from individual experts involved in the test were also analysed. The following SS components were utilised (Worldbank Group, 2015):

- ***Designer Web Interface***: an online tool used to create the WC test questionnaire (http://solutions.worldbank.org),
- ***Headquarters Web Interface:*** an online web interface for user profile creation (SUPERVISORS), uploading household sampling assignment files, centralised survey management, and dashboard reporting.
- ***Supervisor Web Interface:*** an online/offline tool for (INTERVIEWER) creation or (Survey Officers), staff-assignment planning and managing the data-collection process on a team supervisor's level.
- ***Interviewer Tablet Interface:*** Android application software for data capturing on a mobile handheld device used by INTERVIEWERS (see APPENDIX K).

Although the process of navigation to the sampled dwelling was not included on the SS platform, it was a requirement to locate the sampled dwelling accurately. An independent application was used for this purpose, namely, ***Google MyMaps*** (Google, 2015). SS was set up for a test in the Western Cape during August 2015. On the standard SVC for PAPDC, the "Build" phase relates to the piloting and testing of the systems. The "Generic end-to-end-DDC-flow" for SS (see Figure 5.3) was compared with the generic PAPDC SVC on the key components, from designing a questionnaire to accessing the data for analysis.

The speed, cost, and quality of completed information were measured. A sample of 100 geo-referenced points or households were selected. These 100 households were spread across the Western Cape Province. Introducing SS warranted a test of the basic PAPDC household questionnaire design to be translated into a digital questionnaire on the SS platform. The analysis for the WC test focused on the following three parts, namely, 1) the end-to-end DDC process flow (5.4.1); 2) the quality themes tested (5.4.2 to 5.4.7) and 3) the overall debriefing lessons learnt during the process of DDC (5.6.1).

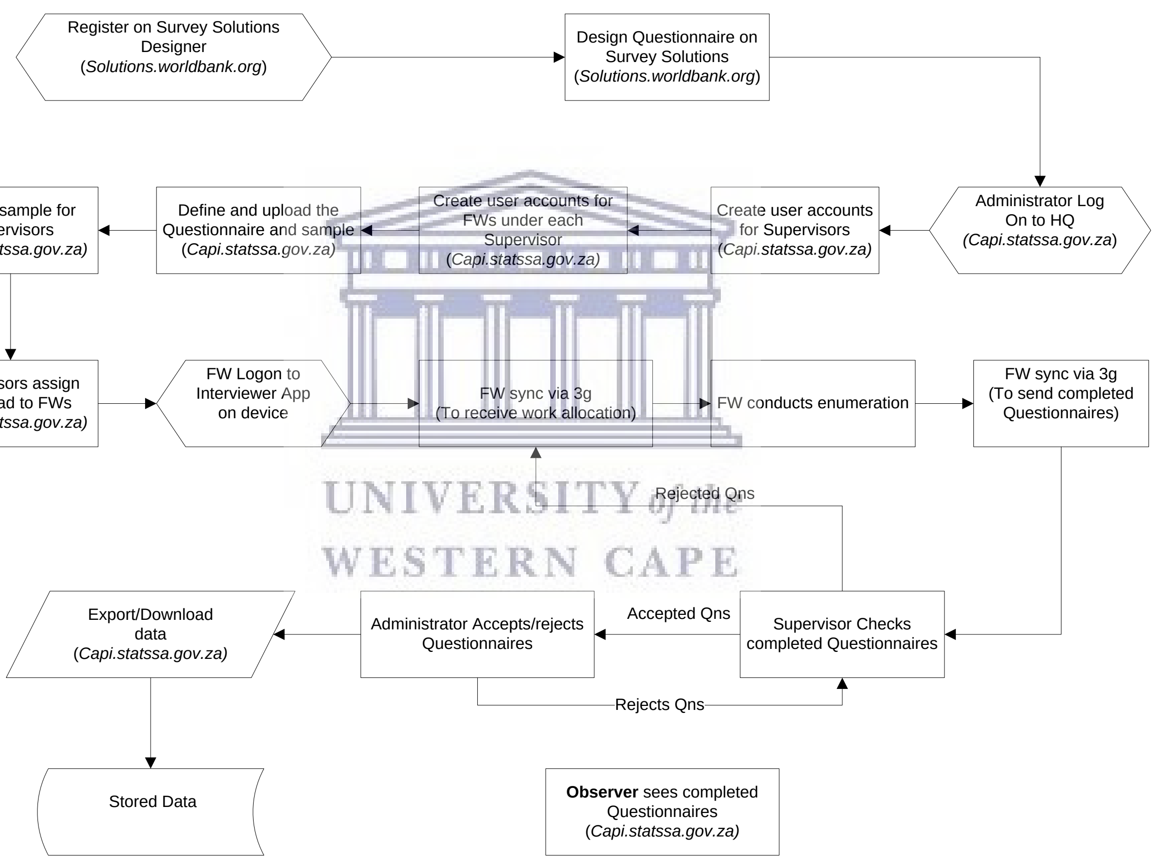


Figure 5.3: Generic Digital Data Collection SVC for Survey Solutions (Statistics South Africa, 2015)

Quantitative secondary data was collected by INTERVIEWERS on the SS platform at 100 households i.e. (Sample Size: n = 100). To ensure the end-to-end process was completed successfully, test case responses were designed by the researcher. INTERVIEWERS were tasked to provide their respective experiences on the processes related to them. The researcher was responsible for the test cases [01, 02, 03, 07, 08] conducted within the office with a computer connected to the internet, whilst the INTERVIEWERS completed the test cases [04, 05, 06] in the field during collection of data using DDC (see APPENDIX M). Figure 5.3 above describes the process involved using the SS software when conducting DDC from start to end. The generic DDC SVC for SS described the basic high-level processes, tools, and resources required for DDC.

The analysis of qualitative primary data responses from the INTERVIEWERS participating during the WC test was conducted using the "*PhD / CS2016 Research Survey Questionnaire"* (see APPENDIX B). Focus-Group data responses from the INTERVIEWERS' were recorded using the "*Focus-Group Guide"* (see APPENDIX C).

5.2.2 Iteration 1 Test Objectives

The primary objective of the WC DDC exercise was to test the questionnaire, the data transmission, and other downstream DDC processes (see Table 5.1). The results of these test are found in APPENDIX M.

Table 5.1: WC CAPI Test Descriptors

TEST CASE NO	TEST	LOCATION
01	USER-PROFILE CREATION	[OFFICE]
02	SURVEY SAMPLE FILE UPLOAD	[OFFICE]
03	ASSIGNMENT PLANNING	[OFFICE]
04	SAMPLE LOCATION VERIFICATION	[IN FIELD]
05	DATA COLLECTION	[IN FIELD]
06	CAPTURED DATA SYNCHRONISATION	[IN FIELD]
07	REAL-TIME REPORTING	[OFFICE]
08	DATA EXTRACTION	[OFFICE]

(Source: Author)

As mentioned above, all INTERVIEWERS in the field conducted test cases [04-06].

In total, 25 test staff participants and support staff were involved during the process, although only 13 responses were solicited from the collecting staff within the NSO. Logistical requirements for the test staff team members (as INTERVIEWERS) included a vehicle for each team. Teams consisted of groups of 3-4 members per team / (INTERVIEWERS) sharing a vehicle. A digital handheld (Android) device per team member was provided to each person and a data bundle for transmission of the data was included (See APPENDIX K2 - device specifications). A power bank per digital handheld device was allocated per INTERVIEWER. Fieldwork consisted of INTERVIEWER's visiting sample dwellings in one week. On completion of the fieldwork , i.e. visiting 100 sample dwellings, a four hour debriefing session was held to collect the primary data for Iteration 1 of this study. The debriefing included presentation summaries of all 5 teams of their respective experiences. INTERVIEWERS were tasked to complete a standardised presentation template on Tuesday 01 September 2015, which included completed test specifications from all the INTERVIEWERS. The secondary (quantitative) sample spread for the 100 households across the Western Cape Province were as follows (see Table 5.2):

Table 5.2: Iteration 1 - Western Cape Sample Spread

Settlement Types per District	**Sample Household Count per enumeration area (EA) Type**
Cape Winelands	**14**
Collective living quarters	6
Formal residential	3
Industrial	5
City of Cape Town	**37**
Commercial	7
Formal residential	8
Industrial	3
Informal residential	10
Parks and recreation	2
Smallholdings	7
Eden	**5**
Collective living quarters	5
Overberg	**25**
Farms	17
Informal residential	8
West Coast	**19**
Formal residential	7
Other	12
Total Western Cape Test Sample	**100**

The results post data collection indicated that a total of **88%** of all the completed questionnaires were submitted to the SS CAPI server. The zoomed-in maps (Figures 5.4 – 5.6) below displayed the satellite images of the collected information within the borders of the Western Cape. Figure 5.4 and Figure 5.5 depicted by the red icon indicated a single location, where the sample point (or dwelling) resided. INTERVIEWERS conducted data collection at these identified locations. Figure 5.6 displays the “current” status of the sampled dwelling, viz. “Approved by Supervisor”.

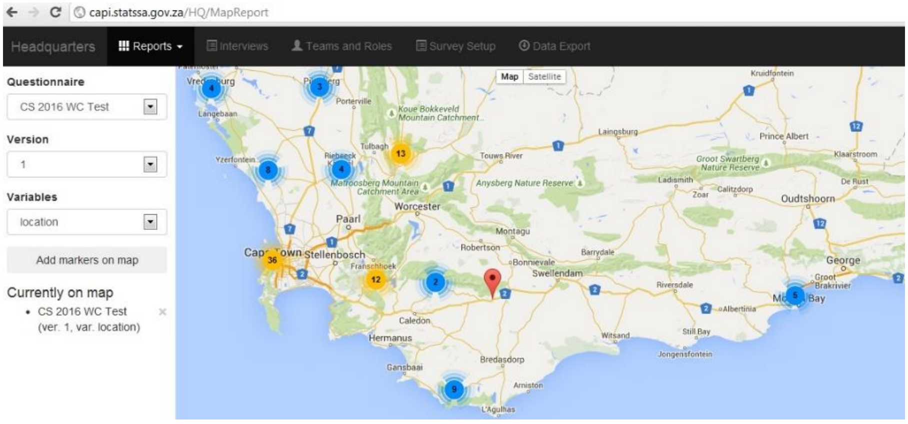


Figure 5.4: WC Test map report

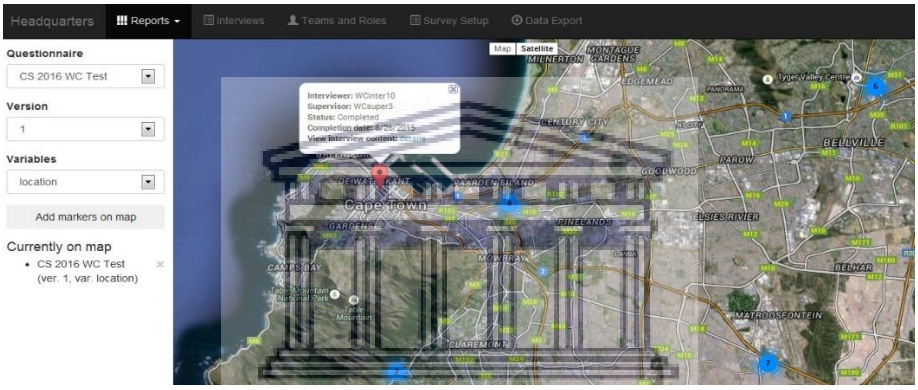


Figure 5.5: WC Test single sample point

Figure 5.6: WC Test sample status

The outcome of this pilot test was successful although the content of the information captured was not analysed. Iteration 1 provided useful information/variables to consider during the data collection phases of DDC. The length of time required to complete a questionnaire took longer than a PAPDC method. At times, it resulted in non-responses from certain respondents. Members in the field highlighted security concerns for navigation, and a correlation existed between the numbers of persons captured vs. the performance speed of the handheld device during the actual data collection. The more respondents required to capture per household (i.e. more than four persons per household), the slower the performance of the device during the data collection. The teams in their respective cases noted various issues, including navigation and methodological issues (See section 5.5.1).

5.3 PART B: DO [Apply INPUTS1]

This section follows on from “PART A: PLAN” phase for Iteration 1 and describes the ground-level work that was performed during the iteration. The “DO” section of the PDSA refers to “how was data gathered?”

First, an interview instrument and a focus group instrument was developed by the researcher to capture primary data for this iteration. As discussed in Chapter 3, this qualitative primary data constituted the perceptions of NSO workers regarding the DDC process and the quality of data DDC produces.

Second, an instrument was developed by the NSO for conducting the actual household survey. This household survey instrument was developed using the SS software package, discussed in section 5.2.1.

Third, data were gathered using the instruments described above. Thirteen NSO workers were interviewed and were involved in a focus group exercise using the primary data collection instruments (see Appendix B and C). A 100 households were surveyed using the household survey tool developed by the NSO in SS.

The data gathered in the “DO” phase of Iteration 1 is presented according to DDC quality themes in the sections to follow.

5.4 PART C: STUDY [Determine OUTPUTS1]

There are seven qualitative DDC dimensions/themes analysed during Iteration 1, both primary and secondary data is discussed per theme. The STUDY phase of Iteration 1 contained the following objectives to:

- determine and study the resultant output primary and secondary data,
- analyse the data captured from 13 INTERVIEWERS (Qualitative) including their experiences from participating in Iteration 1 (security, technology adoption, accuracy and perception of quality),

- analyse the data captured from 100 sampled household RESPONDENTS (Quantitative). The actual data captured from the households to measure the quality (speed and common household variables distributions),
- ensure all data (including the test case responses) is analysed using relevant analytical tools such as SPSS, SAS, Excel and Test Cases Responses
- determine the overall survey costs

Developing DDC quality dimensions based on the SASQAF and the SVC was conducted through the analysis of both qualitative primary data and quantitative secondary data. In Chapter 4, the SASQAF and SVC mapping for PAPDC indicated that there might be potential gaps within the framework when conducting digital household surveys. There was a need to identify those gaps through testing the DDC processes from end-to-end and testing the broader dimensions, or the themes related to quality. The key themes/dimensions considered were (see Table 5.3):

Table 5.3: Iteration 1 - WC Test DDC Themes

SECTION	LITERATURE CHAPTER	DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME	Identification of GAPS or Improvements
5.4.1	(2.5.8)	**PROCESS FLOW AND STRUCTURAL CHANGES**	**5.4.1.1 WCT1a. GENERIC END-TO-END DDC PROCESS FLOW**	
			5.4.1.2 WCT1b. ACCESS TO DATA (EXTERNAL)	
5.4.2	(2.5.3)	**COST**	AVERAGE PAPDC COST PER HH	*Improvement1 - (PRER6(7); De8) Adequate budget
			AVERAGE DDC COST PER HH	
			COST RATIO PAPDC VS. DDC	*Improvement2 - (PRER7(4); De8) Reviewed/Audited budget
5.4.3	(2.5.4)	**ACCURACY, ERROR RATE and PERCEPTION OF QUALITY**	PERCEPTION OF QUALITY	*Gap1 - (ACCU3(2); Co2) ACCURACY, and Perception- How accurately does the SO find the correct sampled Dwelling?
			ACCURACY	
5.4.4	(2.5.5)	**SECURITY**	THEFT	*Improvement3 - (PRER4; Ne3) Security only discusses respondent confidentiality, need Improvements (Plans for theft (Theft, Asset Loss) and Data Loss)
			DATA LOSS	*Gap2 - (INTE5, Co2) Theft high-security concerns for the safety of INTERVIEWERS. Data Loss is critical in terms of confidentiality (Statistics Act, 1999)
			5.4.5.1a **WCT2a. USER INTERFACE:** USER & RESPONDENT ADOPTION OF DDC	

5.4.5	(2.5.9)	**TECHNOLOGY ADOPTION AND ACCEPTANCE**	**WCT2a. USER INTERFACE:** BEST PRACTICES OPINIONS	
			5.4.5.1b **WCT2b. TECHNOLOGY ACCEPTANCE:** UNDERSTANDING DIGITAL TECHNOLOGY	
			WCT2b. TECHNOLOGY ACCEPTANCE: APPROVAL OF DIGITAL TECHNOLOGY USE	
			WCT2b. TECHNOLOGY ACCEPTANCE: EXPECTATION OF WHEN DIGITAL TECHNOLOGY SHOULD HAVE BEEN INTRODUCED	
			WCT2b. TECHNOLOGY ACCEPTANCE: USERS BELIEVE IN ADVANTAGES OF DHDT VS PAPDC	
			5.4.5.1c **WCT2c. ACCESS TO DIGITAL DEVICES:** USER ACCESS TO CELL PHONES	***Gap3** Access to digital devices - Needs to be included in the development of DDC Quality Assurance Framework.
			WCT2c. ACCESS TO DIGITAL DEVICES: OWNERSHIP OF OTHER MOBILE DEVICES	
			WCT2c. ACCESS TO DIGITAL DEVICES: USE OF MOBILE DEVICES FOR WORK PURPOSES	
5.4.6	(2.5.10)	**WCT3. RESPONSE RATE**	EFFECTIVENESS OF DDC VS PAPDC IN RELATION TO RESPONSE RATES	
			ACTUAL RESPONSE RATES	
5.4.7	(2.5.11)	**WCT4. SOUND METHODOLOGIES AND COMMON FRAMEWORK**	OPINION ON BEST DDC BEST PRACTICES - METHODOLOGIES	
			OPINION ON BEST DDC BEST PRACTICES - COMMON FRAMEWORK	
			OPINION ON BEST DDC BEST PRACTICES - ADOPTABILITY TO SASQAF	

5.4.1 Process flow and structural changes

When conducting DDC, a process flow was required to determine the processes, systems, and staff required to collect data using a DDC approach. Internally a generic end-to-end DDC flow was developed for Iteration 1. Gaining access to the data and determining the availability of data once collected to another external component for post-processing of data processes within the NSO was also studied. See APPENDIX N: Table N.1.

5.4.1.1 WCT1a. Generic end-to-end DDC flow: Test 2015 Results

In this section, the speed, cost, accuracy and the quality of the completed information were analysed. A sample of 100 geo-referenced points or households spread across the Western Cape Province was selected. Introducing SS warranted a test of whether a standard PAPDC household questionnaire could be designed on a digital questionnaire form also be used to capture data securely or not.

5.4.1.1a Navigation (Table 5.4):

Table 5.4: Iteration 1 - DDC Navigation Process

Process	Sub-Process	Percentage (%)				
NAVIGATION	Signal Strength(B)	3G	4G	EDGE	GPRS	H+
		89	1	3	6	1
	Located (%)	Located	<5m	<5m		
		73	17	10		
	Mymaps Login	Yes	No			
		100	0			
	Select Sample to navigate	Yes	No			
		100	0			
	Arrived at sample	Yes	No			
		91	9			

The signal strength of the devices indicated that a 3G signal was found to be the predominant cellular signal strength during the data collection process, in the various areas where the sample resided (Table 5.4). Using Google My Maps, 73% of selected sample points were identified successfully. As many as 17% were located within a range of 5 meters or less; while 10% were located within a range of more than 5 meters. All the selected sample points (100%) were able to login into MyMaps on the mobile-handheld devices. Using the Google MyMaps software all INTERVIEWERS could successfully navigate the application (100%). From the 100% successful selection of the sampled points in MyMaps, in terms of navigation, 91% arrived at the sampled point (household) successfully.

5.4.1.1b Data Collection (Table 5.5):

Table 5.5: Iteration 1 - DDC Data Collection Process

Process	Sub-Process	Percentage (%)	
DATA COLLECTION	Log in to Mobile Device	Yes	No
		100	0
	Log in to Survey Solutions	Yes	No
		100	0
	Select Assigned point	Yes	No

		83	17
	Complete Questionnaire	**Yes**	**No**
		71	29

All INTERVIEWERS (100%) could successfully log onto the mobile-device system as well as on the SS application (Table 5.5). INTERVIEWERS indicated that they could select the relevant assignment point (83%); while 17% indicated the opposite. In terms of completing the questionnaire, 71% of all the INTERVIEWERS indicated that they could complete the questionnaire; while 29% indicated they could not.

5.4.1.1c Data Synchronisation (Table 5.6):

Table 5.6: Iteration 1 - DDC Data Synchronisation Process

Process	Sub-Process	Percentage (%)			
DATA SYNCHRONISATION	**Synchronisation to server speed (MM: SS):**	**00:00 - 00:10**	**00:11 - 01:00**	**01:01 - 03:00**	**Did not sync at point**
		60	20	14	6
	Synchronised at the sample point	Yes	**No**		
		86	14		

The synchronisation to server speed (MM:SS) results indicated that a total of 60% of all completed questionnaires were synchronised within 10 seconds from the handheld devices in the field to the server (Table 5.6). It was found that 6% of the completed questionnaires were not synchronised at that point of data collection. Majority (86%) of the INTERVIEWERS indicated that they could synchronise and send completed questionnaires from the actual sample location point successfully; while 14% did not synchronise at the sample location point. In total, 88% of all completed questionnaires were submitted to the SS central server.

5.4.1.2 WCT1b. Access to data (internal):

All data synchronised from the INTERVIEWER tablets were accessible in real-time on the SS Headquarters web platform. The online reports, as well as the facility to export the dataset in (tab), assisted the field management group to monitor progress, and the analysis team in analysing the data in realtime. The secondary data variable used included the exported data set (Primarydata.tab) to determine the NSO stakeholders data access. In the section above, both the internal end-to-end process and external access to data ensured that it is possible to collect data for DDC, and ensure the required access to the data is allowed in realtime.

The following dimensions and themes were also analysed (see APPENDIX N: Table N.2) and yielded some significant results:

5.4.2 Cost

The budget is a key indicator for the NSO when determining the overall budgetary and resource allocation to any household survey. Given that the sample is spatially distributed across a province or the area of interest, the collective cost for collecting the data is based on numerous factors. The sample size of a survey significantly affects its cost. In general, the overall cost of a survey is a function of the fixed overhead costs and the variable costs. The variable costs are associated with the selection and processing of each sample unit during each stage of the sample selection process. Therefore, the larger the sample, the higher the overall cost of the survey implementation (United Nations Department of Economic and Social Affairs, 2005).

> For the Western Cape test, DDC was up to **20** times cheaper when comparing the costs between PAPDC and DDC

Cost covers the following aspects of the SVC (Need, Design, and Archive) and SASQAF (Pre-requisites of quality); see Table 5.7:

Table 5.7: Cost mapping (SASQAF vs SVC) - WCT

SASQAF Quality Dimensions			Statistical Value Chain		
Num	Dimension	SASQAF Code	Need Ne5	Design De8	Archive Ar2 - Cost
3	Pre-requisites of quality	PRER6	COST - PRER6(7) - Resources are commensurate with the needs of statistical programmes.	MAPPED – Improvement1 - COST Adequate budget - 1.6 Resources are commensurate with the needs of statistical programmes. (enhance to indicate components of DDC and related costing) (Plans for initial costs for all downstream SVC processes)	
		PRER7 - COST - Reviewed/Audited Budget		MAPPED – Improvement2 - COST Reviewed/Audited budget - 1.7 Measures to ensure efficient use of resources in 1.6 are implemented. - (enhance to indicate components of DDC and related costing) + (Plans for Reviewed/Audited for all downstream SVC processes)	
4	Accessibility	ACCE1 - COST - 5.11 “Does the pricing policy governing dissemination exist? and is it accessible”			

(Source: Author)

On the existing SASQAF framework and SVC, PRER6 and PRER7 required improvements to be made concerning cost:

- ***IMPROVEMENT 1** - (PRER6(7); De8)
- ***IMPROVEMENT 2** - (PRER7(4); De8)

The SASQAF only refers to the cost of reviewing and auditing a budget (PRER7[4]) as well as the pricing policy (ACCE1(11); Ne11). SVC only mentions the cost in the detailed project plan De8 and Archive2 phases respectively. Improvements are required to enhance and indicate the components of the DDC and the related costing. These include plans for the initial costs and all the downstream SVC processes. Furthermore, improvements are also required to cost under the Reviewed/Audited section (see Table 5.8).

Table 5.8: Cost table - WCT

THEME	SUB THEME	IMPACT	NATURE OF IMPACT	VARIABLE	MAPPED	SASQAF QD	SVC	IMPROVEMENT
COST	Reducing Survey Costs	Internal	Project, Programme, Organisational Savings	AVERAGE PAPDC COST PER HH	(PRER6(7); Ne5)	PRER7	Ar2	**(PRER6(7); De8) (PRER7(4); De8)**
			More or other relevant surveys can be conducted (as a value add to the public) due to cost savings	AVERAGE PAPDC COST PER HH	(PRER6(7); De8)	ACCES11		
			Continue producing quality statistical products with increasingly lesser funding received from National Treasury (Government fiscus)	Derived Variable: COST RATIO PAPDC VS. DDC	(PRER7(4); De8)			
		External	Reduced cost to user-paid surveys					
			Although most data is made available free, where there is data that attracts a cost, a reduction in costs aids wider spread and accessibility of statistical products to the public					

(Source: Author)

In summary, the following variables were analysed in relation to cost:

Secondary Data Variables:

- AVERAGE PAPDC COST PER HH

- What is the average cost to collect the data per household using PAPDC?

- AVERAGE DDC COST PER HH
  - What is the average cost to collect the data per household using DDC?
- COST RATIO PAPDC VS. DDC
  - Derived variable: Is Total Survey Cost per HH (DDC) (is less than) < Total Survey Cost per HH (PAPDC)?

The cost ratio is a significant reduction in terms of the cost of collecting data per household for PAPDC vs. DDC. In the next section (5.4.3), the accuracy and perception of quality are analysed.

5.4.3 Accuracy and perception of quality

The perception of quality and the accuracy of the data is a critical indicator of overall data quality and relates to the "INTEGRITY" dimension of SASQAF. The choices of source data, techniques and dissemination decisions, are informed by statistical considerations only. The primary data collected through the PhD Questionnaire regarding "the perception of quality" found that 100% of the INTERVIEWERS felt that the data collected were good quality. INTERVIEWERS themselves were collecting the data (see Table 5.9).

Table 5.9: Accuracy and perception of quality mapping (SASQAF vs SVC) - WCT

SASQAF Quality Dimensions			Statistical Value Chain		
			Collect	Analyse	Disseminate
Number	Dimension	SASQAF Code	Co2	An4	Di3
2	Accuracy	ACCU2	GAP1 - ACCURACY, and Perception- How accurately does the SO find the correct sampled Dwelling?	ACCURACY, Perception	ACCURACY, Perception - 7.3 Produce 'quality statement'

(Source: Author)

A ***Gap1** has been identified for (ACCU3 (2); Co2), whereby the importance of enumerating the correct sampled dwelling should be emphasised; as it could create bias in the results of the surveys (see Table 5.10).

Table 5.10: Accuracy and perception of quality table - WCT

THEME	SUB THEME	IMPACT	NATURE OF IMPACT	VARIABLE	MAPPED	SASQAF QD	SVC	GAP
ACCURACY AND PERCEPTION OF QUALITY	Accuracy of data collected at sampled dwelling unit/household	Internal	A breach in sampling methodology (bias, weighting, estimations, etc.)	Sample Location	(ACCU3(2.12); An4)	ACCU3(2.12)	An4	(ACCU3(2); Co2)
			Useless data	Get Location	(ACCU3(3); Di3)	ACCU3(3)	Di3	
		External	Negative Organisational Reputation	Accuracy of Sampled visit				
	Perception of quality		Lack of Trust by respondents and users of data	8.2				
				11.1				
				11.2				
				11.3				
				11.4				
				2.5				

(Source: Author)

Primary Data Variables: The primary data variables are derived from the interview instrument (PHD RESEARCH SURVEY QUESTIONNAIRE) administered to the 13 NSO staff participating in the study.

- PERCEPTION OF QUALITY
    - 8.2. Using DHDT in the field would result in an improvement in the quality and accuracy of the household survey-data collection.
    - 11.1 In terms of the survey-data collection quality, I believe DHDT could improve the overall quality of the data collected.
    - 11.2 In terms of the survey data collection quality, I believe DHDT could result in poor data quality during each the response of each item.
    - 11.3 In terms of the survey data collection quality, I believe DHDT could initiate an increased response interaction from the respondents.
    - 11.4 In terms of the survey data-collection quality, I believe DHDT could result in an increase in the trust in the data from the users.
    - 2.5 Perceptions of the quality of data.

The Secondary data collected indicated that INTERVIEWERS could utilise the digital handheld devices effectively, to be able to locate the correct sampled dwelling units for the data collection. This is a key indicator of accuracy for the data collected vs. the sample frame design.

Secondary Data Variables: The secondary data variables are derived from the interview instrument (WC Test Survey Solutions Questionnaire) administered to the 100 households across the Western Cape Province.

- ACCURACY
    - Select Sample point to navigate to (Y/N) % (Quantitative)
    - Arrive at sample point destination (Y/N) % (Quantitative)
    - Derived variable : Accuracy of Sampled visit --> is Sample Location = Get Location

The “Perception of quality” and the “accuracy” (whether or not the INTERVIEWERS were able to locate the correct sampled dwellings) are important factors when considering the precision and reliability of the data collected. In the next section (5.4.4), the security of the overall DDC process was analysed.

5.4.4 Security

Regarding device security, initial security concerns were of the safety of staff and devices being at risk. Being robbed in dangerous areas where data collection was to occur was a serious threat. This risk was mitigated through training, as well as the following the correct fieldwork protocol (i.e. by visiting the gatekeepers of the area before the collection commences, which often included the local South African Police Stations). No devices were stolen through theft, damaged or lost, during the test.

Concerning the possibility of data loss, the results indicated that 100% of the data were received on the central server, thereby indicating a 0% data loss rate. Even though the test yielded a positive result, it was later found during Iteration 3 of this study, for a larger sample, to be conducted by contract staff, device theft increased. This yielded a prompt to address the issue as it was the root cause for some resignations in certain areas during Iteration 3.

The areas in which “Security” intersects with both the SASQAF Quality Dimensions and the SVC are highlighted below (See Table 5.11):

Table 5.11: Security mapping (SASQAF vs SVC) - WCT

SASQAF Quality Dimensions		**Statistical Value Chain**						
Dimension	SASQAF Code	Need	Collect		Process	Analyse	Disseminate	
		Ne3	Co1	Co2	Pr1	An6	Di2	Di3
Pre-requisites of quality	PRER4	Improvement3 - Security only discusses respondent confidentiality, need Improvements (Plans for theft (Theft, Asset Loss) and Data Loss)						
Accessibility	ACCE4						Security	
Integrity	INTE1							Security
	INTE5	X	Security	GAP2 - Data Loss, Theft	Security	Security	X	X

Security	MAPPED
Security	Only SVC
Security	Only SASQAF QD

Legend

(Source: Author)

It was found that regarding this dimension/ theme, there is a ***Gap2** that needs to be included in the development DDC QA Framework. In the test, there was a 100% synchronisation rate of the full data to the server.

- (INTE5, Co2)
  - **Theft** (in the Western Cape there are large areas affected by gang-related theft that poses high-security concerns for the safety of the INTERVIEWERS, as was found during the debriefing focus-group discussions).
  - **Data Loss** is critical in terms of confidentiality (Statistics Act, 1999); and this needs to be addressed from a physical and cyber perspective.

On the existing SASQAF framework and SVC, PRER4 requires an ***Improvement3**:

- (PRER4; Ne3)

SASQAF only refers to respondent confidentiality; and NSOs need plans upfront to mitigate or prevent the possibility of theft (Theft, Asset Loss) and (Data Loss) (see Table 5.12).

Table 5.12: Security table - WCT

THEME	SUB THEME	IMPACT	NATURE OF IMPACT	VARIABLE	MAPPED	SASQAF QD	SVC	IMPROVE MENT	GAP
	THEFT	Internal	Negatively affect staff safety during collection	% Devices Stolen					
	DATA LOSS		Increased costs for asset loss (device)	% Synchronisation Success Rate (Data Loss)					
		External	Increased costs for returning Survey Officers						
SECURITY			Negative Organisational Reputation					**(PRER4; Ne3)**	**(INTE5, Co2)**
			Lack of Trust by respondents and users of data						
			Respondent Burden						
			Respondent **Confidentiality** Breach		(ACCE4; Di1)	PRER4	Co1		
					(INTE1; Di3)	ACCE4	Pr1		
						INTE1	An6		

(Source: Author)

Secondary Data Variables related to Security:

- THEFT
  - Number of Stolen devices
  - Total Number of Survey Officers
  - Derived variable: Device Loss through theft (Number of Stolen devices/ Total Number of Survey Officers)
- DATA LOSS
  - % Synchronisation Success Rate (Data Loss)

This section found 1 gap and 1 improvement that needs to be considered for DDC. The loss of devices through theft and potential data loss during the actual collection needs to be included from the planning phases of any DDC process. The impact of security relates to the trust an NSO displays with both its users and respondents. In the next section (5.4.5), technology adoption and acceptance are analysed.

5.4.5 Technology adoption and acceptance

This theme relates to the "ACCURACY" dimension of the SASQAF and refers to the effects of INTERVIEWERS on the process of collecting data using DDC. Three sub-DDC dimensions were analysed. These include the user interface, technology acceptance and access to digital devices.

5.4.5.1a WCT2a. User interface:

The visual display of the SS software on the tablet varies greatly, in terms of the screen size of the devices. The 7-inch Lenovo tablets were used during Iteration 1 (see APPENDIX K2). The ease of navigating the questionnaire contents on a device by the INTERVIEWER was studied. The experience of the respondent was also noted. The latter was determined through the INTERVIEWER's perception of how comfortable the respondent was during answering a questionnaire administered through a handheld device. Overall, the results indicated that the DDC is very user-friendly.

The following variables were analysed:

Primary Data Variables:

- USER & RESPONDENT ADOPTION OF DDC
  - 13.5 During your participation in DDC pilots thus far, has DHDT been user-friendly for both the interviewer and the interviewee?
    - A total of 59% (42% Agreed, 17% Strongly Agreed) agreed that *"during their participation in DDC pilots thus far, DHDT has been user-friendly for both the interviewer and the interviewee"*. While 8% disagreed (8% Disagree, 0% Strongly Disagree). The proportion of 33% responses were undecided.

- BEST-PRACTICE OPINIONS
  - 12.1 Opinions on best practices in the official household-survey collection
    - Approximately 41% (33% Agreed, 8% Strongly Agreed) agreed that *"DHDT has best practice methodologies defined in South Africa or sub-Saharan Africa"* as opposed to 25% (17% Disagreed, 8% Strongly Disagree). A total of 33% of all the responses were undecided.

5.4.5.1b WCT2b. TECHNOLOGY ACCEPTANCE

INTERVIEWERS, in general, felt that DDC should have been introduced earlier. They cited examples of digital technology uptake in other sectors of the economy, or during their daily interactions with digital technology. Their understanding of the technology is very good, as the variables below indicate.

The following variables were analysed:

Primary Data Variables:

- UNDERSTANDING DIGITAL TECHNOLOGY

- 8.1 I sometimes feel that I have been left behind when it comes to using DHDT
    - The majority of the respondents (77%) disagreed that they have been left behind with DHDT. Only 8% agreed in general that they had been left behind. 15% of all the responses were undecided.
- APPROVAL OF DIGITAL TECHNOLOGY USE
  - 8.6 “I am excited about using DHDT in the household-survey collection.”
    - The majority of the respondents (92%) agreed that they were excited in using DHDT in household surveys, as opposed to only (8%) who were not.
- EXPECTATION OF WHEN DIGITAL TECHNOLOGY SHOULD HAVE BEEN INTRODUCED
  - 12.5 Regarding best practices in the official survey-data collection, I believe DHDT should have been in existence a while back already, given that mobile devices have been increasingly growing w.r.t power and speed.
    - The majority of the respondents (55%) agreed that DHDT should have already been in existence. Perhaps it should have been used as an alternative to PAPDC, as opposed to only (27%), who disagreed; while the remaining (18%) were still undecided.
- USERS BELIEVE IN THE ADVANTAGES OF DHDT VS PAPDC
  - 13.3 “During my participation in DDC pilots thus far, I found that there are significant advantages to DDC vs. PAPDC.”
    - The majority of the respondents (85%) agreed that DDC delivers more benefits (advantages) than PAPDC, as opposed to only (8%) who believed otherwise; while the remaining (8%) were undecided.

5.4.5.1c WCT2c. ACCESS TO DIGITAL DEVICES

All of the INTERVIEWERS had access to cellphones, smartphones, and other mobile devices, before participating in the test (100%). It was found that in this dimension/ theme there is a ***GAP3** that needs to be included in the development of the DDC Quality Assurance Framework. There is no linkage between the SASQAF and SVC, hence no mapping could have been done.

- USER ACCESS TO CELL PHONES:
  - Ensuring that permanent staff members from NSOs or contract workers have access to cell phones is a key component during the collection phase. Reporting to supervisors, as well as contacting team members regarding

pickup location, etc. These are all responsibilities that require telephonic contact and is required for DDC

- OWNERSHIP OF OTHER MOBILE DEVICES:
  - These determine the level of general understanding around maintenance and the operation of mobile-handheld devices for DDC
- USE OF MOBILE DEVICES FOR WORK PURPOSES:
  - These indicate whether INTERVIEWERS understand the definition of utilising a mobile-handheld device for work purposes during DDC

The following variables were analysed:

Primary Data Variables:

6. Do you own a cellular phone?
   - 100% indicated they owned a cellular phone

9. Do you currently own a smartphone, tablet or mobile-computing device?
   - 100% indicated they owned a smartphone, tablet or mobile computing device

10. Do you currently use a smartphone, tablet or mobile computing device for work-related purposes, other than calling, sms'ing, surfing the internet, or for social media-related activities?
    - The majority (75%) responded yes to using a device for work-related purposes.

This section indicated that respondents participating in Iteration 1 felt positively towards the DDC process and also indicated favourably to the ease of use and ownership to related digital devices. The next section (5.4.6) discusses the resultant response rates obtained from the data collected for the 100 households.

5.4.6 WCT3. Response rate

The response rate is subjected to the "ACCURACY" dimension within SASQAF and relates to the effects of INTERVIEWERS on the process of collecting data through DDC. Two sub-DDC dimensions were analysed. These include the "effectiveness of DDC vs. PAPDC concerning response rates" and the "actual calculated response" after data collection has been completed using a DDC process.

Primary Data Variables:

- EFFECTIVENESS OF DDC VS PAPDC IN RELATION TO RESPONSE RATES

  "3.2.13 the effects of the interviewers must be determined and reported" (Statistics South Africa, 2010b).

The primary data collected indicated that actual response codes from a sample of 13 experts equalled 100%.

Secondary Data Variables:

- ACTUAL RESPONSE RATES
  - "3.2.24 The unit non-response rate must be within acceptable levels" (Statistics South Africa, 2010b)

5.4.7 WCT4. Sound methodologies and common framework

Experienced NSO staff involved during Iteration 1 of this study indicated that, regarding sound methodological practices, they believed the following:

Primary Data Variables:

- OPINIONS ON DDC BEST PRACTICES – METHODOLOGIES
  - 12.1 Regarding best practices in official survey data collection, "I believe DHDT has the best practice methodologies defined in South Africa, or in sub-Saharan Africa."
    - As many as 42% Agree; while 25% disagreed. But, 33% remained undecided.

- OPINION ON DDC BEST PRACTICES – COMMON FRAMEWORK
  - 12.3 Regarding best practices in official survey data collection, "I believe DHDT requires a common framework, in order to progress and promote the use of DDC amongst NSOs."
    - The majority (82%) agreed the development of a common DDC framework is required. Only 18% believed otherwise.
- OPINION ON DDC BEST PRACTICES – ADOPTABILITY TO SASQAF
  - 12.4 Regarding best practices in official survey-data collection, "I believe DHDT can be adopted easily; since there currently exists a quality framework (SASQAF) for PAPDC.
    - The majority (82%) agreed that DHDT can easily be adopted, given that there currently exists a PAPDC quality-assessment framework (SASQAF); while 9% believed otherwise, and the remaining 9% remained undecided.

5.5 SECONDARY QUANTITATIVE DATA ANALYSIS

In terms of the secondary quantitative data, the quality of the data collected through Iteration 1 was assessed. Table 5.14 below indicates the metrics for the output data, as captured by the permanent fieldstaff. A total of 249 measurements or variables were developed; while 32% of these measurements were not completed. 2 of the total sample of 100 were not visited, due to the short time period (UNIT); non-response = 2%. In addition, 10 out of 98 households (HHs) displayed low (ITEM) non-responses; there were 15 out of 98 HHs medium, and 73 out of 98 HHs high item non-response items, respectively. The proportion of true responses was 60%; while the remaining 40% included the non-response (non-contacts, refusals) and out-of-scopes (vacant dwelling, demolished, status change).

Table 5.13: OUTPUT 1 - Iteration 1

METRICS	FIGURES	COMMENTS
# Measurements	**249**	Number of Variables in the questionnaire
Proportion column Non-Response	**32%**	Percentage of measurements or variables not completed
Proportion (UNIT) row Non-Response	**2%**	2/100 TOTAL UNIT NON RESPONSE (2 out of 100 units or HH's that was not visited completely by the fieldworkers due to time or other reason)
Proportion (ITEM) row Non-Response (LOW)	**10%**	10/98 (Number of units or HH's that had Low item non-response) - where between [27.7% to 45.0%] of the questions/measurements were answered
Proportion (ITEM) row Non-Response (MEDIUM)	**15%**	15/98 (Number of units or HH's that had Medium item non-response) - where between [8.4% to 16.1%] of the questions/measurements were answered
Proportion (ITEM) row Non-Response (HIGH)	**74%**	73/98 (Number of units or HH's that had High item non-response) - where between [0.4% to 8.0%] of the questions/measurements were answered
The proportion of TRUE Response	**60%**	RC11, RC12
Proportion TRUE Non-Response	**40%**	RC21, RC22, RC31, RC32, RC33, RC34, RC35

Table 5.15 indicates that 53% responses, 17% non-responses, 18% out-of-scopes were assigned as result codes; while 12% were not coded. The true response categories excluded the unassigned sampled HHs.

Table 5.14: OUTPUT 1 - Responses, non-responses, and out of scopes

Response Category	Household Result Code	Result Code (RC)	Count of HHs	Percentage %	True Response Category
RESPONSE (53%)	Completed	11	41	41%	
	Partly completed	12	12	12%	60%
NON-RESPONSE (17%)	Non-contact	21	7	7%	
	Refusal	22	10	10%	19%
	Unoccupied dwelling	31	0	0%	20%

OUT OF SCOPE (18%)	Vacant dwelling	32	1	1%	
	Demolished	33	2	2%	
	New Dwelling Under Construction	34	0	0%	
	Status change	35	15	15%	
NOT CODED (12%)	No Household Result Captured	99	12	12%	
		Grand Total	**100**	**100%**	

Table 5.16 indicates a respectable overall survey response rate of 76% for Iteration 1.

Table 5.15: OUTPUT 1 - Response rate

Response Rate Formula	RC 11 + RC 12 / (total HH) - sum(OOS)
Completed + Partly completed	53
Initial Sample HH's	100
Total HHs with coded Result Code	88
Sum (Out Of Scopes)	18
Response Rate	**76%**

Table 5.17 provides the measurements (variables) studied in Iteration 1 and INPUTS for Iteration 2. The focus of Iteration 1 was to test the general application of the handheld technologies deployed from the navigation application to the questionnaire capturing application. The results of the secondary data captured displayed a high proportion of non-responses, and they indicated the need to relook the training approaches and the training outcomes when conducting DDC. After studying the outcome of the data from Iteration 1, the questions were as follows: “Will DDC complicate the household data-collection processes?” and “How will the organisation transition adapt to DDC?” These questions informed Iteration 2 of the study.

5.6 PART D: ACT [Make CHANGES1]

In this section, the final phase of the PDSA’s Iteration 1 is completed, which is the “ACT” phase. The following actions are reported on:

Identify the SASQAF changes required for additional/further analyses to become the "Adjusted BASE1". Data were collected from 13 INTERVIEWERS using a qualitative primary questionnaire and focus group. The quantitative secondary data was collected from 100 households across the Western Cape Province in South Africa. The analysis tools included a focus group guide (see APPENDIX C), debriefing reports and presentations on issues to take forward. The key features included the feedback from the “STUDY” phase within Iteration 1 (Chapter 5) which feeds into the “PLAN” phase of Iteration 2 in Chapter 6 (i.e. the development of inputs). During the "ACT" phase, the researcher actively tried to incorporate the lessons learned in Iteration 1 into the NSO DDC processes before commencing with Iteration 2.

5.6.1 Qualitative overall debriefing lessons learnt

Classically, after any PDSA iteration, researchers ask design related questions such as: “What Worked?” and “What did not work?” during the PDSA cycle. We conclude the PDSA cycle by summarizing the findings in two ways, namely:

1) What worked and did not work?

2) A summary of SASQAF changes (improvements and gaps) to be considered

Also, INTERVIEWERS were tasked to debrief their findings throughout the process of training, publicity and data collection for the DDC test. Table 5.16 is an example of a debriefed report of experiences. The overall experiences were very positive to the End-to-end process flow, and the identification of areas to improve or gaps identified throughout concerning the PAPDC processes. INTERVIEWERS also felt that process standardisation and methodological issues are still required to be developed.

Table 5. 16: WC Test Debriefing Notes (example)

Group no: 2: **Test Case 04 – Sample location verification (Navigation)**		
What worked	**What did not work**	**Improvements**
• Logging into MyMaps app. • Sampled points per interviewer were loaded on MyMaps. • Select the sample point. • Google maps and MyMaps directions in terms of navigation on the road to get to the point was good.	• The sampled points were not always visible on the MyMaps app due to **signal failure in some remote areas** (mainly farm EA types). • Navigation tool does not navigate you to the sampled point. It **navigates you to the nearest location on a road to the sampled point**.	• The Navigation tool should have the maps with the sampled points per interviewer loaded on the device so that they **can be accessed offline for navigation**. • The tool should incorporate some sort of navigation to get to the sampled point once you have reached the end of the road. Example: "proceed 5 meters in a NE direction". A blue dot on the map will show the current location and which direction to proceed until the sampled coordinates are reached.

Table 5.17: Iteration 1 - Descriptive statistics on QA measurements

		time_elapsed	Cost	MainDwellType	SubsDwell	TenureStat	WaterSource	DistanceWater	WaterSupplier	Toilet	ToiletLocation	ToiletShared	Refuse	ElectrSupplier	EnergyCook	EnergyLight	EnergyWaterHeat	EnergySpaceHeat
N	Valid	100	100	12	98	13	11	0	9	10	9	9	9	4	10	10	10	10
	Missing	0	0	88	2	87	89	100	91	90	91	91	91	96	90	90	90	90
Mean		2:51:41	150.00	1.92	0.22	3.00	2.18		1.33	1.80	1.44	1.67	2.56	1.75	3.10	4.10	3.20	6.00
Median		0:04:01	150.00	1.00	0.00	3.00	1.00		1.00	1.00	1.00	2.00	3.00	1.50	1.00	2.50	2.50	6.50
Std. Deviation		10:19:16	0.000	2.109	.635	1.633	2.401		0.707	1.317	0.726	0.500	1.667	0.957	3.281	3.542	2.530	3.651
Skewness		3.916		2.643	2.742	0.407	1.859		2.121	1.183	1.501	-0.857	0.501	0.855	1.399	0.422	0.646	-0.205
Std. Error of Skewness		0.241	0.241	0.637	0.244	0.616	0.661		0.717	0.687	0.717	0.717	0.717	1.014	0.687	0.687	0.687	0.687
Kurtosis		14.458		7.080	6.383	-0.747	1.868		4.000	-0.577	1.467	-1.714	-1.275	-1.289	0.445	-1.904	-0.717	-1.599
Std. Error of Kurtosis		0.478	0.478	1.232	0.483	1.191	1.279		1.400	1.334	1.400	1.400	1.400	2.619	1.334	1.334	1.334	1.334
Range		50:10:52	0	7	3	5	6		2	3	2	1	4	2	8	8	7	10
Percentiles	25	0:02:13	150.00	1.00	0.00	1.50	1.00		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.50
	50	0:04:01	150.00	1.00	0.00	3.00	1.00		1.00	1.00	1.00	2.00	3.00	1.50	1.00	2.50	2.50	6.50
	75	0:13:00	150.00	1.75	0.00	4.50	2.00		1.50	3.25	2.00	2.00	4.00	2.75	5.25	7.50	5.00	9.00

5.6.2 Qualitative overall DDC gaps and improvements identified

The results from the primary questionnaires yielded positive results regarding *the user interface, technology acceptance, access to digital devices, access to data, response rate, sound methodologies, and a common framework*. Although the content of the information (DDC questionnaire) captured was not analysed, the test provided useful items to consider during the data collection phases. For example, the length of time required in completing a questionnaire takes longer than PAPDC, and this may result in respondent refusals. Members in the field highlighted security concerns for navigation, and a correlation exists between the numbers of persons captured vs. the speed of the device during the data collection (See Table 5.18).

Table 5.18: WC Test Changes to SASQAF

DDC THEME	IMPROVEMENT REQUIRED	GAPS
COST	***Improvement1** - (PRER6(7); De8) : Adequate budget ***Improvement2** - (PRER7(4); De8) Reviewed/Audited budget	
ACCURACY and PERCEPTION OF QUALITY		***Gap1** - (ACCU3(2); Co2): ACCURACY, and Perception- How accurately does the SO find the correct sampled Dwelling?
SECURITY	***Improvement3** - (PRER4; Ne3) Security only discusses respondent confidentiality, need Improvements (Plans for theft (Theft, Asset Loss) and Data Loss)	***Gap2** - (INTE5, Co2) -Theft high-security concerns for the safety of INTERVIEWERS - Data Loss is critical in terms of confidentiality (Statistics Act, 1999)
TECHNOLOGY ADOPTION AND ACCEPTANCE		***Gap3** Access to digital devices - Needs to be included in the development of DDC Quality Assurance Framework.

Nevertheless, gaps and improvements to the PAPDC SASQAF framework were identified in this chapter, when considering the DDC process.

5.6.3 Quantitative questionnaire changes

The questionnaire in the next iteration (Iteration 2) were kept consistent and contained some comparable variables with Iteration 1 to monitor the quality distribution of the responses between iterations (inter-iteration validity of data is tested in section 8.2). In total, 15 variables (see Table 5.17) were retained as a comparable set of indicators across the three iterations. The design of the NSO household questionnaire was also done using the same questionnaire designer on the SS platform across the three iterations of this study.

5.7 CHAPTER SUMMARY

The primary purpose of Iteration 1 was to test and conduct an end-to-end DDC process to measure the key processes required when conducting DDC.

The PDSA cycle started with the first phase i.e. the "PLAN" phase which determined the objectives and plans for setting up the "BASE1". In addition to the quality themes studied in Iteration 1, three cross-cutting dimensional themes were also analysed. These included COST, ACCURACY AND THE PERCEPTION OF QUALITY and SECURITY. The results of Iteration 1 indicated that COST, SECURITY, and the CONFIDENTIALITY OF DATA turned out to be key themes that emerged.

In the next chapter, we present Iteration 2, which highlighted TRAINING (as a theme), and other related NSO-staff variables. The teams in their respective cases noted secondary objectives at the end of Iteration 1, including navigation and methodological issues; although these were briefly summarised in the earlier sections in this study. Google MyMaps could at least navigate the fieldworker to a sample point, viz. to the nearest specific named road. For the most part, the navigation application worked well in the Western Cape; however, it may not be so successful in the farming or rural landscape territories which cellular signal for mobile handheld devices is limited.

In the following chapter, cross-iterative themes, such as SPEED and TRAINING will be analysed, as well as other specific dimensions.

CHAPTER 6: ITERATION 2: KWAZULU-NATAL CITIZEN-SATISFACTION SURVEY 2015/2016

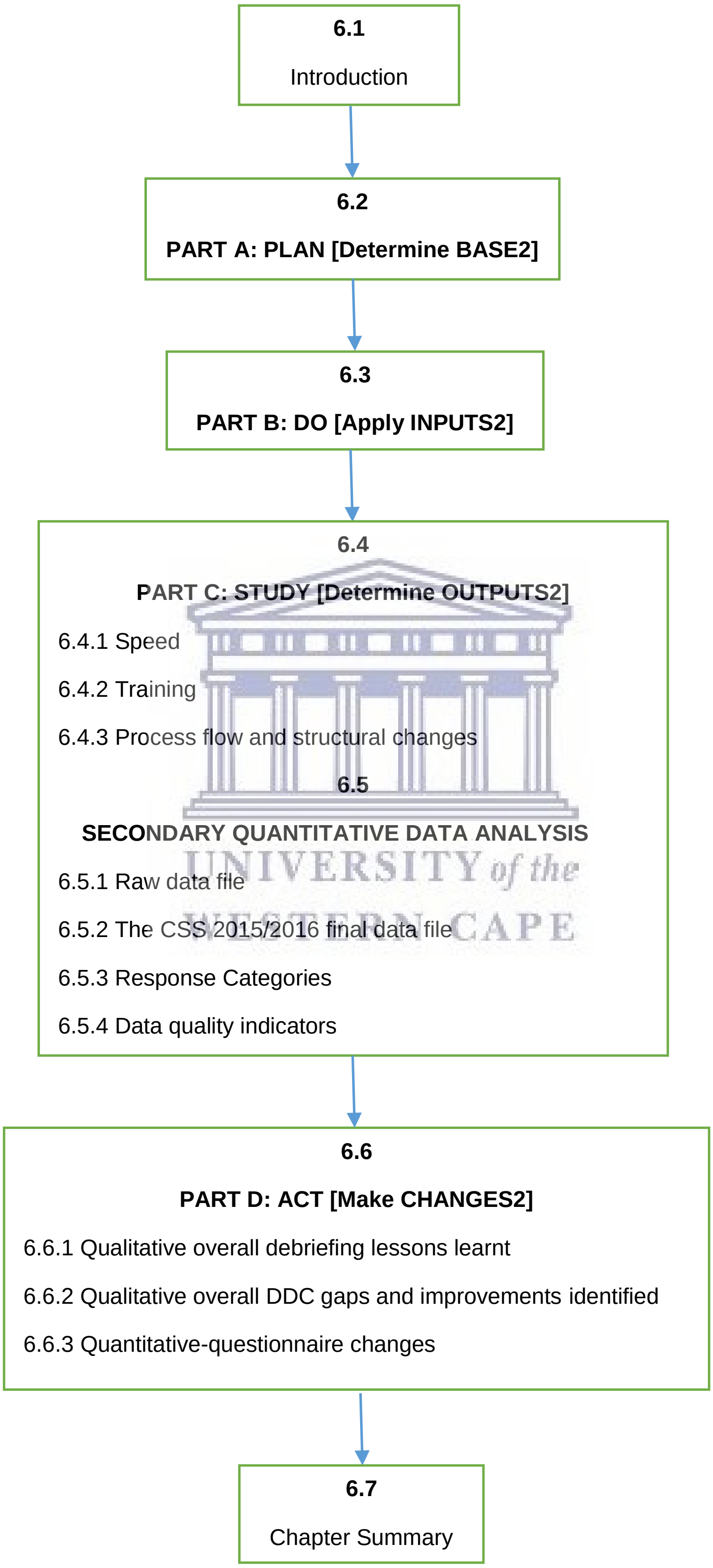


Figure 6.1: Chapter 6 layout

CHAPTER 6: ITERATION 2 - KWAZULU-NATAL CITIZEN-SATISFACTION SURVEY 2015/2016

6.1 INTRODUCTION

In this chapter, the researcher continues to apply the PDSA design approach. As in Chapter 5, this chapter is structured into “planning”, “doing”, “studying” and “acting” sections, with the majority of Iteration 2’s data being presented in the “DO” section (see Section 6.4).

To comply with the PDSA approach, the BASE starting point for the previous iteration (Iteration 1) is adjusted taking into account changes identified from Iteration 1. It is used as the practical and theoretical starting point for Iteration 2 (referred to as BASE2). The data for Iteration 2 was collected during the “DO” phase and is presented as OUTPUT2. The initial secondary data analysed comprised a 50% random sample of (10 409) out of a total of 20 819 households enumerated. The final sample for quantitative analysis included 384 households. Once BASE2 were established and OUTPUTS 2 studied according to quality dimensions, changes to the SASQAF framework are presented (CHANGES 2).

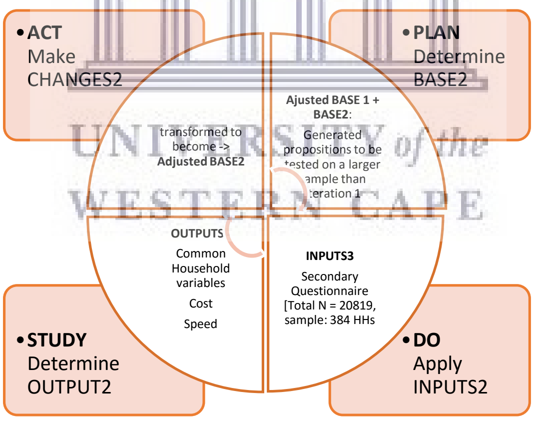


Figure 6.2: Iteration 2 - Plan-Do-Study-Act cycle

In the previous chapter, Iteration 1 identified potential improvements and gaps to be addressed for the development of a DDC Quality Assurance Framework. The full digital generic end-to-end process flow was explored on a smaller scale, concerning the sample size. The necessity for a larger DDC sampled survey was identified, to further analyse the relevant thematic quality dimensions that were to be addressed. Although this Iteration 2 survey was touted as the first official household-based digital survey conducted by the NSO, the standardisation of the quality dimensions for future related mobile DDC surveys was still to be developed in accordance to the SASQAF.

6.2 PART A: PLAN [Determine BASE2]

For Iteration 2, the actual sample size for DDC was scaled up 1:2082 from Iteration 1. The analysis in this research was based on a final sample size of 384 dwellings (see Table 6.1). The primary data survey instrument was not administered in Iteration 2 but only the focus group debriefing. Most of the initial 13 NSO experts involved during the Iteration 1 were also involved in the project as either content specialists for training or support during Iteration 2's collection process. There was no need to repeat the primary data survey instrument for them or other staff members involved in Iteration 1, but instead incorporate their feedback along with other participating members of Iteration 2, through the debriefing reports.

The qualitative data applicable to this iteration were primarily reported data from the KwaZulu-Natal (KZN) Citizen-Satisfaction Survey's (CSS), which includes reports or publication data, debriefing information and related articles by the NSO for Iteration 2. Primary data dominated all the DDC Dimensions / Themes, except for the quantitative secondary data for speed and common QA measurements.

Table 6.1: Iteration 2 - Primary and secondary data sources

Data Source	Title	Appendix	Origin	Sample size	Method
Iteration 2 - Primary data	Articles Publication data Debriefing information	N/A	NSO archives	N/A	Post Iteration 2 reports
Iteration 2 - Secondary data	KZN Citizen Satisfaction Survey 2015 Survey Solutions Questionnaire	N/A	Captured by 217 Contract fieldworkers	Initial Dwelling Sample N = 20 819 A sample of n= 384 dwellings was analyzed	Digitally collected using a handheld device

6.2.1 Iteration 2 Background

As in Iteration 1, the concurrent mixed-method design strategy (quantitative and qualitative) data collected simultaneously highlighted the key improvements to be made to the quality framework, as well as the identification of potential improvements or gaps to the SASQAF. The KZN CSS was conducted in the second semester of 2015, the results of which were released on the 04th February 2016 (Statistics South Africa, 2016d). On the 12th October 2015, the province of KZN commenced the journey into the world of DDC. The province trained fieldworkers for one month, including both provincial and district training. Data collection spanned across the entire province which included 11 districts, enumerating approximately 20 819 actual sampled dwelling units during the six weeks. The project was well received by both the public gatekeepers of the selected sampled dwellings and the general community (see Figure 6.3).

6.2.2 Iteration 2 Test Objectives

The KZN CSS provided provincial and local government alike an opportunity to engage citizens on their views regarding government services in general. The Office of the Premier in KwaZulu-Natal approached the NSO's provincial office in KZN, to conduct a citizen-satisfaction survey amongst the residents within the KZN province. This proposal was welcomed and it was inclusive of the mandate of the South African National Statistical System (SANSS). The mandate of SANSS includes the coordinates of statistics amongst the data producers within the country, as outlined in section 14 of the Statistics Act 6 of 1999 (Statistics Act, 1999).

Figure 6.3: Fieldworkers in the KZN CSS

The survey aimed to measure the attitudes and perceptions of residents in KwaZulu-Natal regarding the performance of provincial and local governments respectively. In terms of a range of criteria, these include the quality of service delivery and government support of citizen engagement and accountability. The KZN CSS data were made accessible to all the citizens on the levels of satisfaction concerning service delivery by government departments in the province. It served as an indicator of governance efficiency for Goal 6 (i.e. Governance) of the Provincial Growth and Development Plan (PGDP). The KZN CSS also provided data at the local municipal level, reflecting the views, opinions, and perspectives on service delivery in the province. The questionnaire was administered to adult members of households residing in selected dwelling units (DUs), drawn from a representative sample on a municipal level. The target population comprised of all persons aged 15 years and above, who are residents in private dwellings within the province of KwaZulu-Natal.

The NSO delivered a statistical report on the level of citizen satisfaction, which reflected the perceptions and the perspectives for service delivery improvement to the Office of the Premier. The survey priorities included the following, measurement of citizens' (Statistics South Africa, 2016d):

- rating of satisfaction with the overall performance of the provincial government;
- rating of satisfaction with the governance of the provincial government;
- ranking of provincial priorities and the performance of the provincial government departments;
- rating of the KwaZulu-Natal provincial government, according to the Batho Pele principles and the assessment of the provincial government's consultative processes, as well as the general awareness among the residents of the provincial government's programmes;
- rating of satisfaction with the overall performance of their local municipality;
- rating of importance of the municipal services and programmes; and
- rating of their satisfaction with the level and quality of the selected municipal services.

DDC was used to interview the individuals across the province. The three primary objectives of the survey concerning this research were to test and refine them:

- Use of digital handheld technology for the official survey-data collection,
- Electronic selection and the recruitment of the survey staff and
- The use of a GIS-enabled sampling frame.

6.3 PART B: DO [Apply INPUTS2]

Similar to the process flow described in Iteration 1 , the digital questionnaire was uploaded onto seven-inch Lenovo tablets. The instrument for gathering the secondary data was the actual digital version of the NSO's SS household survey questionnaire. The primary benefits of these mobile handheld devices included the real-time quality checking of the incoming data, the rapid production, and analysis of the data, as well as the accurate identification of the sampled dwelling units with an online/offline navigation app installed on each device. The manner in which the survey was conducted in the organisation represented a major shift from the standard SVC operational processes found in PAPDC. For example, the new processes included the implementation of DDC and the usage of a geo-referenced dwelling frame for sampling purposes. These initiatives, amongst others, resulted in the survey being conducted in less than six months from the planning to the release phase. This was a first for the NSO in terms of releasing the results quicker to the public.

6.4 PART C: STUDY [Determine OUTPUTS2]

There are three primary themes analysed during Iteration 2. These themes will be discussed separately to the secondary data analysis of the outcome dataset thereafter. The themes covered during Iteration 2 included: *speed*, *training* and *process flow and structural changes*. Iteration 2 with secondary variables looked at the following themes (see Table 6.2):

Table 6.2: Iteration 2 - KZN CSS DDC Themes

SECTION	LITERATURE CHAPTER	DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME
6.4.1	(2.5.6)	**SPEED**	DDC DATA COLLECTION SPEED IN BETWEEN ASSIGNMENT OF DATA TO APPROVAL BY HEADQUARTERS
			DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE
6.4.2	(2.5.7)	**TRAINING**	DDC TRAINING APPROACHES
			TRAINING OUTCOMES
6.4.3	(2.5.8)	**PROCESS FLOW AND STRUCTURAL CHANGES**	**KZNCSS1. PUBLIC ACCESS TO DATA:** PUBLIC STAKEHOLDERS ACCESS TO DATA
			KZNCSS2. ORGANISATIONAL STRUCTURE: DDC ORGANOGRAM
			KZNCSS3. ORGANISATIONAL PROCESSES: WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?
			KZNCSS3. ORGANISATIONAL PROCESSES: HOW DOES THE ORGANISATION TRANSITION TO ADAPT?

6.4.1 Speed

Speed covers all aspects of the SVC from “Need” to “Evaluate”. In terms of SASQAF, it referred to the quality dimension “TIMELINESS”. During the data-collection period, SS contained a survey-process tracking flow system. The tracking flow system assists project owners or survey implementers to monitor the progress of any given DDC survey in terms of time and speed. The SS survey workflow is summarised below (see Figure 6.4):

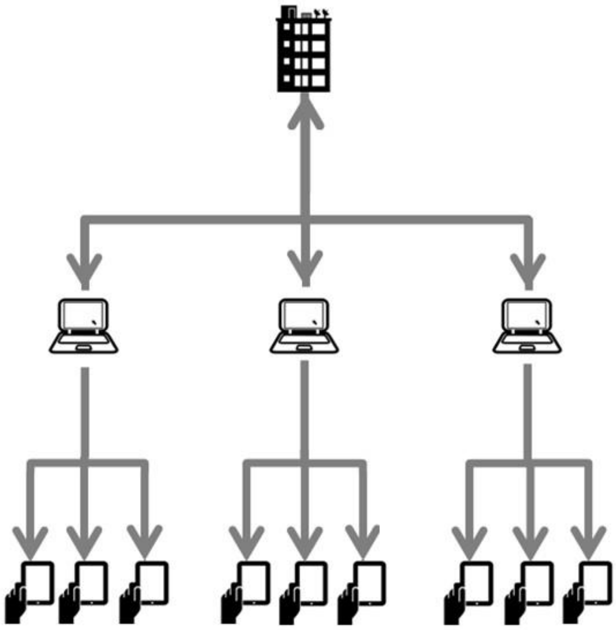


Figure 6.4: Overview of the survey workflow

HEADQUARTERS, designated by the building at the top of the figure, determine the households that need to be interviewed and their assignments to be distributed across the various team supervisors.

TEAM SUPERVISORS, denoted by the laptops in the middle of the figure first received these survey assignments (see Figure 6.5); and then allocated them to the INTERVIEWERS of their respective teams (see Figure 6.6).

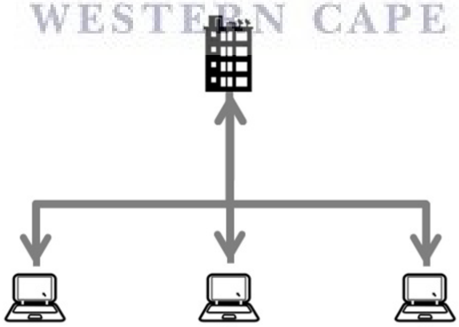


Figure 6.5: Supervisors receive assignments

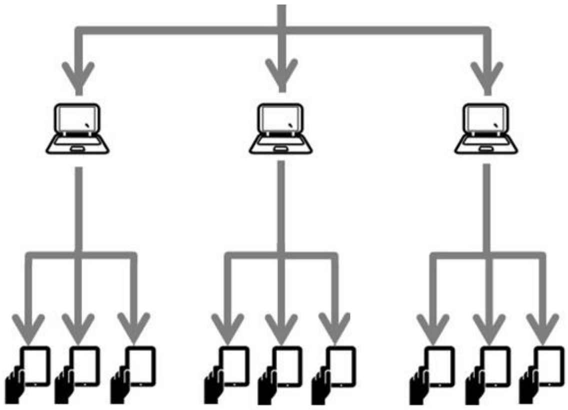


Figure 6.6: Supervisors allocate assignments

INTERVIEWERS, represented by the hands holding a tablet, receive the assignments from their team supervisors. They collect the data for those assignments, and they send the completed assignments back to their supervisors for review (see Figure 6.7).

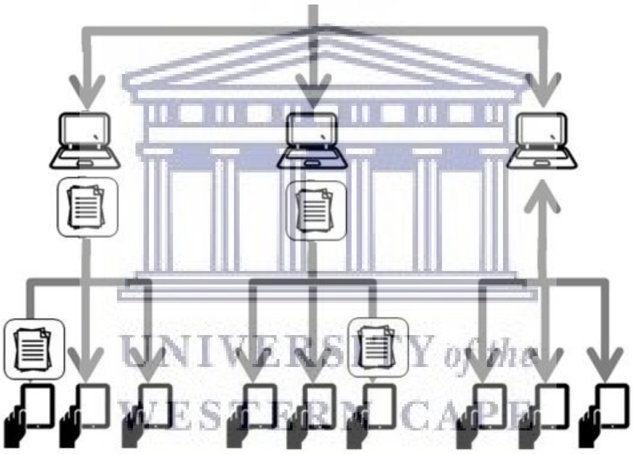


Figure 6.7: Interviewers send completed questionnaires to supervisors

TEAM SUPERVISORS, having received the completed questionnaires, review these questionnaires to confirm that all the questions have been answered accurately, coherently, and that they are plausible. After reviewing the completed assignments, the team supervisors either “approve” or “reject” these completed assignments (see Figure 6.8).

If a TEAM SUPERVISOR approves a completed assignment received from an INTERVIEWER, the assignment is sent to headquarters, as represented by the paper with a checkmark in Figure 6.8. If a TEAM SUPERVISOR rejects a completed assignment received from an interviewer, the assignment is returned to the INTERVIEWER initially responsible for completing it, as represented in Figure 6.9.

Receiving the rejected assignment, the INTERVIEWER must either correct it or provide explanatory notes on the inconsistent or implausible answers. When assignments are corrected, the INTERVIEWER sends them back to the TEAM SUPERVISOR for approval or rejection. This process continues until the assignments are completed with the highest level of quality, according to the TEAM SUPERVISOR.

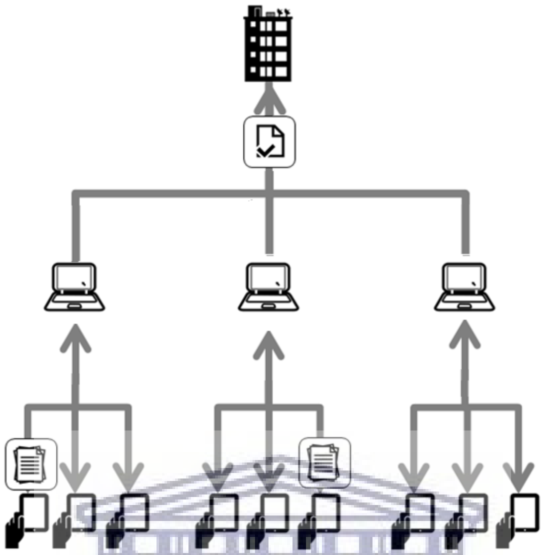


Figure 6.8: Supervisors reject or approve the questionnaires

As an assignment moves through the survey workflow, it takes on a different status at each stage, as illustrated in Figure 6.9. When an assignment is sent from headquarters to the supervisor, the assignment is assigned to the supervisor. This is assigned the status in SS as *"Supervisor Assigned*". When the supervisor allocates an assignment to a member of the team, that assignment is considered as*: "Interviewer Assigned*". When an assignment is completed by an interviewer and sent to the supervisor, the assignment is marked as completed*. The status becomes "Completed*". If the supervisor approves a completed assignment, the assignment is approved by the supervisor – *ApprovedBySupervisor*. If the supervisor rejects the completed assignment, that assignment takes on the status of ‘rejected by the supervisor’ – *RejectedBySupervisor*.

In the SS workflow below, the time taken in minutes and seconds between survey statuses was found to be reasonable considering the timeframes allocated and set by the field management in terms of speed. The *complete* action statuses with timestamps for each unique questionnaire are based on the following statuses:

- ApprovedByHeadquarters
- ApprovedBySupervisor

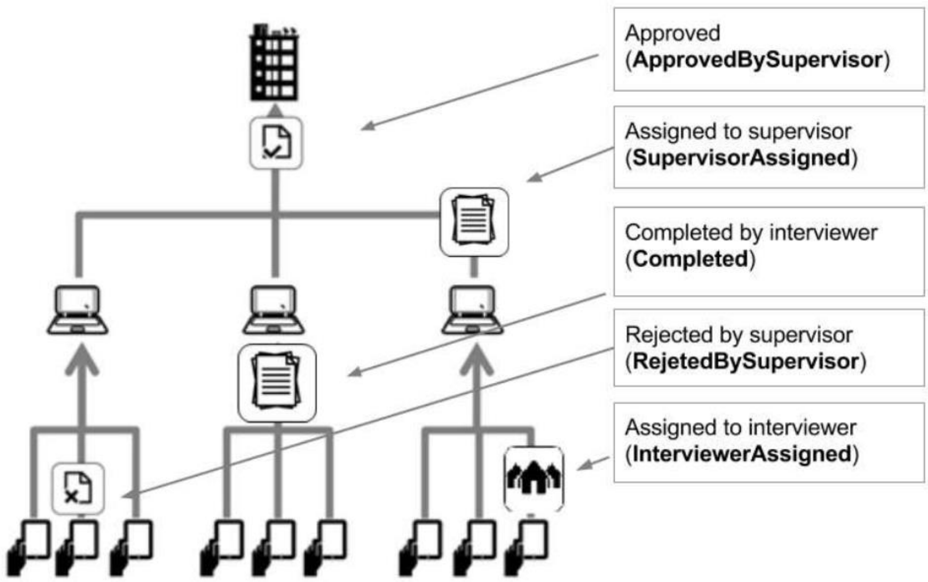


Figure 6.9: Status of an assignment in each stage of the survey workflow

- *****Completed[3]
- FirstAnswerSet
- InterviewerAssigned
- RejectedByHeadquarters
- RejectedBySupervisor
- Restarted
- *****SupervisorAssigned

During the quantitative analysis of the data for speed between the status levels when the initial assignment was created on the device by the Supervisor, "*SupervisorAssigned*"; and the period in between when the Survey Officer collected and completed the assignment. "*Completed*", (Date, Time) timestamps indicates that, on average, the interviewers initially took longer to complete the surveys (i.e. completing the questionnaires during the data collection); but they improved over the survey period; as was also the case with Iteration 1.

In terms of the speed or time taken to release the data, the release date to the public was as scheduled as 04th February 2016 @ 10:00. The secondary data variables include:

- NSO STAKEHOLDERS DATA ACCESS
  - DDC Data-Collection Speed In Between the Assignment of Data to Approval By the Headquarters.
  - Data Dissemination, According to the Planned Schedule.

[3]The actions denoted by * were analysed

The data were released, as per the planned scheduled date to the public (Statistics South Africa, 2016b). Although Table 6.3 shows no "Improvements" or "Gaps" in the existing PAPDC SASQAF Framework, it should be noted that the speed of collecting and disseminating data to the users decreased significantly. Normally statistical releases of this magnitude is released a year or more later when using a PAPDC process, however the data for the KZN CSS was released within 3 months after the completion of data collection. When comparing DDC with PAPDC, DDC released the data more quickly than PAPDC surveys.

Table 6.3: Speed mapping table - KZN CSS

THEME	SUB THEME	IMPACT	NATURE OF IMPACT	VARIABLE	MAPPED	IMPROVEMENT	GAP
Speed	Reducing Survey Time	Internal	More timeous downstream processes of SVC	SupervisorAssigned, Completed, Date, Time	(TIME1; Di4)		
			Meeting data collection deadlines	Planned Release Date	(TIME2; Di4)		
				Actual Release Date	(TIME3; ALL)		
		External	Outdated data, the usefulness of data	Timely dissemination	(TIME4; Ne4)		
			Satisfy user needs				

(Source: Author)

6.4.2 Training

A 3-tier cascade training approach was used for the project. The training team consisted of Head Office, Provincial Office, and District Office personnel. Subject matter specialists were identified from the relevant areas within the NSO to be capacitated on Train-the-Trainer skills as national trainers. Subject matter specialists were responsible for training national trainers. National trainers cascaded training to the Assistant District Survey Co-ordinators (ADSCs) within the KZN districts. The ADSCs finally trained the contract INTERVIEWERS at the District Offices. Training was characterized by content theory, demonstration, and practical exercises (Statistics South Africa, 2016e). The following secondary variables were analysed for the training:

- DDC TRAINING APPROACHES
  - What training techniques were employed for DDC vs. PAPDC?
- TRAINING OUTCOMES
  - The effectiveness of training for DDC vs. PAPDC

Training covered the following aspects of the SVC (Design, Build, and Collect) and SASQAF (Methodological Soundness), see Table 6.4:

Table 6.4: Training mapping - KZN CSS

SASQAF Quality Dimensions			Statistical Value Chain		
			Design	Build	Collect
Number	Dimension	SASQAF Code	De4	Bu5 - Training	Co1 - Training
	Methodological soundness	**METH3 - Improvement to this QD to include Training Approaches and Outcomes (*Gap4)**	"Training only discusses the most appropriate data collection method(s) and instrument(s)" (Statistics South Africa, 2010b). Require "Training Approaches and Outcomes" based on data collection method(s) and instrument	When finalising production systems, "Training Approaches and Outcomes" should be included in technical reports for all users	Execution of training outcomes and approaches needs to be monitored and measured in terms of desired outcomes of trained staff

(Source: Author)

In this dimension/ theme, there is a ***Gap4** that should be included in the development of the DDC Quality Assurance framework.

- ***Gap4** **(DDC TRAINING APPROACHES – development of the DDC training approaches):** Standard PAPDC training processes could not be deployed. DDC training requires more practical training, and the modes of training had to be assessed in terms of the training's effectiveness. There was the need to add a quality indicator "Methodological Soundness 3.10.

In addition to the gap identified on the existing SASQAF framework and SVC, METH3 requires an ***Improvement** to include the "Training Outcomes" and the "Training Approaches", where it intersected with the SVC phases (De4, Bu5, and Co1), respectively:

- ***Improvement4 -** (METH3(10); De4)
- ***Improvement5 -** (METH3(10); Bu5)
- ***Improvement6 -** (METH3(10); Co1)

The tools used for the application of the training methods for the DDC differ from the PAPDC. DDC training reports indicated that DDC catered for mirror-casting of the digital questionnaire with an additional projector displaying the content concurrently. While training the interviewers on the questionnaire content on the one screen, the second screen should display the digital device navigation of the questionnaire concurrently. The SASQAF does not refer to the data-collection staff training within any of its Quality Dimensions. For the SVC, training is mentioned in both the "Build" and "Collection" phases, respectively. The gap is in the SASQAF Quality Dimension Method 3(10), which should be mapped to the respective SVC: Bu5 (Build phase), Co1, (Collect phase) with an additional SVC expansion in the De4 (Design phase) (See Table 6.5.)

Table 6.5: Training mapping table – KZN CSS

THEME	SUB-THEME	IMPACT	NATURE OF IMPACT	VARIABLE	SVC	IMPROVEMENT	GAP
Training	Training Data Collection Staff to collect data	Internal	Poorly collected statistics (data for surveys)	DDC Training Approaches	Bu5	**(METH3(10); De4) (METH3(10); Bu5) (METH3(10); Co1)**	**METH3 (10)**
			Lack of credibility by NSO	DDC Training Outcomes	Co1		
			Bias data (not fit for any purpose)				
		External	Negative Organisational reputation				

(Source: Author)

6.4.3 Process flow and structural changes

6.4.3.1 KZNCSS1. PUBLIC STAKEHOLDER ACCESS TO DATA:

Access to the KZN CSS report and the metadata was available from the NSO, as per the dissemination schedule. It was reported on the NSO that the public had access to the results of the survey (Statistics South Africa, 2016c).

The following were analysed:

Secondary Data Variable:

- PUBLIC STAKEHOLDER ACCESS TO DATA
    - Are the data available to the South African Public?

6.4.3.2 KZNCSS2. ORGANISATIONAL STRUCTURE:

For DDC, during this project, the organisation deployed professional staff from across the organisation. Working divisions included Information Technology, Business Modernisation, Data Processing, Fieldwork Operations, Content and Analysis, Methodology, Geography, Human Resources, Finance, Monitoring and Evaluation, and Provincial Staff, who were led by a project director with support from the national Project Management Office. Contract staff was employed to collect the data as

INTERVIEWERS; while some permanent staff was used as the field supervisory support. Given that this project was the first DDC project, a change to the traditional PAPDC organisational structure was required.

The following were analysed:

Secondary Data Variable:

- DDC ORGANOGRAM
  - How would the NSO be affected in terms of staffing and the relevant functions/roles?

6.4.3.3 KZNCSS3. ORGANISATIONAL PROCESSES

Newly developed systems include the: “Learner-Management System”, used to conduct assessments for the recruitment of the staff and "SASBI", a customised SAS reporting system was used for survey monitoring. "Oryx Maps" was used as a navigation tool to the selected sampled dwellings. Not all of these systems were functional for PAPDC processes previously. Although the SASQAF mentions that “*The prerequisites of quality refer to the institutional and organisational conditions that have an impact on the data quality. These include the institutional and legal environment, and the availability of human, financial and technological resources”,* There are ***Gaps** that need to be included in the development of the DDC Quality Assurance Framework.

- ***Gap5 (WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?):**
  - It was reported that the DDC process is not complicated, in comparison to PAPDC processes. However this hypothesis for the rest of the staff within the NSO has not been tested.
- ***Gap6 (HOW DOES THE ORGANISATION MANAGE TO ADAPT TO DDC?):**
  - Develop a DDC transition strategy for all the affected staff within the organisation.
  - This strategy constantly changes, based on the progressive uptake of all the staff to progress to the best practices sought by the NSO.
  - Best practices is a continuous improvement and developmental process conducted by organisations through constant researching, testing and deploying solutions, processes or methodologies.
  - This should be incorporated in the “pre-need” phase on the standard SVC. The NSO needs to prepare the structural human resources (staff)

beforehand through internal organisational work assessments to adjustment to the new DDC-related processes.

The following were analysed:

Secondary Data Variable:

- WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?
  - Develop and train DDC process flows
- HOW DOES THE ORGANISATION TRANSITION TO ADAPT?
  - Develop a DDC transition strategy for all the affected staff

6.5 SECONDARY QUANTITATIVE DATA ANALYSIS

The secondary data source information in this section was reported directly from the dataset and the 2015 KwaZulu-Natal Citizen Satisfaction Survey: Technical Report (Statistics South Africa, 2016d). Table 6.6 displays the measurements analysed during Iteration 2.

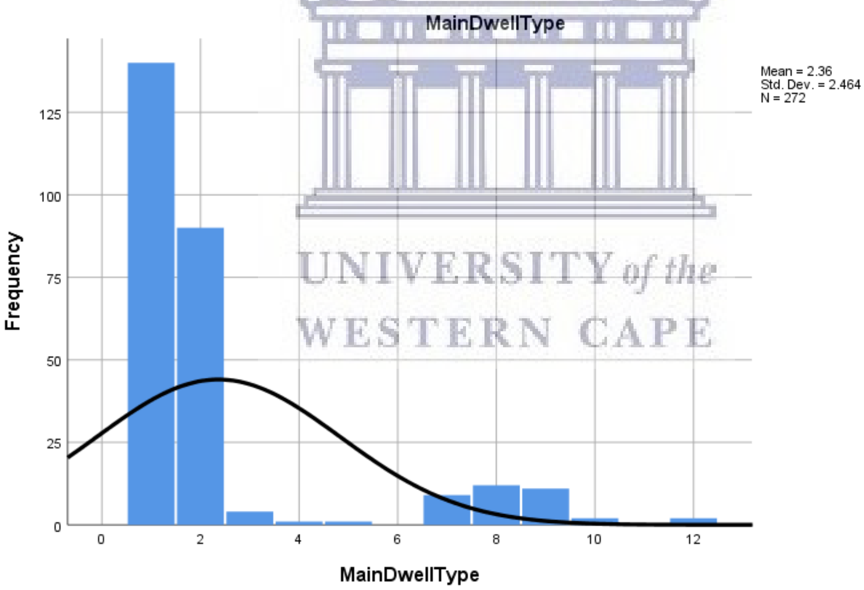


Figure 6.10: Iteration 2 - Main Dwelling Type

Figure 6.10 displays the measurement of “Main Dwelling Type” and the distribution of the responses.

Table 6.6: Iteration 2 – Secondary quantitative descriptive statistics on QA measurements

		time_elapsed	Cost	MainDwellType	SubsDwell	TenureStat	WaterSource	Distance Water	WaterSupplier	Toilet	ToiletLocation	ToiletShared	Refuse	ElectrSupplier	Energy Cook	EnergyLight	EnergyWaterHeat	EnergySpaceHeat
N	Valid	384	384	272	384	272	272	135	272	272	260	260	272	215	272	272	272	272
	Missing	0	0	112	0	112	112	249	112	112	124	124	112	169	112	112	112	112
Median		1:56:14	250.00	1.00	.00	4.00	4.00	2.00	1.00	4.00	2.00	2.00	5.00	3.00	1.00	1.00	1.00	5.00
Mode		0:00:15[a]	250	1	0	4	2	1	1	4	2	2	5	3	1	1	1	1
Skewness		3.694		2.139	0.607	-1.233	0.451	. 0829	1.202	. 0814	-0.104	-0.939	-1.121	-0.339	0.858	1.543	1.197	0.724
Std. Error of Skewness		0.125	0.125	0.148	0.125	0.148	0.148	0.209	0.148	0.148	0.151	0.151	0.148	0.166	0.148	0.148	0.148	0.148
Kurtosis		15.171		3.334	-1.485	1.704	-1.014	-0.473	-0.269	1.179	0.214	-1.126	-0.467	0.050	-0.545	0.521	1.137	-0.706
Std. Error of Kurtosis		0.248	0.248	0.294	0.248	0.294	0.294	0.414	0.294	0.294	0.301	0.301	0.294	0.330	0.294	0.294	0.294	0.294
Range		895:18	0	11	3	6	12	4	5	9	2	1	5	5	9	9	10	10
Percentiles	25	0:06:41	250.00	1.00	0.00	4.00	2.00	1.00	1.00	2.00	2.00	1.00	4.00	2.00	1.00	1.00	1.00	1.00
	50	1:56:14	250.00	1.00	0.00	4.00	4.00	2.00	1.00	4.00	2.00	2.00	5.00	3.00	1.00	1.00	1.00	5.00
	75	16:34:31	250.00	2.00	2.00	5.00	7.00	3.00	3.00	4.00	2.00	2.00	5.00	3.00	5.00	1.00	5.00	5.00

a. Multiple modes exist. The smallest value is shown

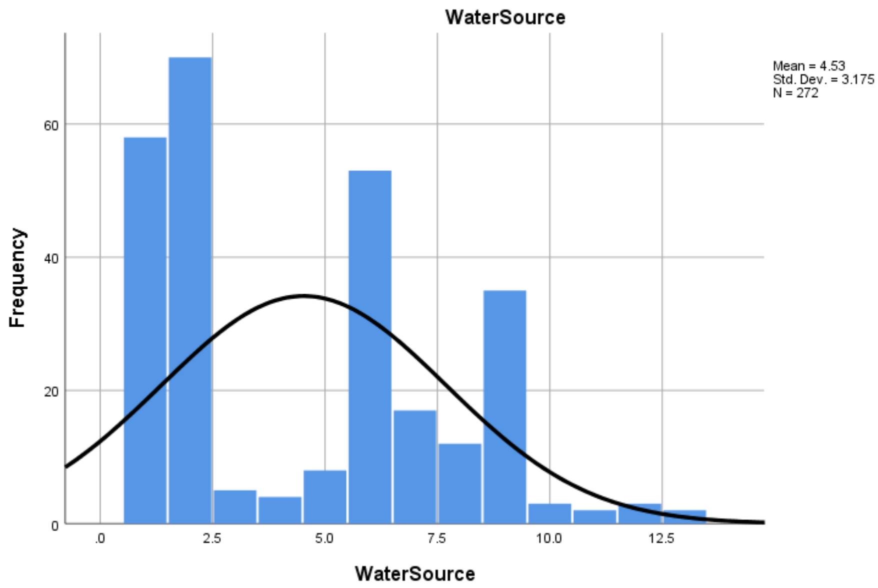


Figure 6.11: Iteration 2 - Water Source

Figure 6.11 displays the measurement of the Water Source and the distribution of the responses.

6.5.1 Raw data file

The raw data file, as captured from SS contained all the personal and household data for the survey. These include those records that were excluded during the editing and the imputation process. The file contained 64 602 records with all the demographic information complete and valid after imputation, amongst which there were 10 duplicated records.

Table 6.7: Distribution of person result codes for unique records

Person Result Code	Label	Number of Unique Persons	Percentage
1	Response	38 845	60,1%
2	Non-Response	104	0,2%
3	Initially missing Age, imputed	649	1,0%
6	Age less than 15 years	22 564	34,9%
7	Non-Usual Resident	2 430	3,8%
Total		**64 592**	**100,0%**

The design of the survey required person level enumeration; therefore, personal level result codes were defined. Table 6.7 above shows the distribution of the records by the result codes for all the personal records. The Full Person file contained 64 592 unique personal records from 16 955 households within 15 204 DUs.

6.5.2 The CSS 2015/2016 final data file

The relationship between the personal records, the households, and the dwelling units in the final file is such that the 39 233 persons were from 16 488 households within 15 085 DUs. The KwaZulu-Natal CSS 2015 compared well with other data sources such as the previous censuses (1996, 2001, and 2011) and the 2014 General Household Survey. This was predominantly in terms of demographics and household services.

6.5.3 Response Categories

The final result codes were mapped to the three response status categories showed in Table 6.7 below, where 1=Respondent, 2=Non-respondent, and 3=Out-of-scope. New final result codes were defined during the out-of-scope verification process. Code ‘24' was assigned to records verified to be general non-response from the initial out-of-scope file and code ‘38’ was assigned to those records with navigation errors, or where the image had shifted. Table 6.8 compared the distribution of the result codes on the final household file between the original result codes and the final result codes, after re-classification. It is clear from the Table 6.8 that the result codes 31 to 36 were mainly affected, and these were reclassified as either 24 or 38. A total of 2,232 (9.72%) records were re-classified as general non-response (24), decreasing the percentage of records classified as out-of-scope. 101 records were re-classified as 38, but the remaining were out-of-scope.

Table 6.8: Distribution of the household final result codes

			Pre-reclassification		Post-reclassification		
	Household Result	**Final Result Code**	**Frequency**	**Percentage**	**Frequency**	**Percentage**	
RESPONSE	Completed and Partly Completed	11, 12	16 446	71,6%	16 446	71,6%	71,6%
NON-RESPONSE	Non-contact, Refusal and Other Non-response	21,22, 23	1 587	6,9%	1 587	6,9%	
	Records verified to be non-response from the initial OOS file	24	0	0,0%	2 232	9,7%	16,6%
OUT OF SCOPE	Unoccupied dwelling	31	1 425	6,2%	370	1,6%	
	Vacant dwelling	32	636	2,8%	234	1,0%	
	Demolished	33	672	2,9%	433	1,9%	
	New Dwelling Under Construction	34	115	0,5%	58	0,3%	
	Status change	35	150	0,7%	87	0,4%	
	Listing (Classification) Error	36	1 543	6,7%	1 026	4,5%	
	Dwelling Unit Result Code 1 was Missing	37	390	1,7%	390	1,7%	
	Navigation Error / Image Shift	38	0	0,0%	101	0,4%	11,8%
		Grand Total	**22 964**	**100%**	**22 964**	**100%**	**100%**

Table 6.9: Distribution of the household response codes

	Pre-reclassification		Post-reclassification	
Response Code	**Frequency**	**Percentage**	**Frequency**	**Percentage**
1 - Respondent	16 446	71,6%	16 446	71,6%
2 - Non Respondent	1 120	4,9%	3 352	14,6%
3 - Out of Scope	5398	23,5%	3 166	13,8%
Grand Total	**22 964**	**100%**	**22 964**	**100%**

Table 6.9 above displays the final distribution of the household responses between the 3 categories. Table 6.10 below indicates the overall survey response rate of 83.1% for Iteration 2.

Table 6.10: OUTPUT 2 - Response rate

Response Rate Formula	RC 11 + RC 12 / (total HH) - sum(OOS)
Completed (RC11) + Partly completed (RC12)	16446
Total HH	22964
Sum (OOS)	3166
Response Rate	**83,1%**

6.5.4 Data quality indicators

The statistical precision of a sample statistic can be defined as the closeness with which it can be expected to estimate the relevant population values. The precision can be estimated by using standard errors and/or coefficients of variation, which estimate the amount of variability that can be expected from the estimates. Several factors can affect the accuracy of the survey estimates; the sample size is one of them. To ascertain the accuracy of the CSS estimates, the standard error (SE), the coefficient of variation (CV) and the confidence intervals for selected variables were calculated. The coefficient of variation is the ratio of the standard error of a survey estimate to the value of the estimate itself. It is a measure of the relative variability of the estimator (Minister of Industry, 2017).

The smaller the CV of an estimate, the more precise the estimate. Figure 6.12 illustrates a model that is generally used to determine the reliability of the survey estimates, based on the CVs obtained for the survey estimates.

It was clear, based on the cut-off levels as illustrated in Figure 6.12 and Figure 6.13, that the estimates at the provincial level were reliable and thus acceptable for publication.

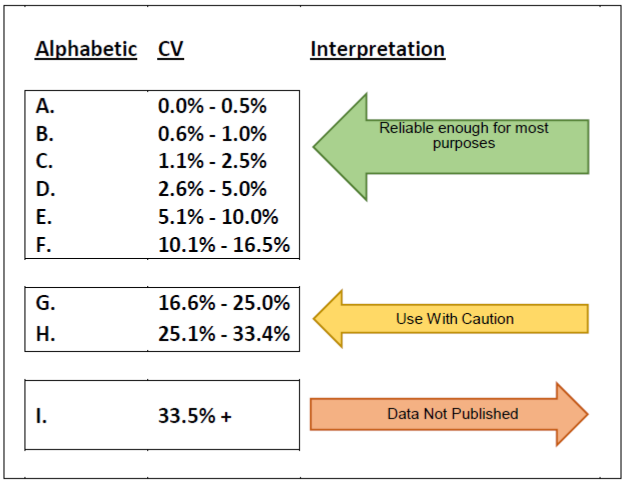


Figure 6.12: Indicators of sampling variability

When cross-tabulating the satisfaction with the overall performance of the provincial government with a sub-population, such as population group, several estimates move into an area of 'use with caution' or 'do not publish' based on the cut-off margins above. This was most likely because of the low number of persons of a particular population group in the sample having the relevant level of satisfaction (see Figure 6.14).

Overall Performance Rating	Raw count of persons	Weighted count of persons	Percentage	Standard error of percentage	Lower confidence limit of percentage	Upper confidence limit of percentage	Coefficient of variation of percentage
Very Dissatisfied	3 599	452 341	6,36	0,32	5,73	6,99	5,05
Dissatisfied	12 167	2 108 317	29,64	1,01	27,66	31,62	3,41
Somewhat Satisfied	10 937	2 224 656	31,28	0,76	29,78	32,77	2,44
Satisfied	11 426	2 236 467	31,44	1,12	29,25	33,64	3,56
Very Satisfied	492	91 127	1,28	0,18	0,93	1,63	13,79
Total	**38 621**	**7 112 909**	**100,00**				

Figure 6.13: Level of satisfaction with the overall performance of the KwaZulu-Natal provincial government, quality indicators at the provincial level

Because the quality indicators are largely dependent on the sample size, this implies that the smaller the sample size on which the estimates are based, the higher the standard errors and the coefficients of variation.

Population Group	Overall performance rating	Raw count of persons	Weighted count of persons	Percentage	Standard error of percentage	Lower confidence limit of percentage	Upper confidence limit of percentage	Coefficient of variation of percentage
Black/ African	Very Dissatisfied	3 494	433 703	6,87	0,36	6,17	7,58	5,25
	Dissatisfied	11 627	1 935 154	30,67	1,10	28,51	32,83	3,59
	Somewhat Satisfied	10 309	1 889 672	29,95	0,72	28,53	31,37	2,41
	Satisfied	10 858	1 975 935	31,32	1,19	28,98	33,66	3,81
	Very Satisfied	462	74 702	1,18	0,13	0,93	1,44	11,00
	Total	**36 750**	**6 309 166**	**100,00**				
Coloured	Very Dissatisfied	10	1 641	1,95	0,94	0,05	3,84	48,45
	Dissatisfied	60	11 935	14,17	4,90	4,35	23,99	34,56
	Somewhat Satisfied	75	30 212	35,87	6,18	23,46	48,27	17,24
	Satisfied	98	38 825	46,09	4,60	36,87	55,31	9,97
	Very Satisfied	3	1 624	1,93	0,95	0,02	3,83	49,22
	Total	**246**	**84 237**	**100,00**				
Indian/ Asian	Very Dissatisfied	29	6 622	1,49	0,55	0,40	2,59	36,98
	Dissatisfied	215	82 395	18,57	2,66	13,27	23,86	14,35
	Somewhat Satisfied	338	216 565	48,80	4,76	39,34	58,26	9,76
	Satisfied	288	136 221	30,70	4,67	21,42	39,97	15,21
	Very Satisfied	8	1 964	0,44	0,28	0,00	1,00	63,06
	Total	**878**	**443 766**	**100,00**				
White	Very Dissatisfied	66	10 375	3,76	1,00	1,79	5,73	26,45
	Dissatisfied	265	78 834	28,59	3,53	21,61	35,57	12,33
	Somewhat Satisfied	215	88 207	31,99	3,50	25,05	38,93	10,96
	Satisfied	182	85 486	31,00	4,16	22,77	39,23	13,41
	Very Satisfied	19	12 838	4,66	2,24	0,22	9,09	48,09
	Total	**747**	**275 740**	**100,00**				

Figure 6.14: Level of satisfaction with the overall performance of the KwaZulu-Natal provincial government, quality indicators at the provincial level by population group

Figure 6.15 below illustrates the impact of collapsing some of the variable categories, according to their similarities. To illustrate the said impact of collapsing the variable categories, the satisfaction with the overall performance of the provincial government was re-tabulated with collapsed rating categories. It is, therefore, clear that collapsing the response categories improves the realised sample size; and therefore, this improves the measure of the standard errors and the coefficients of variation.

Local Municipality Name	Overall Performance Rating	Raw count of persons	Weighted count of persons	Percentage	Standard error of percentage	Lower confidence limit of percentage	Upper confidence limit of percentage	Coefficient of variation of percentage
Dannhauser	Dissatisfied	122	9 541	15,25	2,00	11,15	19,35	13,09
	Somewhat Satisfied	486	39 738	63,52	4,07	55,16	71,88	6,41
	Satisfied	167	13 283	21,23	4,35	12,30	30,16	20,5
	Total	**775**	**62 561**	**100,00**				
KwaDukuza	Dissatisfied	205	50 074	26,45	3,61	19,19	33,72	13,66
	Somewhat Satisfied	273	67 029	35,41	2,35	30,67	40,14	6,65
	Satisfied	302	72 194	38,14	3,57	30,97	45,31	9,35
	Total	**780**	**189 298**	**100,00**				
Ulundi	Dissatisfied	274	73 621	61,16	6,44	48,11	74,21	10,52
	Somewhat Satisfied	153	34 120	28,34	4,42	19,39	37,30	15,58
	Satisfied	58	12 634	10,50	2,81	4,80	16,19	26,76
	Total	**485**	**120 375**	**100,00**				
Umzimkhulu	Dissatisfied	224	35 939	32,01	3,40	25,13	38,88	10,63
	Somewhat Satisfied	234	36 098	32,15	2,53	27,03	37,27	7,88
	Satisfied	259	40 246	35,84	2,88	30,02	41,66	8,03
	Total	**717**	**112 283**	**100,00**				

Figure 6.15: Level of satisfaction with the overall performance of the KwaZulu-Natal provincial government, quality indicators at the local municipal level

Furthermore, similar observations were made in the standard errors and the coefficients of variation for the other subpopulations and the variables of interest.

6.6 PART D: ACT [Make CHANGES2]

During the final phase of the PDSA, there were key data quality improvement items that needed to be carried forward into Iteration 3. The training methods needed to be refined and scaled up to cater for the increasing sample size. The scope for Iteration 2 was confined to one province only. The next iteration (Iteration 3) covered 9 provinces across South Africa, which increased the risk of recruited staff having little to zero experience on DDC significantly. The training, process flow and structural changes done in Iteration 2 were important considerations to take forward to Iteration 3.

6.6.1 Qualitative overall debriefing lessons learnt

Individuals involved in the entire KZN CSS process (all phases of the SVC) including the INTERVIEWERS were tasked to debrief their findings throughout the process of training, publicity and data collection for the survey. Table 6.11 is an example of a debriefing report of experiences during the survey:

Table 6.11: KZN CSS debriefing notes (example)

What Worked (Good)?	What Didn't work (Bad)?	Challenges encountered
•It was easy to access the CAPI site •Accessing Questionnaires and navigation •The line up of sections and Questions (was easy to follow) •Synching/sending of completed questionnaires •CAPI indication of invalid or unanswered questions •CAPI also makes provision for spelling checks. •The system also allowed for administration of multiple households •The option of rejecting questionnaires allowed FWs to rectify errors committed during data collection	•Display of gadget (device) to the public made FWs / (INTERVIEWERS) to vulnerable to robberies. •The screen is dark outdoors especially when the sun is shining •Navigation causes delays (long route, no street names) •Losing or deleting of information during synching •Repetitions of some questions •HH (Household) final result Code vs Person result code created some confusion to some of the Interviewers	•Device froze in some cases during the interview •Making corrections on the rejected questionnaires sometimes did not allow those questionnaires to be synched for the second time. •Age limit on perceptions questions

6.6.2 Qualitative overall DDC gaps and improvements identified

Thematic areas contained in the SASQAF were expanded on and, in some areas, gaps and improvements were identified. These were recommended to be included in the development of the DDC Quality Assurance Framework (see Table 6.12). In the following chapter (Chapter 7), an even larger sample-sized DDC implementation will be discussed. The primary focus of Iteration 2 was to apply DDC on a larger scale and test other key themes. Secondly, to also identify the resources required across the organisation to adapt to a DDC process.

Table 6.12: KZN CSS test changes to SASQAF

DDC THEME	IMPROVEMENT REQUIRED	GAPS
TRAINING	***Improvement4** - (METH3(10); De4): Training only discusses the most appropriate data collection method(s) and instrument(s) (Statistics South Africa, 2010b). Require "Training Approaches and Outcomes" based on data collection method(s) and instrument ***Improvement5** - (METH3(10); Bu5): When finalising production systems, "Training Approaches and Outcomes" should be included in technical reports for all users ***Improvement6** - (METH3(10); Co1): Execution of training outcomes and approaches needs to be monitored and measured in terms of desired outcomes of trained staff	***Gap4** - DDC TRAINING APPROACHES - develop DDC training approaches
PROCESS FLOW AND STRUCTURAL CHANGES		***Gap5** (WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?): It was reported that DDC is not complicated in comparison to PAPDC processes. ***Gap6** (HOW DOES THE ORGANISATION TRANSITION TO ADAPT TO DDC?):

(Source: Author)

6.6.3 Quantitative questionnaire changes

As in the previous iteration (Iteration 1), the actual household questionnaire was developed by the NSO on SS. The standard DDC questionnaire for Iteration 2 contained the same 15 variables as in Iteration 1. For Iteration 3, these variables were kept constant.

6.7 CHAPTER SUMMARY

Iteration 2 contained a significantly larger sample size than in Iteration 1, i.e. from a sample of 100 dwellings to 20 819 dwellings. From this dataset, a sample of 384 dwellings was analysed for the quantitative analysis of Iteration 2 of this study. The rationale for selecting 384 dwellings as secondary data for Iteration 2 is explained in Chapter 3 (3.3.1). The survey sample size in Iteration 2 was larger than in Iteration 1. It was deemed relevant; as the pilot study sample sizes are typically based on the pragmatics of the recruitment and the necessity for examining the feasibility (Leon, Davis, & Kraemer, 2011). The KZN CSS was conducted in the KwaZulu-Natal province, and it covered a larger spatial area. It was earmarked as the first full digital implementation of DDC conducted by the NSO.

In terms of DDC themes, Iteration 2 identified three improvements areas to be unpacked in the existing SASQAF regarding TRAINING and one gap. Two gaps were listed for the PROCESS FLOW AND STRUCTURAL CHANGES. The results of the secondary data captured displayed a higher proportion response than did Iteration 1; while the overall quality indicators on the measurements (variables) displayed very satisfactorily, as indicated. After studying the outcome of the data, the questions to be taken forward to the next iteration (Iteration 3) were as follows:

- Scaling and increasing the sample size of DDC across a broader geographical space (entire South Africa)
- Identifying the related costs, the perception of quality, security, speed, training and the maturity of the organisation towards accepting the technology for collecting household survey data

CHAPTER 7: ITERATION 3: COMMUNITY SURVEY 2016

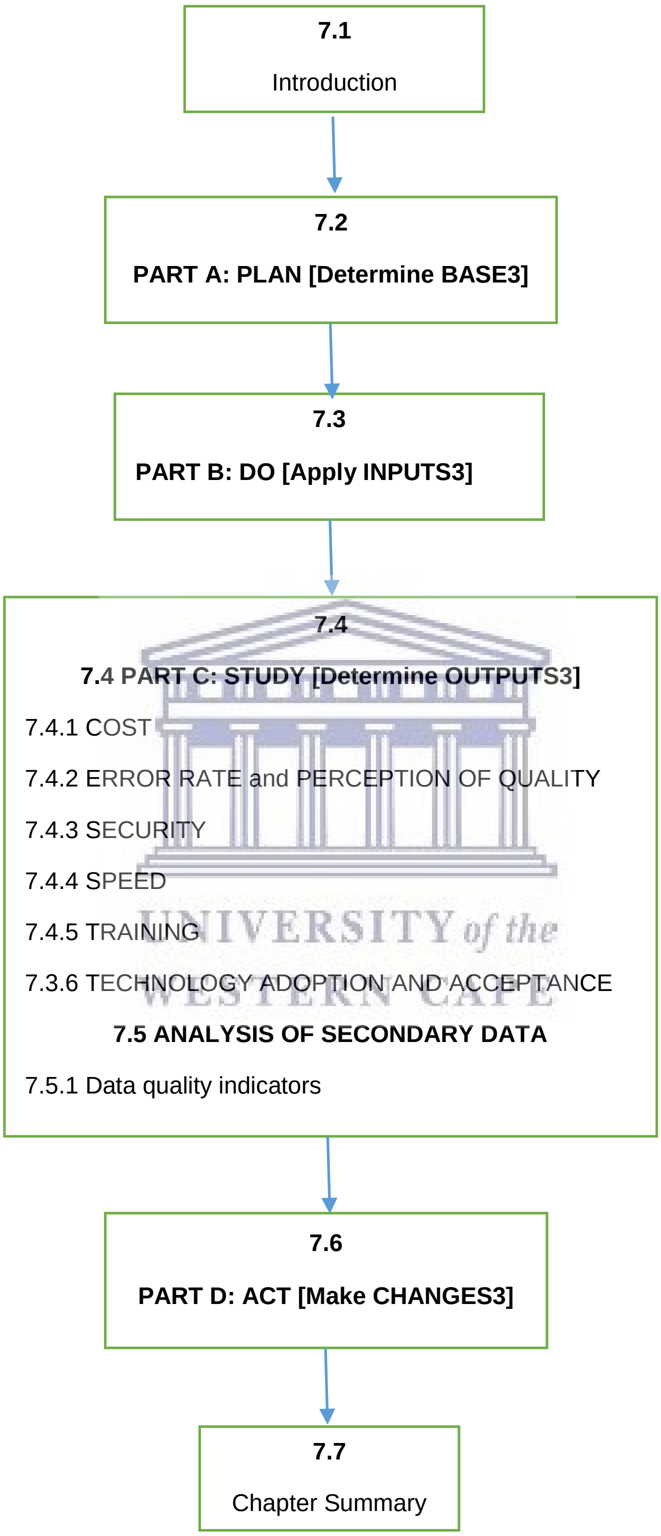


Figure 7.1: Chapter 7 layout

CHAPTER 7: ITERATION 3 - COMMUNITY SURVEY 2016

7.1 INTRODUCTION

Iteration 3, the third and final iteration of the study is characterised by the largest sample size spread spanning the entire country of South Africa.

Using the PDSA cycle (see Figure 7.2), Iteration 3 or “adjusted BASE 2” dataset for DDC is determined in conjunction with “BASE 3” to form the PLAN phase (see section 7.2). Applying the “INPUTS 3” based on the themes from the literature related to quality aspects, which are to be considered and applied for the official household survey collection results in the “DO” phase discussed in section 7.3. The net result of applying the inputs is the “OUTPUTS 3”, or the outcomes of the captured data as reflected in section 7.4.

These secondary data are once more studied to determine the level of quality with respect to the “Responses” and “Non-Responses”. The secondary data (384 records) is further analysed from a 10% random sample of (137081) out of the 1370809 households enumerated. Once the “adjusted BASE 2” was established and the “OUTPUTS 3” studied, based on the resultant data quality, the final “CHANGES” were made and discussed in section 7.6.

In addition to analysing the secondary data, primary data were collected from four experts. They participated in Iteration 3 as well as during Iteration 1. The analysed data, provided an indication of their procedural experiences, based on the newly introduced DDC process.

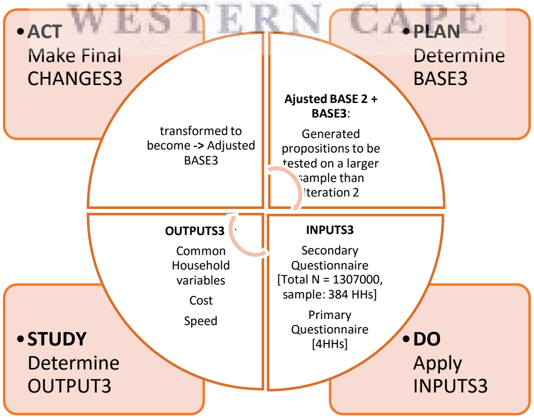


Figure 7.2: Iteration 3 - Plan-Do-Study-Act cycle

Quality assurance measures are embedded in all the phases of censuses and surveys. These include the questionnaire development and the design, the data collection, the data capturing (both manual and automated), coding and analysis. All the quality procedures and measures are aimed at promoting consistency and the improvement of the data quality, making them “fit for use”.

7.2 PART A: PLAN [Determine BASE3]

7.2.1 Iteration 3 Background

The Community Survey (CS) 2016 Quality Assurance processes were applied in two phases. The first phase involved creating validation rules and enabling conditions during the electronic questionnaire development and design. Several validation rules were used to avoid inconsistencies between those variables that were linked. For example, age is linked to the date of birth; sex is linked to fertility; employment questions are linked to the working-age etc. Phase two of the quality assurance involved the application of pre-determined editing specifications/rules, aimed at identifying the invalid values, and the missing values. Errors relating to the structure were resolved, by using a structural-query language (SQL) in an Oracle dataset. The SAS software was used programmatically to resolve content-related errors.

7.2.2 Iteration 3 Test Objectives

The goal of CS 2016 was to provide indicators that would inform the implementation, monitoring and the evaluation of development programmes for communities at the local municipality level in South Africa (Statistics South Africa, 2016a).

The key objectives of CS 2016 were to:

- provide an estimate of the population count by the local municipality.
- provide an estimate of the household count by the local municipality.
- provide the measurement of the demographic factors, such as fertility, mortality, and migration.
- provide the measurement of the socio-economic factors, such as employment, unemployment, and the extent of poverty in these households.
- produce a measurement of the access to facilities and services, such as piped water, sanitation, and electricity for lighting.

7.3 PART B: DO [Apply INPUTS3]

An interview instrument was developed by the researcher to capture primary data for Iteration 3. This instrument excluded the variable "Are you currently working fulltime / parttime" given that the NSO experts were mostly involved in previous iterations (Iteration 1 and/or Iteration 2). As discussed in Chapter 3, this primary data constitute perceptions of NSO workers regarding the DDC process and the quality of data DDC produces. Second, an instrument was developed by the NSO for conducting the actual household survey. This household survey instrument was developed using the SS software package, discussed in section 5.2.1. Thirdly, data was gathered using the instruments described above. Four NSO workers were interviewed using the primary data collection instruments (see Appendix B). A total of 1370810 households were surveyed using the household survey tool developed by the NSO in SS. A sample of 384 records was used to analyse the data (see Table 7.1). The data gathered in the "DO" phase of Iteration 3 is presented according to DDC quality themes in the sections to follow.

Table 7.1: Iteration 3 Primary and secondary data sources

Data Source	Title	Appendix	Origin	Sample size	Method
Iteration 3 - Primary data	CS 2016 RESEARCH SURVEY QUESTIONNAIRE	B (excluding variable: Are you currently working fulltime / part-time)	Compiled by author	n = 4	Paper-based
Iteration 3 - Secondary data	CS 2016Household_QN Survey Solutions Questionnaire	N/A	Captured by 11 245 Contract fieldworkers	Initial Dwelling Sample N = 1 370 810 A sample of n = 384 dwellings was analysed	Digitally collected using a handheld device

During this phase, it was a requirement to test the secondary data distributions from all 3 iterations. The Kruskal-Wallis (KWt) test was used as a statistical method to determine the relationships between the three iterations (see Chapter 8 and APPENDIX Q). For KWt, all groups should have the same shape distributions. SPSS was used to test for this condition as part of the test. For all three iterations the distribution is positively skewed (skewness), hence not normally distributed but contains the same shape distribution.

7.4 PART C: STUDY [Determine OUTPUTS3]

When developing DDC quality dimensions based on the SASQAF and the SVC, Iteration 3 with secondary variables looked at the following themes (see Table 7.2 and APPENDIX P). All the cross-cutting themes were measured, firstly the cost differed significantly from Iteration 1, however, is similar to Iteration 2. In terms of comparing DDC to PAPDC, the cost proved to be a significant variable. Secondly, the error rate and perception of quality were discussed in terms of the primary qualitative data responses. However, Iteration 3's data record acceptance quality processes are explained in terms of the Minimum Acceptability Rules (MAR). Thirdly, the security-related aspects found that pockets of areas exists in the City of Cape Town which proved to be the most dangerous for fieldwork using handheld digital devices for DDC. Fourthly in terms of speed, no improvements were found, however, data was released much faster than PAPDC. Fifthly, the training DDC theme resulted in one gap and three improvements to be considered for DDC. Lastly, technology adoption and acceptance indicated that two gaps were identified.

Table 7.2: Iteration 3 - CS2016 DDC themes

SECTION	LITERATURE CHAPTER	DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME
7.4.1	(2.5.3)	COST	AVERAGE PAPDC COST PER HH
			AVERAGE DDC COST PER HH
			COST RATIO PAPDC VS. DDC
7.4.2	(2.5.4)	ERROR RATE and PERCEPTION OF QUALITY	PERCEPTION OF QUALITY
			ERROR RATE
7.4.3	(2.5.5)	SECURITY	THEFT
			DATA LOSS
7.4.4	(2.5.6)	SPEED	DDC DATA COLLECTION SPEED IN BETWEEN ASSIGNMENT OF DATA TO APPROVAL BY HEADQUARTERS
			DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE
7.4.5	(2.5.7)	TRAINING	DDC TRAINING APPROACHES
			TRAINING OUTCOMES
7.4.6	(2.5.9)	TECHNOLOGY ADOPTION AND ACCEPTANCE	**CS1. DDC COMMUNICATIONS:** INTERNAL OR EXTERNAL DDC COMMUNICATION
			CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT): EDUCATION LEVEL
			CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT): EXPERIENCE
			CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT): ABILITY USING DDC

(Source: Author)

7.4.1 Cost

The primary purpose for re-analysing the cost theme during Iteration 3 was to attain a better estimate (than in Iteration 1) of survey cost during a standard project life cycle of the NSO on a larger scale (United Nations Department of Economic and Social Affairs, 2005).

The following variables were analysed (See APPENDIX P: CHAPTER 7 – Iteration 3 Tables):

Secondary Data Variables:

- AVERAGE PAPDC COST PER HH
  - What is the average cost to collect the data per household, using Paper and Pen Data Collection (PAPDC)?
- AVERAGE DDC COST PER HH
  - What is the average cost to collect the data per household by using the Digital Data Collection (DDC)?
- COST RATIO PAPDC VS. DDC
  - Derived variable: Is the Total Survey Cost per HH (DDC) < Total Survey Cost per HH (PAPDC)?

CS 2016 used the CAPI methodology, whereby the data collection occurred digitally through electronic handheld devices, rather than the traditional PAPDC method. Due to this mode change, the overall cost has been reduced from an average cost per household for a PAPDC from R 2 000 to R200 for DDC; a tenfold reduction in cost, because of the use of digital technology (Statistics South Africa, 2016b). The cost ratio of 1:10 is a significant reduction per sampled household, given the large sample surveys usually conducted by an NSO. The cost difference of DDC vs. PAPDC between Iteration 1 (1:20) and Iteration 3 (1:10) respectively is primarily due to the additional expenditure for exceptional cases. Where staff was recruited far from the sample location and the number of repeat visits based on a methodology to attain favourable result codes of completed questionnaires.

In other words, for each PAPDC household visited, ten DDC households instead could have been visited for enumeration.

7.4.2 Error rate and perception of quality

The perception of quality from an interviewer's perspective and the actual error rate for each record were determined for DDC.

The following variables were analysed:

Primary Data Variables:

- PERCEPTION OF QUALITY
  - 8.2. Using DHDT in the field would result in an improvement in the quality and accuracy of the household survey-data collection
    - All strongly agreed (100%) that DHDT would improve the quality and accuracy of collection for the household survey data collection

- 11.1 In terms of the survey data-collection quality, I believe DHDT can improve the overall quality of the data collected
  - All agreed (100%) that DHDT could improve the overall survey data collection quality

- 11.2 In terms of the survey-data collection quality, I believe DHDT could result in poor data quality during each item response
  - There is a total disagreement (100%) that DHDT would result in poor data quality on each question related to DDC, theeby confirming the previous results in 11.1

- 11.3 In terms of the survey-data collection quality, I believe DHDT can initiate an increased response interaction from the respondents
  - There was a 100% undecided response to this; as the respondents were not yet certain regarding Iteration 1, where there was a stronger agreement (50%) than disagreement (25%) prevalent.

- 11.4 In terms of the survey-data collection quality, I believe DHDT can result in an increase in trust in the data from the users
  - There was a (100%) undecided response to this question; as the respondents were not yet certain regarding Iteration 1, where there was stronger agreement (58%), disagreement (17%) and (25%) undecided.

- 2.5 Perceptions of the quality of the data
  - The overall perception from the staff involved in the CS2016 was positively optimistic; as reported through the newsletters and the focus-group debriefings.

Secondary Data Variables:

- ERROR RATE
  - MAR (Minimum Acceptability Rules)
  - RejectedbyHeadquarters

The minimum acceptability rule (MAR) of DDC records is a procedure attempting to remove illogical and structurally incomplete records. These include household, persons or bad questionnaires that have been falsely captured by the interviewer. The Content and Development Division Team of the NSO developed MAR. Each day, an export of all the records in CAPI was loaded into a SQL-server database. The MAR procedure checked if there was minimum information allowing the identification of the questionnaire (EA, map reference number, dwelling unit, household number). Most importantly, the map reference number and the dwelling unit had to match the sampled units. Also, the procedure ascertained whether the final result code was for a responding household. In the case of non-respondents, the procedure returned the questionnaire to the field for a follow-up attempt to try to reduce the non-response rate. In the case of out-of-scope dwellings, the procedure returned the questionnaire to the fieldwork supervisor (FWS) for confirmation of the out-of-scope status of the questionnaire (Statistics South Africa, 2016a).

For each personal record, there is a variable called P_MINIMUM_FLAG, which could be either zero (GOOD) or one (BAD). For each household, there is a variable called H, the_MINIMUM_FLAG, which can be either zero (GOOD) or one (BAD). The following is the example procedure on how the minimum acceptability rule was applied:

A. For any CAPI questionnaire (Cover page):

- Check whether there is any value (meaning different from a null value or zero) for the total number of households (HH) in the dwelling. This means, check whether the HH count > one.
- Check if there is any value for the HH number.
- Check if there is any value for the total number of persons in the household
- If the total number of HH=0 and latitude and longitude coordinates have radius>50m then:
    - The MINIMUM_FLAG is given the value of one (BAD); otherwise, the MINIMUM_FLAG is given the value of zero (GOOD);
- If the total number of HH=zero and latitude and longitude coordinates have a radius<50m then:
    - If the final result code (DU result code) is UNOCCUPIED DWELLING (31), VACANT DWELLING (32), DEMOLISHED DWELLING (33), NEW DWELLING UNDER CONSTRUCTION (34), STATUS CHANGE (35) or CLASSIFICATION ERROR (36), the MINIMUM_FLAG = zero (GOOD);
    - If not, the final results code captured and the CAPI transition are complete (enumerated), flag the questionnaire MINIMUM_FLAG = one (BAD)

B. For each personal record in the DEMOGRAPHICS SECTION

Check whether there are valid responses for at least three variables out of the following demographic variables: (PERSON NAME, SEX, DATE OF BIRTH OR AGE, RELATIONSHIP TO HEAD OF HOUSEHOLD, MARITAL STATUS, POPULATION GROUP). Furthermore check that there are at least two responses in either of the subsequent sections (Migration, Parental survival, Education) concerning the personal records.

I. If there is a value for each of the fields above and any additional two values in the subsequent sections of the questionnaire, compute a count. This is a count of the number of completed fields that are there for a person with value one, otherwise, zero value is added. This means that if all fields above have a completed value, the sum of completed fields would yield the value of eight (i.e. the sum of eight).

II. If the sum of the count is between two and eight, check all the other remaining fields of the PERSON INFORMATION from "language" variable until the "date of death of the last child born" variable within the questionnaire. Compute a new count based on those fields. If the sum of the new count is > one, consider the record as GOOD. Update P_MINIMUM_FLAG field for the person with zero (i.e. GOOD or an acceptable record).

This means that if the sum of the first count >= three and the sum of the second count > one, the record is acceptable. Otherwise, if the sum of the first count <three, and the sum of the second count = zero, the record is not acceptable. Update P_MINIMUM_FLAG field for the person with one BAD or not acceptable record.

C. For each Household record

Check whether there are responses for each of the following variables:

- TYPE OF MAIN DWELLING
- TENURE STATUS
- MAIN SOURCE OF WATER
- TOILET FACILITY- ACCESS TO ELECTRICITY

I. If there is a value at each field above, then compute the account of the household fields (count number of completed fields for the household) with value one; otherwise zero. This means if all fields above have completed valid value, the sum of the completed fields should be = four (sum of count fields = four).

II. If the sum of the count of household fields is between two and four, check all the other remaining fields of the household section. Compute a new count of the household field, based on those additional fields. If the sum of the new count of household field > one consider, the record as GOOD. Update H_MINIMUM_FLAG field for a household with zero (i.e. acceptable record).

This means that if the sum of the first count of household field >= two; and the sum of the second count of household field > one, the record is acceptable. Otherwise, if the sum of the first count of household field < two; and the sum of the second count of household field = zero, the record is not acceptable. Update H_MINIMUM_FLAG field for a household with one BAD record.

D. Overall MAR implementation

Each questionnaire in CAPI was subjected to the automated process, where the personal record and/or the household records were flagged, as BAD or GOOD. Any questionnaire having a record marked as BAD was rejected, and sent back to the interviewers (mobile handheld device) for further verification. The supervisor took note of the flagged questionnaires and assisted in correcting the mistakes. The QA automated procedure ran each time a questionnaire was completed, and it was synchronised by the interviewer. However, the rejection was applied only once (i.e. the first time).

Distance formula	Distance = Arccos($\sin \varphi_1 \cdot \sin \varphi_2 + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \cos \Delta\lambda$) $\cdot$ R
$X = (x_1, x_2)$ =latitude in degrees $Y = (y_1, y_2)$ =longitude in degrees	$\cos^{-1}\left(\sin\left(\frac{x_1 * \pi}{180}\right) * \sin\left(\frac{x_2 * \pi}{180}\right) + \cos\left(\frac{x_1 * \pi}{180}\right) * \cos\left(\frac{x_2 * \pi}{180}\right) * \cos\left(\frac{y_2 * \pi}{180} - \frac{y_1 * \pi}{180}\right)\right) * 6371000$

Figure 7.3: Distance formula between interview and sampled co-ordinates (Statistics South Africa, 2016a)

Another key function of MAR was the calculation of the distance function. There were exceptional cases where an interview was observed to have taken place far from the initial sampled geographical reference coordinates. For these cases, where the distance where more than 30-meter in radius; the fieldworker was requested at least 4 times to confirm whether the interview was completed at the correct sampled DU. This was done by using two sources of information: i.e. the "Go-Survey" navigation tool and SS CAPI captured geo-co-ordinates against the sampled geo-co-ordinates. The distance was calculated by using the Spherical Law of Cosines, which provides accurate measures of distances as small as a few meters on the earth's surface (Figure 7.3). A set of decision rules was used by choosing the distance within the 30-meter radius from either source of the distance measured as being acceptable. For a dwelling unit situated in a new informal settlement with a large number of dwellings, the distance is acceptable. Provided it is within a 1 km radius (Figure 7.4) (Statistics South Africa, 2016a):

	Go survey	CAPI	Decision
Distance	Distance within 30 m (or within 1 km for Segment)	Distance within 30 m (or within 1 km for Segment)	Accept the questionnaire using both distance
	Distance NOT within 30 m (or NOT within 1 km for Segment)	Distance within 30 m (or within 1 km for Segment)	Accept the questionnaire using CAPI
	Distance within 30 m (or within 1 km for Segment)	Distance NOT within 30 m (or NOT within 1 km for Segment)	Accept the questionnaire using GO survey
	Distance NOT within 30 m (or NOT within 1 km for Segment)	Distance NOT within 30 m (or NOT within 1 km for Segment)	Reject the questionnaire

Figure 7.4: Decision table on acceptable distance (Statistics South Africa, 2016a)

7.4.3 Security

Digital handheld devices are viewed as assets; therefore, the security of these handheld devices is important. The security of collecting confidential information from the respondents are also important. All data collected were encrypted on the handheld device before being synchronised to the central server by the INTERVIEWER. The information was therefore not accessible when devices were lost or stolen. Each INTERVIEWER was issued with a unique identifier (ID) and a password to access the data-collection application. Once the questionnaire was synchronised to the server, the data was automatically removed, or wiped from the device.

The following variables were analysed:

Secondary Data Variables:

- THEFT
  - Number of Stolen devices
  - Total Number of Survey Officers
  - Device Loss through Theft
- DATA LOSS
  - % Synchronisation Success Rate (Data Loss)

In relation to the theft of devices, most of the reported theft occurred in the City of Cape Town as apposed to other areas across South Africa. It was reported that no data loss occurred during the survey.

7.4.4 Speed

The data were released, as per the planned scheduled date to the public. **N**o "Improvements" or "Gaps" to the existing PAPDC SASQAF Framework were required. It should be noted that the speed of collecting and disseminating the data to users increased significantly when comparing DDC vs. PAPDC.

7.4.5 Training

Training approaches were refined for CS2016. District training venues across the country conducted training to contract staff. Similar to Iteration 2, the ***GAP** and **Improvements** in Iteration 3, regarding training, were once more identified (as in Iteration 2, see ***Gap4** and **Improvement4-6**).

7.4.6 Technology adoption and acceptance

7.4.6.1 CS1. DDC COMMUNICATIONS:

It was found that in this dimension/theme there is a ***Gap7** that needs to be included in the development of the DDC Quality Assurance Framework.

- INTERNAL OR EXTERNAL DDC COMMUNICATION
  - Ensuring permanent staff members from NSOs or the public at large are informed by educating staff on the transition processes related to the adoption of DDC from PAPDC.

The following variable was analysed:

Primary Data Variable:

- INTERNAL OR EXTERNAL DDC COMMUNICATION
  - 12. Regarding best practices in official survey data collection, I believe DHDT methodology and standards exist; but they are not well communicated internally to the South African NSO, or externally.
    - There was a 100% undecided response to this, since respondents were not yet certain in during this iteration. There was a stronger agreement (67%) that DHDT is not well-communicated either internally or externally than the disagreement of (25%).
    - Permanent staff (involved as supervisors) felt that communications related to DDC transition could be explained more effectively both internally and externally by the NSO.

7.4.6.2 CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT):

It was found that in this dimension/theme there is a ***Gap8** that needs to be included in the development of the DDC QA Framework.

- EDUCATIONAL LEVEL
  - The adoption of innovations by firms and workers is an important part of the process of technological change. Many earlier studies found that highly educated workers tend to adopt new technologies faster than those do with less education. Positive correlations as such between the level of education and the rate of technology adoption, however, do not necessarily reflect the true causal effect of education on the adoption of new technology (Riddell & Song, 2011).
  - Uematsu and Mishra (2010) attempted to test the widely-held belief that the effect of education has a positive impact on technology adoption; however, results were inconclusive (Uematsu & Mishra, 2010).

    The following primary data variables were analysed:

    - 3. What is the highest level of education that you have successfully completed?
- EXPERIENCE
  - In the NSO, age distribution plays a factor in terms of technology uptake from PAPDC to DDC. Firms with a higher share of younger employees are more likely to adopt new technologies. According to literature, the impact of the age of employees increases the probability of technological change and acceptance of innovations in the manufacturing industries. The results reveal that firms with a higher share of employees younger than 30 years of age have a greater probability of adopting new technologies. Firms with a higher share of employees older than 55 years of age on the otherhand have a reduced likelihood of introducing new technologies, or software (J. Meyer, 2008).

    The following primary data variables were analysed:

    - 4. How many years of working experience do you have at Statistics South Africa in the related areas responsible for the collection of household surveys?
    - 5. Are you currently working Full-time or Part-time?

- 7.1 No experience with DHDT.

- ABILITY USING DDC
  - Table 7.3 summarises the abilities of staff regarding DDC:

Table 7.3: Primary data variable result (7)

Unfamiliar	No experience with DHDT.	**0%**
Newcomer	Attempted to use DHDT, but I still require help on a regular basis.	**0%**
Beginner	Able to perform basic functions in a limited number of DHDT.	**8%**
Average	Demonstrate a general competency in a number of DHDT.	**33%**
Advanced	Acquired the ability to competently use a broad spectrum of DHDT.	**58%**
Expert	Extremely proficient in using a wide variety of DHDT.	**0%**

  - The uptake of digital technologies by large proportions of the population is occurring much more rapidly in comparison with other technological transitions.
  - In Table 7.3 above, at least 81% of the staff indicated that they "Demonstrated a general competency in a number of DHDT"; or they had "Acquired the ability to competently use a broad spectrum of DHDT."

The following primary data variables were analysed:

- 7. How would you describe yourself with respect to the various levels of digital handheld device technology (DHDT) usage? Mark an "X" next to the level that best describes you.
  - No experience with DHDT (unfamiliar)
  - I have attempted to use DHDT, but I still require help on a regular basis (Newcomer)
  - I am able to perform basic functions in a limited number of DHDT areas (Beginner)
  - I demonstrate a general competency in a number of DHDT functions (Average)

    - I have acquired the ability to competently use a broad spectrum of DHDT (Advance)
    - I am extremely proficient in using a wide variety of DHDT functions (Expert)
- 13.1 During your participation in DDC pilots thus far:
    - I have not been involved in any DDC pilots

7.5 SECONDARY QUANTITATIVE DATA ANALYSIS

The final sample size for CS 2016 was 1 370 809 dwellings sampled from a total of 93 427 enumerated areas in South Africa. Table 7.4 provides the distribution of the CS 2016 DU sample by province (Statistics South Africa, 2016a).

Table 7.4: Distribution of CS 2016 DU sample by province

Province	Number of In-scope EAs	Number of Sampled DUs
Western Cape	9 851	149 100
Eastern Cape	15 742	195 301
Northern Cape	2 742	36 125
Free State	5 595	83 645
KwaZulu-Natal	15 719	219 182
North West	6 726	102 120
Gauteng	19 022	331 125
Mpumalanga	7 197	105 058
Limpopo	10 833	149 153
South Africa	**93 427**	**1 370 809**

Table 7.5 provides all the measurements (variables) studied in Iteration 3.

Table 7.5: Iteration 3 - Descriptive statistics on QA measurements

		time_el apsed	Cost	MainDw ellType	SubsD well	Tenure Stat	WaterSo urce	DistanceW ater	WaterSu pplier	Toile t	ToiletLo cation	ToiletS hared	Refuse	ElectrSup plier	EnergyC ook	EnergyLi ght	EnergyWater Heat	EnergySpace Heat
N	Valid	384	384	330	384	329	330	96	329	329	316	317	328	289	330	329	330	330
	Missing	0	0	54	0	55	54	288	55	55	68	67	56	95	54	55	54	54
Mean		57:56:53	200	2.62	1.49	3.85	2.97	1.69	1.60	2.91	1.65	1.75	2.62	2.27	1.63	1.50	1.84	3.73
Std. Error of Mean		7:59:58	0.000	0.166	0.038	0.069	0.153	0.099	0.072	0.136	0.031	0.025	0.110	0.062	0.075	0.091	0.110	0.215
Skewness		3.676		1.593	-0.944	-0.665	1.700	1.456	2.036	1.239	0.110	-1.053	0.532	0.045	1.826	3.224	2.842	1.080
Std. Error of Skewness		0.125	0.125	0.134	0.125	0.134	0.134	0.246	0.134	0.134	0.137	0.137	0.135	0.143	0.134	0.134	0.134	0.134
Kurtosis		15.505		0.916	-0.340	1.045	1.785	1.460	2.743	0.987	-0.755	-0.564	-1.545	-0.718	1.606	8.979	8.762	-0.484
Std. Error of Kurtosis		0.248	0.248	0.268	0.248	0.268	0.268	0.488	0.268	0.268	0.273	0.273	0.268	0.286	0.268	0.268	0.268	0.268
Range		1199:33	0	11	3	6	12	4	5	9	2	2	6	5	5	8	10	10
Sum		22252:05	76800	865	574	1266	980	162	528	956	521	555	860	655	539	492	606	1232
Percentiles	25	0:20:32	200.	1.00	1.00	4.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	0:40:26	200	1.00	2.00	4.00	2.00	1.00	1.00	1.00	2.00	2.00	1.00	3.00	1.00	1.00	1.00	1.00
	75	10:34:57	200	2.00	2.00	4.00	3.00	2.00	1.00	5.00	2.00	2.00	5.00	3.00	1.00	1.00	1.00	5.00

Table 7.6 below indicates that 14.1% of the data were missing in the measurement “Main Dwelling Type”.

Table 7.6: Iteration 3 - Main Dwelling Type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	231	60.2	70.0	70.0
	2	26	6.8	7.9	77.9
	3	4	1.0	1.2	79.1
	4	2	0.5	0.6	79.7
	5	2	0.5	0.6	80.3
	6	2	0.5	0.6	80.9
	7	13	3.4	3.9	84.8
	8	16	4.2	4.8	89.7
	9	27	7.0	8.2	97.9
	10	3	0.8	0.9	98.8
	12	4	1.0	1.2	100.0
	Total	330	85.9	100.0	
Missing	System	54	14.1		
Total		384	100.0		

Table 7.7 displays the measurement of the Water Source and the distribution of the responses.

Table 7.7: Iteration 3 - Water Source

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	118	30.7	35.8	35.8
	2	110	28.6	33.3	69.1
	3	35	9.1	10.6	79.7
	4	4	1.0	1.2	80.9
	5	2	0.5	0.6	81.5
	6	10	2.6	3.0	84.5
	7	16	4.2	4.8	89.4
	8	7	1.8	2.1	91.5
	9	5	1.3	1.5	93.0
	10	20	5.2	6.1	99.1
	12	1	0.3	0.3	99.4
	13	2	0.5	0.6	100.0
	Total	330	85.9	100.0	
Missing	System	54	14.1		
Total		384	100.0		

Table 7.8 displays the measurement of the Toilets Shared and the distribution of the responses.

Table 7.8: Iteration 3 - Toilets Shared

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	80	20.8	25.2	25.2
	2	236	61.5	74.4	99.7
	3	1	0.3	0.3	100.0
	Total	317	82.6	100.0	
Missing	System	67	17.4		
Total		384	100.0		

7.5.1 Data quality indicators

The key variables, as listed in Figure 7.5 below, were considered for determining the quality of the indicators produced from CS 2016. These variables covered the priority indicators, as determined by the CS 2016 analysis team at household levels and personal levels. The quality indicators were produced for the survey estimates of percentages and incorporated in the complex sample design features of CS 2016 (Statistics South Africa, 2016a).

Dataset	Variables Analysed (as listed on the dataset)	Description of the variables
Person	*Attendance*	**Attendance at an educational institution** • At national level, ○ By age group (4 broad age groups) ○ By gender ○ By population group • At province level, ○ By age group (4 broad age groups) ○ By gender • At municipal level, ○ By gender
	EducLevel	**Highest Level of Education Completed (No schooling, Primary education, Secondary education, Bachelor's degree)** • At national level, ○ By age group (4 broad age groups) ○ By gender ○ By population group • At province level, ○ By age group (4 broad age groups) ○ By gender • At municipal level, ○ By gender
Household	*MainDwellType*	**Type of main dwelling** • At national level, ○ By age group of household head ○ By gender of household head ○ By population group of household head • At province level, ○ By age group of household head ○ By gender of household head • At municipal level, ○ By gender of household head
	WaterSource	**Main source of water for drinking** • At national level, ○ By age group of household head ○ By gender of household head ○ By population group of household head • At province level, ○ By age group of household head ○ By gender of household head • At municipal level, ○ By gender of household head
	Variables for cross-classification (as listed on the dataset)	**Description of the variables**
	Sex	Gender of person
	Population_Group	Population group
	Age_Broad_Groups	4 Broad Age Groups: 0-14, 15-34, 35-64, 65+
	MN_Code_2011	Local municipality/Metro name

Figure 7.5: Key variables used in determining the data quality for CS 2016 (Statistics South Africa, 2016a)

The CVs at national, provincial and municipal levels obtained for the following key indicators at the household level:

- Type of main dwelling
- Water source

Province	Type of main dwelling	Raw count of persons	Weighted count of persons	Percentage	Standard error of percentage	Lower confidence limit of percentage	Upper confidence limit of percentage	Coefficient of variation of percentage
Western Cape	Formal	70 419	1 593 891	82,43	0,15	82,14	82,72	0,18
	Informal	14 976	330 324	17,08	0,15	16,8	17,37	0,85
	Traditional	453	9 401	0,49	0,03	0,43	0,54	5,62
Eastern Cape	Formal	80 117	1 154 843	65,13	0,15	64,83	65,42	0,23
	Informal	9 297	146 713	8,27	0,1	8,08	8,47	1,2
	Traditional	44 742	471 699	26,6	0,13	26,34	26,86	0,5
Northern Cape	Formal	18 838	295 318	83,5	0,29	82,93	84,08	0,35
	Informal	2 952	50 104	14,17	0,28	13,62	14,72	1,98
	Traditional	609	8 245	2,33	0,11	2,12	2,54	4,66
Free State	Formal	50 440	791 485	83,62	0,16	83,29	83,94	0,2
	Informal	8 757	139 585	14,75	0,16	14,44	15,06	1,08
	Traditional	1 125	15 509	1,64	0,05	1,54	1,74	3,18
KwaZulu-Natal	Formal	126 132	2 090 067	72,68	0,11	72,46	72,9	0,16
	Informal	15 318	265 334	9,23	0,08	9,08	9,38	0,83
	Traditional	37 926	520 244	18,09	0,09	17,91	18,28	0,52
North West	Formal	58 630	977 031	78,26	0,2	77,87	78,64	0,25
	Informal	12 959	248 342	19,89	0,19	19,51	20,27	0,98
	Traditional	1 513	23 146	1,85	0,05	1,75	1,96	2,9
Gauteng	Formal	195 753	4 029 069	81,39	0,09	81,22	81,57	0,11
	Informal	46 004	910 375	18,39	0,09	18,22	18,56	0,48
	Traditional	503	10 763	0,22	0,01	0,2	0,24	5,15
Mpumalanga	Formal	64 716	1 048 973	84,68	0,16	84,36	85	0,19
	Informal	7 743	149 786	12,09	0,15	11,79	12,39	1,26
	Traditional	2 427	39 992	3,23	0,07	3,08	3,37	2,28
Limpopo	Formal	99 869	1 423 523	88,92	0,11	88,7	89,14	0,13
	Informal	6 011	95 675	5,98	0,09	5,8	6,16	1,53
	Traditional	6 094	81 746	5,11	0,07	4,97	5,24	1,36
South Africa	**Formal**	**764 914**	**13 404 199**	**79,22**	**0,05**	**79,12**	**79,31**	**0,06**
	Informal	**124 017**	**2 336 239**	**13,81**	**0,04**	**13,72**	**13,89**	**0,30**
	Traditional	**95 392**	**1 180 745**	**6,98**	**0,03**	**6,93**	**7,03**	**0,36**

Figure 7.6: National and Provincial estimates of the type of main dwelling including measures of precision (Statistics South Africa, 2016a)

The national and provincial estimates for this variable are well within the thresholds (see Figure 7.6 and Figure 7.7 respectively) and; they are therefore considered to be reliable. The analysis provided shows that survey estimates for all categories of the key variables at the national and provincial level were reliable. At the municipal level, however, there were some categories, where caution needed to be exercised when interpreting the survey estimates, due to the relatively small number of units within these cells. (Statistics South Africa, 2016a)

Province	Water source	Raw count of persons	Weighted count of persons	Percentage	Standard error of percentage	Coefficient of variation of percentage
Western Cape	Piped water inside dwelling	64 896	1 487 774	76,93	0,16	0,21
	Piped water inside yard	11 485	232 892	12,04	0,12	1,01
	Piped water from access point outside the yard	9 085	204 414	10,57	0,12	1,12
	No access to piped water	394	8 797	0,45	0,03	5,99
Eastern Cape	Piped water inside dwelling	35 324	592 428	33,41	0,17	0,49
	Piped water inside yard	32 650	433 028	24,42	0,14	0,56
	Piped water from access point outside the yard	35 726	429 037	24,19	0,13	0,55
	No access to piped water	30 466	318 902	17,98	0,11	0,63
Northern Cape	Piped water inside dwelling	9 429	154 529	43,69	0,38	0,88
	Piped water inside yard	8 047	129 197	36,53	0,37	1,02
	Piped water from access point outside the yard	4 585	64 272	18,17	0,29	1,58
	No access to piped water	340	5 710	1,61	0,14	8,83
Free State	Piped water inside dwelling	20 957	357 926	37,81	0,23	0,61
	Piped water inside yard	35 101	519 086	54,83	0,23	0,42
	Piped water from access point outside the yard	3 776	62 136	6,56	0,12	1,76
	No access to piped water	491	7 490	0,79	0,04	4,99
KwaZulu-Natal	Piped water inside dwelling	61 544	1 076 667	37,44	0,13	0,34
	Piped water inside yard	61 430	965 066	33,56	0,12	0,36
	Piped water from access point outside the yard	39 772	616 299	21,43	0,11	0,49
	No access to piped water	16 642	217 811	7,57	0,06	0,85
North West	Piped water inside dwelling	15 739	300 221	24,04	0,20	0,82
	Piped water inside yard	36 340	616 426	49,36	0,22	0,45
	Piped water from access point outside the yard	19 897	310 170	24,84	0,18	0,72
	No access to piped water	1 140	21 950	1,76	0,06	3,56
Gauteng	Piped water inside dwelling	140 404	2 972 973	60,05	0,11	0,19
	Piped water inside yard	81 372	1 568 039	31,67	0,11	0,34
	Piped water from access point outside the yard	19 885	396 334	8,00	0,06	0,79
	No access to piped water	647	13 792	0,28	0,02	5,40
Mpumalanga	Piped water inside dwelling	18 238	359 033	28,98	0,21	0,73
	Piped water inside yard	40 937	630 078	50,86	0,21	0,42
	Piped water from access point outside the yard	12 960	207 087	16,72	0,15	0,92
	No access to piped water	2 758	42 663	3,44	0,08	2,24
Limpopo	Piped water inside dwelling	13 216	210 302	13,14	0,12	0,91
	Piped water inside yard	50 589	721 579	45,07	0,16	0,36
	Piped water from access point outside the yard	40 230	562 950	35,16	0,16	0,45
	No access to piped water	7 951	106 251	6,64	0,08	1,19
South Africa	**Piped water inside dwelling**	**379 747**	**7 511 853**	**44,39**	**0,06**	**0,14**
	Piped water inside yard	**357 951**	**5 815 391**	**34,36**	**0,05**	**0,16**
	Piped water from access point outside the yard	**185 916**	**2 852 700**	**16,86**	**0,04**	**0,25**
	No access to piped water	**60 829**	**743 366**	**4,39**	**0,02**	**0,47**

Figure 7.7: National and provincial estimates of the type of water in the main dwelling including measures of precision (Statistics South Africa, 2016a)

7.6 PART D: ACT [Make CHANGES3]

During the final phase of the PDSA for Iteration 3, the researcher acknowledged that iterative improvements could continue to the Nth iteration based on the continuous nature of the PDSA cycle. However, after 3 iterations spanning across a number of years, the research process needed to conclude the study and present the findings. Operationally conducting a large scale DDC was one of the highlights of the research and highlighted to possibilities, benefits, and challenges to the NSO when considering quality factors for official DDC.

7.6.1 Qualitative overall debriefing lessons learnt

From the account of staff participating during this survey, the majority of them felt that they had few or no major difficulties in using the digital device for data collection. Although some staff with poor eyesight reportedly had difficulty reading the screen in direct sunlight. Previous survey experience was not related to the ease-of-use of digital handheld devices. As a result of DDC, no failures were reported in terms of handheld devices or the transmission of data. Therefore, no data were lost during the survey.

From a "user acceptability" perspective, staff felt empowered when conducting the interviews using the device. The respondents were intrigued and often complimented the interviewers on the technology being used for the data collection when compared with the standard PAPDC processes in previous surveys. Even though data collection at the household initially took longer for DDC, interviewers gained more confidence and experience. This increased the acceptance of technology from the user (interviewer) and the respondent's perspective.

In a similar study conducted in Tanzania (Kizito et al., 2007), user acceptability explained by the respondents revealed the positive stance they had towards DDC with PDAs.

7.6.2 Qualitative overall DDC gaps and improvements identified

The gaps found during Iteration 3 are in addition to those tested for "COST", "SECURITY" and "TRAINING" (see Table 7.9):

Table 7.9: CS 2016 Changes to SASQAF

DDC THEME	IMPROVEMENT REQUIRED	GAPS
TECHNOLOGY ADOPTION AND ACCEPTANCE		*Gap7 - **CS1. DDC COMMUNICATIONS:** - INTERNAL OR EXTERNAL DDC COMMUNICATION *Gap8 - **CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT)**: EDUCATION LEVEL

(Source: Author)

There was a very high level of acceptance by the community and unanimous approval of DDC by interviewers, supervisors and field coordinators alike. Several components assisted with the data-collection phase, some of which included the various supporting apps, forms and help manuals on the device. These were used to assist interviewers uniformly to explain concepts and definitions better to the respondents. Interviewers were encouraged to show the device during the formal collection process by the respondent when requested to do so. A major feature of the DDC approach was the increased need for more careful pre-training preparation as opposed to a PAPDC survey. Sufficient time was required to pre-program the acceptable ranges for numeric variables, to create skip patterns, to check for logical consistencies in responses; and to test thoroughly that all routines work. Once completed, however, there were marked reductions in the extent of checking required by supervisors during the survey, freeing up time for other quality-control and field-activities.

7.7 CHAPTER SUMMARY

This survey was the largest DDC implementation to date for South Africa in terms of sample size (i.e. more than 1.37 million households). The sample spread across the entire country. Eventhough permanent staff was used to train and supervise, contract workers were employed to collect the data from all the sampled households across South Africa. Most of the trainees did not have formal mobile handheld survey experience before. This was primarily due to the selection criteria of recruiting contract staff in the areas where they resided. The CS 2016 experience proved to have no major hardware or software problems during the survey. The completeness of the data was very high; and ensuring a fully-completed dataset for analysis took significantly less time than in previous surveys, based on PAPDC on a similar sample scale.

Although the largest part of the questionnaire and programming were finalised, pre-training, and some minor modifications to questions were still made during training. For DDC this was simpler, and not as time-consuming or costly as PAPDC would have been with respect to rerunning the print runs. Furthermore, DDC resulted in reducing the logistics for transporting large quantities of paper questionnaires to the central data-processing center. The time required to train interviewers did not increase because of DDC; and the observation that the time required to complete the interviews did not reduce dramatically after the pilot period.

The need to recharge the handheld mobile device battery was a concern in remote rural settings without any mains electricity supply. However, the additional power banks ensured staff encountering power challenges were catered for. Otherwise, there were no catastrophic failures of electronic devices, despite the adverse climatic conditions in some cases, and the initial unfamiliarity of most interviewers with digital handheld devices. The CS 2016 experience represented a cost-effective and timesaving approach to enhance the quality and completeness of the data.

The following chapter summarises the research findings and expands on interpreting the analysis outcomes of the previous three iterations (chapters 5 – 7).

CHAPTER 8: A PROPOSED DIGITAL DATA COLLECTION QUALITY FRAMEWORK

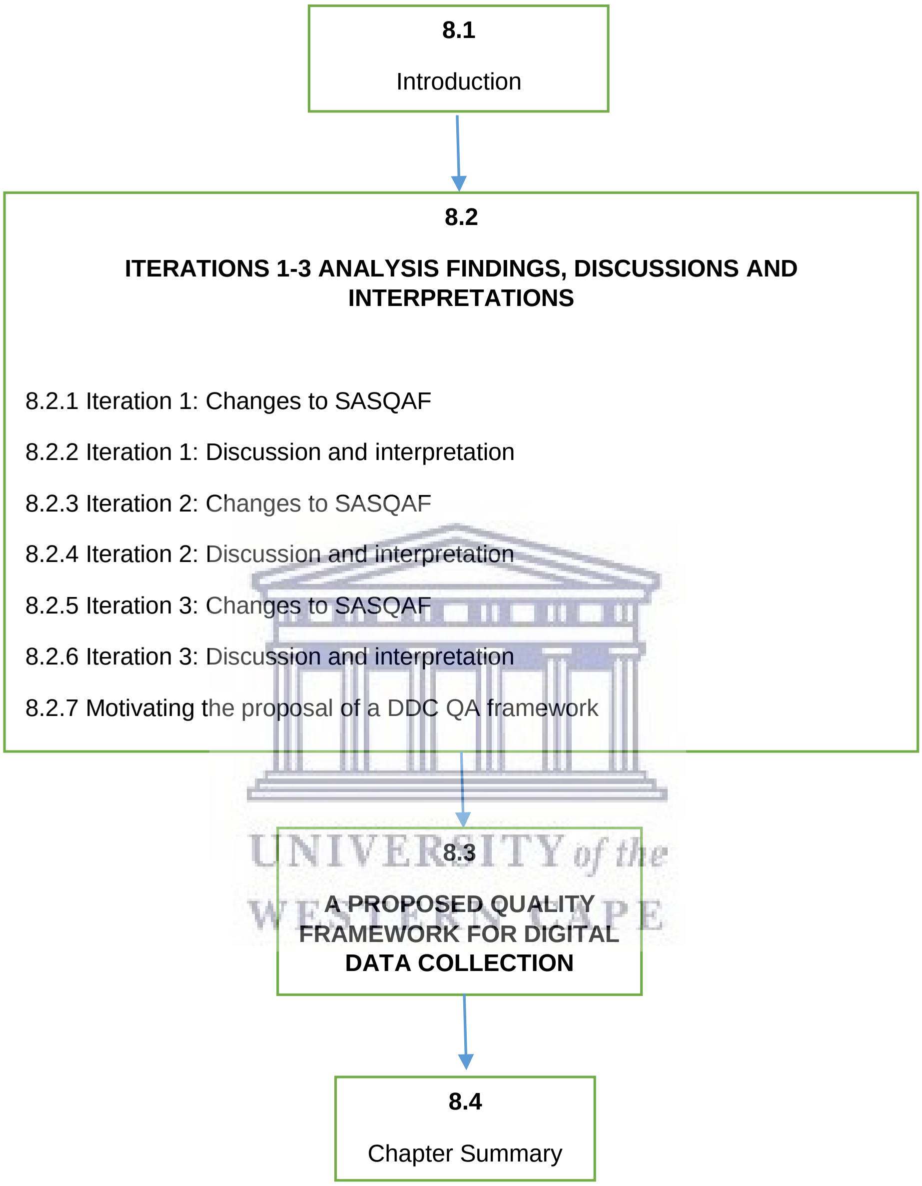


Figure 8.1: Chapter 8 layout

CHAPTER 8: A PROPOSED DIGITAL DATA-COLLECTION QUALITY FRAMEWORK

8.1 INTRODUCTION

After three iterations (chapters 5 - 7) of testing DDC in different environments, user groups and increasing sample size(s) across the South African population, chapter 8 provides a summary of the research findings. This chapter discusses the changes to the SASQAF as suggested from all three iterations.

When analysing three iterations of varying sample sizes, it becomes important to measure the effects on the resultant data over time. This is to determine whether any significant relationships exist between or across the iterations. The generally accepted statistical tool to measure these type of effects is the Kruskal–Wallis test (KWt). In section 8.2, an abbreviated note on KWt and thereafter the corresponding detailed analysis in Appendix Q is explained. The gaps and improvements identified throughout the study are discussed, summarised and detailed in sections 8.2.1 – 8.2.6 respectively, culminating in motivating a proposal to develop a new developmental DDC QA framework in section 8.2.7. Section 8.3 concludes the chapter by proposing the actual changes required for a DDC quality framework.

8.1.1 The public sector and total quality management (TQM)

As a government department or public institution, the NSO for South Africa utilised a Quality Management System (QMS) that is functional, operational, methodologically sound, transparent and easily understood by the public sector. Donald Murphy (1999) from the NSO in Ireland looked at the challenges of implementing a QMS, such as total quality management (TQM) in the public sector. He noted the following: *"The nature of TQM itself inhibits its application to the public sector"* (Eurostat et al., 1999). Public sector institutions exist in a limited competition environment; and unlike the private sector and QMS implementation, they may not immediately translate to the delivery of services on the ground. A proposed DDC developmental QA framework is not generally seen as a direct public service benefit however; it is deemed an internal performance and quality-related mechanism for collecting outcome data for the public.

Examples of public good outcomes in other public government institutions are the Department of Housing, whose mandate is to organise more tangible outputs of ensuring houses are developed for the public; or the Department of Social Development, responsible for various types of public social grants, etc. However, within the broader QMS, any proposed DDC developmental QA framework indirectly has a bearing on all planning components within planning departments. In turn, this directly benefits the public good concerning costs and improved efficiencies if utilised in this respect.

The public sector also suffers from other problems, for example, resistance to change; resources not linked to performance; no reward system for management on achieving quality; a constrained management atmosphere; increasing service quality with reduced costs is difficult and the work culture of professionals, etc. The transition processes for an NSO include factors both within an organisation's control (internally) and externally. The phrase: "doing more with less" in the public sector is often a coined termed as the basis for promoting the need to seek to adopt alternative or more streamlined processes to conduct current core business with less state funding resources (The South African Presidency, 2009).

Change is inevitable, and as tight as budgets are, for government agencies, the vice is likely to squeeze even tighter in the coming years. As a result, public sector leaders will no doubt continue to search for more efficient and cost-effective ways to meet the growing demand for information, albeit with a more cost-effective approach. In terms of cloud technologies, agencies will need to address organisational, change management and cultural issues. At the same time as they deploy analytical tools, or move workloads to the cloud. The good news is that considerable opportunities still exist to gain efficiencies and to reduce costs in agency operations. The result can be not only leaner, but more effective for organisations (Deloitte Development LLC, 2013).

"The public sector environment is more complicated than the private sector" (Eurostat et al., 1999). Given the complexities arising from political, public and legal accountability. Accountability for the NSO is detrimental on all levels of the proposed DDC developmental QA framework. In terms of public spending, reprioritisation of public needs is transparent. It should be based on the independent Statistics Council, the NSO head of the department, and the staff being politically and legally accountable to the public at large. In the South African context, the Statistics Council is appointed in terms of the Statistics Act (6 of 1999). The role of the Statistical Council is to advise the Minister and the Statistician-General on issues concerning the production and use of official statistics. This includes (Statistics Act, 1999):

- Strategic advice on statistical policy and priorities
- Technical advice, in general, or on specific statistical data collections and methodological issues
- Co-ordination of statistical activities.

The SASQAF over the years has been the primary quality tool used and is well understood by members within the NSS in South Africa. The primary concepts, standards, and principles within the SASQAF provide the reader with a direct link to understanding the core quality dimensions that are to be affected for a proposed DDC quality assurance framework. The following section (8.2) refers to the changes required in SASQAF to accommodate a DDC process.

8.2 ITERATIONS 1-3 ANALYSIS, FINDINGS AND DISCUSSIONS

The Kruskal–Wallis test and its equivalent test for two-samples (generally called the Manne-Whitney U test) are probably the two most widely used non-parametric statistical tests conducted in behavioural sciences. It tests whether the mean ranks are the same in all the groups (John H. McDonald, 2014). The KWt is used when you have one independent variable with two or more levels and an ordinal dependent variable. In other words, it is the non-parametric version of ANOVA and a generalised form of the Mann-Whitney test method; since it permits two or more groups. The KWt merely indicated that the 3 iteration groups differ in some way; therefore we needed to inspect the group means or medians to decide precisely how they differ (Hole, 2011; Ruxton & Beauchamp, 2008). The findings of the KWt indicated the similarities and the differences between the three iterations across different population samples within South Africa. Iterations 2 and 3 were found to be more coherent to each other, due to larger total sample size selections; while this is not so much the case with Iteration 1.

For full analysis results, refer to *APPENDIX Q: CHAPTER 8* – *Kruskal-Wallis test – Quantitative analysis across all three iterations.*

8.2.1 Iteration 1 Changes to SASQAF

In Iteration 1, the proposed changes to SASQAF are predominantly within the themes of cost (5.4.2); accuracy, error rate and perception of quality (5.4.3); security (5.4.4) and technology adoption and acceptance (5.4.5) see Table 8.1 below.

In relation to the cost theme, two improvements were identified during Iteration 1 for the developmental QA DDC framework. Firstly, to ensure there is adequate budget, there is a need to specify all DDC resources in terms of components, applications and related resources required for DDC. These includes costs for: the initial DDC setup costs, costs of ensuring the availability of devices, cellular connectivity to a secure centralised server costs and all other downstream survey processes costs on the SVC. Secondly, reviewing and auditing the budget is an ongoing process that need to occur to ensure the surveys are conducted within the relevant costing limits as specified through the project planning documents during the “Prerequisites of Quality” dimension in the SASQAF and the “Design” phase of the SVC.

Table 8.1: Iteration 1: Gaps and improvements

DDC THEME	IMPROVEMENT REQUIRED	GAPS
COST (5.4.2)	***Improvement1** - (PRER6(7); De8) : ***Improvement2** - (PRER7(4); De8)	
ACCURACY and PERCEPTION OF QUALITY (5.4.3)		***Gap1** - (ACCU3(2); Co2):
SECURITY (5.4.4)	***Improvement3** - (PRER4; Ne3)	***Gap2** - (INTE5, Co2)
TECHNOLOGY ADOPTION AND ACCEPTANCE (5.4.5)		***Gap3**

For the accuracy and perception of quality theme, a gap exists in indicating the importance of collecting data at the correct sampled dwelling. During the "Collect" phase of the SVC, INTERVIEWERS are required to visit the sampled dwelling using the relevant navigational tools available. The "Accuracy" dimension in SASQAF should include the methodological processes and procedures for the accurate collection of data using the relevant tools and processes. The mapping, (ACCU3(2); Co2) defines the gap1 (see Table 5.10)

In terms of "security", data security, staff safety, device safety (stealing of devices), and respondent confidentiality is both a "gap" and an "improvement" that is required on the proposed DDC QA Developmental Framework. The confidentiality of data collected is crucial especially the respondent's information provided to the NSOs. The Protection of Personal Information Act (POPI) is a South African public law related to confidentiality of personal information. The POPI Act, No. 4 of 2013 (POPI) has been set up as a law, to regulate, in harmony with international standards, the processing of personal information by public and private bodies. POPI gives effect to the right of privacy and is subject to justifiable limitations that are aimed at protecting the public's rights and important interests (Law Society of South Africa, 2018; Parliament of the Republic of South Africa, 2013).

The implications are indeed huge since the act determines that it is the responsibility of the "holder of the data or information" to maintain them, to protect them, and to ensure it cannot be accessed and/or misused. This relates to data sticks, flash drives, and other portable storage means, including cell phones, to being clean of any individual's data. Especially if the individual has not consented to the storage of the data over a period. All of this data/information may need to be stored safely. The day when personal contact information may no longer be stored on mobile phones may well be nigh if the law were to be applied to the letter (BIZCOMMUNITY, 2018). NSO's are responsible for storing personal data and ensuring the confidentiality of such data. This concern now applies to NSO's who are using handheld devices to store personal information of survey participants. Even though methods of statistical disclosure and risk assessments are well established in statistical-science literature, the integration of these methods by agencies into general statistical collection practice still proves difficult (Skinner, 2012). Another SASQAF change suggested from Iteration 1 is the possible loss of equipment; the security of respondent data and risking staff safety. The risks mentioned prior needs to be “planned” for, mitigated and executed for well in advance when conducting DDC. The SASQAF quality dimension: "integrity" should incorporate security as a quality indicator (see Table 8.1):

The analysis in Iteration 1 indicated that staff slowly showed a more positive attitude towards DDC for official household-survey collection. When studying the quality theme of technology adoption and acceptance, a significant proportion of users (59%) agreed, “during their participation in DDC pilots thus far, DHDT has been user-friendly for both the interviewer and the interviewee". It became easier the more they used the devices for DDC.

8.2.2 Iteration 1: Discussion and interpretation

The researcher conducted a cost comparison analysis between Iteration 1 (DDC) and a standard PAPDC process. The average cost to collect data per household between the two methods (DDC vs. PAPDC) was significantly lower for DDC. The cost comparison ratio indicated that DDC was 20 times cheaper than a standard PAPDC process. The key cost reduction drivers for DDC included the lack of appointing dedicated data capturers to capture completed questionnaires, savings in terms of the non-procurement of boxes, printed questionnaires and courier costs.

The accuracy and perception of quality theme focused on achieving a high standard of precision and accuracy in locating sampled dwellings. The Google Mymaps application installed on the handheld devices was used to navigate to the selected sampled dwellings. The test results (APPENDIX M – Test Case 04 and Table 5.18) listed a satisfactorily outcome in terms of navigation to the selected sampled dwellings. Navigating on known roads on the Google platform is easy while connected online (i.e. cellular signal is present), however when not connected (i.e. offline mode), it becomes a challenge given no maps exists for navigation. Google Mymaps directs the INTERVIEWERS only on the nearest location of the road and the sampled dwelling. This is challenging for INTERVIEWERs to navigate especially if the road network is far from the selected sampled dwelling, since Google Mymaps will not be functional online in those scenarios.

The security theme needs to be expanded, given the importance of data security on all phases of the SVC (Hattas & Eloff, 2011; Pfleeger C., 2007). This theme includes the security of physical devices used for data collection. When looking at technology adoption and acceptance, NSOs, including African NSOs, should monitor advances in technology and the adoption of digital tools for the statistical collection of official household surveys. These include collection systems used by staff to foster greater uptake and the effective use of digital devices when conducting DDC.

For technology adoption and acceptance, access and ownership to digital handheld devices and the internet are continuously on the rise. As the world becomes increasingly interconnected both economically and socially; technology adoption remains one of the defining factors in human developmental progress. In the case of th e NSO, there are instances where a noticeable rise in the use of the internet was evident. A higher proportion of people in the emerging and developing nations use the internet and own at least one smartphone (Poushter, 2016) (see Figure 8.2 and 8.3 respectively).

Figure 8.2: Worldwide internet usage - 2016 (Poushter, 2016)

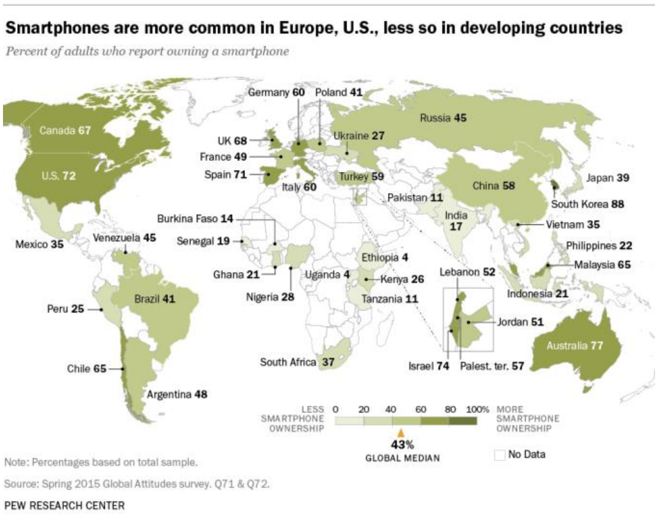


Figure 8.3: Access to Smartphones – 2016 (Poushter, 2016)

When considering DDC, the SASQAF needs to be adapted due to emerging digital technologies and the adoption thereof are key driving forces for DDC acceptation. Even though people in advanced economies still use the internet more, and own more high-tech gadgets, the rest of the emerging world is catching up fast.

8.2.3 Iteration 2: Changes to SASQAF

In Iteration 2, the proposed gaps to SASQAF revolves around the themes of staff training (Gap 4), how complicated DDC is (Gap5) and the transitioning from DDC to PAPDC processes within the NSO (Gap6) respectively. Training identified three improvements during Iteration 2, see Table 8.2 below.

PAPDC training methods incorporated content knowledge of survey definitions, standards, and operations and were usually presented using a combination of tools and assessments. Before deciding to use DDC instead of PAPDC, management should keep the training, cost, quality, staff, and infrastructural issues in mind (Development Outlook, 2012).

Table 8.2: Iteration 2: Gaps and improvements

DDC THEME	IMPROVEMENT REQUIRED	GAPS
TRAINING	***Improvement4** - (METH3(10); De4): Training only discusses the most appropriate data collection method(s) and instrument(s) (Statistics South Africa, 2010b). Require "Training Approaches and Outcomes" based on data collection method(s) and instrument ***Improvement5** - (METH3(10); Bu5): When finalising production systems, "Training Approaches and Outcomes" should be included in technical reports for all users ***Improvement6** - (METH3(10); Co1): Execution of training outcomes and approaches needs to be monitored and measured in terms of desired outcomes of trained staff	***Gap4** - DDC TRAINING APPROACHES - develop DDC training approaches
PROCESS FLOW AND STRUCTURAL CHANGES		***Gap5** (WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?): It was reported that DDC is not complicated in comparison to PAPDC processes. ***Gap6** (HOW DOES THE ORGANISATION TRANSITION TO ADAPT TO DDC?):

An important consideration that relates to cost, quality, staff, and infrastructure is training. Training requirements for DDC differ to those of PAPDC. In terms of training, Table 8.3 illustrates the levels of experience, instrument staff needs training on and the recruitment considerations between PAPDC and DDC concerning training processes:

Table 8.3: Gap3 - DDC vs. PAPDC staff and training

Process	DDC / Digital Survey	PAPDC / Paper Survey
Staffing and Training	- Staff should have basic field experience and basic computer knowledge. - Staff should be trained on both the questionnaire and device operations. - Recruitment and training typically take longer.	- Staff should have basic field experience. - Staff should be trained to address the questionnaire. - Recruitment is simpler in comparison to electronic surveys.

Source: (Development Outlook, 2012)

Staff training relates to organisational change management. Staff should be trained so that they can navigate the changes required to move from PAPDC to DDC. It is critical, from a change-management perspective, to assess the areas of change brought about by the adoption and the implementation of DDC. The NSO, as an employer, has a responsibility to assist employees to deal with this change. Careful planning and consultations with all the stakeholders affected by this change are essential, to minimise the associated risks and the challenges about changes in the statistical collection processes. This proposed change to SASQAF also takes into consideration the preparatory training/coaching work that needs to be done to stimulate a positive perception of DDC under internal staff. This section also addresses the perception of internal staff towards the transition from PAPDC to DDC. This transition, albeit focusing on the initial stages of the process flow, also sought to address ongoing strategies required for a smoother transition from PAPDC to DDC.

The NSO should adopt an incremental and phased approach to the transition that is anchored in its long-term digital-strategy implementation. The following steps are recommended, as the organisation seeks to implement an “Integration-by-Design” model (see Table 8.3):

Table 8.4: Integration by design model (Statistics South Africa, 2017a)

• Refine the strategy and the implementation plan
• Institutionalise stakeholder engagement (theme-based approach)
• Develop and institutionalise the enterprise architecture
• Integrate planning across the organisation (vertical and horizontal)
• Standardise methodologies
• Review the structure for Integration by Design
• Reskill staff and define new roles and responsibilities

•	The data revolution: research integration of administrative sources, as part of the business process (SVC)
•	Interrogate the SVC and identify changes, and discover how they can be addressed

Reskilling staff and defining new roles and responsibilities relate clearly to the proposed changes to SASQAF suggested from Iteration 2.

8.2.4 Iteration 2: Discussion and interpretation

Insufficient staff training and time allocated for training (among other factors) indicated that data-collection interfaces must be kept simple. Unless staff can be adequately trained to use new software, such as SS (refer to section 5.2.1). Training is an essential component of all DDC efforts. In the case of practice-based research networks (PBRNs), Pace and Staton (2005) indicated that both pros and cons need to be considered when conducting training for DDC (Pace & Staton, 2005). The execution and approaches to training differ from PAPDC, and the training outcomes can determine the eventual success or failure of DDC projects.

This section seeks to address the promotion of effective change management approaches and processes during the transition phase of DDC. These approaches and processes, once adopted, should ensure minimal disruptions to service delivery, and should alleviate the potential low staff morale. The change-management process emphasises the following (see Table 8.5 and Table 8.6 respectively) (Statistics South Africa, 2017a):

Table 8.5: Change-management process

• The adoption of shared and consistent approaches to change management during the PAPDC to DDC transition period; • Empowering and enabling management in the organisation to manage change effectively; • The impact of the change and transformation implementation on stakeholders; • Involvement of all stakeholders in the change process and managing their expectations, requirements, and concerns;	• Gaining support for the "new way of doing things"; managing resistance to change; • Ensuring that expectations are managed during the implementation process and post-implementation; • Supporting the leadership of the NSO to ensure visible sponsorship of the change through communication and involvement; and • Ensuring that there is frequent communication on the issue of project status.

• The preparation for and the monitoring of the impact of the change efforts	

Table 8.6: Critical success factors for change management

• Commitment at all levels. • Clear roles and responsibilities documented as part of a committee, job or role description, and underpinned by key performance indicators. • **Appropriate resourcing, engagement, and training (refer to Table 6.1).**	• Clear timelines and reporting structures. • Information required or desired determined in focus groups or other forms of communication with the key stakeholders. • Appropriate funding and budget processes to enable the strategies to be implemented.

8.2.5 Iteration 3: Changes to SASQAF

The most important changes to SASQAF suggested from Iteration 3 revolves around the themes of technology adoption and acceptance (Table 8.7). It has been proven that implementing new processes and especially IT related projects in organisations is extremely challenging (D'Ortenzio, 2012). This requires continuous communication on the reasons behind the change; how it may affect individuals or units internally and externally and how it can improve activities in the future. Significant change creates uncertainty, as employees seek to come to terms with new processes and new skills that is required. Employees need to be willing to adapt to new tools and participate in their training thereof. It is very important throughout all stages of major change to encourage innovation and eventual acceptance of all aspects of the work. This is done by involving staff at all levels in the organisation as much as possible and addressing their concerns. People react differently to change; some align naturally whilst others resist passively. Change needs to be understood and managed in a way that enables people to cope with it effectively. Employees need to understand the need for change and to have a chance to be involved in the planning and implementation of such change.

Table 8.7: Iteration 3: Gaps and improvements

DDC THEME	IMPROVEMENT REQUIRED	GAPS
TECHNOLOGY ADOPTION AND ACCEPTANCE		***Gap7** - **CS1. DDC COMMUNICATIONS:** - INTERNAL OR EXTERNAL DDC COMMUNICATION ***Gap8** - **CS2. EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT)**: EDUCATION LEVEL

The effect of new business processes, changes in organisational structure, or cultural changes must be appropriately managed (Alvesson & Sveningsson, 2007). Technology acceptance is also about people, and how they accept, and adapt to new technologies. Organisational change directly affects all areas of work from the entry-level employee to senior management. The entire organisation must learn how to transition to embrace the change. Extensive communication is essential when implementing technological changes, and it requires a multi-faceted communication strategy that includes key messages tailored to specific audiences.

This section has identified the need to address the experience, the educational level and the ability of staff to collect quality data. The well-documented rise in the use of mobile-handheld devices brings with it many opportunities for survey researchers to enhance and extend their measurement. However, using these technologies to improve survey measurement also presents its challenges. Debates surrounding the use of digital devices for business are becoming more prevalent. Some of these debates are related to coverage, the differential access to, or the use of these technologies within the workplace. With the increasing use of mobile devices, this has become more nuanced than the standard "digital divide" of the haves and the have-nots. The second-level digital divide distinguishes people based on how they use the technologies, rather than just whether or not they have the technology.

In a study done by Jäckle et.al (2017), suggests that the use of mobile technologies for personal purposes are increasing. This is increasing especially among older groups in the population where participation in survey activities by using these technologies is likely to increase (Jäckle, Burton, Couper, & Lessof, 2017).

8.2.6 Iteration 3: Discussion and interpretation

In organisations there is greater space between senior managers and staff, which slows down communication and interaction between these groups (Churchill, 1997). This gap can create confusion and tension; unless attempts are made to improve communication channels and opportunities for all staff (Churchill, 1997). The need is for competent (science-based), empowering and caring leadership that promotes competence, excellence, ownership, communication, and accountability. Through fair and dignified ways of managing and leading staff, to build good staff morale.

An efficient organisational structure is vitally important and should effectively delineate the spheres of responsibility, allow for fast, effective, efficient decision-making and the sharing of ideas across the organisation. The structure should also allow for easy and effective communication flows in pursuance of the organisation's goals and democratic outcomes. Finally, a structure that recognises and rewards the contribution of staff members; such a structure is important. The education level of staff together with the experience and ability to use digital handheld devices to collect the data are key factors to consider when developing DDC training programmes. In the context of modernisation, the existing training programs need to be reviewed. In Iteration 3, the central DDC theme of technological adoption and acceptance by staff towards DDC in terms of communications and the increasing abilities to use current mobile technologies necessitated the need for training material to be updated. This is required to reflect any new business processes, changes to the governance model, and other relevant information. If necessary, training programs on new skills should be developed. Training should include project management, common tools, technical, methodological and IT expertise.

The performance-management framework should include the recognition of excellence in employee performance including participation in modernisation initiatives. It should promote positive morale and ensure that performance issues are resolved quickly and effectively. Senior managers should be made accountable for the successful implementation of modernisation projects. This can be reinforced through their performance agreements. Senior managers will have to demonstrate greater leadership in guiding their teams, and they should be aware that the organisational culture might need to change, to realise modern outcomes.

8.2.7 Motivating the proposal of a DDC QA framework

Adhering to the guidelines and the high-level processes within SASQAF for manual household survey-data collection ensures that best practices are applied to the official statistics. Applying the same concepts by using digital handheld devices during the data collection phase of household surveys provides the same effect on user trust in the statistics collected when using this method. Modifying a manual process of data collection to custom-fit a digital process of data collection entails more than just the mapping or overlaying of current best practices for data collection. It requires the ability to adapt to the situation under which the household survey is to be conducted and ensuring the security of the data collected.

The QA process is a specialised domain, which covers the broader statistical survey life-cycle; and its importance and relevance cannot be undermined or underestimated. In mathematics, the Borromean rings consist of three linked topological circles. Removing any ring leaves the other two rings unconnected (see Figure 8.4). All three iterations are interlinked. The quantitative analysis of the secondary data proved that DDC remained consistent over increasing sample sizes. The diagram depicts the remaining DDC themes as they were studied during the respective iterations. The proposed DDC QA Framework is located in the core of Figure 8.4 below.

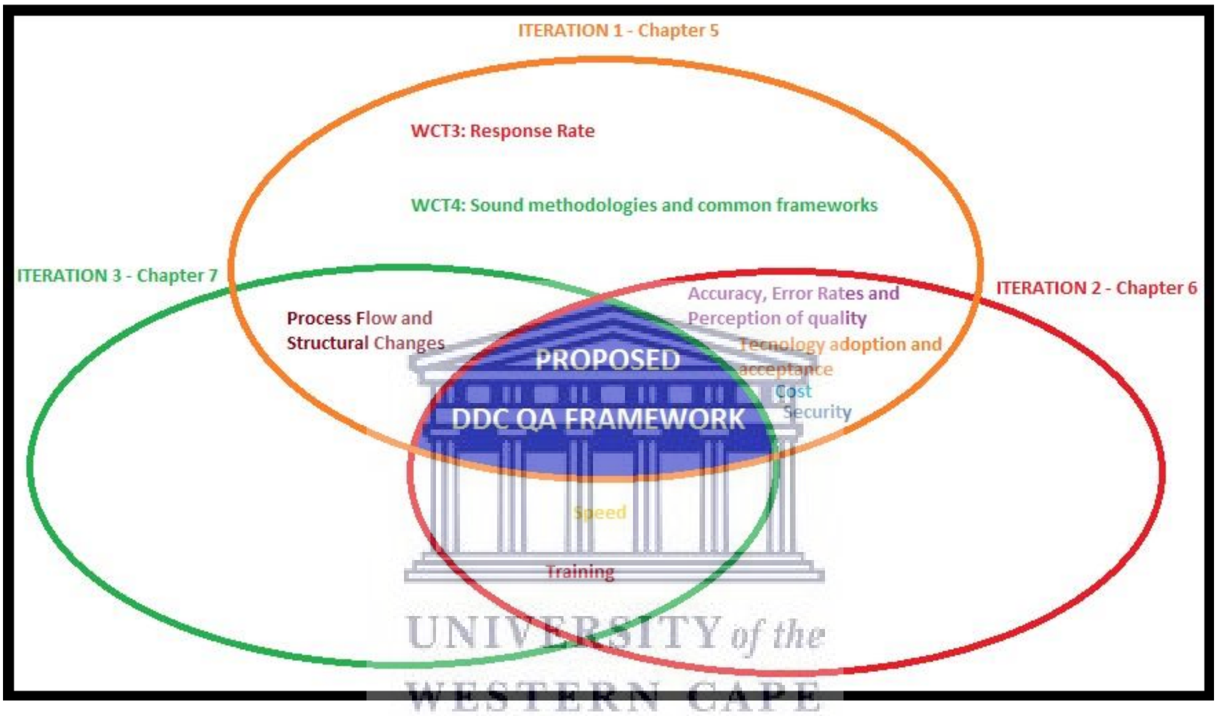


Figure 8.4: Proposed DDC QA Framework - Three Circle Model (Source: Author)

8.3 A PROPOSED QUALITY FRAMEWORK FOR DIGITAL DATA COLLECTION

The quality assurance framework for PAPDC is already developed and is currently being used by the NSO in South Africa as the standard for quality data collection for both household surveys and administrative data-collection. For DDC, quality standards or higher-level dimensions or themes need to be developed through the identification of the quality gaps on the standard survey life cycle and the broader quality framework as a whole. The three iterations in the previous chapters highlighted these gaps and improvements.

Table 8.8 highlights the additions that must be included in the proposed DDC developmental QA framework. All high-level quality dimensions based on the PAPDC SASQAF remain important; however, within these high-level dimensions, additional quality standards are required for the DDC.

Table 8.8: Proposed DDC developmental QA framework

DDC vs PAPDC SASQAF/SVC		
Proposed DDC SASQAF DIMENSIONS	**IMPROVEMENTS**	**Gaps Identified within SASQAF and SVC**
1. Pre-requisites of quality	***Improvement1** - (PRER6(7); De8) COST Adequate budget (Refer to Iteration 1)	***Gap8** CS2.EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT): (EDUCATION LEVEL) (Refer to Iteration 3)
	***Improvement2** - (PRER7(4); De8) COST Reviewed/Audited budget (Refer to Iteration 1)	
	***Improvement3** - (PRER4; Ne3) Improvements (Plans for theft (Theft, Asset Loss) and Data Loss) (Refer to Iteration 1)	
2. Relevance		
3. Accuracy		***Gap1** - (ACCU3(2); Co2) : How accurately does the SO find the correct sampled Dwelling? (Refer to Iteration 1)
4. Timeliness		
5. Accessibility		***Gap3** needs to be included in the development DDC Quality Assurance Framework. • USER ACCESS TO CELL PHONES • OWNERSHIP OF OTHER MOBILE DEVICES • USE OF MOBILE DEVICES FOR WORK PURPOSES (Refer to Iteration 1)
		***Gap7** CS1.DDC COMMUNICATIONS : (INTERNAL OR EXTERNAL DDC COMMUNICATION) (Refer to Iteration 3)
6. Interpretability		
7. Comparability and Coherence		
8. Methodological soundness	***Improvement4** - (METH3(10); De4) Training only discusses the most appropriate data collection method(s) and instrument(s). Require "Training Approaches and Outcomes" based on data collection method(s) and instrument (Refer to Iteration 2)	***Gap4** (DDC TRAINING APPROACHES - develop DDC training approaches) (Refer to Iteration 2)
	***Improvement5** - (METH3(10); Bu5) When finalising production systems, "Training Approaches and Outcomes" should be included in technical reports for all users (Refer to Iteration 2)	

	*Improvement6 - (METH3(10); Co1) Execution of training outcomes and approaches needs to be monitored and measured in terms of desired outcomes of trained staff (Refer to Iteration 2)	
9. Integrity		*Gap2 - (INTE5, Co2) - Data Loss (Refer to Iteration 1)
		*Gap5 WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES? It was reported that DDC is not complicated, in comparison to PAPDC processes. (Refer to Iteration 2)
		*Gap6 HOW DOES THE ORGANISATION TRANSITION TO ADAPT TO DDC? (Refer to Iteration 2)

(Source: Author)

8.4 CHAPTER SUMMARY

The development of a DDC Quality Assurance framework depend on the core principles of quality management. The approach to quality should be extended beyond the “fitness for use”,it should include the conformity to design specifications and value for money (Feo, 2017; Grandon Gill & Hevner, 2011). DDC costing has proven to be significantly more viable and cheaper than PAPDC. Both private and public sectors demand quality data for strategic planning, public-wide planning/monitoring, or attaining market dominance for financial gain, in the case of the private sector.

When transitioning from a PAPDC to a DDC method, a phased-in approach is recommended for the design and implementation of the proposed DDC developmental QA framework. These emanate primarily from lessons learned during Iteration 2 and 3, which dealt with staff training and technology acceptance. The standards and principles require both core and support area domains, internally and externally, these need to be equally covered. Both core and support areas should be prioritised simultaneously, to attain the most efficient results from the proposed DDC developmental QA framework implementation.

The work towards the establishment of an integrated generic process flow for DDC, incorporating the developmental DDC quality framework (DDCQF), should be based on the following activities:

- the establishment of a Quality-Management System (QMS) project team, which compiles a comprehensive project (survey) plan along with all the relevant stakeholders;

- the proposed DDC developmental QA framework highlighted the need to address the gaps in the pre-requisites of quality dimension. These include:
    - The experience and ability to use digital handheld device technology by internal staff or the end-users of DDC;
    - DDC communications;
    - Organisational processes;
- other high-level quality-dimension themes/areas include:
    - Methodological Soundness, Accessibility, and Integrity
- determining the current status of quality-management infrastructure
- adopting TQM, as the underlying base for the DDCQF
- developing a continuous process-improvement plan in relation to DDCQF

The following chapter concludes the overall study and assesses the contributions made towards answering the research questions.

CHAPTER 9: CONCLUSION AND RECOMMENDATIONS

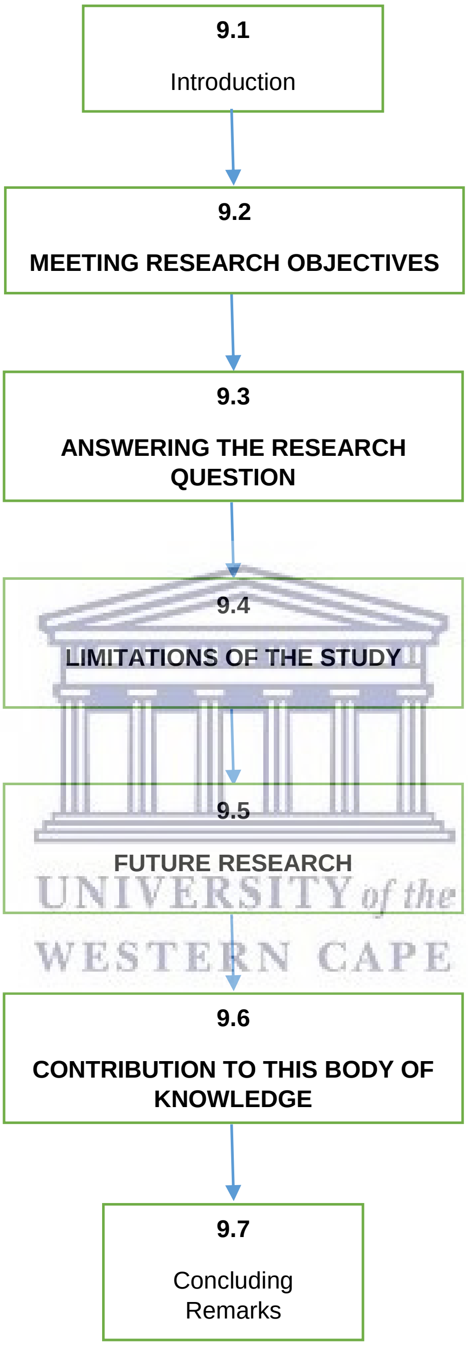


Figure 9.1: Chapter 9 layout

CHAPTER 9: CONCLUSION AND RECOMMENDATIONS

9.1 INTRODUCTION

Firstly this chapter concludes the overall research by revisiting and answering the research questions. Secondly, the chapter confirms that the research objectives were met. Thirdly the limitation of the study and future research is discussed. In conclusion, the study's contribution to the IS body of knowledge is formulated.

9.2 MEETING RESEARCH OBJECTIVES

This study sought to capacitate NSO's data-collection practitioners and management on how to improve the quality of their household-based survey collection. By incorporating the use of digital handheld technologies, while simultaneously identifying some of the direct potential challenges created by these technologies through the implementation of three iterations of DDC.

In addition to the problem statement of the study, the primary research objectives (PROs) of the study were:

- **PRO 1:** *To identify the shortfalls of SASQAF (if any) related to DDC processes, and to make suggestions for the enhancement of the existing PAPDC framework.*
    - o This was attained by reviewing and mapping the SASQAF and SVC for PAPDC in Chapter 4 and thereafter during the iteration; phases of chapters 5 and chapter 6 (refer to tables 5.7, 5.9, 5.11, 6.2, 6.3 and 6.4 respectively). Key gaps or shortfalls, improvements to work phases, responsibilities, timeframes, and processes were identified for DDC.

- **PRO 2:** *Investigate the current "best practices" applied in official household survey collection within a global and South African context.*
    - o Chapter 2 reviewed the literature related to "best practices for official household-survey collection" amongst NSOs across the globe (section 2.3.1, Table 2.3). Gaining an understanding of PAPDC Quality Assurance frameworks, and determining whether NSOs have already created or adopted a framework for DDC, or not.

- **PRO 3**: *Design and describe a quality framework for adopting an effective digitised household survey process in South Africa.*

- After determining the shortfalls and investigating the best practice from both PRO 1 and PRO 2, it was evident that a need existed for a newly proposed DDC quality assurance framework was required.
- Thematic groupings were developed, based on the literature described in Chapter 2 (Section 2.5, Table 2.7) by identifying a set of key themes linked to quality standards/indicators for PAPDC. Thereafter testing these through the empirical setting of the study throughout Chapter 4 and iteration Chapters 5 - 7. This produced a proposed list of new or adaptable quality standards/indicators that need to be considered. When conducting official DDCs as presented in Chapter 8 (refer to table 8.3) and is proposed as a developmental DDC QA Framework.

9.3 ANSWERING THE RESEARCH QUESTION

This study highlighted the fact that although there is a rapid change in available digital technologies, careful considerations are required when transitioning from a paper-based collection method for official household surveys to a digital one utilising handheld devices. In recent years, NSOs have found themselves in a position where processes are changing to incorporate elements of DDC. For South Africa as a country, the SASQAF framework does not cater for these changing processes within the generic PAPDC SVC components, however not fully towards DDC. Given the objectives of the study, the main research question (MRQ) and the sub-research questions (SRQs) were as follows:

MRQ:

How can the South African Statistical Quality Assurance Framework (SASQAF) be enhanced to support digitised household-survey processes in South Africa?

The MRQ was answered primarily in SRQ3, through the development of a proposed DDC developmental QA framework. SRQ3 required both SRQ1 and SRQ2 to be answered, before its realisation. Although the pilots and the tests were conducted primarily within the borders of South Africa, they should be tested more broadly in alternative spatial settings by NSOs across the African continent, to further solidify or generalise the findings for other African countries' NSOs.

The primary research question (MRQ) is furthermore subdivided into three sub-research questions (SRQs), and these are linked directly and chronologically to their respective PROs above:

SRQ 1: What are the shortfalls of the SASQAF when related to digital-data collection?

- Chapter 2's literature review identified the global-quality frameworks used by NSOs for official survey data collection, mainly using a PAPDC method. Chapter 4 unpacked the SASQAF used in South Africa, including the SVC, whereby linkages to DDC were derived. Chapters 5-7 displayed the analytical results from three iterations of DDC at varying sample size scales. Each iteration identified the need to either redefine or adjust the standard collection processes, methodologies or criteria when implementing DDC for official household survey data collection. The gaps and improvements (shortfalls) were summarised in Chapter 8.

SRQ 2: Which components of the household-survey process require changes for ensuring the creation of an effective developmental QA framework for DDC?

- Current "best practices" in quality frameworks across NSOs vary for official household survey data collections. However, there are key quality dimensions and standards that are prevalent across most of the NSOs related quality frameworks.
- From an implementation perspective, this study described the end-to-end household collection survey processes for DDC through the three iterations (Chapters 5 – 7).
- The themes, mapping and linkages discussed in Chapter 4 described which areas of the household-survey process needed to be enhanced, and how this should be done for DDC

SRQ 3: What would an effective framework entail to ensure high quality DDC?

- Chapter 8 introduced the proposed DDC developmental QA framework. The findings, discussions, and interpretations of the preceding chapters formed the core reasoning for the newly developed proposed framework.

9.4 LIMITATIONS OF THE STUDY

The primary limitation of the research included developing a QA framework for DDC within the South African context. The scope of the research is limited to official household-based social surveys conducted by the official national supplier of statistics (Stats SA) within South Africa. Due to the nature of the case study focussing only on one case, it does not include extensive research into all the other NSOs on the African continent and abroad. The statistical evaluation of the primary questionnaire could have been further analysed by using techniques, such as factor analysis, or structural-equation analysis.

9.5 FUTURE RESEARCH

Modern-day business and system applications often require rapid evolutionary approaches to be adopted before they are more widely utilised. Similarly, informed quality knowledge-based or information or data remain a pre-requisite for answering the ever-evolving challenges in the social and economic spheres locally, as well as globally. Given the findings of the current research, there are future related-research areas, on which academia could focus on. Research measuring the impact of **para-data** captured during the collection of data from digital handheld devices could be used to study behavioural patterns displayed by respondents responding to the questions and the techniques employed by the interviewer when asking the questions. Furthermore, in web surveys, researchers are making increasing use of para-data; such as keystrokes, clicks, and timestamps. This to evaluate and improve survey instruments, but also to understand respondents and how they answer surveys (Mick P Couper & Singer, 2013). An example of para-data, which is characterised by timestamps and actions, assists trainers of DDC surveys with substantial real-time evidence (data) when training INTERVIEWERS. This is based on their respective patterns for collecting primary data in order to improve the quality or performance of the data collected from the respondents.

Big data is starting to be incorporated or developed through an all-inclusive adaptable DDC – QA framework by the NSS public / private partnerships within countries. While big data potentially holds great promise in terms of their value for official statistics, the changes required to the practices and the ways of working for statistical authorities and the statistical techniques used are significant. In particular, statistical authorities should assume greater and different responsibilities and functions. Some of these changes will demand greater flexibility and adaptability of approach (Khan et al., 2014). Big data or transactional digital data integration across various suppliers of data fall within the ambit of the NSS. NSOs strategic planning should include partnerships to further explore this domain.

Today, many NSOs face the dual challenge of declining response rates and budgetary cuts. Using **mixed-mode data-collection** techniques, as well as switching between modes between CAPI, CATI and CAWI might improve survey response rates and reduce survey costs. However, achieving both simultaneously is difficult (Johnson & Onwuegbuzie, 2004). Research into Digital Quality Frameworks encompassing all **modes switching** and/or mode selection via parallel, or otherwise, could assist in the most efficient or economical collection of quality data by any NSO. These should include the response-rate analysis between DDC and PAPDC, as well as the relevant error imputation rates.

SASQAF Lite was introduced for PAPDC for NSOs or other NSS partners to self-assess the quality of their respective statistical collections. For DDC, the question remains whether a similar product (i.e. a self-assessment tool) could also be developed to cater for the needs of NSS partners seeking to comply with the requisite quality measures to be considered when conducting DDC for household surveys.

9.6 CONTRIBUTION TO THE BODY OF KNOWLEDGE

Within the discipline of IS, this study presents a *proposed DDC developmental QA framework*. As a point of reference for NSO management planning to adopt handheld mobile DDC for collecting official household surveys. More importantly, for using it as a guideline or benchmark when determining the quality of data collected. The development of a QA framework for DDC contributes to the body of IS knowledge, by taking into cognisance its ability to improve the quality of an official survey-data collection at an NSO in South Africa. The framework is easily comprehensible and adaptable, given that a PAPDC framework called the SASQAF currently exists within the NSO, and similarly across all NSOs within Africa.

There is no single framework, model or theory that can be applied to DDC in SA. While this framework was developed for the NSO in South Africa, portions of the proposed framework can also applied to any NSO in Africa collecting official household-survey data. This is because NSO environments across countries have very similar or homogeneous characteristics, this in relation to the outcomes for dealing with collecting data for the state (government) of the country or for public consumption. For this reason, the proposed DDC developmental QA framework has a certain degree of flexibility and generalisation. Equally important, the *proposed DDC developmental QA framework* contributes to the emerging body of knowledge, in terms of the digital mobile-handheld implementation of official surveys in a developing country, such as South Africa; specifically highlighting the components of DDC implementation in a complex NSO domain.

Organisational transitional changes towards quality DDC within an NSO can be iteratively assessed through DDC implementations. They can be adapted over time through the requisite qualitative and quantitative measures. This study also contributes to NSOs seeking to adopt digital-handheld technologies to collect quality data more frequently at a lower cost. Given that the benefits of DDC are clear at the levels of being, amongst others, economical, secure, and cost- and time-effective. Most African NSOs, as well as other conduits or entities using information systems of population or migration studies, are increasingly using handheld mobile devices as a viable option, to collect the data and respond to their user needs within today's information-hungry societies. (Baig et al., 2014)

Furthermore, a QA framework for NSOs seeking to collect household survey data using handheld devices is vital to the continuous collection of information facilitating those fast-growing country democracies and their respective economies. The contribution to the body of knowledge is also necessary on a practical level for NSOs, when considering the transitioning from a manual paper-based data collection mode to a digital one. Another major contribution (theoretically and methodologically) of this thesis, is the use of the PDCA/PDSA cycle or Deming cycle (Shewhart & Deming, 1986; Weinstein & Vasovski, 2004) as a theoretical underpinning in IS, to execute a research project.

To date, there is no developmental QA framework for DDC; although many quality frameworks exist for PAPDC modes. This research has produced a proposed developmental QA framework (or QA product) to consider, when conducting small or large-scale DDC within the South African, or the broader African context.

9.7 CONCLUDING REMARKS

The conversion from PAPDC to DDC in South Africa, and also in other developing country contexts has presented researchers with both benefits and challenges. It highlighted some of the significant issues that remain to date, unresolved in the implementation of DDC. Other studies have indicated that the advantages gained in DDC have highlighted the potential for the use of DDC in the collection from smaller to larger data sample size collections.

In conclusion, the future of DDC will be based on the ever-changing and adaptable processes taken by NSOs. These are based on a well-refined SVC, from the need, design, build and rapid development of questionnaire design, to operationalising an effective operational process of collection, to the field of collection and ultimately the tabulation of weighted quality data for dissemination. Free, open-source websites offering tools for building, collecting and aggregating data digitally are readily available and relatively easy to use. With the increasing use of DDC and cloud-based platforms, it will become increasingly important, to ensure that the public is provided with the necessary information pertinent to the principles of official statistics.

REFERENCES

Adams, G. and Schvaneveldt, J. (1991). *Understanding Research Methods.* (2nd edn.). New York: Longman.

Adesina, A. O., Agbele, K. K., Februarie, R., Abidoye, A. P., & Nyongesa, H. O. (2011). Ensuring the security and privacy of information in mobile health-care communication systems. *South African Journal of Science*, *107*(9–10), 1–7. https://doi.org/10.4102/sajs.v107i9/10.508

African Union Commission. (2015). Agenda 2063, (August), 24. Retrieved from www.agenda2063.au.int

Ahmed, R., Robinson, R., Elsony, A., Thomson, R., Bertel Squire, S., Malmborg, R., … Mortimer, K. (2018). A comparison of smartphone and paper data-collection tools in the Burden of Obstructive Lung Disease (BOLD) study in Gezira state, Sudan. *PLoS ONE*, *13*(3), 1–15. https://doi.org/10.1371/journal.pone.0193917

Al-Azri, A. (2010). The Successful Implementation of E-Government Transformation: A Case Study in Oman. *European, Mediterranean & Middle Eastern Conference on Information Systems*, 1–11.

Alvesson, M., & Sveningsson, S. (2007). *Changing organizational culture: Cultural change work in progress*. *Changing Organizational Culture: Cultural Change Work in Progress*. https://doi.org/10.4324/9780203935965

Alwin, D. (2007). *Margins of error: a study of reliability in survey measurement*. Wiley-Blackwell.

Ameen, N., Willis, R., & Hussain Shah, M. (2018). An examination of the gender gap in smartphone adoption and use in Arab countries: A cross-national study. *Computers in Human Behavior*, *89*(March), 148–162. https://doi.org/10.1016/J.CHB.2018.07.045

Andersson, C., Lindstrom, H. L. and Lyberg, L. (1997). Quality declarations at Statistics Sweden—principles and practices. *Statistical Policy Working Paper 26*.

Andersson, R., Eriksson, H., & Torstensson, H. (2006). Similarities and differences between TQM, six sigma and lean. *TQM Magazine*, *18*(3), 282–296. https://doi.org/10.1108/09544780610660004

Andrade, A. D. (2009). Interpretive Research Aiming at Theory Building : Adopting and Adapting the Case Study Design. *The Qualitative Report*, (14), 42–60.

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of Mobile versus PC Web on Survey Response QualityA Crossover Experiment in a Probability Web Panel. *Public Opinion Quarterly*, *81*(S1), 280–306. Retrieved from http://dx.doi.org/10.1093/poq/nfw088

Australian Bureau of Statistics. (2015). Data Quality Online. Retrieved July 28, 2017, from https://www.nss.gov.au/dataquality/purpose.jsp

Baig, A., Cadena, A., Chui, M., Dua, A., Farrell, D., Law, J., … Woetzel, J. (2014). Innovation in Local Government: Open Data and Information Technology. *McKinsey&Company*. Retrieved from http://www.mckinsey.com/industries/public-sector/our-insights/government-designed-for-new-times-number-2

Baxter, P. and Jack, S. (2008). Qualitative Case Study Methodology: Study Design and implementation for Novice Researchers. *The Qualitative Report*, *13*, 544–559.

Bentahar, O., & Cameron, R. (2014). Design and Implementation of a Mixed Method Research Study in Project Management. *The Electronic Journal of Business*, *13*(1477–7029), 16. https://doi.org/ISSN 1477-7029

Bergdahl, M., Ehling, M., Elvers, E., & Földesi, E. (2007). Quality indicators. *Handbook on Data Quality Assessment Methods and Tools*, 16–20.

Bergman, E. (2000). Information Appliances and Beyond: Interaction Design For Consumer Products. *London, Morgan Kaufmann*.

Bernard, H. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*.

BIZCOMMUNITY. (2018). Five ways POPI will change marketing research as we know it. Retrieved September 5, 2018, from http://www.bizcommunity.com/Article/196/19/148782.html

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291. Retrieved from http://dx.doi.org/10.1093/pubmed/fdi031

Brand South Africa. (2015). Brand South Africa Research Note The African Union & Agenda Summary & Overview Aspirations Vision, (September).

Brix, J. (2017). Exploring knowledge creation processes as a source of organizational learning: A longitudinal case study of a public innovation project. *Scandinavian Journal of Management*, *33*(2), 113–127. https://doi.org/10.1016/j.scaman.2017.05.001

Bukachi, F., & Pakenham-Walsh, N. (2007). Information technology for health in developing countries. *Chest*, *132*(5), 1624–1630. https://doi.org/10.1378/chest.07-1760

BusinessDictionary. (2017). What is Information? definition and meaning - BusinessDictionary.com. Retrieved February 15, 2018, from http://www.businessdictionary.com/definition/information.html

Cao, J., Crews, J. M., Lin, M., Deokar, A., Burgoon, J. K., & Nunamaker Jr, J. F. (2006). Interactions Between System Evaluation And Theory Testing: A Demonstration of the Power of a Mulitfaceted Approach to Systems Research. *Journal of Management Information Systems*, *22*(4), 207–235. https://doi.org/10.2753/MIS0742-1222220408

Carlsson, S. A. (2006). Design Science Research in Information Systems: A Critical Realist Perspective. *17th Australasian Conference on Information Systems*, 40. https://doi.org/10.1007/978-1-4419-5653-8_15

Chen, T., Raeside, R., & Khan, H. (2014). Development and Application of a Survey Quality Assessment Model, *1*(1), 15–29.

Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(3), 197–235.

Choy, L. T. (2014). The Strengths and Weaknesses of Research Methodology: Comparison and Complimentary between Qualitative and Quantitative Approaches. *Journal of Humanities and Social Science*, *19*, 99–104.

Churchill, C. F. (1997). Managing Growth: The Organizational Architecture of Microfinance Institutions. *1997*.

Collis, J., Hussey, R. (2003). *Business research: a practical guide for undergraduate and postgraduate students*. Houndmills, Basingstoke, Hampshire, Palgrave Macmillan.

Cook, S. (2012). Interpretivist Sociology. Retrieved from https://revisesociology.wordpress.com/2012/05/13/3-interpretivist-sociology/

Couper, M. P. (2005). Technology Trends in Survey Data Collection. *Social Science Computer Review*, *23*, 486–501. https://doi.org/10.1177/0894439305278972

Couper, Mick P. (2017). New Developments in Survey Data Collection. *Annual Review of Sociology*, *43*(1), 121–145. https://doi.org/10.1146/annurev-soc-060116-053613

Couper, Mick P, & Singer, E. (2013). Informed Consent for Web Paradata Use, *7*(1), 57–67.

Creswell, J. W. (2003). *Research Design Qualitative Quantitative and Mixed Methods*

Approaches. Thousand Oaks: Sage Publications.

Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process* (London: SA).

D'Ortenzio, C. (2012). UNDERSTANDING CHANGE AND CHANGE MANAGEMENT PROCESSES : A CASE STUDY, (August).

David B Resnik. (2015). What is Ethics in Research & Why is it Important? *National Institutes of Health - U.S. Department of Health and Human Services*. Retrieved from https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13(3)*, 319–340.

Deloitte Development LLC. (2013). Doing_more_with_less_in_contac.PDF.

Development Outlook. (2012). Digital Data Collection Vs. Data Collection on Paper | Development Outlook | An IFMR LEAD Blog. Retrieved February 13, 2018, from http://www.developmentoutlook.org/2012/07/digital-data-collection-vs-data.html

Dobbs, J., Gibbins, C., Martin, J., Davies, P., & Dodd, T. (1991). REPORTING ON DATA QUALITY AND PROCESS QUALITY, 32–40.

Eng, J. L. Vanden, Wolkon, A., Frolov, A. S., Terlouw, D. J., Eliades, M. J., Morgah, K., … Hightower, A. W. (2007). Use of Handheld Computers with Global Positioning Systems for Probability Sampling and Data Entry in Household Surveys, *77*(2), 393–399.

Engela, R., Ajam, T., Banque mondiale, & Banque mondiale. (2010). *Implementing a government-wide monitoring and evaluation system in South Africa*. World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/27574

EUROSTAT. (2009). Statistics in development cooperation - quality in statistics - Statistics Explained. https://doi.org/ISSN 2443-8219

EUROSTAT. (2015). Quality Assurance Framework of the European Statistical System.

Eurostat, European Commission, & Statistika centralbyran. (1999). Quality Work and Quality Assurance within Statistics.

Feo, J. A. De. (2017). Conformance to Specification and Fitness for Use. *Juran's Quality Handbook: The Complete Guide to Performance Excellence, Seventh Edition*. McGraw Hill Professional, Access Engineering. Retrieved from https://www.accessengineeringlibrary.com:443/browse/jurans-quality-handbook-the-complete-guide-to-performance-excellence-seventh-edition/c9781259643613ch24lev1sec03

Fitzgerald, G., & Fitzgibbon, M. (2014). *A Comparative Analysis of Traditional and Digital Data Collection Methods in Social Research in LDCs - Case Studies Exploring Implications for Participation , Empowerment , and (mis) Understandings*. *IFAC Proceedings Volumes* (Vol. 47). IFAC. https://doi.org/10.3182/20140824-6-ZA-1003.02796

Flick, U. (1998). *An introduction to qualitative research*. London: Sage Publications.

Forster, D., Snow, B., Schellenberg, J. A., Manzi, F., John, D., Mushi, A., … Schellenberg, D. (1992). Using microcomputers for rapid data collection in developing countries. *Health Policy and Planning*, *7*(1), 67–71. https://doi.org/10.1093/heapol/7.1.67

Frank, N. (2014). Beyond 2011 - Future Population Statistics for the United Kingdom, 1181–1188.

Gibson, D. G., Pereira, A., Farrenkopf, B. A., Labrique, A. B., Pariyo, G. W., & Hyder, A. A. (2017). Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review. *Journal of Medical Internet Research*, *19*(5), e139. https://doi.org/10.2196/jmir.7428

Gonzalez, R. (2014). A Concept Map for Information Systems Research Approaches, (May).

Google. (2015). Google My Maps. Retrieved October 24, 2015, from https://www.google.com/maps/d/

Graham, J. W., Smith, E. A., Caldwell, L. L., & Mathews, C. (2008). Rates of missing responses in Personal Digital Assistant (PDA) versus paper assessments, 257–272.

Grandon Gill, T., & Hevner, A. R. (2011). A Fitness-Utility Model for Design Science Research The Dependent Variable in Design Science Research. *Lncs*, *6629*, 237–252. Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-642-20633-7_17.pdf

Guba, E. G., & Lincoln, Y. S. (2005). *Paradigmatic Controversies, Contradictions, and Emerging Confluences* (3rd Editio). Sage, Thousand Oaks.

Hattas, M. (2013). *Towards secure digital data collection (sddc) in household surveys for South Africa*. University of South Africa.

Hattas, M., & Breytenbach, J. (2017). Improving the quality of official survey data collection in South Africa through digital data collection. In *XXVIII International Population*

Conference. Cape Town. Retrieved from https://iussp.confex.com/iussp/ipc2017/meetingapp.cgi/Paper/4002

Hattas, M., & Eloff, M. (2011). Secure digital data collection in household surveys - Case study: Statistics South Africa. In *IEEE AFRICON Conference*. https://doi.org/10.1109/AFRCON.2011.6072142

Hirschheim, R. (1992). Information Systems Epistemology: an historical perspective. In Henley-on-Thames (Ed.), *in Galliers, R. (Ed.) Information Systems Research: issues, methods and practical guidelines* (Alfred Wal).

Hole, G. (2011). Kruskal-Wallis handout, *100*, 1–10.

Iacono, J., Brown, A. and Holtham, C. (2009). Research Methods—A Case Example of Participant Observation. *The Electronic Journal of Business Research Methods*, *7*, 39–46.

International Monetary Fund. (2012). IMF -Introduction to the Data Quality Reference Site. Retrieved August 12, 2017, from http://dsbb.imf.org/pages/dqrs/DQAF.aspx

ITU. (2014). The World in 2014 - ICT Facts and Figures. *International Telecommunication Union*, (in 2008), 8. https://doi.org/10.1787/9789264202085-5-en

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2017). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: response rates and response biases. *Understanding Society Working Paper*, *2017–09*.

Jebreen, I. (2012). Using Inductive Approach as Research Strategy in Requirements Engineering. *International Journal of Computer and Information Technology*, (1), 162–173.

John H. McDonald. (2014). *Kruskal–Wallis test - Handbook of Biological Statistics*. *Handbook of Biological Statistics*. Sparky House Publishing. Retrieved from http://www.biostathandbook.com/kruskalwallis.html

John, P. (2016). The policy agendas project : a review The Policy Agendas Project : a review, *1763*(March), 1–24. https://doi.org/10.1080/13501760600923870

Johnson, & Onwuegbuzie. (2004). Mixed methods research: a research paradigm whose time has come. *American Educational Research Association*, *33*(7), 14–26.

Karthikeyan.P, & K.A, M. (2010). Research Framework : Paradigms in Information Science Research, *10*(5).

Kawulich, B. (2015). Selecting a research approach : Paradigm , methodology and methods, (January 2012).

Kelegama, S. (2016, December 20). Role of statistics for the economic and social development of a country. *The 5th Annual General Meeting of the Institute of Applied Statistics Sri Lanka*. Retrieved from http://www.ft.lk/article/587967/Role-of-statistics-for-the-economic-and-social-development-of-a-country

Keller, R. K. (1995). User Interface Tools : a Survey and Perspective. *Lecture Notes in Computer Science, 896*(SOFTWARE ENGINEERING AND HUMAN-COMPUTER INTERACTION), 225–231. https://doi.org/10.1007/BFb0035818

Khan, N., Yaqoob, I., Abaker, I., Hashem, T., Inayat, Z., Kamaleldin, W., … Gani, A. (2014). Big Data : Survey , Technologies , Opportunities , and Challenges, *2014*.

Kilic, T., Serajuddin, U., Uematsu, H., & Yoshida, N. (2017). Costing Household Surveys for Monitoring Progress Toward Ending Extreme Poverty and Boosting Shared Prosperity. *World Bank Policy Research Working Paper*, *7951*(January). Retrieved from http://documents.worldbank.org/curated/en/260501485264312208/pdf/WPS7951.pdf

Kizito, S., Oscar, M., Fatuma, M., Davis, J., Adiel, M., Mwifadhi, M., … David, S. (2007). The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. *Emerging Themes in Epidemiology*, *4*.

Kjeldskov, J., & Stage, J. (2002). Designing the User Interface of a Handheld Device for Communication in a High-Risk Environment. *Proceedings of the ERCIM Conference on Universal Access*. Retrieved from http://www.cs.auc.dk/~jesper/papers/ERCIM02-final.pdf%5Cnpapers2://publication/uuid/8372767C-0709-4F36-AB3D-434108FFA9C8

Klein, H. & Myers, M. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, *Vol. 23*(No. 1), 67–96.

Kniivilä, M. (2007). Industrial Development and Economic Growth: Implications for Poverty Reduction and Income Inequality. *Industrial Development for the 21st Century: Sustainable Development Perspectives*, (1), 295–332. Retrieved from http://www.un.org/esa/sustdev/publications/industrial_development/3_1.pdf Accessed: 8/08/2016, 23:07

Lauren, H. (2003). From PAPI to CAPI: consequences for data quality on the British Household Panel Study. *ISER Working Paper Series, No. 2003-14*. Retrieved from https://www.econstor.eu/bitstream/10419/92209/1/2003-14.pdf

Law Society of South Africa. (2018). Protection of Personal Information for South African Law Firms Protection of Personal Information Guideline 2018. Retrieved from https://www.lssa.org.za/upload/files/Resource documents/Protection of Personal Information for South African Law Firms LSSA Guidelines 2018.pdf

Lee, A. S. (2004). Thinking about social theory and philosophy for information systems. *Social Theory and Philosophy for Information Systems*, 1–26.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008

Lukwago, L., Nanyunja, M., Ndayimirije, N., Wamala, J., Malimbo, M., Mbabazi, W., … Talisuna, A. (2013). The implementation of Integrated Disease Surveillance and Response in Uganda: a review of progress and challenges between 2001 and 2007. *Health Policy and Planning*, *28*(1), 30–40. Retrieved from http://dx.doi.org/10.1093/heapol/czs022

Marais, H. J. (2017). ASSESSING THE DATA QUALITY OF PERFORMANCE INFORMATION GENERATED BY THE HEALTH SECTOR IN THE BREEDE VALLEY SUBDISTRICT FOR EVIDENCE-BASED DECISION- by, (March).

Maritz, J., Van Huyssteen, E., Le Roux, A., Pieterse, A., Ndaba, D., Mans, G., & Ngidi, M. (2016). Are we achieving spatial transformation In South Africa? Can sub-city spatial indicators make a contribution? Retrieved from http://researchspace.csir.co.za/dspace/handle/10204/9244

Mehra, B. (2001). Bias in Qualitative Research: Voices from an Online Classroom. *The Qualitative Report*, *7*(1), 1–19. Retrieved from http://nsuworks.nova.edu/tqr/vol7/iss1/2/?utm_source=nsuworks.nova.edu%2Ftqr%2Fvol7%2Fiss1%2F2&utm_medium=PDF&utm_campaign=PDFCoverPages%5Cnhttp://www.nova.edu/ssss/QR/QR7-1/mehra.html

Merriam-Webster Dictionary. (2018). Definition of information. Retrieved from https://www.merriam-webster.com/dictionary/information

Meyer, C. (2001). A Case in Case Study Methodology. *Field Methods*, 329–352. Retrieved from http://dx.doi.org/10.1177/1525822X0101300402

Meyer, J. (2008). Older Workers and the Adoption of New Technologies, (07). Retrieved from http://ftp.zew.de/pub/zew-docs/dp/dp07050.pdf

Mingers, John; (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. https://doi.org/10.1287/isre.12.3.240.9709

Mingers, John, & Stahl, B. (2018). Philosophy and information systems: where are we and where should we go? *European Journal of Information Systems*, *27*(3), 263–277. https://doi.org/10.1080/0960085X.2018.1470776

Minister of Industry. (2017). *Statistics Canada's Quality Assurance Framework Third Edition - 2017*.

Mitev, N. (2000). Toward Social Constructivist Understanding of IS Success and Failure: introducing a new computerized reservation system. In *Proceedings of the twenty fist international conference on information systems* (pp. 84–93). Brisbane, Queensland, Australia,.

Moussiegt, L., DeStefano, T., & De Backer, K. (2017). Determinants of digital technology use by companies. *OECD Science, Technology and Industry Policy Papers*, (40), 53. https://doi.org/http://dx.doi.org/10.1787/23074957

Munoz, J. (2003). A guide for data management of household surveys. *United Nations, The Analysis of Operating …*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.246&rep=rep1&type=pdf

Nandhakumar, J. & Jones, M. (1997). Too Close for Comfort? Distance and engagement in interpretive information systems research. In *Information Systems Journal* (Vol. No. 7, pp. 109–131). Blackwell Science Ltd.

NATIONAL INSTITUTE OF STATISTICS OF RWANDA. (2014). National Quality Assurance Framework.

National Treasury. (2010). Framework for Strategic Plans and Annual PerformancePlans.

Nederpelt, P. Van. (2011). Standard for statistical processes 2011.

Nicholaas G, Thomson K, L. P. (2000). The feasibility of conducting electoral surveys in the UK by telephone. *Centre for Research into Elections and Social Trends*.

Njie, B. and Asimiran, S. (2014). Case Study as a Choice in Qualitative Methodology. *IOSR Journal of Research & Method in Education*, *4*, 35–40. Retrieved from http://dx.doi.org/10.9790/7388-04313540

North, H. (2015). Citizen Satisfaction Survey 2015.

Oates, B. J. (2006). *Researching Information Systems and Computing* (1st ed.). London: SAGE.

OECD. (2001). OECD Glossary of Statistical Terms - Statistical information system Definition. Retrieved February 15, 2018, from https://stats.oecd.org/glossary/detail.asp?ID=3010

Olsen, W. (2004). Triangulation in social research: Qualitative and quantitative methods can really be mixed. *Developments in Sociology*, 1–334. https://doi.org/10.4324/9781315838120

Onatu, G. O. (2013). Building Theory from Case Study Research : The Unanwered Question in Social Sciences. *1st Global Virtual Conference*, *1*, 170–173.

Organisation for Economic Co-operation and Development. (2007). Glossary of statistical terms.

Osborne, J. (1942). Sampling errors of systematic and random surveys of cover-type areas. *Journal of the American Statistical Association*, *37*(218), 256–264.

Pace, W. D., & Staton, E. W. (2005). Electronic data collection options for practice-based research networks. *Annals of Family Medicine*, *3 Suppl 1*(Suppl 1), S21-9. https://doi.org/10.1370/afm.270

Parliament of the Republic of South Africa. (2013). Protection of Personal Information Act, 2013 Ensuring protection of your personal information and effective access to information, (4), 154. Retrieved from http://www.justice.gov.za/inforeg/docs/InfoRegSA-POPIA-act2013-004.pdf

Pfleeger C., & P. S. L. (2007). *Security in computing* ((4th ed)). Upper Saddle River: NJ: Prentice Hall.

Pistorius, M. (2010). UN Expert Group on National Quality Assurance Frameworks.

Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies | Pew Research Center. Retrieved February 12, 2018, from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Rai, A., Ajay, A., Sharma, S., Pal, S., Arora, A., & Paul, R. K. (2017). Open data kit for diagnostic crop production survey at landscape level in India, 11–17.

Reid, R. D., & Sanders, N. R. (2013). *Operations Management, 5th Edition, Reid & Sanders, 2013: Operations Management*. John Wiley & Sons, Inc. Retrieved from https://books.google.co.za/books?id=ee9lDwAAQBAJ

Remenyi, D. (2002). Research strategies - Beyond the differences. *Electronic Journal of Business Research Methods*, *1*(1), 38–41.

Republic of South Africa: Department of Provincial and Local Government. (2000). Integrated Development Planning System.

Republic of South Africa: The Presidency. (2005). Provincial Growth and Development Strategy Guidelines Provincial Growth and Development Strategy Guidelines, (July).

Republic of South Africa. (2014). Medium Term Strategic Framework Combined 1-14, (June 2014).

Riddell, W. C., & Song, X. (2011). The Role of Education in Technology Use and Adoption : Evidence from the Canadian Workplace and Employee Survey. *CLSRN Working Papers*, (83), 1–34. https://doi.org/10.1177/0019793916687719.

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). *New York, NY: The Free Press.*

Ronald Moen. (2009). Foundation and History of the PDSA Cycle. *Associates in Process Improvement-Detroit (USA)*, 2–10. Retrieved from https://www.deming.org/sites/default/files/pdf/2015/PDSA_History_Ron_Moen.pdf

Rowley, J. (2002). Using Case Studies in Research. *Management Research News*, (25), 16–27. Retrieved from http://dx.doi.org/10.1108/01409170210782990

Ruxton, G. D., & Beauchamp, G. (2008). Some suggestions about appropriate use of the Kruskal-Wallis test. *Animal Behaviour*, *76*(3), 1083–1087. https://doi.org/10.1016/j.anbehav.2008.04.011

Ryan JM, Corry JR, Attewell R, S. M. (2002). A comparison of the electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Qual Life Res*, (11), 19–26.

SANS 27002. (2008). Information technology - Security techniques - Code of practice for information security management.

Sarndal C.E., Swensson B., W. J. (2002). OECD Glossary of Statistical Terms - Non-response Definition. Retrieved October 29, 2018, from https://stats.oecd.org/glossary/detail.asp?ID=3764

Saunders, M., Lewis, P., & Thornhill, A. (2008). *Research Methods for Business Students*. *Research methods for business students*. https://doi.org/10.1007/s13398-014-0173-7.2

Seebregts, C. J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A. J., Seebregts, C., … Klepp, K. I. (2009). Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm??? Pilot interviewing system. *International Journal of Medical Informatics*, *78*, 721–731. https://doi.org/10.1016/j.ijmedinf.2008.10.006

Shanks, G. (2002). Guidelines for conducting positivist case study research in information systems. *Australasian Journal of Information Systems*, *10*(December), 76–85. https://doi.org/10.3127/ajis.v10i1.448

Sharma, G., Bao, X. and Peng, L. (2014). Public Participation and Ethical Issues on E-Governance : A Study Perspective in Nepal. *Electronic Journal of E-Government*, *12*, 82–96.

Shewhart, W. A., & Deming, W. E. (1986). *Statistical Method from the Viewpoint of Quality Control*. Dover Publications. Retrieved from https://books.google.co.za/books?id=ALGbNNMdnHkC

Shezi, L. (2016). SA's 26.8 million internet users spend almost three hours a day on social media - htxt.africa. Retrieved July 20, 2017, from http://www.htxt.co.za/2016/04/29/the-stuff-south-africa-26-8-mil-internet-users-spend-most-their-time-doing-online/

Simon, M. (2011). Validity and Reliability in Qualitative Research. *Curationis*, *16*, 35–38.

Smith, T. W. (2011). Report on the International Workshop on Using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys, 1–29.

Somda, Z. C., Meltzer, M. I., Perry, H. N., Messonnier, N. E., Abdulmumini, U., Mebrahtu, G., … Sow, I. (2009). Cost analysis of an integrated disease surveillance and response system: case of Burkina Faso, Eritrea, and Mali. *Cost Effectiveness and Resource Allocation*, *7*(1), 1. https://doi.org/10.1186/1478-7547-7-1

Somda, Z. C., Perry, H. N., Messonnier, N. R., Djingarey, M. H., Ki, S. O., & Meltzer, M. I. (2010). Modeling the cost-effectiveness of the integrated disease surveillance and response (IDSR) system: Meningitis in Burkina Faso. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0013044

South African Ministry of Planning. (2012). *National Development Plan 2030 - Chapter 8 - Transforming human settlement and the national space economy* (Vol. Chapter 8).

Southern African Development Community (SADC). (2017). Summary of the SADC Revised

Regional Indicative Strategic Development Plan 2015-2020, 1–20.

Statistical Services Centre of the University of Reading. (2009). International Household Survey Network Survey Quality Assessment Framework SQAF, (May).

Statistics Act. (1999). Statistics Act No. 6 of 1999.

Statistics Canada. (1998). Statistics Canada Quality Guidelines. 3rd edition, *12-539-XIE*, 101.

Statistics South Africa. (2008). *South African Statistical Quality Assessment Framework (SASQAF) - First edition* (1st ed.).

Statistics South Africa. (2010a). Concepts and definitions for Statistics South Africa. Retrieved from http://www.statssa.gov.za/?page_id=371

Statistics South Africa. (2010b). *South African Statistics Quality Assessment Framework (SASQAF)*. Retrieved from http://www.statssa.gov.za/standardisation/SASQAF_Edition_2.pdf

Statistics South Africa. (2011). National Statistics Division - NSSD. Pretoria.

Statistics South Africa. (2012). *Census 2011 - Statistical Release - P0301.4*. Pretoria. Retrieved from http://www.statssa.gov.za/Publications/P03014/P030142011.pdf

Statistics South Africa. (2014). Integrated Communication and Marketing Strategy.

Statistics South Africa. (2015). Head Office CAPI Team. Pretoria.

Statistics South Africa. (2016a). *Community Survey 2016: Technical Report*. Retrieved from chrome-extension://oemmndcbldboiebfnladdacbdfmadadm/http://cs2016.statssa.gov.za/wp-content/uploads/2016/06/CS-2016-Technical-report_Web.pdf

Statistics South Africa. (2016b). CS 2016: The biggest CAPI survey to date. *Pulse Magazine*.

Statistics South Africa. (2016c). Improvement Areas for Citizen Satisfaction Survey (CSS)_V1.

Statistics South Africa. (2016d). KwaZulu-Natal citizen satisfaction survey: Analytical report, 2015. Retrieved January 25, 2018, from http://www.statssa.gov.za/?page_id=1854&PPN=Report-03-00-07

Statistics South Africa. (2016e). KZN Citizen Satisfaction Survey. Retrieved from http://www.kznonline.gov.za/images/Downloads/Publications/Report-03-00-072015.pdf

Statistics South Africa. (2016f). *Stats SA Annual Report 2015/2016*. Retrieved from http://cs2016.statssa.gov.za

Statistics South Africa. (2016g). Stats SA ready for Community Survey 2016 | Statistics South Africa. Retrieved July 31, 2017, from http://www.statssa.gov.za/?p=6224

Statistics South Africa. (2017a). An integrated digital transition for household surveys: A Synthesised Paper.

Statistics South Africa. (2017b). Proposal for an Integrated Quality Management System (QMS) at Statistics South Africa.

Statistics South Africa. (2017c). *Work Programme 2017/18*.

Surbhi S. (2016). Difference Between Primary and Secondary Data (With Comparison Chart) - Key Differences. Retrieved January 13, 2017, from https://keydifferences.com/difference-between-primary-and-secondary-data.html

Teganga, S. P., Emukule, G., Wambugu, S., Kabore, I., & Mwarogo, P. (2009). A comparison of paper-based questionnaires with PDA for behavioral surveys in Africa : Findings from a behavioral monitoring survey in Kenya. *Journal of Health Informatics in Developing Countries*, *3*(1), 22–25. https://doi.org/10.1186/s12911-014-0107-5

The South African Presidency. (2009). Improving Government Performance: Our Approach.

Thorne, S., Kirkham, S.R. and O'Flynn-Magee, K. (2004). The Analytic Challenge in Interpretive Description. *The Analytic Challenge in Interpretive Description*, *3*, 1–11.

Tokosi, T. O. (2016). *ELECTRONIC PATIENT RECORD (EPR) SYSTEM IN SOUTH AFRICA : INFORMATION , STORAGE , RETRIEVAL AND SHARE AMONGST CLINICIANS*.

Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., … Jackson, D. (2009). The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC Medical Informatics and Decision Making*, *9*, 51. https://doi.org/10.1186/1472-6947-9-51

Toscano, C. M., Vijayaraghavan, M., Salazar-Bolaños, H. M., Bolaños-Acuña, H. M., Ruiz-González, A. I., Barrantes-Solis, T., … Hyde, T. B. (2013). Cost analysis of an integrated vaccine-preventable disease surveillance system in Costa Rica. *Vaccine*, *31*, C88–C93. https://doi.org/10.1016/J.VACCINE.2013.05.036

Uematsu, H., & Mishra, A. K. (2010). Can Education Be a Barrier to Technology Adoption ? *Agricultural & Applied Economics Association 2010 AAEA*, 38. Retrieved from http://ageconsearch.umn.edu/bitstream/61630/2/11158 AAEA Education Paper.pdf

UN Millennium Project | About the MDGs. (2006). Retrieved July 28, 2017, from http://www.unmillenniumproject.org/goals/

UN system task team. (2013). Statistics and indicators for the post-2015 UN development agenda, (July), 70. Retrieved from http://www.un.org/en/development/desa/policy/untaskteam_undf/UNTT_MonitoringReport_WEB.pdf

UNCTAD. (2017). UN Statistics Quality Assurance Framework, (September 2016), 1–48.

UNDP. (2015). Sustainable Development Goals |UNITED NATIONS DEVELOPMENT PROGRAMME. Retrieved July 28, 2017, from http://www.undp.org/content/undp/en/home/sustainable-development-goals.html

United Nations. (2015). The Millennium Development Goals Report. *United Nations*, 72. https://doi.org/978-92-1-101320-7

United Nations. (2017). *Principles and recommendations for population and housing censuses - Revision 2*. *Statistical Papers* (Vol. 3).

United Nations Department of Economic and Social Affairs. (2005). *Household Sample Surveys in Developing and Transition Countries*. *Studies in Methods* (Vol. F). Retrieved from http://unstats.un.org/unsd/hhsurveys/

United Nations Economic Comission for Europe. (2012). Making Data Meaningful : A guide to improving statistical literacy. *United Nations Economic Comission for Europe*, *Part 4*, 28. https://doi.org/10.1111/j.1600-0897.1998.tb00338.x

United Nations Statistics Division. (2007). Fundamental Principles of Official Statistics.

Üstun, T. B., Chatterji, S., Mechbal, A., & Murray, C. J. L. (2005). Quality assurance in surveys: standards, guidelines and procedures. *Household Sample Surveys in Developing and Transition Countries*, 199–230.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Research essay bridging the qualitative – quantitative divide: Guidelines for conducting mixed methods, *X*(X), 1–34.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178.

Viswanath Venkatesh, Sykes, T. A., & Zhang, X. (2018). ICT FOR DEVELOPMENT IN RURAL INDIA: A LONGITUDINAL STUDY OF WOMEN'S HEALTH OUTCOMES. *MIS Quarterly*, 1–56.

Walliman, N. (2001). Your Research Project: A Step by Step Guide for the First Time Researcher. Sage Publications.

We Are Social. (2016). Digital in 2016. *Http://Wearesocial.Com/*, 1–29. Retrieved from http://wearesocial.com/uk/special-reports/digital-in-2016

Wedawatta, G., Ingirige, B. and Amaratunga, D. (2011). *Case Study as a Research Strategy: Investigating Extreme Weather Resilience of Construction SMEs in the UK*. *7th Annual International Conference of International Institute for Infrastructure*.

Weinstein, J., & Vasovski, S. (2004). *The PDCA Continous Improvement Cycle*.

Wheeldon, & Ahlberg. (2012). Mapping Mixed-Methods Research. *Visualizing Social Science Research Maps, Methods, & Meaning*, 113–148. https://doi.org/http://dx.doi.org/10.4135/9781483384528

Wisdom, J., & Creswell, J. W. (2013). Mixed Methods: Integrating Quantitative and Qualitative Data Collection and Analysis While Studying Patient-Centered Medical Home Models | PCMH Resource Center. https://doi.org/13-0028-EF

Worldbank Group. (2015). Survey Solutions Questionnaire Designer. Retrieved January 24, 2015, from https://solutions.worldbank.org/account/login?ReturnUrl=%2F

Worldbank Group. (2017). THE 2017 JOINT REVIEW OF THE STANDARDS AND CODES INITIATIVE. *Policy Papers*, *17*(49). https://doi.org/10.5089/9781498347105.007

Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A Comparison of Computer-Assisted and Paper-and-Pencil Self-Administered Questionnaires in a Survey on Smoking, Alcohol, and Drug Use. *Public Opinion Quarterly*, *62*(3), 331–353. Retrieved from http://dx.doi.org/10.1086/297849

Wright, N., & Losekoot, E. (2012). Interpretative Research Paradigms: Points of Difference. In *Proceedings of the 11th European Conference on Research Methods in Business Management: ECRM 2012* (pp. 416–422).

Wynn, E. (2001). *Möbius Transitions in the Dilemma of Legitimacy*. *in Trauth, E. (Ed.)*

Qualitative Research in IS: issues and trends. Hershey: Idea Group Publishing.

Yin, R. (2003). *Case Study Research: Design and Methods* (3rd Editio). Sage Publications.

Yin, R. (2009). *Case Study Research: Design and Methods* (4th Editio). Sage Publications.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), 321–332. https://doi.org/10.1177/1356389013497081

Yu, P., de Courten, M., Pan, E., Galea, G., & Pryor, J. (2009). The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. *International Journal of Medical Informatics*, *78*(8), 532–542. https://doi.org/10.1016/j.ijmedinf.2009.03.002

Zachariadis, M., Scott, S. and Barrett, M. (2013). For Mixed methods Research. Management Information Systems Quarterly. *Management Information Systems Quarterly*, *37*, 855–879.

Zachariadis, M., Scott, S., & Barrett, M. (2013). Bridging The Qualitative–Quantitative Divide: Guidelines For Conducting Mixed Methods Research In Information Systems. *MIS Quarterly*, *37*(3), 855–879. https://doi.org/10.25300/MISQ/2013/37.3.09

Zakaria, O. (2004). Understanding Challenges of Information Security Culture: A Methodological Issue. *Proceedings of the 2nd Australian Information Security Management Conference*, 83–93.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S., & Hitzler, P. (2013). Quality assessment methodologies for linked open data. *Submitted to Semantic Web Journal*, *1*, 1–5. https://doi.org/10.3233/SW-150175

Zeng, W., Gheorghe, A., & Nair, D. (2016). A Discussion Paper of Health System Level Approaches to Addressing Quality of Care in Low- and Middle- Income Countries. *Rbfhealth.Org*, 1–39. Retrieved from https://www.rbfhealth.org/sites/rbf/files/Approaches for Adressing QoC.pdf

Zimmermann, R. (2008). Digital Data Collection Demonstration White PaperA Comparison of Two Methodologies: Digital and Paper-based, 0–8.

APPENDICES

APPENDIX A: UWC ETHICAL CLEARANCE

UNIVERSITY *of the* WESTERN CAPE

DEPARTMENT OF RESEARCH DEVELOPMENT

28 April 2016

To Whom It May Concern

I hereby certify that the Senate Research Committee of the University of the Western Cape approved the methodology and ethics of the following research project by:
Mr MM Hattas (Information Systems)

Research Project:	A quality assurance framework for digital household survey processes in South Africa.
Registration no:	15/6/68

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

Private Bag X17, Bellville 7535, South Africa
T: +27 21 959 2988/2948 . F: +27 21 959 3170
E: pjosias@uwc.ac.za
www.uwc.ac.za

A place of quality,
a place to grow, from hope
to action through knowledge

APPENDIX B: PHD RESEARCH SURVEY QUESTIONNAIRE

UNIVERSITY of the
WESTERN CAPE

PHD RESEARCH SURVEY

The purpose of this survey is to examine the use of digital devices for official household survey data collection by Statistics South Africa

Your responses will help develop an understanding of the potential benefits of digital devices used for official household-survey collection. Kindly complete all questions, which should require about 5-10 minutes of your time. All answers will remain confidential and anonymous.

A. BACKGROUND INFORMATION

Respondent particular information (Please mark an "X" where applicable)

1. Gender:

Male		Female	

2. How old are you? (In completed years – In whole numbers)

Age		

3. What is the highest level of education that you have successfully completed?

	01 = Grade 11/Standard 9 and lower
	02 = Grade 12/Standard 10/Form 5/Matric

	03 = Higher Diploma (Technikon/University of Technology)
	04 = Post Higher Diploma (Technikon/University of Technology Masters, Doctoral)
	05 = Bachelors Degree
	06 = Honours Degree
	07 = Higher degree (Masters, Doctorate)
	08 = Other

4. How many years of working experience do you have at Statistics South Africa in the related areas responsible for the collection of household surveys?

1-5	6-10	11-15	16-20	> 21

5. Are you currently working….

Full-time	Part-time

6. Do you own a cellular phone?

Yes	No

B. EXPECTATIONS OF DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT) AND EXPERIENCE THEREOF

7. How would you describe yourself in respect to the various levels of digital handheld-device technology (DHDT) usage? Mark an "X" next to the level that best describes you.

01		Unfamiliar	No experience with DHDT.
02		Newcomer	Attempted to use DHDT, but I still require help on a regular basis.

03		Beginner	Able to perform basic functions in a limited number of DHDT.
04		Average	Demonstrate a general competency in a number of DHDT.
05		Advanced	Acquired the ability to competently use a broad spectrum of DHDT.
06		Expert	Extremely proficient in using a wide variety of DHDT.

8. Please indicate your beliefs about the use of DHDT as expressed in each of the following statements by selecting **one** level of agreement or disagreement.

Where:

SD = Strongly Disagree

D = Agree

U = Undecided

A = Agree

SA = Strongly Agree

SD	D	U	A	SA	
					1. I sometimes feel that I have been left behind when it comes to using DHDT
					2. Using DHDT in the field would result in an improvement in the quality and accuracy of household survey data collection
					3. DHDT will complicate the process of household survey data collection
					4. Using DHDT in the field is difficult since fieldworkers will be more comfortable using PAPDC as opposed to DHDT
					5. DHDT could reduce the number of staff employed in the future, especially data capturers.
					6. I am excited about using DHDT in household survey collection.

C. EQUIPMENT ACCESS

9. Do you currently own a smartphone, tablet or mobile computing device?

Yes	No

10. Do you currently use a smartphone, tablet or mobile computing device for work-related purposes, other than calling, sms'ing, surfing the internet or for social media-related activities?

Yes	No

D. DATA QUALITY

11. In terms of survey data collection quality, I believe DHDT can…

SD	D	U	A	SA	
					1. improve the overall quality of data collected
					2. result in poor data quality during each item response
					3. initiate an increased response interaction from respondents
					4. result in an increase in trust in the data from the users

E. BEST PRACTICES IN OFFICIAL HOUSEHOLD SURVEY DATA COLLECTION

12. Regarding best practices in official survey data collection, I believe DHDT …

SD	D	U	A	SA	
					1. has Best practice methodologies defined in South Africa or in sub-Saharan Africa
					2. methodology and standards exist, but are not well communicated internally to the South African NSO or externally
					3. requires a common framework in order to progress and promote the use of DDC amongst NSO's
					4. can be adopted easily since there currently exists a quality framework (SASQAF) for PAPDC
					5. should have been in existence a while back already given that mobile devices have been increasingly growing w.r.t power and speed

F. PILOT EXPERIENCES OF DDC

13. During your participation in DDC pilots thus far …

SD	D	U	A	SA	
					1. I have not been involved in any DDC pilots
					2. I found DDC very effective in collecting household survey information
					3. I found that there are significant advantages to DDC vs. PAPDC
					4. there was an increase in response rates

					5. DHDT is user-friendly for both the interviewer and the interviewee

THANK YOU

Your efforts and time in completing this questionnaire are most appreciated. Should you require a summary of the results of this research, please email me at mhattas@gmail.com

APPENDIX C: FOCUS GROUP DISCUSSION GUIDE

UNIVERSITY *of the* WESTERN CAPE

Department of Economic and Management Services

FOCUS-GROUP DISCUSSION GUIDE

Consent Process

Consent forms and information sheets for focus-group participants are completed in advance by all those seeking to participate. Below is a summary of the information in the consent form that focus group organisers and facilitators should use to make sure participants understand the information in the consent form.

Thank you for agreeing to participate. We are very interested to hear your valuable opinion on how digital handheld devices can be used effectively by Statistics South Africa for the collection of official household-based surveys.

- *The purpose of this study is to learn how employees of Statistics South Africa view the potential benefits of moving from a paper-based official household-based survey collection towards a digital handheld collection process.*
- *The information you provide us is completely confidential, and we will not associate your name with anything you say in the focus group.*
- *The researcher would like to tape the focus groups so that he can make sure to capture the thoughts, opinions, and ideas he hears from the group. No names will be attached to the focus groups and the tapes will be destroyed as soon as they are transcribed.*
- *You may refuse to answer any question or withdraw from the study at any time.*
- *The researcher understands how important it is that this information be kept private and confidential; and he will ask the participants to respect each other's confidentiality.*
- *If you have any questions now or after you have completed the questionnaire, you can always contact the researcher, or his supervisor, whose names and phone numbers can be located in this form.*
- *Please sign to show you agree to participate in this focus group.*

Introduction:

1. Welcome
   Introduce yourself and the note-taker, and send the Sign-In Sheet with a few quick demographic questions (age, gender, etc.) around to the group while you are introducing the focus group.

   Review the following:

   - Who is the researcher; and what is he trying to do?
   - What will be done with this information?
   - Why did the researcher request your participation?

2. Explanation of the process
   Ask the group if anyone has participated in a focus group before. Explain that focus groups are being used more and more often in social and information research.

   About focus groups

   - The researcher learns from you (positive and negative)
   - Not trying to achieve consensus, we're gathering information
   - No virtue in long lists: looking for priorities
   - In this project, the researcher is using semi-structured interviews, survey questionnaires and focus group discussions. The reason for using these tools is to gather more in-depth information from a smaller group of people in focus groups. This allows the researcher to understand the context behind the answers given in the written survey and helps explore topics in more detail than opposed to using just one method of data collection.

   Logistics

   - Focus group will last about 20 – 35 minutes
   - Feel free to move around
   - Where is the bathroom? Exit?

3. Ground Rules
   Ask the group to suggest some ground rules. After they brainstorm some, make sure the following are on the list.

   - Everyone should participate.
   - Information provided in the focus group will be kept confidential
   - Stay with the group and please don't have side conversations
   - Turn off cell phones if possible
   - Have fun

4. **Switch on the Tape Recorder**

5. Ask the group if there are any questions before we get started, and address those questions.

6. Introductions
   - Go around the table: determine job level, research interests, etc.

The discussion begins; people are allowed time to think before answering the questions; and don't move too quickly. Use the probes to make sure that all issues are addressed, but move on when you feel you are starting to hear repetitive information.

Questions:

1. Let's start the discussion by talking about what makes Statistics South Africa a good place to work. What are some of the positive aspects of working at Statistics South Africa?

2. Other questions to follow…

Probes for Discussion:

- *Current experiences in household survey collection*
- *Perceptions of digital collection*
- *Expectations of digital collection*
- *Access to equipment*
- *Perceptions of the quality of data*
- *Opinion on best practices in official household survey collection*
- *Management and supervision*
- *Is there a sense of ownership of the outcomes here?*
    - *Work content, responsibility*

That concludes our focus group. Thank you so much for coming and sharing your thoughts and opinions with me. There is a short evaluation form that the researcher would like you to fill out; if you have the necessary time. If you have additional information that you did not get to say in the focus group, please feel free to write it on this evaluation form.

Materials and supplies for focus groups

- Sign-in sheet
- Consent forms (one copy for participants, one copy for the team)
- Evaluation sheets, one for each participant
- Name tents
- Pads & Pencils for each participant
- 1 recording device
- Batteries for recording device
- Permanent marker
- Notebook for note-taking

APPENDIX D: INFORMATION SHEET FOR PARTICIPANTS IN FOCUS GROUPS

UNIVERSITY *of the* WESTERN CAPE

Department of Economic and Management Services

INFORMATION SHEET FOR PARTICIPANTS IN FOCUS GROUPS

Dear participant

Mogamat Mahier Hattas (Student no: 2103745) – PhD Information Systems

The title of my thesis is:

A quality assurance framework for digital household survey processes in South Africa

Please take time to read through this information sheet carefully, in order for you to be knowledgeable about what is required of you as a research participant in this study.

As a participant who gave consent for your participation in this study, you will be required to:

1. ***Attend a short presentation on the digital collection for official household survey concepts (taking approximately 20 minutes).***
2. ***Ask clarifying questions about digital collection for official household surveys.***
3. ***Provide your comments, recommendations and rating of digital collection for official household surveys, in order to improve the process. This will be facilitated through the completion of a short focus group "digital collection for official household survey evaluation sheet".***

Your participation in this study is voluntary and no remuneration will be provided in return for your contribution. You remain free not to participate and have the right to withdraw from the study at any time without the need to provide any reason for such withdrawal. Your participation in the process might result in research, which may be published; but your identity will never be revealed. The researcher will ensure your anonymity throughout the research process.

If you have any questions concerning this research, feel free to contact me, or my supervisor:

Mahier Hattas | Cell phone: 0795013249 | (mhattas@gmail.com)

Dr Johan Breytenbach | Cell phone: 0837088444 | (breytenbachj@gmail.com)

I hereby consent voluntarily to participate in this study. I have been given a copy of this form.

_____________________________ _____________________________

Name of participant **Date**

Email:____

Thank you for participating in my study.

APPENDIX E: INFORMATION SHEET FOR PARTICIPANTS IN RESEARCH QUESTIONNAIRE

UNIVERSITY *of the* WESTERN CAPE

Department of Economic and Management Services

INFORMATION SHEET FOR PARTICIPANTS IN RESEARCH QUESTIONNAIRE/SURVEY

Dear participant

Mogamat Mahier Hattas (Student no: 2103745) – PhD Information Systems

The title of my thesis is:

A quality assurance framework for digital household survey processes in South Africa

Please take time to read this information sheet carefully, in order for you to be informed about what is required of you as a research participant in this study.

As a participant who gave consent for your participation in this study, you will be required to:

1. ***Participate in a short survey (taking approximately 20 minutes). The survey questions are about your experience in the field of official statistics and the potential use of digital handheld devices for household survey data collection.***
2. ***If selected to do so, engage in a conversation with the researcher to verify the validity of the survey results for your organisation. Note only key stakeholders in the organisation will be asked to discuss the survey results.***

Your participation in this study is voluntary and no remuneration will be provided in return for your contribution. You remain free not to participate and have the right to withdraw from the study at any time without the need to provide any reason for such withdrawal. Your participation in the survey process might result in research which may be published, but your identity will never be revealed. The researcher will ensure your anonymity throughout the research process.

If you have any questions concerning this research, feel free to contact me, or my supervisor:

Mahier Hattas | Cell phone: 0795013249 | (mhattas@gmail.com)

Dr Johan Breytenbach | Cell phone: 0837088444 | (breytenbachj@gmail.com)

I hereby consent voluntarily to participate in this study. I have been given a copy of this form.

____________________________ ____________________________

Name of participant **Date**

Email:______________________________________

Thank you for participating in my study.

APPENDIX F: INFORMATION SHEET FOR PARTICIPANTS IN SEMI-STRUCTURED INTERVIEWS

RESPICE PROSPICE

UNIVERSITY *of the* WESTERN CAPE

Department of Economic and Management Services

INFORMATION SHEET FOR PARTICIPANTS IN SEMI-STRUCTURED INTERVIEWS

Dear participant

Mogamat Mahier Hattas (Student no: 2103745) – PhD Information Systems

The title of my thesis is:

A quality assurance framework for digital household survey processes in South Africa

Please take the necessary time to read this information sheet carefully, in order for you to be knowledgeable about what is required of you as a research participant in this study.

As a participant who gave consent for your participation in this study, you will be required to:

1. ***Participate in a short semi-structured interview (taking approximately 15 minutes). The questions are about your experience in the field of official statistics and the potential use of digital handheld devices for household survey data collection.***
2. ***If selected to do so, engage in a conversation with the researcher to verify the validity of the survey results for your organisation.***

Your participation in this study is voluntary and no remuneration will be provided in return for your contribution. You remain free not to participate and have the right to withdraw from the study at any time without the need to provide any reason for such withdrawal. Your participation in the process might result in research which may be published, but your identity will never be revealed. The researcher will ensure your anonymity throughout the research process.

If you have any questions concerning this research, feel free to contact me, or my supervisor:

Mahier Hattas | Cell phone: 0795013249 | (mhattas@gmail.com)

Dr Johan Breytenbach | Cell phone: 0837088444 | (breytenbachj@gmail.com)

I hereby consent voluntarily to participate in this study. I have been given a copy of this form.

____________________________ ____________________________

Name of participant **Date**

Email:______________________________________

Thank you for participating in my study.

APPENDIX G: LETTER OF CONSENT FORM

UNIVERSITY *of the* WESTERN CAPE

Department of Economic and Management Services

Letter of Consent

CONSENT FORM FOR FOCUS GROUP PARTICIPANT OR QUESTIONNAIRE OR SEMI- STRUCTURED INTERVIEW

I, ……………………………………………………………………., have had the opportunity to ask questions related to this study and obtained satisfactory answers to my questions.

I have also received any additional information that I may have requested about this research.

I agree to participate in this research.

I understand that my participation in this study is voluntary and that no remuneration will be provided in return for my contribution. I am free not to participate and have the right to withdraw from the study at any time without the need to provide any reason for such withdrawal.

I am aware that the focus group outcomes might result in research which may be published, but that my identity will never be revealed. It is my understanding that the researcher will ensure my anonymity throughout the research process.

I retain the right of refusal to answer any question with which I do not feel comfortable, or able to answer.

Date: ………………………………..………………………

Participant Name: …………………………………………………………

Participant Signature: …………………………………………………………

Interviewer name: **Mogamat Mahier Hattas…..……………**

Interviewer Signature: …………………………………………………………

If you have any questions concerning this research, feel free to contact me: Mogamat Mahier Hattas, Cell phone: 0795103249 or my Supervisor Dr Johan Breytenbach, Cell phone: 0837088444

APPENDIX H: MAPPING THE STATISTICAL VALUE CHAIN TO SASQAF (Statistics South Africa, 2010b)

Activities of the statistical value chain		SASQAF Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
Need	Determine need for information	Prerequisites of quality	1.1 The responsibility for producing statistics is clearly specified.
Need	Determine need for information	Prerequisites of quality	1.2 Standards and policies are in place to promote the consistency of methods and results.
Need	Establish output objectives	Prerequisites of quality	1.3 Data sharing and coordination among data-producing agencies is clearly specified.
Need	Establish output objectives	Prerequisites of quality	1.4 Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.
Need	Determine need for information	Relevance	2.1 Have both internal and external users of the data been identified?
Need	Determine need for information	Relevance	2.2 Is there a process to identify user needs?
Need	Establish output objectives	Relevance	2.3 Changes are made as a result of user needs assessments.
Need	Check data availability	Relevance	2.4 To what extent is the primary data appropriate for the statistical product produced?
Need	Prepare business case	Timeliness	4.4 Periodicity of release.
Need	Check data availability	Accessibility	5.1 Legal arrangements are in place to allow access to administrative records via manual, automated or electronic processes.
Need	Frame and sample methodology, Data collection	Methodological soundness	8.1 The scope of the study is consistent with accepted standards, guidelines or good practices
Need	Check data availability	Integrity	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.
Design	Detailed project plan	Prerequisites of quality	1.5 Resources are commensurate with the needs of statistical programmes. • Staff • Facilities • Computing resources • Financing
Design	Frame and sample methodology	Accuracy	3.5 Register/frame maintenance procedures are adequate. • updates. • quality assurance • data audit.
Design	Data collection	Accuracy	3.6 Are data collection systems sufficiently open and flexible to cater for

APPENDIX I: SASQAF LITE FOR PAPDC SURVEYS (Statistics South Africa, 2011)

SASQAF LITE

QUALITY DIMENSION (QD)	Key Survey Performance Area (KSPA)	QUALITY INDICATOR (QI)	QUALITY INDICATOR STANDARD (QIS)	Assessment Level 4 (Quality Statistics)	Assessment Level 3 (Acceptable Statistics)	Assessment Level 2 (Questionable Statistics)	Assessment Level 1(Poor Statistics)
Pre-requisites of quality	**Responsibility**	**QI1.1** The responsibility for producing statistics is clearly specified	**QIS1.1.1** A law or legal arrangement exists that explicitly mandates the production of statistics	A law or legal arrangement exists that explicitly provides the mandate for the production of statistics	A law or legal arrangement exists that implies that statistical production is part of its mandate.	No law or legal arrangement exists, but an informal agreement exists for statistical production.	No law or legal arrangement exists.
	Data Sharing	**QI1.3** Data sharing and coordination among data-producing agencies are clearly specified and adhered to	**QIS1.3.1** A legal arrangement must exist which allows for the timely and efficient sharing of data between the collection agency and the secondary user.	A legal arrangement exists and is adhered to.	A legal arrangement exists but is not adhered to.	An informal arrangement exists.	No arrangement exists.
	Confidentiality	**QI1.4** Measures are in place to ensure that individual data are kept confidential and used for statistical purposes only	**QIS1.4.1** There must be a law or policy that ensures information collected is kept confidential and used for statistical purposes only.	A law/policy guaranteeing the confidentiality of respondents' information, that it is used for statistical purposes only, exists and is adhered to at all times.	A law/policy guaranteeing the confidentiality of respondents' information, that it is used for statistical purposes only exists but is not adhered to at all times.	Although no law/policy exists, efforts are made to keep respondent information confidential and ensure they are used for statistical purposes only	No law/policy exists

	Resources	**QI1.6** Resources are commensurate with the needs of statistical programmes (***Staff**, Facilities, Computing Resources, Financing)	**QIS1.6.3** Facilities must have the infrastructure to manage the needs of statistical programmes	Facilities are well equipped and have the necessary infrastructure for the production of statistics.	Not Applicable	Not Applicable	Facilities are ill-equipped and do not have the necessary infrastructure for the production of statistics.
			QIS1.6.4 Computer hardware resources must be adequate in terms of data storage, data backup media, power supply (uninterrupted), memory, and other necessary hardware equipment, etc.	The hardware computing resources are sufficient in terms of data storage, data backup media, power supply, memory and computers.	The hardware computing resources are sufficient in terms of data backup media, power supply, memory and computers.	The hardware computing resources are sufficient in terms of power supply, memory and computers.	The hardware resources are entirely insufficient.
			QIS1.6.6 Computer software resources must be adequate in terms of capturing systems, editing systems, coding systems, statistical software, up-to-date licenses, virus protection, and appropriate access rights	Computer software resources are adequate in terms of statistical and other specialised software, up-to-date licenses, virus protection, and appropriate access rights	Computer software resources are adequate in terms of appropriate access rights, virus protection, statistical and other specialised software.	Computer software resources are adequate in terms of appropriate access rights, statistical and other specialised software.	Computer software resources are entirely inadequate
			QIS1.6.7 Budgets must be adequate	Budgets are sufficient for the needs of statistical programmes	Budgets are sufficient for the needs of statistical programmes but not made available timeously.	Not Applicable	Budgets are entirely inadequate

	Quality Management	**QI1.8** Processes are in place to focus on, monitor and check quality	**QIS1.8.1** The agency must have a quality management system in place.	The agency has a quality management system in place.	Not Applicable	Not Applicable	The agency has no quality management system in place.
Relevance	**User Database**	**QI2.1** Have both the internal and external users of the data been identified?	**QIS2.1.1** An up-to-date user database must exist.	An up-to-date user database exists	A user database exists but is not up-to-date.	Users are known but not recorded in a database.	Users have not been identified.
	User Satisfaction	**QI2.5** Is there a process to determine the satisfaction of users with the statistical information?	**QIS2.5.1** A formal process must exist to determine the satisfaction of users with statistical information.	Formal processes exist to determine the satisfaction of users with statistical information. The results are incorporated into the statistical production process.	Formal processes exist to determine the satisfaction of users with statistical information. The results are not incorporated into the statistical production process.	Processes exist to determine the satisfaction of users with statistical information, but they are not formal.	There is no process at all.
Accuracy	**Sampling Errors**	**Metric [QIS 3.1.2]** $SE = \sqrt{Var(\hat{\theta})}$ $CV = \frac{\sqrt{Var(\hat{\theta})}}{E(\hat{\theta})}$ CI : $MSE(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$ Where $\hat{\theta}$ is an estimator of a parameter of interest					

		QI3.1 Measures of sampling errors for key variables are calculated. Amongst others, these are: • standard error • the coefficient of variation (CV) • confidence interval (CI) • mean square error (MSE) • design effect (DEFF)	**QIS3.1.2** Measures of sampling errors must fall within acceptable standards. At a minimum, the following must be calculated: standard error, the coefficient of variation, confidence interval, mean square error. The low accuracy of variables (if these exist) are explained. (See **Metrics** above)	CV< or = 5% and $B^2(\hat{\theta}) > 0$	CV< or = 5% and $B^2(\hat{\theta}) > 0$	5% < CV< or = 30%	CV > 30%
	Non-Sampling Errors	**Metric [QIS 3.2.1]** $a = \frac{\Sigma \lvert \text{ final weights - design weights } \rvert}{\Sigma \text{ design weights}}.$ $b = \max\left[\frac{\lvert \text{final weights - design weights} \rvert}{\text{design weights}}\right].$ where a = accepted standard x,y = agreed thresholds					
		QI3.2 Measures of non-sampling errors are calculated, viz.: • Frame	**QIS3.2.1.** The extent of measures of non-sampling errors must be kept to an acceptable level. (See **Metrics** above)	a. < 15% and b. < 15%.	15% a. < 25%; or 15% b. < 25%.	a.$\geq$ 25% or b.$\geq$ 25%.	a and b are not calculated.

		coverage errors (e.g. duplication in the frame/register used to conduct a survey, number of statistical units out of scope (i.e. the number of ineligible units)) • Misclassification errors • Systematic errors to determine the extent of bias introduced for both administrative records and surveys • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	**QIS3.2.7** The proportion of units which are misclassified must be at an acceptable level.	x<=a	a+y>=x>a	a+z>=x > a+y	x>a+z
			QIS3.2.8 Systematic errors must be identified and reported.	Systematic errors are both identified and reported.	Systematic errors are identified, but not reported.	Not Applicable	Systematic errors are not identified.
			QIS3.2.22 The model assumptions must be stated. All models used in the estimation of statistics must be described.	The model assumptions are stated. All models used in the estimation of statistics are described. The model assumptions are valid, and evidence is provided for this.	The model assumptions are stated. Not all models used in the estimation of statistics are described. The model assumptions are valid, and evidence is provided for this.	The model assumptions are not stated. All models used in the estimation of statistics are described. The model assumptions are not valid, or no evidence is provided for validity	The model assumptions are not stated. All models used in the estimation of statistics are not described. The model assumptions are not valid, or no evidence is provided for this.
			QIS3.2.23 Item non-response rate must be within acceptable levels.	x<=a	a+y>=x>a	a+z>=x > a+y	x>a+z
			QIS3.2.24 Unit non-response rate must be within acceptable levels. where a= accepted standard x,y = agreed thresholds	x<=a	a+y>=x>a	a+z>=x > a+y	x>a+z
	Flexibility	**QI3.6** Are data collection systems sufficiently open and flexible to	**QIS3.6.1** The system must be flexible in all its components. It must be designed to allow for new developments (e.g.	The system is flexible in all its components. It is designed to allow for new developments. There is no need	Not Applicable	Not Applicable	The system is inflexible

		cater for new developments?	changes in definitions, classifications, etc.)	for an ad hoc collection system.			
Timeliness	**First Release Lag**	**QI4.1** Average time between the end of reference period and date of the preliminary results.	**QIS4.1.1** The preliminary results must be released according to the prescribed standard. Metrics (above)	0.9<a<=1 Preliminary results are released in accordance with prescribed timeframes.	1<a ><=1.25 times the prescribed lapse.	a<0.9 or a >=1.25 times the prescribed lapse.	No preliminary results are released.
	Planned Timelines	**Metric [QIS4.3.3]** $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}} + 1.$ a > 1 implies that collections the phase exceeded planned duration. a < 1 implies that collection phase completed within the planned duration.		**Metrics [QIS4.3.4]** a Planned delivery date – actual delivery date $b = \text{Int}\left(\frac{a}{\text{reference period in days}}\right).$		**Metrics [QIS4.3.5],[QIS 4.3.6], [QIS4.3.7]** $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}} + 1.$ a > 1 implies that the processing/analysis/dissemination phase exceeded the planned duration a < 1 implies that the processing/analysis/dissemination phase was completed within the planned duration.	
		QI4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection • data processing • data analysis • dissemination	**QIS4.3.1** Project plan/schedule of key deadlines related to the statistical value chain must be compiled.	A schedule of key deadlines does exist.	Not Applicable	Not Applicable	A schedule of key deadlines does not exist.
			QIS4.3.3 Data collection must follow the project plan/schedule.	a ≤ 1 times the planned duration for data collection.	1< a. <1.25 times the planned duration for data collection.	1.25 ≤ a. <1.5 times the planned duration for data collection.	a ≥ 1.5 times the planned duration for data collection.
			QIS4.3.4 A protocol for the timely delivery of administrative data must	A protocol for the timely delivery of administrative	A protocol for the timely delivery of administrative	Not Applicable	A protocol for the timely delivery of administrative

			exist and must be adhered to.	data exists and is adhered to	data exists and is not adhered to		data does not exist.
			QIS4.3.5 Data processing must follow the project plan. **Metrics** -->	a. ≤ 1 times the planned duration for data processing.	1< a. <1.25 times the planned duration for data processing.	1.25 ≤ a. <1.5 times the planned duration for data processing.	a. ≥1.5 times the planned duration for data processing.
			QIS4.3.6 Data analysis must follow the project plan/schedule.	a. ≤ 1 times the planned duration for data analysis.	1< a. <1.25 times the planned duration for data analysis.	1.25 ≤ a. <1.5 times the planned duration for data analysis.	a. ≥1.5 times the planned duration for data analysis.
			QIS4.3.7 Dissemination must follow the project plan/schedule.	a. ≤ 1 times the planned duration for dissemination	1< a. <1.25 times the planned duration for dissemination	1.25 ≤ a. <1.5 times the planned duration for dissemination	a. ≥1.5 times the planned duration for dissemination
Accessibility	**Accessibility to public**	**QI5.1**Are statistical products (e.g. data, metadata) available to the public?	**QIS5.1.1** The statistical products must be disseminated to the public.	The statistical products are disseminated to the public.	Not Applicable	The statistical products are disseminated to selected users.	The statistical products are not disseminated.
	Pre-announced release schedule	**QI5.5** Statistics are released on a pre-announced schedule	**QIS5.5.1** Statistics must be released on a pre-announced schedule.	Statistics are released on a pre-announced schedule which is made available to users annually and is adhered to.	A pre-announced schedule exists and is made available to users annually but not adhered to.	A pre-announced schedule exists which is made available to users in an ad hoc manner but is not adhered to.	No pre-announced schedule exists.
	Statistical Products	**QI5.6**Statistical products are made available to all users at the same time.	**QIS5.6.1**Statistical products must be made available to all users at the same time.	Statistical products have an embargo date and time and are made available to all users at the same time.	Statistical products do not have an embargo date and time, but are made available to all	Statistical products have an embargo date and time, but are not made available to all users at the same time	Statistical products do not have an embargo date and time, and are not made available to all

					users at the same time.		users at the same time
	Data dissemination policy	**QI5.9**Does a data dissemination policy exists, and is it maintained and accessible?	**QIS5.9.1** A dissemination policy must be accessible.	A data dissemination policy exists, is freely accessible and is adhered to.	A data dissemination policy exists, is freely accessible but is not adhered to.	A data dissemination policy exists but it is not freely accessible	No dissemination policy is in place.
	Metadata	**QI5.12** Metadata are readily accessible to users.	**QIS5.12.1** Minimum metadata required for interpreting the product must be accessible.	Minimum metadata required for interpreting the product is available and readily accessible to users	Minimum metadata required for interpreting the product is available, but not readily accessible to users.	Incomplete metadata is available	No metadata is available at all.
Interpretability	**Statistical releases**	**QI6.3**Statistical releases contain a summary of the key findings.	**QIS6.3.1**Statistical releases must contain a summary of key findings.	Statistical releases contain a summary of key findings	Not Applicable	Not Applicable	Statistical releases are published without a summary of key findings
Comparability and Coherence	**Consistency**	**QI7.2** Statistics are consistent and reconcilable over a reasonable period of time.	**QIS 7.2.1** Statistics must be consistent over time. **Metrics** a = length of time series	a is greater than or equal to 10 years (a>=10 years.)	a is greater than or equal to 5 years and less than 10 years. (5 years <=a<10years)	a is less than 5 years (a<5)	Not Applicable
			QIS7.2.2 The statistics must follow an expected trend established over time. Any inconsistencies in the key variables must be reconciled.	The statistics do follow an expected trend established over time. Inconsistencies in the key variables are reconciled	The statistics do not follow an expected trend established over time. Inconsistencies in the key variables are not reconciled	The statistics do follow an expected trend established over time. Inconsistencies in the key variables are not reconciled	The statistics do not follow an expected trend established over time. Inconsistencies in the key variables are not reconciled.

	Consistency checks	**QI7.4** Statistics are checked for consistency with those obtained through other data sources.	**QIS7.4.1** Statistics must be checked for consistency with a comparable dataset. Inconsistencies must be reconciled.	The data producer demonstrates that the statistics are consistent. Any inconsistencies in the key variables are reconciled	The data producer does not demonstrate that the statistics are consistent. Any inconsistencies in the key variables are reconciled	The data producer demonstrates that the statistics are consistent. Any inconsistencies in the key variables are not reconciled	The data producer does not demonstrate that the statistics are consistent. Any inconsistencies in the key variables are not reconciled.
Methodological soundness	**Concepts, Definitions, and Classifications**	**QI 8.1** Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed)	**QIS8.1.1**The concepts and definitions must satisfy accepted standards, guidelines or good practice in line with national, international and peer-agreed norms and must be documented. Deviations from the standard must be formally approved, and these are fully documented.	The concepts and definitions satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms and are documented. Deviations from the standard are formally approved and fully documented.	The concepts and definitions satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms and are documented. Deviations from the standards are not approved and fully documented.	The concepts and definitions are documented but do not satisfy accepted standards, guidelines or good practice	No documented concepts and definitions exist
			QIS8.1.2 The classifications must satisfy accepted standards, guidelines or good practice in line with national, international and peer-agreed norms and must be documented. Deviations from the standard must be formally approved, and are these fully documented.	Classifications satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms and are documented. Deviations from the standard are formally approved and fully documented	Classifications satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms and are documented. Deviations from the standards are not approved and fully documented	Classifications are documented but do not satisfy accepted standards, guidelines or good practice	No documented classifications exist

	Questionnaire	**QI8.3** Methodologies used follow accepted standards, guidelines or good practices.	**QIS8.3.1** The designing of the questionnaire must follow accepted standards, sets of guidelines or good practices.	The designing of the questionnaire follows accepted standards, sets of guidelines or good practices.	Not Applicable	Not Applicable	The designing of the questionnaire does not follow accepted standards, sets of guidelines or good practices.
			QIS8.3.6 Editing and imputation methods must follow accepted standards, sets of guidelines or good practices.	Editing and imputation methods follow accepted standards, sets of guidelines or good practices.	Not Applicable	Not Applicable	Editing and imputation methods do not follow accepted standards, sets of guidelines or good practices.
			QIS8.3.8 Revision methods used must follow accepted standards, sets of guidelines or good practices.	Revision methods used follow accepted standards, sets of guidelines or good practices.	Not Applicable	Not Applicable	Revision methods used do not follow accepted standards, sets of guidelines or good practices.
Integrity	**Conditions for policy-makers**	**QI9.2** Describe the conditions under which policy-makers, specifically government, may have access to data before release. Are the conditions published?	**QIS9.2.1** A data dissemination policy detailing the conditions under which policy-makers have access to the data must be available.	Policy-makers have controlled access to the data. The conditions, along with reasons for their access, are published and adhered to.	Policy-makers have controlled access to the data. The conditions, along with reasons for their access, are published and not adhered to	Policy-makers have uncontrolled access to the data. The conditions, along with reasons for their access, are not published.	There are no conditions preventing policymakers getting access to the data prior to the release.
	Methodology changes	**QI9.3** Advance notice is given of major changes in	**QIS9.3.1** Advance notice of at least 6 months must be given of major changes in	Advance notice of at least 6 months is given of major changes in	Advance notice of between 3 months to 6 months is given of major changes in	Advance notice of less than 3 months is given of major changes in	No advance notice is given of major changes in

		methodology and source data.	methodology and source data.	methodology and source data.	methodology and source data.	methodology and source data.	methodology and source data.
	Code of conduct	**QI9.6** Ethical guidelines for staff behaviour are in place and are well known to the staff.	**QIS9.6.1** A professional code of conduct must be in place providing ethical guidelines for staff behaviour.	A professional code of conduct is in place providing ethical guidelines for staff behaviour. The code of conduct is well known and enforced.	A professional code of conduct is in place providing ethical guidelines for staff behaviour. The code of conduct is well known but is not enforced.	A professional code of conduct is in place providing ethical guidelines for staff behaviour. The code of conduct is not well known and not enforced.	A professional code of conduct is not in place.

APPENDIX J: LIST OF PUBLICATIONS AND SUBMISSIONS

IN-HOUSE PRESENTATIONS / SUBMISSIONS :

C1. Hattas M.M, 2017, “Position Paper 2.2: What approach to take?”, in proceedings in the CAPI Transition Workshop, Salvokop, Pretoria Available: Head office

CONFERENCE PAPERS:

C2. Hattas M.M., Breytenbach J., 2017, “Improving the quality of official survey data collection in South Africa through digital data collection”, in proceedings in the IUSSP Conference, 29 Oct – 04 Nov 2017, Cape Town International Convention Center, South Africa

APPENDIX K: DEVICE SPECIFICATIONS [SAMSUNG, LENOVO]

K1. Samsung Galaxy Tab 3 V:

3G/4G connectivity module was required for synchronization from the field.

Network	Technology	GSM / HSPA
Launch	Announced	2015, March
	Status	Available. Released 2015, April
Body	Dimensions	193.4 x 116.4 x 9.7 mm (7.61 x 4.58 x 0.38 in)
	Weight	322 g (11.36 oz)
	SIM	Micro-SIM
Display	Type	TFT capacitive touchscreen, 16M colours
	Size	7.0 inches (~61.3% screen-to-body ratio)
	Resolution	600 x 1024 pixels (~170 ppi pixel density)
	Multitouch	Yes
Platform	OS	Android 4.4.4 (KitKat)

	CPU	Quad-core 1.3 GHz
Memory	Card slot	microSD, up to 32 GB (dedicated slot)
	Internal	8 GB, 1 GB RAM
Camera	Primary	2 MP
	Features	Geo-tagging
	Video	480p@24fps
	Secondary	2 MP
Sound	Alert types	MP3, WAV ringtones
	Loudspeaker	Yes
	3.5mm jack	Yes
Comms	WLAN	Wi-Fi 802.11 b/g/n, Wi-Fi Direct, hotspot
	Bluetooth	4.0, A2DP
	GPS	Yes, with GLONASS
	Radio	No
	USB	microUSB 2.0
Features	Sensors	Accelerometer
	Messaging	SMS (threaded view), MMS, Email, Push Mail, IM
	Browser	HTML5
	Java	No
		- MP4/H.264 player - MP3/WAV/eAAC+/Flac player - Photo/video editor
Battery		Non-removable Li-Ion 3600 mAh battery
	Talk time	Up to 8 h (multimedia) (2G) / Up to 18 h (3G)
	Music play	Up to 71 h
Misc.	Colours	White, Black
	SAR US	0.45 W/kg (head) 1.58 W/kg (body)
	SAR EU	0.66 W/kg (head) 0.66 W/kg (body)

K2. Lenovo A7-30 A3300:

Network	Technology	GSM
	2G bands	GSM 850 / 900 / 1800 / 1900
	GPRS	Yes
	EDGE	Yes
Launch	Announced	2014, April
	Status	Available. Released 2014, May
Body	Dimensions	198 x 119.8 x 10.5 mm (7.80 x 4.72 x 0.41 in)
	Weight	327 g (11.53 oz)
	SIM	Micro-SIM
Display	Type	Capacitive touchscreen, 16M colours
	Size	7.0 inches (~58.1% screen-to-body ratio)
	Resolution	600 x 1024 pixels (~170 ppi pixel density)
	Multi-touch	Yes
Platform	OS	Android 4.2.2 (Jelly Bean)
	Chipset	Mediatek MT8382M
	CPU	Quad-core 1.3 GHz Cortex-A7

	GPU	Mali-400MP2
Memory	Card slot	microSD, up to 64 GB (dedicated slot)
	Internal	8 GB, 1 GB RAM
Camera	Primary	2 MP
	Features	Geo-tagging
	Video	Yes
	Secondary	VGA
Sound	Alert types	Vibration; MP3, WAV ringtones
	Loudspeaker	Yes
	3.5mm jack	Yes
Comms	WLAN	Wi-Fi 802.11 b/g/n, hotspot
	Bluetooth	3.0
	GPS	Yes, with A-GPS
	Radio	FM radio
	USB	microUSB 2.0
Features	Sensors	Accelerometer
	Messaging	SMS (threaded view), MMS, Email, Push Mail, IM
	Browser	HTML
	Java	No
		- MP3/WAV/WMA/AAC player - MP4/H.264 player - Document viewer - Photo viewer/editor
Battery		Non-removable Li-Ion 3500 mAh battery
Misc.	Colours	Black, Silver

APPENDIX L: STATSSA IT INFORMATION SECURITY POLICY

Relevant Procedures:

To be read in conjunction with related legislation, policies and standards listed below:

- Minimum Information Security Standards (MISS)
- State Information Technology Agency Act (Act no. 88 of 1998)
- SACSA/090/1(4) "Communication Security in the RSA"
- Information Act (Act no. 70 of 2002)
- Promotion of Access to Information Act (Act no. 2 of 2000)
- Electronic Communication and Transaction Act (Act no. 25 of 2002)
- Copyright Act (Act no. 98 of 1978)
- National Strategic Intelligence Act (Act no. 39 of 1994)
- National Archives of SA Act (Act no. 43 of 1996)
- Public Service Act (Act no. 103 of 1994)
- State the procurement policy. At an appropriate level of management of Stats SA
- Change Control Management Policy
- Backup Policy
- Electronic Communications – Acceptable use policy
- Server and Server room policy
- Anti-virus policy

Informative References:

SANS 27002:2008 REFERENCES:

5.1.1 Information security policy document

5.1.2 Review of the information security policy

6.1.1 Management commitment to information security

7.1.1 Inventory of assets

7.1.2 Ownership of assets

7.1.3 Acceptable use of assets

8.1.2 Screening

8.2.2 Information security awareness, education and training

8.3.3 Removal of access rights

9.1.1 Physical security perimeter

9.1.4 Protection against external and environmental threats

9.1.5 Working in secure areas

9.2.2 Supporting utilities

9.2.4 Equipment maintenance

9.2.6 Secure disposal or re-use of equipment

APPENDIX M: WESTERN CAPE COMPLETED TEST CASES

Test Case 01: User Profile Creation

Test Case ID	CS2016_TS_DEV_WCMini_01a			
Test Case:	**User profile creation** - SUPERVISOR			
Description:	Test whether Survey Solutions HEADQUARTERS user can create a SUPERVISOR(s) (singular and plural, viz. 1: many) for the WC Mini Test			
Important Notes:	**Statistics South Africa HEADQUARTERS server** **http://capi.statssa.gov.za (internal / external link)**			
Expected Result:	The tester will be able to create **(N)** SUPERVISORS on the Survey Solutions HEADQUARTERS (SSH) server			
Prerequisites:	Internet or Intranet access to Statistics South Africa			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to SSH, select “Teams and Roles” → click “Add Supervisor”. Complete following: • User Name • Password • Confirm Password • Email Click “Save changes”	SUPERVISOR **N** is created (N = 1 case) SUPERVISOR **N** is created (N > 1 case)		**P** **P** **P**
Alternative Flow				
2	Log in to SSH, select “Teams and Roles” → click “Add Supervisor”. Complete following: • User Name	SSH error message.		

	• Password • Confirm Password • Email Click “Save changes”			

Test Verification:	**Procedural Steps**	**P/F**
1	Review the HEADQUARTERS SUPERVISOR list- To Ensure that all SUPERVISOR(s) are added on SSH	**P**

Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**21/08/2015**	**Pass**	Yes **X**/No ☐
Test Comments:	**Successfully created 3 Supervisors on the Survey Solutions HEADQUARTERS (SSH) server (WCsuper1, WCsuper2 and WCsuper3)**				

Test Case ID	CS2016_TS_DEV_WCMini_01b			
Test Case:	**User profile creation** - INTERVIEWER			
Description:	Test whether Survey Solutions HEADQUARTERS user can create an INTERVIEWER (singular and plural, viz. 1: many) for the WC Mini Test			
Important Notes:	**Statistics South Africa HEADQUARTERS server** http://capi.statssa.gov.za **(internal / external link)**			
Expected Result:	The tester will be able to create **(N)** INTERVIEWERS on the Survey Solutions HEADQUARTERS (SSH) server			
Prerequisites:	Internet or Intranet access to Statistics South Africa			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to SSH, select “Teams and Roles” → click “INTERVIEWERS” next to the relevant SUPERVISOR”. Select “Create”	INTERVIEWER **N** is created (N = 1 case)		**P** **P**

	Complete following: • User Name • Password • Confirm Password • Email Click “Save changes”	INTERVIEWER **N** is created (N > 1 case)		**P**
Alternative Flow				
2	Log in to SSH, select “Teams and Roles” → click “INTERVIEWERS” next to the relevant SUPERVISOR”. Select “Create” Complete following: • User Name • Password • Confirm Password • Email Click “Save changes”	SSH error message.		

Test Verification:	**Procedural Steps**	**P/F**
1	Review the HEADQUARTERS INTERVIEWER list- To Ensure that all INTERVIEWER(s) are added on SSH	**P**

Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**21/08/2015**	**Pass**	Yes **X**/No ☐
Test Comments:	**Successfully created 10 INTERVIEWERS on the Survey Solutions HEADQUARTERS (SSH) server (WCinter1 – WCinter10)**				

Test Case 02: Survey sample file upload

Test Case ID	CS2016_TS_DEV_WCMini_02			
Test Case:	**Survey sample file upload** - HEADQUARTERS			
Description:	Test whether Survey Solutions HEADQUARTERS can upload a sample file received by Methodology			
Important Notes:	**Statistics South Africa HEADQUARTERS server** **http://capi.statssa.gov.za (internal / external link)**			
Expected Result:	The tester will be able to upload a sample file to the Survey Solutions HEADQUARTERS (SSH) server			
Prerequisites:	Internet or Intranet access to Statistics South Africa Sample file is prepared and saved as (.tab) file containing the pre-loaded fields			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to SSH, select "Survey Setup" → click "Import template" → Login to World-Bank DESIGNER → Click questionnaire to import. Select "Survey Setup" Select "Batch Upload" from actions dropdown box → Select "Browse" → search for file and select → Select "Upload".	Sample file is uploaded		P
Alternative Flow				
2	Log in to SSH, select "Survey Setup" → click "Import template" → Log in to WorldBank DESIGNER → Click questionnaire to import. Select "Survey Setup" Select "Batch Upload" from	SSH error message.		

	actions dropdown box → Select “Browse” → search for file and select → Select “Upload”.			

Test Verification:	Procedural Steps	P/F
1	Review the HEADQUARTERS Interviews list- Select “Interviews”	P
	Select relevant questionnaire “Template”	P
	View ***uploaded*** sample	P

Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**22/08/2015**	**Pass**	Yes **X**/No ☐
Test Comments:	**Successfully loaded the WC mini test sample of 100 sampling points on the Survey Solutions HEADQUARTERS (SSH) server**				

Test Case 03: Assignment planning

Test Case ID	CS2016_TS_DEV_WCMini_03
Test Case:	**Assignment Planning**
Description:	Check if the WC Mini Test sample workload can be assigned/ distributed amongst the SUPERVISORs and INTERVIEWERs in preparation for data collection
Important Notes:	**Statistics South Africa HEADQUARTERS server** **http://capi.statssa.gov.za (internal / external link)** Methodology to include all relevant fields: • Province Code (P_PROVINCE, 1 digit) • Municipality Code (P_MUNIC, 3 digits) • Main Place (MainPlace, 6 digits) • Sub-Place (SubPlace, 9 digits)

	• Enumeration Area Number (EAQN, 8 digits) • Dwelling Unit Number (DUNumber, 1 digit) • Map Reference Number (MapRefNo, 3 digits)			
Expected Result:	The sample on CAPI HEADQUARTERS is assigned successfully to SUPERVISORS and INTERVIEWERS participating in the WC Mini Test			
Prerequisites:	Internet or Intranet access to Statistics South Africa Sample file is prepared and uploaded as (.tab) file containing the pre-loaded fields mentioned above SUPERVISORS and INTERVIEWERS have been created			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to SSH, select “Interviews” → select template questionnaire “CS2016 WC test” Select from sample file in the Interviews panel and assign the sampled dwelling unit to the relevant INTERVIEWER	Sample file is assigned to all SUPERVISORs Sample file is assigned to all INTERVIEWERs		P P
Alternative Flow				
2	Log in to SSH, select “Interviews” → select template questionnaire “CS2016 WC test” Select from sample file in the Interviews panel and assign the sampled dwelling unit to the relevant INTERVIEWER	SSH error message.		
Test Verification:	**Procedural Steps**			**P/F**
1	Review the HEADQUARTERS Interviews list- Select “Interviews”			P

	Select relevant questionnaire "Template" View ***assigned*** sample	**P** **P**

Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**24/08/2015**	**Pass**	Yes X/No ☐
Test Comments:	**Successfully completed the Assignment planning on the Survey Solutions HEADQUARTERS (SSH) server**				

Test Case 04: Sample location verification

Test Case ID	CS2016_TS_DEV_ WCMini_04			
Test Case:	**Sample location verification**			
Description:	Check if the INTERVIEWERs can locate the sampled dwelling using Google MyMaps and identify the actual feature found at this point. Viz. (Test sample point description vs. infield sample point description located)			
Important Notes:	INTERVIEWERs to use the device for navigation to ***EACH SAMPLED POINT***			
Expected Result:	The INTERVIEWER successfully navigates to the sample location using Google MyMaps			
Prerequisites:	INTERVIEWERs Vehicle Device enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWERs to install the ***MyMaps*** application from Google Playstore on an Android device prior to sample location			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to device Select MyMaps app	Successful login to device		

	Select Assigned Sample point Select Navigate to Sample Point Drive to destination	Successful login to MyMaps View selected sample point GPS Direction to sample point on the device Arrive at Destination		
Alternative Flow				
2	Log in to device Select MyMaps app Select Assigned Sample point Select Navigate to Sample Point Drive to destination	Error login (username/password) MyMaps not available / not functional Error: No sample Error: No sample Error: MyMaps app failed		
Test Verification:	**Procedural Steps**			**P/F**
1	Log in to Mymaps			

	Select Sample point to navigate to Arrive at sample point destination				
Tested By:	INTERVIEWER **(See INTERVIEWER name list)**	**Date Tested**		**Pass**	Yes☐/No ☐
Test Comments:	Ensure comments are provided for EACH SAMPLE POINT: 1. Located successfully 2. Location < 5 meters from sample point 3. Location > 5 meters from sample point 4. Cannot locate sample point (**NB:** Provide reasons for cases 2-4) **// See Section 3.1 above**				

Test Case 05: Data collection

Test Case ID	CS2016_TS_DEV_ WCMini_05
Test Case:	**Data Collection**
Description:	Check if CAPI application opens at the sample point Check if CAPI "Get Collection" button can save the GPS location coordinates Navigate through the ***CS 2016 WC Test (Version 1)*** questionnaire and complete
Important Notes:	INTERVIEWERs to open and save GPS Location at ***EACH SAMPLE POINT*** INTERVIEWERs should ***NOT contact actual respondents***

	INTERVIEWERs to complete a fictitious questionnaire in the field (after GPS location coordinates captured on CAPI) **NB:** 1. If the area does not allow (i.e. danger, safety or otherwise); INTERVIEWERs to complete the questionnaire in a safe location or in the office **NB:** 2. For ***EACH SAMPLE POINT*** capture comments in relation to the structure found at the ***SAMPLE POINT*** on CAPI in the comments section **NB:** 3. The questionnaire skips and logic of the ***CS 2016 WC Test (Version 1)*** questionnaire will **not** be tested in this TEST CASE EXERCISE (this is done by the CS CONTENT DIVISION HO), since this questionnaire version is ***not the final*** version of CS 2016.			
Expected Result:	The INTERVIEWER successfully captures the completed questionnaire on CAPI			
Prerequisites:	INTERVIEWERs Devices enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWER application Assignment Plan completed by SUPERVISOR for INTERVIEWER			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to device Select INTERVIEWER app and log in Select Assigned Sample point	A successful login to the device Successful login to INTERVIEWER app Opens the CAPI questionnaire for data collection		

	Complete the questionnaire and select complete when done	Saved completed questionnaire on CAPI		
Alternative Flow				
2	Log in to device Select INTERVIEWER app and log in Select Assigned Sample point Complete the questionnaire and select complete when done	Error login (username/password) Error: Cannot log in to INTERVIEWER app Error: Cannot open assigned sample point Error: cannot save the questionnaire		

Test Verification:	**Procedural Steps**	**P/F**
1	Log in to CAPI INTERVIEWER app	
2	Select INTERVIEWER app and log in	
3	Select Assigned Sample point	
4	Complete the questionnaire and select complete when done	

Tested By:	INTERVIEWER **(See INTERVIEWER name list)**	**Date Tested**		**Pass**	Yes☐/No ☐
Test Comments:	Ensure comments are provided for EACH SAMPLE POINT: **// See Section 3.2 above**				

Test Case 06: Captured data synchronisation

Test Case ID	CS2016_TS_DEV_ WCMini_06			
Test Case:	**Captured data synchronisation**			
Description:	Check if CAPI application synchronises captured or completed data			
Important Notes:	INTERVIEWERs to save and synchronise data captured to SSH for ***EACH SAMPLE POINT***			
Expected Result:	The INTERVIEWER successfully synchronises completed questionnaire on CAPI SSH			
Prerequisites:	INTERVIEWERs Devices enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWER application Assignment Plan completed by SUPERVISOR for INTERVIEWER Data collection completed by INTERVIEWER			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to the device as an INTERVIEWER Select "Synchronisation" tab	Successful login to the device Activate Synchronisation		

	Select “Sync”	Data Synchronised to CAPI SSH successfully		
Alternative Flow				
2	Log in to the device as an INTERVIEWER Select “Synchronisation” tab Select “Sync”	Error login (username/password) Error: inactivate Synchronisation Error Message: Data is NOT synchronised to CAPI SSH successfully		

Test Verification:	**Procedural Steps**	**P/F**
1 **2**	Log in to CAPI INTERVIEWER app Select Synchronisation tab and sync data	

Tested By:	INTERVIEWER **(See INTERVIEWER name list)**	**Date Tested**		**Pass**	Yes☐/No ☐
Test Comments:	Ensure comments are provided for EACH SAMPLE POINT: **// See Section 3.3 above**				

Test Case 07: **Real-time reporting**

Test Case ID	CS2016_TS_DEV_ WCMini_07a
Test Case:	**Real-time reporting**
Description:	Check if CAPI application reports are available in real-time
Important Notes:	INTERVIEWERs to save and synchronise data captured to SSH for ***EACH SAMPLE POINT***
Expected Result:	SUPERVISOR to successfully view reports in real time (after synchronisation from assigned INTERVIEWER)
Prerequisites:	INTERVIEWERs Devices enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWER application Assignment Plan completed by SUPERVISOR for INTERVIEWER Data collection completed by INTERVIEWER INTERVIEWER synced data to the server

Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to the device as a SUPERVISOR	Successful login to the device		P
	Select “Reports” tab	Activate Reports		P
	Select one of 3 types of reports “a. Surveys and Statuses” “b. Team member and Statuses” and “c. Quantity” tab	Activate relevant report		P

	Select relevant INTERVIEWER to view report	View report (survey status) in real-time			P
Alternative Flow					
2	Log in to the device as a SUPERVISOR Select “Reports” tab Select one of 3 types of reports “a. Surveys and Statuses” “b. Team member and Statuses” and “c. Quantity” tab Select relevant INTERVIEWER to view report	Error login (username/password) Error: inactivate Reports			
Test Verification:	**Procedural Steps**				**P/F**
1 **2**	Log in to CAPI INTERVIEWER app as a SUPERVISOR Select “Reports” tab and one of 3 reports (see above)				**P** **P**
Tested By:	SUPERVISOR **(See supervisor name list)**	**Date Tested**	**26/08/2015**	**Pass**	Yes **X** /No ☐
Test Comments:	**Successfully viewed all reports from Lenovo 7 inch tablet.**				

Test Case ID	CS2016_TS_DEV_ WCMini_07b			
Test Case:	**Real-time reporting**			
Description:	Check if CAPI application reports are available real-time			
Important Notes:	INTERVIEWERs to save and synchronise data captured to SSH for ***EACH SAMPLE POINT***			
Expected Result:	HEADQUARTERS to successfully view reports real time (after synchronisation from assigned INTERVIEWER)			
Prerequisites:	INTERVIEWERs Devices enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWER application Assignment Plan completed by SUPERVISOR for INTERVIEWER Data collection completed by INTERVIEWER INTERVIEWER synced data to server			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to HEADQUARTERS on computer	Successful login to HEADQUARTERS on computer		P
	Select “Reports” tab	Activate Reports		P
	Select one of **5** types of reports “a. Surveys and Statuses” “b. Team member and Statuses” “c. Map Report” “d. Cumulative Interview Chart”	Activate relevant report		P

	"e. Quantity" tab Select relevant Survey Template to view report	View report (survey status) real-time		P
Alternative Flow				
2	Log in to the device as a HEADQUARTERS Select "Reports" tab Select one of 5 types of reports "a. Surveys and Statuses" "b. Team member and Statuses" "c. Map Report" "d. Cumulative Interview Chart" "e. Quantity" tab Select relevant Survey Template to view report	Error login (username/password) Error: inactivate Reports		

Test Verification:	**Procedural Steps**	**P/F**
1	Log in to CAPI INTERVIEWER app as HEADQUARTERS	P
2	Select "Reports" tab and one of 5 reports (see above)	P

Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**28/02/15**	**Pass**	Yes **X** /No ☐
Test Comments:	**All 5 reports successfully loaded and displayed the WC Mini test data.** **A normal connection on Stats SA's intranet and the 3G connection proved to be sufficient for all the report viewing.**				

Test Case 08: Data extraction

Test Case ID	CS2016_TS_DEV_ WCMini_08			
Test Case:	**Data Extraction**			
Description:	Check if CAPI application can export data collected			
Important Notes:	INTERVIEWERs to save and synchronise data captured to SSH for ***EACH SAMPLE POINT***			
Expected Result:	HEADQUARTERS successfully extracts or downloads data collected in the field from CAPI SSH			
Prerequisites:	INTERVIEWERs Devices enabled with: Data Bundle Internet / 3G / GPRS INTERVIEWER application Assignment Plan completed by SUPERVISOR for INTERVIEWER Data collection completed by INTERVIEWER			
Test Steps:	**Action**	**Expected Result**	**Notes**	**P/F**
1	Log in to HEADQUARTERS on computer	Successful login to HEADQUARTERS on computer Activate Data Export		**P** **P**

	Select "Data Export" tab Select relevant questionnaire "Template" Select "All data" → "Data.tab" Select "Save" to extract /download data to the computer	View buttons to download selected data Data.tab visible Data saved to the computer successfully		P P P
Alternative Flow				
2	Log in to HEADQUARTERS on computer Select "Data Export" tab Select relevant questionnaire "Template" Select "All data" → "Data.tab" Select "Save" to extract /download data to the computer	Error login (username/password) Error: "Data Export" tab Error: questionnaire not visible Error: data.tab not visible Error: data not saved		
Test Verification:	**Procedural Steps**			**P/F**
1	Log in to HEADQUARTERS on computer			P

2 3 4	Select “Data Export” tab Select “All data” → “Data.tab” Select “Save”				P P P
Tested By:	HEADQUARTERS **Mahier Hattas**	**Date Tested**	**28/02/15**	**Pass**	Yes **X** / No ☐
Test Comments:	**All test data was exported successfully and saved to local computer.**				

APPENDIX N: CHAPTER 5 – Iteration 1 Tables

Table N.1: WC Quality Themes [WCT1a. Generic end-to-end DDC process flow]

DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME	SASQAF Dimension	SASQAF CODE	SASQAF Indicator (Statistics South Africa, 2010b)	SASQAF Standard (Statistics South Africa, 2010b)	SVC CODE	Data Source	Research Choice	Source	Variables	Questions / Test
PROCESS FLOW AND STRUCTURAL CHANGES (2.5.8)	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu1	Secondary Data (WC TEST Case 01)	Quantitative	Survey Solutions Production Data (Anonym)	User Profile Creation	Successfully created users for DDC (Y/N) %
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu1	Secondary Data (WC TEST Case 02)	Quantitative	Survey Solutions Production Data (Anonym)	Survey Sample File Upload	Successfully uploaded sample file for DDC (Y/N) %
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu1	Secondary Data (WC TEST Case 03)	Quantitative	Survey Solutions Production Data (Anonym)	Assigned to Interviewer	Successfully assigned sample workload to Interviewer for DDC (Y/N) %
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Sample Location	Select Sample point to navigate to (Y/N) %

	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Get Location	Arrive at sample point destination (Y/N) %
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Accuracy of Sampled visit	Derived variable : Accuracy of Sampled visit --> is Sample Location = Get Location
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 05)	Quantitative	Survey Solutions Production Data (Anonym)	Data Collection	Successfully assigned sample workload to Interviewer for DDC (Y/N) %
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 06)	Quantitative	Survey Solutions Production Data (Anonym)	Synchronisation to Central Server	% data synchronised to the server
	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 07)	Quantitative	Survey Solutions Production Data (Anonym)	Real-time reporting	% Data available for real-time reporting to field management

	WCT1a. Generic end to end DDC process flow	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Computing resources**	1.6.6 Computer software resources must be adequate in terms of · capturing systems;	Bu4	Secondary Data (WC TEST Case 08)	Quantitative	Survey Solutions Production Data (Anonym)	Data Extraction	Available to NSO
PROCESS FLOW AND STRUCTURAL CHANGES (2.5.8)	WCT1b. Access to data (external) NSO STAKEHOLDERS DATA ACCESS	Pre-requisites of quality	PRER6 (6)	1.6 Resources are commensurate with the needs of statistical programmes. **Staff (access)**	1.6.2 There must be a statistics unit or component responsible for compiling statistics.	De7	Secondary Data (WC TEST Case 08)	Quantitative	Survey Solutions Production Data (Anonym)	Exported Data Set	Available to NSO

Table N.2: WC Quality Themes [Other]

DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME	SASQAF Dimension	SASQAF CODE	SASQAF Indicator (Statistics South Africa, 2010b)	SASQAF Standard (Statistics South Africa, 2010b)	SVC CODE	Data Source	Research Choice	Source	Variables	Questions / Test
COST	AVERAGE PAPDC COST PER HH	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes. • Financing.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	Researcher Cost Estimate	Average PAPDC Cost (per HH surveyed)	What is the average cost to collect data per household using Paper and Pen Data Collection (PAPDC)?
	AVERAGE DDC COST PER HH	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes. • Financing.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	Researcher Cost Estimate	Average DDC Cost (per HH surveyed)	What is the average cost to collect data per household using Digital Data Collection (DDC)?

	COST RATIO PAPDC VS. DDC	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes. • Financing.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	Researcher Cost Estimate	Cost Efficiency	Derived variable: Is Total Survey Cost per HH (DDC) < Total Survey Cost per HH (PAPDC)?
ACCURACY, and PERCEPTION OF QUALITY	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	8.2	8.2. Using DHDT in the field would result in an improvement in the quality and accuracy of household survey data collection
	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	11.1	11.1 In terms of survey data collection quality, I believe DHDT can improve the overall quality of data collected
	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	11.2	11.2 In terms of survey data collection quality, I believe DHDT can result in poor data quality during each item response
	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	11.3	11.3 In terms of survey data collection quality, I believe DHDT can initiate an increased response interaction from respondents

	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	11.4	11.4 In terms of survey data collection quality, I believe DHDT can result in an increase in trust in the data from the users
	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2	Primary Data	Qualitative	Focus Group Guide	2.5	2.5 Perceptions of the quality of data
	ACCURACY	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Misclassification errors	3.2.7 The proportion of units that are misclassified must be at an acceptable level.	De7	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Sample Location	Select Sample point to navigate to (Y/N) %
	ACCURACY	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Misclassification errors	3.2.7 The proportion of units that are misclassified must be at an acceptable level.	De7	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Get Location	Arrive at sample point destination (Y/N) %
	ACCURACY	*GAP1 Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Misclassification errors	3.2.7 The proportion of units that are misclassified must be at an acceptable level.	De7	Secondary Data (WC TEST Case 04)	Quantitative	Survey Solutions Production Data (Anonym)	Accuracy of Sampled visit	Derived variable : Accuracy of Sampled visit --> is Sample Location = Get Location
SECURITY	THEFT	*GAP2 (Security - Theft of devices /				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Number of Stolen devices	

		data at source)									
	THEFT	*GAP2 (Security - Theft of devices/ data at source)				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Total Number of Survey Officers	
	THEFT	*GAP2 (Security - Theft of devices/ data at source)				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Device Loss through Theft	Derived variable: Number of Stolen devices/ Survey Officers / Total Number of Survey Officers
	DATA LOSS	*GAP2 (Security - Data Loss)				Co1	Secondary Data (WC TEST Case 06)	Quantitative	Survey Solutions Production Data (Anonym)	% Synchronisation Success Rate (Data Loss)	
	WCT2. **USER INTERFACE** USER & RESPONDENT ADOPTION OF DDC	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.14 Respondent effects must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	13.5	13.5 DHDT is user-friendly for both interviewer and interviewee

TECHNOLOGY ADOPTION AND ACCEPTANCE	**WCT2.** **USER INTERFACE** BEST PRACTICES OPINIONS	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.14 Respondent effects must be determined and reported.	Bu4	Primary Data	Qualitative	Focus Group	2.6	2.6 Opinion on best practices in official household survey collection
	WCT2b. TECHNOLOGY ACCEPTANCE UNDERSTANDING DIGITAL TECHNOLOGY	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.13 The effects of the interviewers must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	8.1	8.1 I sometimes feel that I have been left behind when it comes to using DHDT
	WCT2b. TECHNOLOGY ACCEPTANCE APPROVAL OF DIGITAL TECHNOLOGY USE	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.13 The effects of the interviewers must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	8.6	8.6 I am excited about using DHDT in household survey collection.
	WCT2b. TECHNOLOGY ACCEPTANCE	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g.	3.2.13 The effects of the interviewers must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	12.5	12.5 Regarding best practices in official survey data collection, I believe DHDT should have been in existence a while back already given that mobile devices have been

	EXPECTATION OF WHEN DIGITAL TECHNOLOGY SHOULD HAVE BEEN INTRODUCED			questionnaire effects, data collection effects, interviewer effects and respondent effects)							increasingly growing w.r.t power and speed
	WCT2b. TECHNOLOGY ACCEPTANCE USERS BELIEVE IN ADVANTAGES OF DHDT VS PAPDC	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.13 The effects of the interviewers must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	13.3	13.3 During your participation in DDC pilots thus far I found that there are significant advantages to DDC vs. PAPDC
	WCT2c. ACCESS TO DIGITAL DEVICES USER ACCESS TO CELL PHONES	*GAP3 (ACCESS TO DIGITAL DEVICES - Staff access to mobile equipment)				Bu4	Primary Data	Quantitative	PHD / CS2016 Research Survey Questionnaire	6	6. Do you own a cellular phone?
	WCT2c. ACCESS TO DIGITAL DEVICES OWNERSHIP OF OTHER	*GAP3 (ACCESS TO DIGITAL DEVICES - Staff access to mobile				Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	9	9. Do you currently own a smartphone, tablet or mobile computing device?

	MOBILE DEVICES	equipment)									
	WCT2c. ACCESS TO DIGITAL DEVICES USE OF MOBILE DEVICES FOR WORK PURPOSES	***GAP3** (ACCESS TO DIGITAL DEVICES - Staff access to mobile equipment)				Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	10	10. Do you currently use a smartphone, tablet or mobile computing device for work-related purposes, other than calling, sms'ing, surfing the internet or for social media-related activities?
WCT3. RESPONSE RATE	EFFECTIVENESS OF DDC VS PAPDC IN RELATION TO RESPONSE RATES	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)	3.2.13 The effects of the interviewers must be determined and reported.	Bu4	Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	13.2	13.2 During your participation in DDC pilots thus far I found DDC very effective in collecting household survey information
	ACTUAL RESPONSE RATES	Accuracy	ACCU (2)	3.2 Measures of non-sampling errors are calculated, viz.: • Non-response errors (e.g. item non-response rates, unit non-response rates and overall response rates)	3.2.24 Unit non-response rate must be within acceptable levels.	Bu4	Secondary Data	Quantitative	CS 2016 Report / WC Test Report	Response Rate	Reported response rates

WCT4. SOUND METHODOLOGIES AND COMMON FRAMEWORK	OPINION ON BEST DDC BEST PRACTICES - METHODOLOGIES						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	12.1	12.1 Regarding best practices in official survey data collection, I believe DHDT has best practice methodologies defined in South Africa or in sub-Saharan Africa
	OPINION ON BEST DDC BEST PRACTICES - COMMON FRAMEWORK						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	12.3	12.3 Regarding best practices in official survey data collection, I believe DHDT requires a common framework in order to progress and promote the use of DDC amongst NSO's
	OPINION ON BEST DDC BEST PRACTICES - ADOPTABILITY TO SASQAF						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	12.4	12.4 Regarding best practices in official survey data collection, I believe DHDT can be adopted easily since there currently exists a quality framework (SASQAF) for PAPDC

APPENDIX O: CHAPTER 6 – Iteration 2 Tables

Table O.1: KZN CSS Quality Themes [A - E]

DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME	SASQAF Dimension	SASQAF CODE	SASQAF Indicator (Statistics South Africa, 2010b)	SASQAF Standard (Statistics South Africa, 2010b)	SVC CODE	Data Source	Research Choice	Source	Variables	Questions / Test
SPEED	6.3.1.1 DDC DATA COLLECTION SPEED IN BETWEEN ASSIGNMENT OF DATA TO APPROVAL BY HEADQUARTERS	Timeliness	TIME3 (3)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection	4.3.3 Data collection must follow the project plan/schedule		Secondary Data	Quantitative	Survey Solutions Production Data (Anonym)	Supervisor Assigned, Completed, Date, Time	CAPI Interview Actions Table
	6.3.1.2a. DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Qualitative	KZN CSS Report (Summary Notes)	Planned Release Date	Available to NSO
	6.3.1.2b. DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Qualitative	KZN CSS Publication	Actual Release Date	

	6.3.1.2c. DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Qualitative	KZN CSS Publication	Timely dissemination	Derived variable: is Planned Release Date = Actual Release Date
TRAINING	DDC TRAINING APPROACHES	*GAP4 (DDC TRAINING APPROACHES - develop DDC training approaches)				Co1	Secondary Data	Qualitative	CS 2016 Report / KZN CSS Report / Debriefing Focus Groups	Training Process	What training techniques were employed for DDC vs. PAPDC
	TRAINING OUTCOMES					Co1	Secondary Data	Qualitative	KZN CSS Report / Debriefing Focus Groups	Training Outcomes	Effectiveness of training for DDC vs. PAPDC
PROCESS FLOW AND STRUCTURAL CHANGES	**KZNCSS1. PUBLIC ACCESS TO DATA:** PUBLIC STAKEHOLDERS	Accessibility	ACCE1 (1)	5.1 Are statistical products (e.g. data, metadata) available to the public?	5.1.1 The statistical products must be disseminated to the public.		Secondary Data	Qualitative	KZN CSS Report (Release Notes) / Debriefing Focus Groups		Available to South African Public
	KZNCSS2. ORGANISATIONAL STRUCTURE : DDC ORGANOGRAM						Secondary Data	Qualitative	KZN CSS Report (Summary Notes) / Debriefing Focus Groups		Available to NSO

	KZNCSS3. ORGANISATION AL PROCESSES: WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?	***GAP5** (WILL DDC COMPLICATE HOUSEHOLD DATA COLLECTION PROCESSES?- develop and train DDC process flows)		Secondary Data	Qualitative	KZN CSS Report (Summary Notes) / Debriefing Focus Groups		Available to NSO
	KZNCSS3. ORGANISATION AL PROCESSES: HOW DOES THE ORGANISATION TRANSITION TO ADAPT?	***GAP6** (HOW DOES THE ORGANISATION TRANSITION TO ADAPT?- develop DDC transition strategy for all affected staff)		Secondary Data	Qualitative	KZN CSS Report (Summary / Debriefing Notes) / Debriefing Focus Groups		Available to NSO

APPENDIX P: CHAPTER 7 – Iteration 3 Tables

Table P.1: COMMUNITY SURVEY 2016 Quality Themes

DDC DIMENSION / THEME	Sub DDC DIMENSION / THEME	SASQAF Dimension	SASQAF CODE	SASQAF Indicator (Statistics South Africa, 2010b)	SASQAF Standard (Statistics South Africa, 2010b)	SVC CODE	Data Source	Research Choice	Source	Variables	Questions / Test
COST	AVERAGE PAPDC COST PER HH	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes. • Financing.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	CS 2016 Report	Average DDC Cost (per HH surveyed)	What is the average cost to collect data per household using Paper and Pen Data Collection (PAPDC)?
	AVERAGE DDC COST PER HH	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes. • Financing.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	CS 2007 Report / Other Report	Average PAPDC Cost (per HH surveyed)	What is the average cost to collect data per household using Digital Data Collection (DDC)?
	COST RATIO PAPDC VS. DDC	Pre-requisites of quality	PRER6 (7)	1.6 Resources are commensurate with the needs of statistical programmes.	1.6.7 Budgets must be adequate and timely.		Secondary Data	Quantitative	CS 2016 Report	Cost Efficiency	Derived variable: Is Total Survey Cost per HH (DDC) < Total Survey Cost per HH (PAPDC)?

				• Financing.							
	PERCEPTION OF QUALITY								PhD / CS2016 Research Survey Questionnaire	8.2	8.2. Using DHDT in the field would result in an improvement in the quality and accuracy of household survey data collection
	PERCEPTION OF QUALITY								PhD / CS2016 Research Survey Questionnaire	11.1	11.1 In terms of survey data collection quality, I believe DHDT can improve the overall quality of data collected
ERROR RATE and PERCEPTION OF QUALITY	PERCEPTION OF QUALITY								PhD / CS2016 Research Survey Questionnaire	11.2	11.2 In terms of survey data collection quality, I believe DHDT can result in poor data quality during each item response
	PERCEPTION OF QUALITY								PhD / CS2016 Research Survey Questionnaire	11.3	11.3 In terms of survey data collection quality, I believe DHDT can initiate an increase response interaction from respondents
	PERCEPTION OF QUALITY						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	11.4	11.4 In terms of survey data collection quality, I believe DHDT can result in an increase in trust in the data from the users
	PERCEPTION OF QUALITY	Integrity	INTE(5)	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	Di2		Qualitative	Focus Group Guide	2.5	2.5 Perceptions of the quality of data

	ERROR RATE	Accuracy	ACCU2 (16)	3.2 Measures of non-sampling errors are calculated	3.2.16 Data entry must average an acceptable accuracy rate		Secondary Data	Quantitative	Survey Solutions Production Data (Anonym) - **MAR**	Error rate	% of REJECTEDBYHQ
SECURITY	THEFT	GAP (Security - Theft of devices/data at source)				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Number of Stolen devices	
	THEFT	GAP (Security - Theft of devices/data at source)				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Total Number of Survey Officers	
	THEFT	GAP (Security - Theft of devices/data at source)				Co1	Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report / WC Test Report	Device Loss through Theft	Derived variable: Number of Stolen devices/ Survey Officers / Total Number of Survey Officers
	DATA LOSS	GAP (Security - Data Loss)				Co1	Secondary Data	Quantitative	CS 2016 Report	% Synchronisation Success Rate (Data Loss)	
SPEED	DDC DATA COLLECTION SPEED IN BETWEEN ASSIGNMENT OF DATA TO APPROVAL BY HEADQUARTERS	Timeliness	TIME3 (3)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection	4.3.3 Data collection must follow the project plan/schedule		Secondary Data	Quantitative	Survey Solutions Production Data (Anonym)	Supervisor Assigned, Completed, Date, Time	CAPI Interview Actions Table

	DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Quantitative	CS 2016 Report / KZN CSS Report (Summary Notes)	Planned Release Date	Available to NSO
	DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Quantitative	KZN CSS Publication	Actual Release Date	
	DATA DISSEMINATION ACCORDING TO PLANNED SCHEDULE	Timeliness	TIME3 (7)	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • dissemination.	4.3.7 Dissemination must follow the project plan/schedule.		Secondary Data	Quantitative	KZN CSS Publication	Timely dissemination	Derived variable: is Planned Release Date = Actual Release Date

TRAINING	DDC TRAINING APPROACHES	***GAP4** (DDC TRAINING APPROACHES - develop DDC training approaches)				Co1	Secondary Data	Qualitative	CS 2016 Report / KZN CSS Report	Training Process	What training techniques were employed for DDC vs. PAPDC
	TRAINING OUTCOMES					Co1	Secondary Data	Qualitative	CS 2016 Report / KZN CSS Report	Training Outcomes	Effectiveness of training for DDC vs. PAPDC
TECHNOLOGY ADOPTION AND ACCEPTANCE	***GAP7** **CS1.DDC COMMUNICATIONS :** (INTERNAL OR EXTERNAL DDC COMMUNICATION)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	12.2	12.2 Regarding best practices in official survey data collection, I believe DHDT methodology and standards exist but are not well communicated internally to the South African NSO or externally
	***GAP10** **CS2.EXPERIENCE AND ABILITY OF USING DIGITAL HANDHELD DEVICE TECHNOLOGY (DHDT):** (EDUCATION LEVEL)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	3	3. What is the highest level of education that you have successfully completed?
	(EXPERIENCE)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	4	4. How many years of working experience do you have at Statistics South Africa in the related areas responsible for the

											collection of household surveys?
	(EXPERIENCE)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	5	5. Are you currently working Fulltime or Part time
	(EXPERIENCE)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	7.1	7.1 No experience with DHDT.
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	7.2	7.2 Attempted to use DHDT, but I still require help on a regular basis.
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	7.3	7.3 Able to perform basic functions in a limited number of DHDT.
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	7.4	7.4 Demonstrate a general competency in a number of DHDT.
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey	7.5	7.5 Acquired the ability to competently use a broad spectrum of DHDT.

									Questionnaire		
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	7.6	7.6 Extremely proficient in using a wide variety of DHDT.
	(ABILITY USING DDC)						Primary Data	Quantitative	PhD / CS2016 Research Survey Questionnaire	13.1	13.1 During your participation in DDC pilots thus far I have not been involved in any DDC pilots

APPENDIX Q: CHAPTER 8 – Kruskal-Wallis Test (KWt)

Q1. KWt – Quantitative analysis across all three iterations [time_elapsed]

The **assumptions** for the KWt are that the variables should have:

- One independent variable (Iteration No) with two or more levels (independent groups) three independent sample groups (1,2,3). The test is more commonly used when you have three or more levels. For two levels, consider using the Mann Whitney U Test instead.
- Ordinal scale, Ratio Scale or Interval scale dependent variables.
- Your observations should be independent. In other words, there should be no relationship between the members in each group or between groups.
- All groups should have the same shape distributions. SPSS was used to test for this condition as part of the test. For all three iterations the distribution is positively skewed (skewness), hence not normally distributed but contains the same shape distribution (see Table Q.1 and Figures Q.1 – Q.3]:

Table Q.1: Iteration [1,2,3] Frequency Statistics [time_elapsed]

Iteration1WCTest	N	Valid	100
		Missing	0
	Std. Deviation		10:19:16
	Skewness		**3.916**
	Kurtosis		14.458
Iteration2KZNCSS	N	Valid	384
		Missing	0
	Std. Deviation		133:03
	Skewness		**3.694**
	Kurtosis		15.171
Iteration3CS2016	N	Valid	384
		Missing	0
	Std. Deviation		156:45
	Skewness		**3.676**
	Kurtosis		15.505

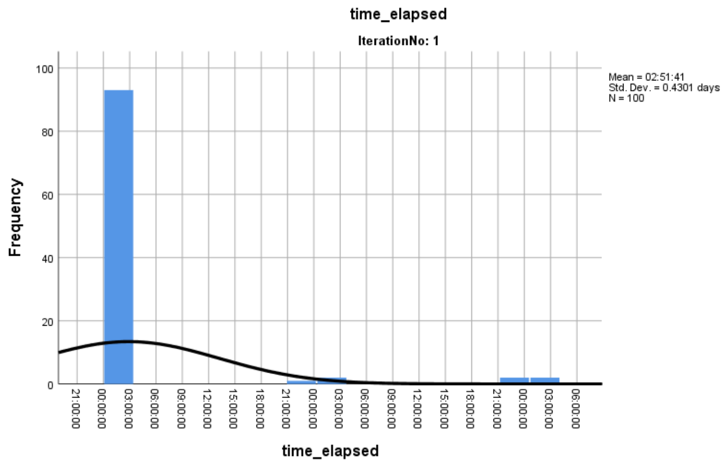


Figure Q.1: Iteration 1 Distribution (Skewness Test)

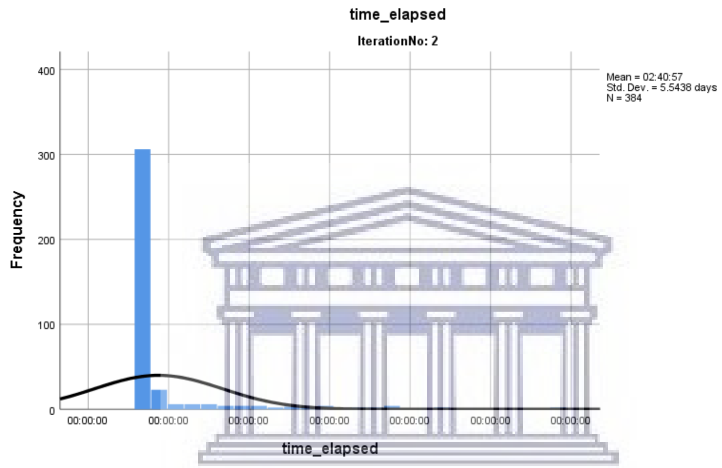


Figure Q.2: Iteration 2 Distribution (Skewness Test)

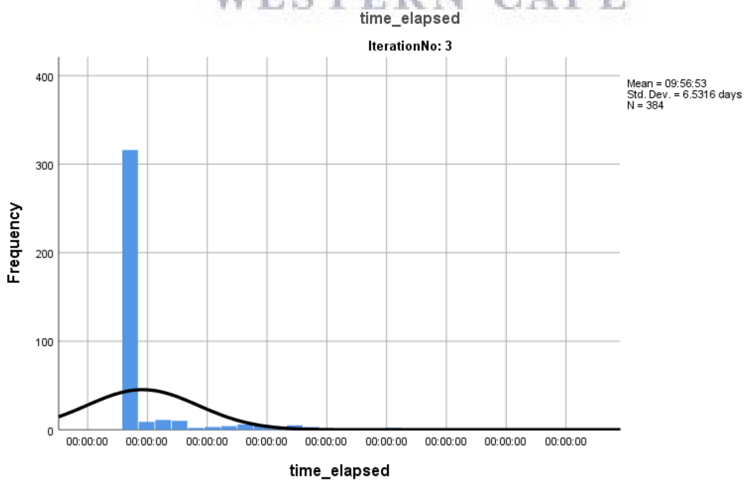


Figure Q.3: Iteration 3 Distribution (Skewness Test)

For KWt, the **hypotheses** are:

- H0: there is no statistically significant difference between the median ***elapsed_time*** taken to conduct an interview using DDC between three different iterations of DDC (medians are equal)
- H1: there is a statistically significant difference between the median ***elapsed_time*** taken to conduct an interview using DDC between three different iterations of DDC (the medians are not all equal)

Table Q.2: Iteration [1,2,3] Ranks of the non-parametric test for equality

Ranks			
	Iteration No	N	Mean Rank
time_elapsed	1	100	233.14
	2	384	470.19
	3	384	451.24
	Total	868	

The KWt ranks the iteration elapsed number of_times for the whole sample; and then compares the mean rank for each iteration group. Table Q.2 above indicates the mean ranks of Iteration 2 and 3 are double that of Iteration 1 respectively (233.14). This already provides an indication that there may be a stronger relationship with Iteration 2 and 3 as opposed to Iteration 1 with the other iterations for this variable.

Table Q.3: Iteration [1,2,3] Hypothesis Test for time_elapsed

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of time_elapsed is the same across categories of IterationNo.	Independent-Samples Kruskal-Wallis Test	.000*	**Reject** the null hypothesis.

Asymptotic significances are displayed. The significance level is .050.

Table Q.3 displays the hypothesis test for time_elapsed to conduct a DDC interview. It indicates that the null hypothesis is rejected at a significance level (0.05) between the three iterations.

Table Q.4: Iteration [1,2,3] Independent-Samples Kruskal-Wallis Test

Total N	868
Test Statistic	74.004[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000

a. The test statistic is adjusted for ties.

Table Q.4 shows that the result of the KWt: Chi-Square (74.004) value; Degrees of freedom (2) and the significance level (an exact p-value): 0.000 (rejected null hypothesis).

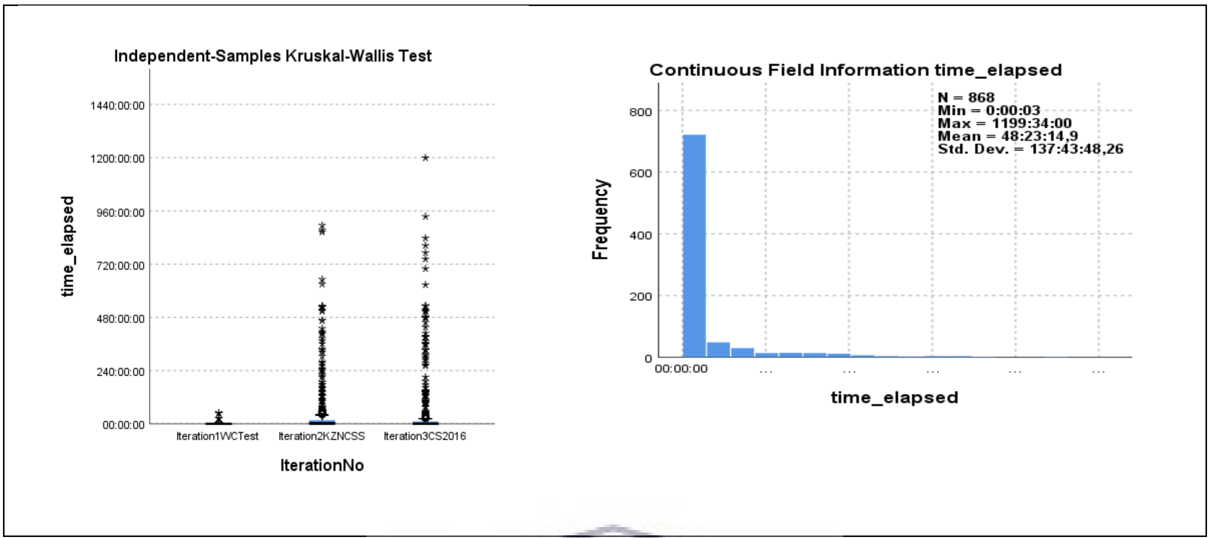


Figure Q.4: Iteration [1,2,3] Boxplot and frequency of time_elapsed

The boxplot (Figure Q.4) compares the medians and spread of the data for time_elapsed by iteration group. The frequency chart indicates the average time taken between initially opening a questionnaire and completing it was 48 hours, 23 minutes and 14 seconds. It should be remembered the data is positively skewed and the majority of observations are much closer to less than 30 minutes, for example in Iteration 2 KZN CSS, 72% of the observations had an average elapsed_time of below 2 hours.

Table Q.5: Iteration [1,2,3] Pairwise Comparisons for Time_elapsed

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Iteration1WCTest-Iteration3CS2016	-218.110	28.147	-7.749	.000	**.000***
Iteration1WCTest-Iteration2KZNCSS	-237.059	28.147	-8.422	.000	**.000***
Iteration3CS2016-Iteration2KZNCSS	18.949	18.094	1.047	.295	.885

Each row tests the null hypothesis that Sample 1 and Sample 2 distributions are the same.

Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

The Bonferroni *post hoc* test indicated significant values after adjustments were made between [Iteration 1 and Iteration 2] and between [Iteration 1 and Iteration 3]; however, there is no statistical significance between [Iteration 3 and Iteration 2] (see Table Q.5).

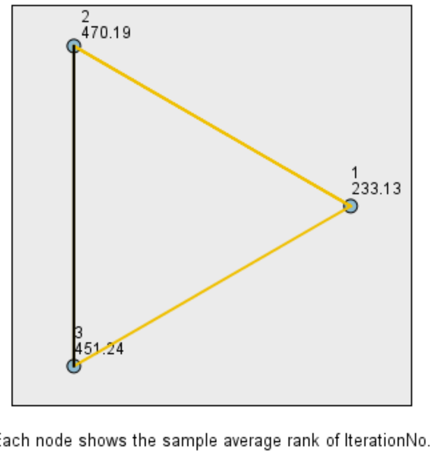


Figure Q.5: Iteration [1,2,3] Pairwise comparison of iteration numbers

Dunn's post hoc tests are carried out on each pair of the iteration groups. As multiple tests are being carried out, SPSS makes an adjustment to the p-value. The Bonferroni adjustment is to multiply each Dunn's p-value by the total number of tests being carried out. The pairwise comparisons above (Figure Q.5) show the results of the Dunn-Bonferroni tests on each pair of iteration groups. Orange line = significant difference between pairs [Iteration 1 and Iteration 2] and [Iteration 1 and Iteration 3] respectively. In summary, the KWt test provided very strong evidence of a difference ($p < 0.001$) between the mean ranks of at least two pairs of iteration groups. Dunn's pairwise tests were carried out for the three pairs of iteration groups. There was very strong evidence ($p < 0.001$, adjusted using the Bonferroni correction) of a difference between the groups [Iteration 1 and Iteration 2] and [Iteration 1 and Iteration 3]. The median elapsed_time for the groups differed. There was no evidence of a difference between [Iteration groups 2 and 3].

Q2. KWt – Quantitative analysis across all three iterations [cost]

A portion of the Stats SA budget was utilised for all operating costs of the three respective iterations with the following key cost drivers for Iteration 2 and Iteration 3 being:

- procurement of digital devices and accessories,
- hiring of vehicles to support fieldwork activities,
- accommodation for project planning sessions and training,
- subsistence and travelling costs,
- venues and catering service

Key cost drivers for the Iteration2KZNCSS were (North, 2015):

- **Enumeration**: the rate of pay was finalised at R8054.80 for the 6 week period (R6125.13 for 1 month subject to applicable tax provisions). The rate of pay was finalised during the first week of training, this impacted projected budget estimations and assumptions made in the planning phase. M&E monitors rate of pay was set at supervisor level R10155.32.
- **Vehicles**: the cost of hiring vehicles was more than the cost of paying staff, and the use of the GG fleet adds an even higher cost, making transport the single largest cost driver.
- **Accommodation**: accommodation expenditure related mostly to the CSS Imbizo and Extended EXCO sessions as accommodation was minimal for CS training and monitoring in comparison.
- **Devices**: the number of devices was limited by the ceiling of R500 000 to avoid a tender process which would have added significant delays to procurement.

A high-level analysis of expenditure is reported on page 14, while it is noted that this summary does not include all expenditure related to the project, total expenditure was approximately (See Figure Q.6) R5 million –the survey cost approximately R250 per dwelling unit dwelling.

Planned high-level budget

Code	Item	Requirement	Total budgeted
587C	Salaries	Salaries for 6 weeks estimated @ R10 000 per month	R 2 100 000.00
	Transport allowance	54 Fieldworkers estimated @R 1 000	R 54 000.00
	Planning Training	Accommodation estimated @R 900 per night	R1 066 000.00
		S&T	
		Catering estimated @R50 per head for 3740	
	Flights	11 ICT support staff, training support	R63 000.00
	Field Gear	T Shirts, Caps, Bibs, Bags, Lanyards	R78 000.00
	Vehicles	20 sedans rental	R 168 000.00
		30 bakkies rental	R 1 008 000.00
		Kilometres, fuel and oil	R 750 000.00
	Devices	250 Lenovo tablets	R 500 000.00
	Airtime	30 vouchers @900	
	Network provider	250 SIM cards, APN and MDM system	
	Data bundles		
	M&E		
	Debriefing		R 400 000.00
		TOTAL	**R 6 187 000.00**

Actual budget expenditure

Actual	Spent	
208 salaries fieldworkers	R 1 585 000.00	R 1 608 620.00
Transport allowance	R 23 620.00	
Accommodation	R 144 695.00	
S&T	R 83 000.00	
Catering	R 300 000.00	
Venues	R 96 000.00	
Air transport	R 421 800.00	
210 caps, bibs	R35 568.00	
57 vehicles rental	R 1 060 000.00	R 1 650 200.00
GG vehicles	R 134 186.00	
Kilometres, fuel and oil	R 225 000.00	
263 Lenovo devices	R 446 731.00	R 578 652.00
263 device covers, powerbanks	R 131 920.00	
Airtime	R 27 000.00	
VPN, router, fibre connectivity	R 193 285.00	
202 GB data usage	R 59 955.00	
6 salaries M&E	R 60 930.00	
TOTAL	**R 5 028 690.00**	

Figure Q.6: Iteration 2 KZN CSS Total Cost in Rands - (North, 2015)

The survey cost per dwelling unit dwelling for Iteration3 CS 2016 amounted to R200 (Statistics South Africa, 2016g), whilst the approximate cost for Iteration1 WC Test amounted to R150 per dwelling (due to the smaller sample size and corresponding resources used over a shorter period). The KWt was conducted on the cost between the three iterations with the following hypothesis (Table Q.6) indicates that the null hypothesis is rejected at a significance level (0.05) between the three iterations:

Table Q.6: Iteration [1,2,3] Null Hypothesis for the cost to conduct a DDC survey per dwelling

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Cost is the same across categories of IterationNo.	Independent-Samples Kruskal-Wallis Test	.000*	**Reject** the null hypothesis.

Asymptotic significances are displayed. The significance level is .050.

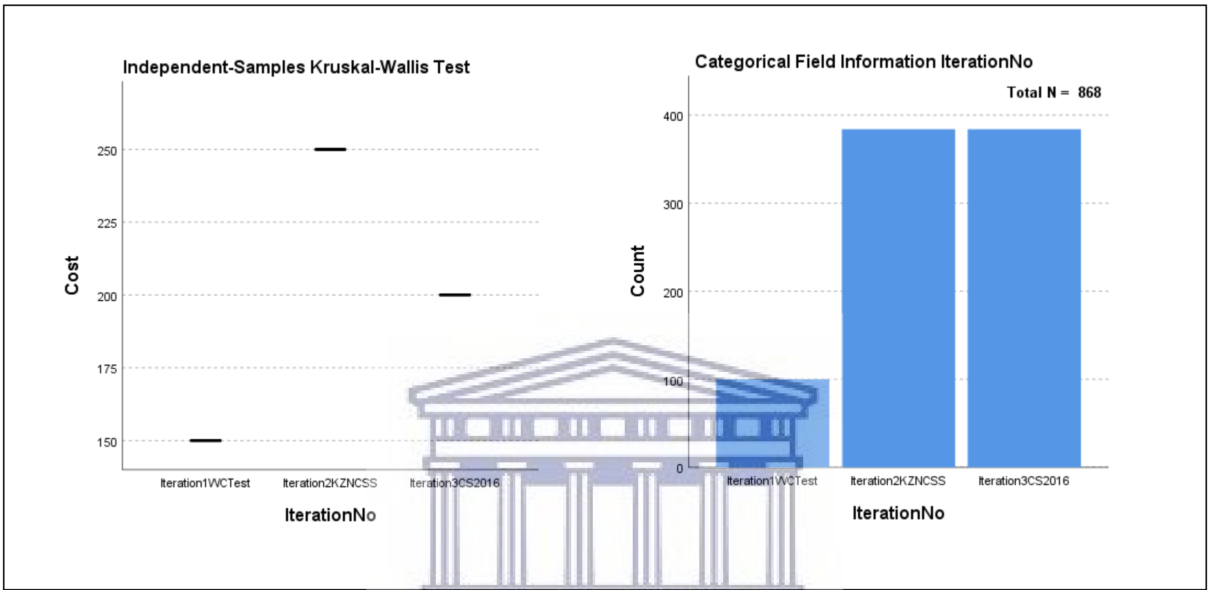


Figure Q.7: Iteration [1,2,3] Cost and sample sizes per iteration

The boxplot (Figure Q.7) compares the medians and spread of the data by iteration group. The values contain one constant value per iteration in rands.

Table Q.7: Iteration [1,2,3] Pairwise Comparisons of Iteration No for cost

Pairwise Comparisons of Iteration Nos.

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Iteration1WCTest-Iteration3CS2016	-218.110	28.147	-7.749	.000	.000
Iteration1WCTest-Iteration2KZNCSS	-237.059	28.147	-8.422	.000	.000
Iteration3CS2016-Iteration2KZNCSS	18.949	18.094	1.047	.295	.885

Each row tests the null hypothesis that Sample 1 and Sample 2 distributions are the same.

Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

The Bonferroni post hoc test indicated significant values after adjustments were made between [Iteration 1 and Iteration 2] and between [Iteration 1 and Iteration 3]; however, there is no statistical significance between [Iteration 3 and Iteration 2] (see Table Q.7).

Q3. KWt – Quantitative analysis across all three iterations [common household availability or perception measurements]

The ordinal and nominal measurements across the three iterations were analysed using the KWt. These measurements were standard questions posed to respondents, which included the same coding categories (except for the difference in variables names listed in Table Q.8 below):

Table Q.8: Iteration [1,2,3] 15 Common Household Measurements

Common Household Variables		
Iteration1-WcTest[100]	**Iteration2-KZN CSS [384]**	**Iteration3-CS2016[384]**
Perception Questions	**Availability**	**Perception Questions**
H01_TypeMainDwell	***H1_dwelling***	***MainDwellType***
H03_SubsidisedDwell	***H2_rdp***	***SubsDwell***
H02_TenureStat	***H3_tenure***	***TenureStat***
H07a_DrinkWaterSource	***H4_main_water***	***WaterSource***
H08a_DistancetoDrinkWater	***H5_dist_water***	***DistanceWater***
H09_WaterSupply	***H6_supplier***	***WaterSupplier***
H16_Toilet	***H7_toilet_facility***	***Toilet***
H17_ToiletLocation	***H8_location***	***ToiletLocation***
H19_ToiletShared	***H9_shared***	***ToiletShared***
H28_RefuseDisposal	***H10_disposal***	***Refuse***
H24_ElectrSupply	***H12_supply***	***ElectrSupplier***
H27a_EnergyCook	***H13_cook_A+ H13_cookB***	***EnergyCook***
H27b_EnergyLight	***H13_light_A+ H13_light_B***	***EnergyLight***
H27c_EnergyWaterHeat	***H13_wheat_A + H13_wheat_B***	***EnergyWaterHeat***
H27d_EnergySpaceHeat	***H13_sheat_A+ H13_sheat_B***	***EnergySpaceHeat***

Two of the fifteen measurements indicated that the null hypothesis was retained; whilst the remaining thirteen were rejected at a significance level (0.05) between the three iterations (see Table Q.9):

Table Q.9: Iteration [1,2,3] Kruskal-Wallis Test Hypothesis Summaries of all Common Household Measurements

	Null Hypothesis	Test	Sig.	Decision
(**8.3.2.1**)	The distribution of **Main Dwell Type** is the same across categories of Iteration No.	Independent-Samples Kruskal-Wallis Tests	.009*	**Reject** the null hypothesis.
(**8.3.2.2**)	The distribution of **Subs Dwell** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.3**)	The distribution of **Tenure Stat** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.4**)	The distribution of **Water Source** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.5**)	The distribution of **DistanceWater** is the same across categories of Iteration No.		.053	**Retain** the null hypothesis.
(**8.3.2.6**)	The distribution of **Water Supplier** is the same across categories of Iteration No.		.007	**Reject** the null hypothesis.
(**8.3.2.7**)	The distribution of **Toilet** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.8**)	The distribution of **Toilet Location** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.9**)	The distribution of **Toilet Shared** is the same across categories of Iteration No.		.534	**Retain** the null hypothesis.
(**8.3.2.10**)	The distribution of **Refuse** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.11**)	The distribution of **Electr. Supplier** is the same across categories of Iteration No.		.005	**Reject** the null hypothesis.
(**8.3.2.12**)	The distribution of **Energy Cook** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.13**)	The distribution of **Energy Light** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.14**)	The distribution of **Energy Water Heat** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.
(**8.3.2.15**)	The distribution of **Energy Space Heat** is the same across categories of Iteration No.		.000*	**Reject** the null hypothesis.

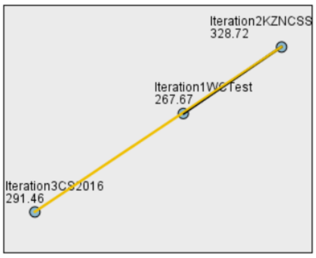
Asymptotic significances are displayed. The significance level is 0.050.

Q3.1 Main Dwelling Type:

For the measurement ***"Which of the following describes the MAIN dwelling that this household currently lives in?"***

From the total N=614 responses, 61,9% indicated that they currently live in a formal dwelling/house or brick/concrete block structure on a separate stand or yard or on a farm, whereas 19,1% indicated they reside in a Traditional dwelling/hut/structure made of traditional material. The remaining option accounted for the other responses. Figure Q.8 indicates that Iteration 3 and Iteration 2 are statistically significant (Reject null hypothesis) in terms of the pairwise comparison analysis.

Pairwise Comparisons of IterationNo

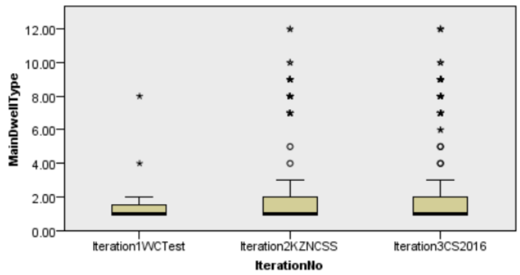


Each node shows the sample average rank of IterationNo.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Iteration1WCTest-Iteration3CS2016	-23.795	45.316	-.525	.600	1.000
Iteration1WCTest-Iteration2KZNCSS	-61.048	45.485	-1.342	.180	.539
Iteration3CS2016-Iteration2KZNCSS	37.253	12.628	2.950	.003	.010

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

Independent-Samples Kruskal-Wallis Test



Total N	614
Test Statistic	9.519
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.009

1. The test statistic is adjusted for ties.

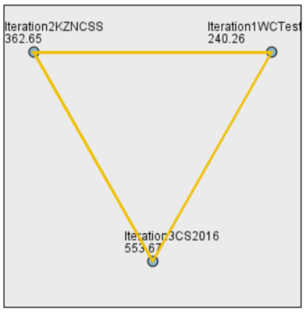
Figure Q.8: Iteration [1,2,3] Main Dwelling Type

Q3.2 Government Subsidised Dwelling:

For the measurement, ***"Is the MAIN dwelling that the household currently lives in an RDP, or a government-subsidised dwelling?"***

From the total N=866 responses, 42,7% did not respond; whereas 14,5% indicated option 1 (Yes), 42,0% indicated option 2 (No); and only 0,7% indicated option 3 (Do not know). Figure Q.9 indicates that all three iterations are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Pairwise Comparisons of IterationNo

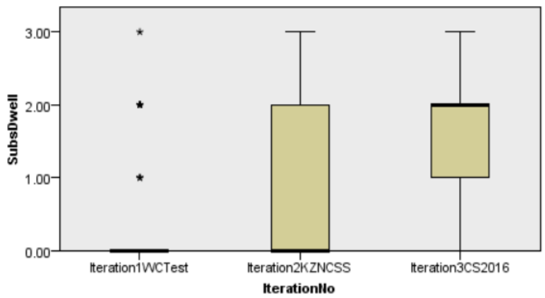


Each node shows the sample average rank of IterationNo.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Iteration1WCTest-Iteration2KZNCSS	-122.393	26.018	-4.704	.000	.000
Iteration1WCTest-Iteration3CS2016	-313.414	26.018	-12.046	.000	.000
Iteration2KZNCSS-Iteration3CS2016	-191.021	16.591	-11.514	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

Independent-Samples Kruskal-Wallis Test



Total N	866
Test Statistic	210.645
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

1. The test statistic is adjusted for ties.

Figure Q.9: Iteration [1,2,3] Government Subsidised Dwelling

Q3.3 Tenure Status:

For the measurement ***"What is the tenure status of the MAIN dwelling that this household currently occupies?"***

From the total N=614 responses, 55,9% indicated that they Owned and fully paid off the dwelling they currently occupy whereas 18,4% indicated they are occupant rent-free. The remaining five categories constitute the difference. The KWt indicates that the two pairs [Iteration1 – Iteration2] and [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.4 Water Source:

For the measurement ***"What is the household's MAIN source of water for drinking?"***

From the total N=613 responses, 30,0% indicated that they have Piped (tap) water inside the dwelling/house whereas 29,5% indicated they have Piped (tap) water inside the yard. The remaining 11 categories constitute the difference. The KWt indicates that the two pairs [Iteration1 – Iteration2] and [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.5 Distance to Water Source:

For the measurement, ***"How far is the main source of water for drinking from the dwelling or yard?"***

From the total N=231 responses, 50,6% indicated that they have access to drinking water Less than 200 metres whereas 25,5% indicated that it is within the range 201-500 metres. The remaining 11 categories constitute the difference. Iteration 1 had no responses. The KWt indicates the null hypothesis should be retained (not statistically significant).

Q3.6 Main source of drinking water:

For the measurement, ***"Is the household's main source of drinking water supplied by...?***

From the total N=610 responses, 73,8% indicated that they receive water from the municipality whereas the remaining 4 categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.7 Main type of toilet facility:

For the measurement, ***"What is the MAIN type of toilet facility used by this household?"***

From the total N=611 responses, 39,9% indicated that they have a flush toilet connected to a public sewerage system; whereas the remaining 9 categories constitute the difference. The KWt indicates that the two pairs [Iteration1 – Iteration2] and [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pair-wise comparison analysis.

Q3.8 Toilet location:

For the measurement, ***"Is the MAIN toilet facility which the household has access to in the dwelling, in the yard, or outside the yard?"***

From the total N=585 responses, 61,7% indicated that they have a toilet in the dwelling/house whereas the remaining 2 categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.9 Shared Toilet:

For the measurement ***"Is the MAIN toilet facility shared with other households?"*** From the total N=586 responses, 72,9% indicated that they have a main toilet facility that is shared with other households located in the yard, whereas 27,0% indicated in the dwelling/house. Figure Q.10 indicates the null hypothesis should be retained (not statistically significant).

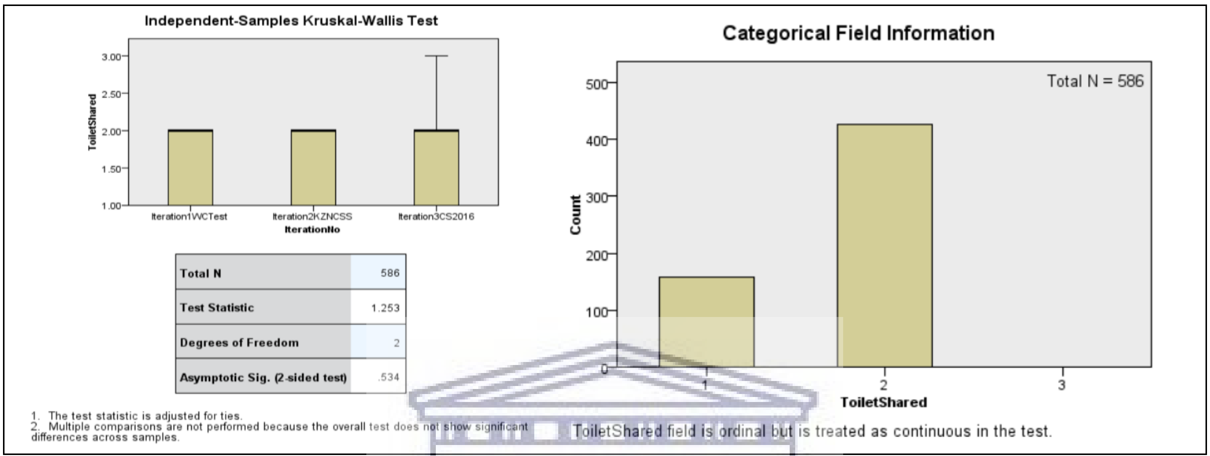


Figure Q.10: Iteration [1,2,3] Shared toilet

Q3.10 Refuse:

For the measurement, ***"How is the refuse or rubbish of this household MAINLY collected or removed?"***

From the total N=609 responses, 45,2% indicated that they use their own refuse dump whereas the remaining 5 categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.11 Electricity Supply:

For the measurement ***"Is this household's electricity supplied by...?"***

From the total N=508 responses, 57,7% indicated that their supply of electricity is from Eskom-Prepaid meters whilst 31,7% indicated their supply is from the Municipality-Prepaid. The remaining four categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.12 Main source of energy used for cooking:

For the measurement, ***“What is this household's MAIN source of energy for cooking?”***

From the total N=612 responses, 72,2% indicated that they used ELECTRICITY FROM MAINS for cooking purposes, whilst the remaining 8 categories constituted the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.13 Main source of energy used for lighting:

For the measurement ***“What is this household's MAIN source of energy for lighting?”***

From the total N=611 responses, 84,5% indicated that they used the ELECTRICITY FROM MAINS for lighting purposes, whilst the remaining 6 categories constitute the difference. The KWt indicates that all three pairs [Iteration3 – Iteration2], [Iteration3 – Iteration1] and [Iteration2 – Iteration1] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.14 Main source of energy used for heating water:

For the measurement ***“What is this household's MAIN source of energy for water heating?”***

From the total N=612 responses, 71,4% indicated that they used the ELECTRICITY FROM MAINS for heating water purposes, whilst the remaining nine categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

Q3.15 Main source of energy used for space heating:

For the measurement ***“What is this household's MAIN source of energy for space heating?”***

From the total N=612 responses, 50,8% indicated that they used the ELECTRICITY FROM MAINS for heating water purposes, whilst the remaining nine categories constitute the difference. The KWt indicates that the pair [Iteration3 – Iteration2] are statistically significant (i.e. reject null hypothesis) in terms of the pairwise comparison analysis.

APPENDIX R: PROOF READING AND EDITING CERTIFICATION

Language Quality Assurance Practitioner
(Professional language editor)

Dr PJS Goldstone
22 Invicta Ave.
Berea,
Durban.

South Africa

Cell: 073-196-0087

pat@pemail.co.za

26th February 2019

TO WHOM IT MAY CONCERN

We hereby certify that we have language-edited the doctoral thesis of Mr Mogomat Hahier Hattas entitled: A QUALITY-ASSURANCE FRAMEWORK FOR DIGITAL HOUSEHOLD SURVEY PROCESSES IN SOUTH AFRICA.

We are satisfied that, provided the changes we have made are effected to the text, the language is of an acceptable standard, and is fit for publication.

Kate Goldstone	**Dr Patrick Goldstone**
BA (Rhodes)	BSc (Stell.)
SATI No: 1000168	DEd (UPE)
UPE Language Practitioner (1975-2004)	
NMMU Language Practitioner (2005)	

Language Quality Assurance – Certification Statement

APPENDIX S: STATISTICIAN GENERAL (HEAD OF DEPARTMENT) APPROVAL GRANTED FOR DATA USAGE

STATS SA
STATISTICS SOUTH AFRICA

19 October 2018

Mr Mahier Hattas
Director: Field Operations
Western Cape Provincial Office
Cape Town

Dear Mahier

With regard to your letter requesting access to the Kwa-Zulu Natal Citizen Satisfaction Survey 2015 (KZN CSS 2015) and Community Survey 2016 (CS 2016) files to facilitate your PhD Information Systems at the University of the Western Cape, I am happy to inform you that after much review and discussion with the relevant internal stakeholders, I have decided to grant your request.

Attached you will find the official Confidentiality Undertaking form that outlines the framework and legal conditions that you must adhere to when working with the KZN CSS 2015 and CS 2016 unit records. As an employee of Stats SA, you must be well aware of the strict confidentiality that accompanies such datasets, especially the KZN CSS 2015 and CS 2016 data. Therefore, strict security protocols will be put in place to ensure the integrity of the data you access. These protocols will include the use of the data within the Western Cape provincial office (Stats SA's provincial office) or using a Stat SA's barcoded laptop under the supervision of Population Statistics staff. To maintain the confidentiality of individual records, all individual records names, surnames, contact details and sample location information may not be divulged.

Stats SA reserves the right to amend and/or expand on these security protocols when deemed necessary to ensure compliance with the Statistics Act (Act No. 6 of 1999).

Dr Christine Khoza (Chief Director for Population Statistics) will be the coordinator of this process on Stats SA's side and will contact you to make the necessary arrangements for when you can begin the process of working with the unit records.

Yours sincerely

Risenga Maluleke
Statistician-General

Private Bag X44, Pretoria, 0001, South Africa
ISIbalo House, Koch Street, Salvokop, Pretoria, 0002 T +27 12 310 8911 F +27 12 310 8500 www.statssa.gov.za
info@statssa.gov.za