Ensuring High Quality Public Safety Data in Participatory Crowdsourcing used as a Smart City Initiative

By

Bhaveer Bhana

Ensuring High Quality Public Safety Data in Participatory Crowdsourcing used as a Smart City Initiative

by

Bhaveer Bhana

200803573

Dissertation

submitted in fulfilment of the requirements for the degree

Masters of Commerce

in

Information Systems

in the

Faculty of Management and Commerce

of the

University of Fort Hare

Supervisor: **Prof. Stephen Flowerday**

December 2013

Abstract

The increase in urbanisation is making the management of city resources a difficult task. Data collected through observations of the city surroundings can be used to improve decision-making in terms of manage city resources. However, the data collected must be of quality in order to ensure that effective and efficient decisions are made. This study is focused on improving emergency and non-emergency services (city resources) by using Participatory Crowdsourcing as a data collection method (collect public safety data) utilising voice technology in the form of an advanced IVR system known as the Spoken Web.

The study illustrates how Participatory Crowdsourcing can be used as a Smart City initiative by illustrating what is required to contribute to the Smart City, and developing a roadmap in the form of a model to assist decision-making when selecting the optimal Crowdsourcing initiative. A Public Safety Data Quality criteria was also developed to assess and identify the problems affecting Data Quality.

This study is guided by the Design Science methodology and utilises two driving theories: the characteristics of a Smart City, and Wang and Strong's (1996) Data Quality Framework. Five Critical Success Factors were developed to ensure high quality public safety data is collected through Participatory Crowdsourcing utilising voice technologies. These Critical Success Factors include: Relevant Public Safety Data, Public Safety Reporting Instructions, Public Safety Data Interpretation and Presentation Format, Public Safety Data Integrity and Security, and Simple Participatory Crowdsourcing System Setup.

Keywords: Data Quality, Public Safety, Participatory Crowdsourcing, Smart City, IVR, Spoken Web.

Declaration

I, Mr Bhaveer Bhana, hereby declare that:

- The work in this dissertation is my own work.
- This dissertation has not previously been submitted in full or partial fulfilment of the requirements for an equivalent or higher qualification at any other recognised educational institution.
- I am fully aware of the University of Fort Hare‘s policy on plagiarism and I have taken every precaution to comply with the regulations.
- I am fully aware of the University of Fort Hare‘s policy on research ethics and I have taken every precaution to comply with the regulations. Ethical clearance certificate number: **FLO011SCIL01**

Name: .....

Signature: .....

Date: .....

Acknowledgements

I would firstly like to thank my supervisor, Professor Stephen Flowerday, for all his advice, guidance and motivation throughout the progress of this study. I would like to express special thanks to him for involving me in the CSI Participatory Crowdsourcing project, allowing me to gain added insight, knowledge and experience.

I would like to thank IBM for their support and input throughout my involvement in the CSI Participatory Crowdsourcing project.

I would like to express my appreciation to all members of the expert review group for taking time out of their busy schedule to participate in my study and providing me with valuable feedback in a timely manner.

I am grateful to the proof reader for reviewing my writing skills and providing me with helpful and informative feedback.

I would like to thank all my friends and family for their support, encouragement and patience through my Masters degree.

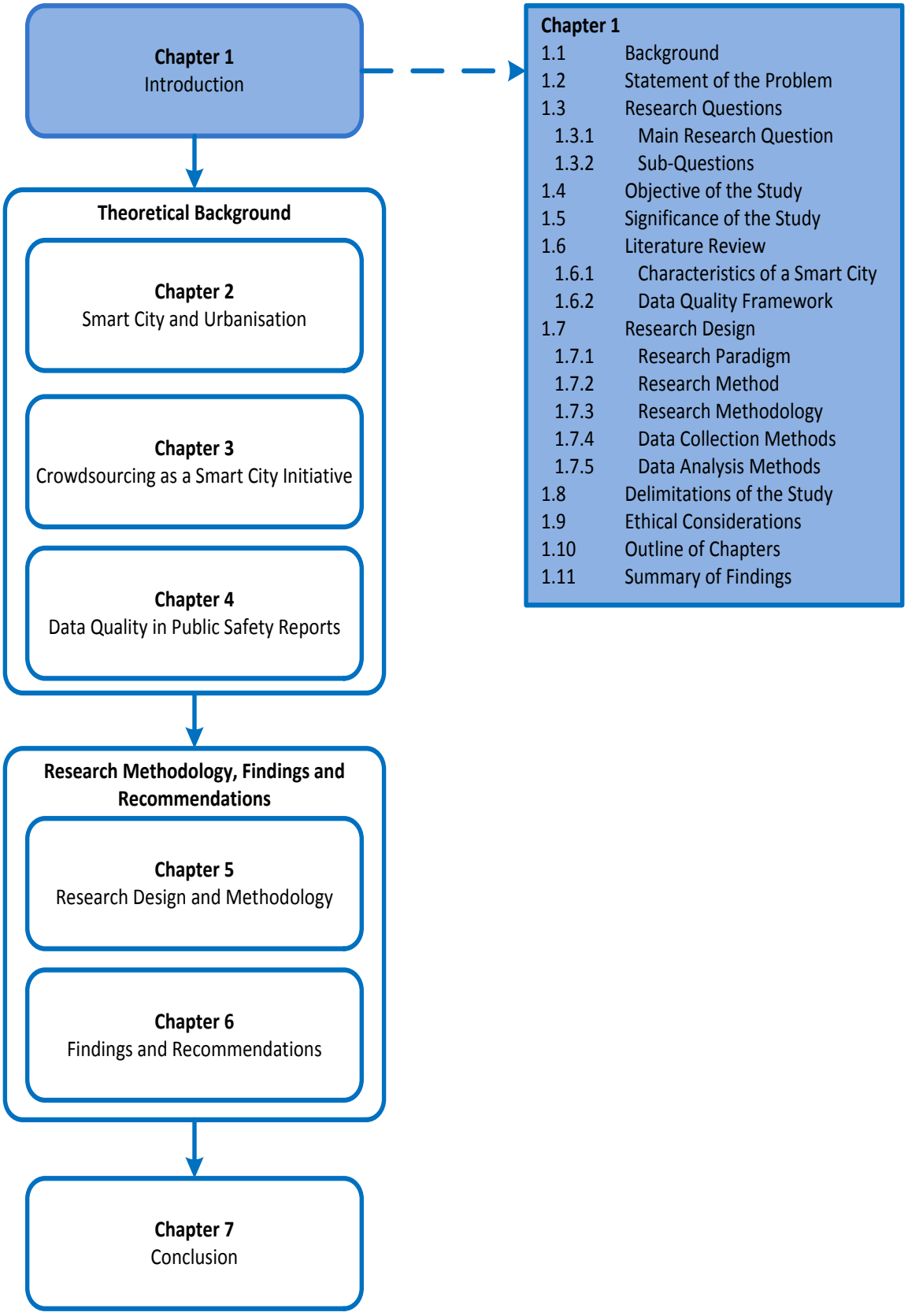
I would like also like to thank my Gods for helping me cope with all the stress and giving me strength to complete my dissertation to the best of my ability.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction



1.1 Background

Local Government accepts the responsibility of maintaining the city's infrastructure, as well as for providing a safe living environment (Fuzile, 2011). Problems which affect these responsibilities occur within a city on a daily basis, such as pollution, public safety, sanitation, electricity outages and road maintenance. However, many of these problems are not brought to the attention of local Governments. Crowdsourcing is an initiative which may improve local Government's awareness of certain problems occurring in the city, by allowing citizens to report on these problems.

Crowdsourcing is a term which refers to the collection of large volumes of data, or reports on certain events by utilising the geographical dispersion of people (Väätäjä, Sirkkunen, & Salo, 2011). Data can be collected and/or reported through the use of applications, using images collected from a camera phone; by voice, through a phone call; and text, through Instant Messaging, e-mail, or social network (Christin, Reinhardt, Kanhere & Hollick, 2011). Two types of sourcing exist, namely, participatory sourcing and opportunistic sourcing. Participatory sourcing requires a high level of human involvement in the task of collecting data and/or information, as opposed to opportunistic sourcing which focuses more on the mobile device, securing minimum user involvement throughout the collection process (Ganti, Ye & Lei, 2011). Note that when this study mentions crowdsourcing, it refers to all types of crowdsourcing in general, both participatory sourcing and opportunistic sourcing. The aim of crowdsourcing is to implement a project where random or selected participants use an appropriate data capturing device to collect data or report on an event or provide data.

This study focuses on a specific participatory crowdsourcing project (the CSI participatory crowdsourcing project) which is currently running in East London, South Africa. The CSI (Crowdsourcing Safety Initiative) participatory crowdsourcing project is run by the University of Fort Hare, East London, South Africa and IBM (International Business Machines), New Delhi, India. The primary objective of the project is to provide citizens with the opportunity to create awareness to the local Government about city issues. In general, participants can report issues relating to transportation, pollution, public safety, electricity issues and health services. However, for the purpose of the CSI participatory crowdsourcing project, the focus is on public safety issues. An advanced IVR (Interactive Voice Response) system was used to collect these public safety reports.

IVR systems are an interactive telephonic interface (Greeff, Coetzee & Pistorius, 2008) where pre-coded messages are provided to the user, who supplies an audio input to the system (Dawes & Rowley, 1998). The Spoken Web is an advanced IVR system which serves as a substitute to the World Wide Web used to supply information to people through their telephones (Agarwal, Kumar, Nanavati & Rajput, 2011). The sophisticated IVR system used in the CSI participatory crowdsourcing project allows for the transmission of voice messages from participants' mobile or landline phones to a database used independently to record audio data. The data provided by the participants is a means to bring attention to local government about reported events or problems, and overall, contributing to the Smart City.

The Smart City is a concept in which a city utilizes the use of data and communication technology in order to improve the city's socio-economic development and quality of life (Schaffers *et al*., 2011). The CSI participatory crowdsourcing project utilises information and communication technology in the form of an advanced IVR system (Spoken Web), which provides the public with the opportunity to contribute to the city's development and create a safer environment, by reporting public safety issues. The collaborative adoption of an advanced IVR system and crowdsourcing creates an opportunity for this project to contribute to the Smart City. However, it should be noted that reported reactions are dependent on the data provided, as one cannot effectively respond to a report unless the data provided contains a certain level of quality.

1.2 Statement of the Problem

Urban areas offer sophisticated education facilities, job opportunities and a higher quality of life attracting rural dwellers to relocate to urban environments. This has caused urbanisation to increase over time and is expected to do so in the future. Although urbanisation is a positive phenomenon, it is attached to a number of problems, specifically related to the scarcity of resources required to run a city, for example, water and electricity, infrastructure construction and various Government services (police, ambulance services and fire fighters). This is not because city resources are scarce but because they are not being managed effectively and efficiently (Exner, Zeile & Streich, 2011). The collection of data on a city's environment will assist decision-making on resource management by identifying patterns in the data. Crowdsourcing is seen as an effective method of collecting this type of data as it takes advantage of citizen dispersion allowing a broad range of data to be collected in a short span of time. As mentioned above, this study will

focus on a specific participatory crowdsourcing project (CSI participatory crowdsourcing project) which is focused on the collection of public safety data from citizens. This data will be utilised by the city's emergency and non-emergency units to improve service delivery. However, the success of the CSI participatory crowdsourcing project relies heavily on the participants' abilities to provide high quality data (Burke *et al*., 2006).

Participants may provide inaccurate or misleading data (Yang, Zhang & Roe, 2011) which will negatively affect the quality of data. For example, if ambiguous or incomplete data is provided, an ineffective or incorrect action will be taken. Srivastava, Abdelzaher and Szymanski (2012) support this statement by claiming that before one can effectively act on the data collected, a certain level of quality is required before a logical and appropriate action/s can be taken. When low quality data is collected, resolving an issue requested by the caller is not possible (Solomon & Uchida, 2003). Therefore, the problem is that attending to a caller's request is not possible if the caller provides low quality data. When low quality data is collected, it could endanger the public and response team, lead to respondents appearing in the incorrect location or wasting time searching for the correct location, applying a diminutive work force to a problem requiring a large workforce, and appearing once the problem has been resolved or grown to the extent that the current solution becomes ineffective. The reviewing of low quality data will also lead to inaccurate identification of patterns which can be used to improve the distribution of emergency and non-emergency units.

1.3 Research Questions

1.3.1 Main Research Question

- **What factors are required to ensure high quality public safety data in participatory crowdsourcing used as a Smart City initiative?**

To answer the main research question, the following sub-questions are addressed.

1.3.2 Sub-Questions

- **How can the Smart City concept be used to reduce problems associated with urbanisation?**

  The first sub-question educates the reader on the Smart City concept's influence on urbanisation, by firstly identifying the problems created from the increase in urbanisation. The various contribution areas of the Smart City are then discussed

after explaining the requirements for a successful Smart City initiative to contribute to urbanisation.

- **How can participatory crowdsourcing be used as a Smart City initiative to reduce public safety issues?**

  The second sub-question is intended to inform the reader about how crowdsourcing can contribute to resolving problems associated with urbanisation. This is about familiarising the reader of the context in which data quality is applied to this study and emphasising the importance of data quality within such a context. Once context familiarisation is accomplished, the next sub-question is introduced.

- **What factors constitute data quality within a public safety participatory crowdsourcing context?**

  The last sub-question is aimed at constructing a Public Safety Data Quality criteria to identify any weaknesses in the quality of data provided by citizens. This criteria assesses the primary data collected from citizens (public safety reports) to identify any weaknesses in the reports, which would assist the development of Critical Success Factors required to allow for high quality public safety data in a crowdsourcing initiative, using an advanced IVR system.

1.4 Objective of the Study

Critical Success Factors are established to allow for high quality public safety data in a participatory crowdsourcing initiative, using a sophisticated IVR system. This monolithic objective is divided into smaller sub-objectives to achieve this study's main objective; the sub-objectives included:

1. The establishment of Smart City contributions to problems created from urbanisation, which will be addressed in the first sub-question.
2. Determine if crowdsourcing can be used as a Smart City initiative to resolve problems associated with urbanisation, as highlighted in the second sub-question.
3. Determine the assessment of data quality attributes existence in public safety data in response to the third sub-question.

Addressing the three sub-objectives allows for the development of Critical Success Factors which can be used to ensure that factors affecting public safety data quality are mitigated. The conclusion of the expressed objective will introduce the significance of the study.

1.5 Significance of the Study

This study discovered a means of mitigating data quality problems experienced in participatory crowdsourcing by focusing on a specific crowdsourcing project (CSI participatory crowdsourcing project) related to the reporting of public safety issues in East London. This participatory crowdsourcing project aims at providing an environment for the public to communicate safety issues to the Buffalo City Municipality (responsible for East London) that is data provided by the participants related to public safety issues. The collection of high quality public safety data from reports, through participatory crowdsourcing, will result in an increased ability to resolve the issues raised by the public (Solomon & Uchida, 2003). The project's success will provide an additional technique for collecting data from the public in an effective and efficient manner. This technique will also be applicable to other facets of Government and commercial industries, providing a significant contribution to the Smart City.

1.6 Literature Review

This section will provide a theoretical base for the study. Firstly the characteristics of a Smart City will be discussed, followed by the Data Quality Framework. These theories will be used to answer the three sub-questions mentioned in section 1.3.2. The characteristics of a Smart City will be used to determine how the Smart City (first sub-question) and crowdsourcing (second sub-question) can be used to contribute to the problems brought forward by urbanisation. The Data Quality Framework will assist the construction of a Public Safety Data Quality criteria which will be used to identify any weakness in the quality of data.

1.6.1 Characteristics of a Smart City

The CSI participatory crowdsourcing project is aimed at facilitating the reporting of public safety issues through the use of information and communication technology, and by doing so, contribute to the Smart City initiative. Giffinger *et al*. (2007) have used the characteristics of a Smart City to rank medium-sized cities in Europe according to the smartness of the city. These Smart City characteristics are constructed by considering a number of theories on urban development and growth. The characteristics of a Smart City consist of thirty one factors and seventy four indicators (Giffinger *et al*., 2007). The seventy four indicators are categorised into thirty one factors, which are organised into six characteristics. The characteristics of a Smart City include: Smart Economy, Smart

People, Smart Governance, Smart Mobility, Smart Environment and Smart Living (Giffinger *et al*., 2007).

The consideration of these characteristics will assist the researcher in determining how Smart City initiatives can be used to address the problems related to urbanisation, as well as how crowdsourcing can be used as a Smart City initiative. However, the success of the CSI participatory crowdsourcing project relies on the quality of data collected from the public. The Data Quality Framework was found to be the most appropriate theoretical background in determining the quality of data in the CSI participatory crowdsourcing project.

1.6.2 Data Quality Framework

The Data Quality Framework organises data quality attributes into four categories namely: 1. Intrinsic Data Quality, 2. Contextual Data Quality, 3. Representational Data Quality and 4. Accessibility Data Quality. Note that the data quality attributes organised into the four categories do not include all data quality attributes but only those deemed important to a data consumer (Wang & Strong, 1996). A data consumer is described as a person or organisation that accesses or uses the data (Wang & Strong, 1996); therefore, the data consumer within the CSI participatory crowdsourcing project would be the entity responsible for acting on the data collected (in this case, emergency and non-emergency services). The Data Quality Framework is used to construct a Public Safety Data Quality criteria to assess the presence of the data quality attributes within the data collected from the citizens through the CSI participatory crowdsourcing project.

1.7 Research Design

The research methodology section will be divided into five sub-sections namely: research paradigm, research method, research design, data collection methods and data analysis methods.

1.7.1 Research Paradigm

Oates (2006) describes the clear difference between two research paradigms, positivism and interpretivism. Taking into consideration the separated characteristics of both positivism and interpretivism, conducting this study under the strict characteristics of positivism or interpretivism is extremely unlikely. Collis and Hussey (2009) express a more realistic alternative by providing research paradigms which combine characteristics of both the positivism and the interpretivism approach. This allows one to take a more

extensive view of the research context. Figure 1-1 illustrates the research paradigms mentioned by Collis and Hussey (2009).

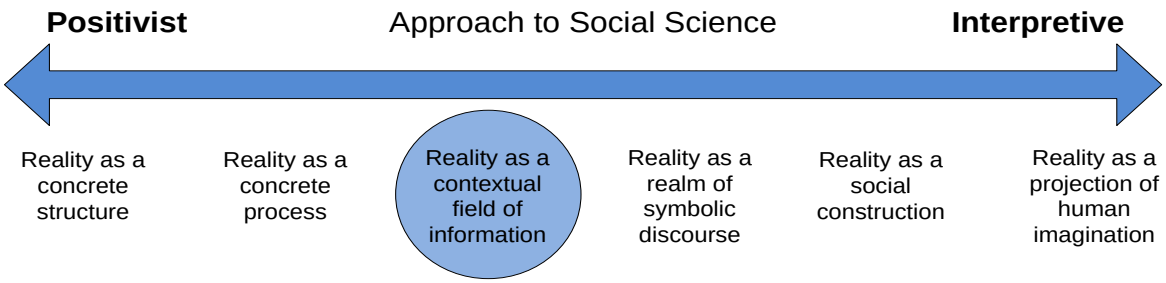


Figure 1-1: Continuum of Core Ontological Assumptions (Collis & Hussey, 2009)

This research study determines the research subject's cognitive interpretation of situations and information regarding those situations. Theories and experts in the relevant fields are used to verify all logical assumptions. The reflection of cognitive interpretations and verification from theories and experts emphasises that the *reality as a contextual field of information* research paradigm will be most suitable for this study. The research method will follow the research paradigm.

1.7.2 Research Method

This study has adopted a mixed method approach which utilises qualitative research and quantitative research. This was considered the most optimal research method when considering the process required when developing the Critical Success Factors for public safety data quality. The qualitative research approaches used in this study include case studies on the Smart City and crowdsourcing initiatives, theories on data quality, and an expert review (see Design Science methodology) to assess the Public Safety Data Quality criteria (see Table 5-4) and the proposed Critical Success Factors. The quantitative research approach involves the assessment of the public safety reports collected from the CSI participatory crowdsourcing project. The Public Safety Data Quality criteria developed from relevant theories and literature on data quality, is used to assess 100 public safety reports in order to identify the areas jeopardising the quality of data. The next section will discuss the research methodology adopted for this study.

1.7.3 Research Methodology

The Design Science guidelines are followed throughout this study, leading towards the development of the Critical Success Factors. Design Science is a problem solving paradigm which "... seeks to create innovations that define the ideas, practices, technical

capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished" (Hevner, March, Park & Ram, 2004, p. 76). The Design Science paradigm consists of seven guidelines which must be considered to effectively utilise this research methodology (Hevner, March, Park & Ram, 2004). The seven guidelines of Design Science are displayed in Table 1-1, which includes a description of each guideline. The application of the guidelines to the research study will follow the presentation of Table 1-1.

Table 1-1: Design Science Research Guidelines (Hevner, March, Park & Ram, 2004)

Design Science Guidelines	
Guideline	**Description**
Guideline 1: Design as an Artefact	Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.
Guideline 6: Design as a Search Process	The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Note that the Design Science guidelines presented by Hevner, March, Park and Ram (2004) have been rearranged for the purpose of this study.

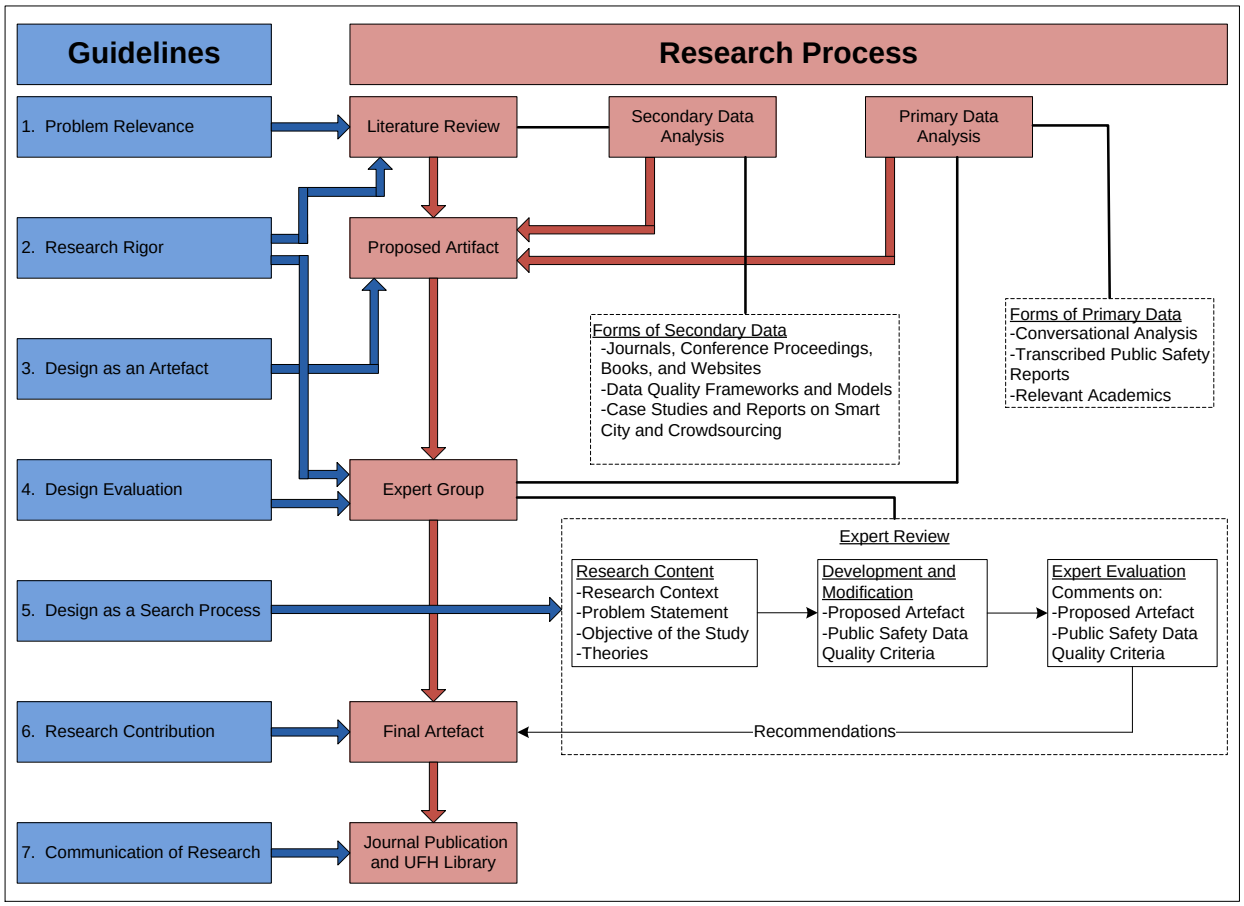


Figure 1-2: Applied Design Science Guidelines

- **Guideline 1: Problem Relevance**

  The problem identified in this study refers to the review of low quality data, which would result in ineffective decisions on the management of city resources. This study will focus specifically on the data quality of public safety reports provided by citizens.

- **Guideline 2: Research Rigor**

  Secondary data were used to create and support all logical conclusions towards the development of the Critical Success Factors. Theories are used to guide the research and validate any assumptions. The expert reviews are conducted for the purpose of assessing the Public Safety Data Quality criteria (expert review 1) and the Critical Success Factors (expert review 2).

- **Guideline 3: Design as an Artefact**

  This study intends to produce the Critical Success Factors required during the collection of high quality public safety data through participatory crowdsourcing, utilising voice technologies.

- **Guideline 4: Design Evaluation**

  The Public Safety Data Quality criteria and Critical Success Factors were evaluated by applying expert review (Hsu & Sandford, 2007). The Public Safety Data Quality criteria and the proposed artefact were developed after considering the content of the research. The material was then presented to six experts for evaluation (the same experts were used for both expert reviews), with sufficient and relevant educational background and experience. These experts commented on the presented material and provided feedback in the form of comments and recommendations. These comments and recommendations were considered and the material was appropriately modified.

- **Guideline 5: Design as a Search Process**

  The Public Safety Data Quality criteria and the proposed artefact were constructed from related primary data and secondary data. The comments and recommendations from the expert group allowed both the Public Safety Data Quality criteria and the proposed artefact to be refined.

- **Guideline 6: Research Contributions**

  The design artefact is in the form of Critical Success Factors which address the issue of receiving poor quality public safety data because decisions based on low quality data will provide ineffective results.

- **Guideline 7: Communication of Research**

  This study will provide the management audience with an awareness of the importance of data quality, as well as the way in which data can be assessed to discover factors which produce low data quality. From a technological perspective, an illustration of how high quality data can be achieved through the use of technology will be realised. The findings will be published in academic journals and conferences and made available to the public for future research. The dissertation will also be accessible through the library at the University of Fort Hare. IBM will also be provided with access to this research for future research purposes.

By ensuring that the seven Design Science guidelines were accounted for confidently reinsured that credible conclusions were reached. The next section of the research design refers to data collection methods.

1.7.4 Data Collection Methods

The primary data collection included: conversational analysis, transcribed reports, recommendations from the expert reviews and comments from relevant academic experts.

- Primary Data Collection Methods
    - Data was gathered from continuous meetings with the IBM team and the UFH team regarding CSI participatory crowdsourcing project matters. This form of data is commonly referred to as conversational analysis, and was the phrase used in this study when comments and recommendations were collected from these meetings.
    - The public safety reports from citizens (provided in audio format as voice was used) were transcribed into text. These transcribed responses were used to identify data quality problems.
    - The expert review was used to assess the Public Safety Data Quality criteria and the Critical Success Factors by collecting recommendations from 6 experts with sufficient and relevant educational background and experience.
    - All published material and this dissertation will be reviewed by academic experts in the respected field. The comments provided by them will be taken into consideration.
- Secondary Data Collection Methods
    - This study took advantage of related literature in journals, conference proceedings, books, websites, theories and methodologies.
    - The frameworks and models used related to data quality and case studies and reports referred to Smart City and various crowdsourcing initiatives.

The above section discussed what data were sourced for this study. The next section discusses how this data will be used.

1.7.5 Data Analysis Methods

All data collected was analysed appropriately. This included all relevant data collected through conversational analysis, transcribed reports, the expert review and recommendations from relevant academic experts.

- Primary Data Analysis
    - Conversational analysis is a Social Sciences method of data collection through social interactions, usually including verbal and non-verbal cues

(Goldkuhl, 2003). The conversational analysis based on the meetings with the IBM team was taken into consideration when developing the Critical Success Factors.

  - The transcribed reports were assessed using the Public Safety Data Quality criteria constructed in Chapter 4. A sample of 100 reports was deemed appropriate for analysis.
  - All relevant feedback from the expert review was used to refine the Public Safety Data Quality criteria and the proposed Critical Success Factors.
  - All comments received from academic experts were considered before publication and before the final dissertation was submitted.

- Secondary Data Analysis
  - Secondary data was analysed in an inductive manner to determine whether the logical conclusions agreed or disagreed with supported theories.

This section discussed the research paradigm, research design and the data collection and analysis methods of the study. The next section will provide delimitations of the study.

1.8 Delimitations of the Study

The data collected from the callers were only related to public safety issues occurring in East London. The public safety data referred to issues such as assaults, accidents, thefts and other hazardous issues. The IVR system is restricted to English speaking candidates, as the system cannot recognize other local languages at this stage. The data collected from the callers were in an audio format. At this stage, all the data collected from callers' reports were used strictly for research purposes; therefore the callers still had to report the issues to the authorities responsible for acting on the data. Analytics will be conducted on the public safety data collected; however, it will not be considered for this study. The boundaries of this study have been emphasised through delimitations, which will lead to the ethical considerations of the study.

1.9 Ethical Considerations

Punch (2006) states that it is the author's responsibility to comply with academic integrity and honesty, as well as to respect other people. Punch (2006) continues by listing categories of ethical issues which should be considered by researchers. Informed consent (Punch, 2006) will be obtained by sharing all necessary information about the research to all participants. Participants will be guaranteed anonymity (Punch, 2006) by constructing

the questionnaires in a manner which ensures no personal information is captured. The misuse of results (Punch, 2006) will be mitigated by ensuring information gathered from participants will only be used for the indented and expressed purpose of the research. Participants will also be allowed to refuse participation in the research at anytime. It should also be noted that the information collected from participants will be used for research purposes only, as instructed by IBM's terms and conditions (see Appendix A). All participants will be aware of the use of the information but if necessary it is suggested they report any issues to the relevant authorities first and then report the public safety issue on the IVR system. Additionally, ethical clearance will be obtained before any candidates participate in this research study. All these ethical issues have been considered and the research undertaken will comply with these considerations. The next section provides an outline of the chapters.

1.10 Outline of Chapters

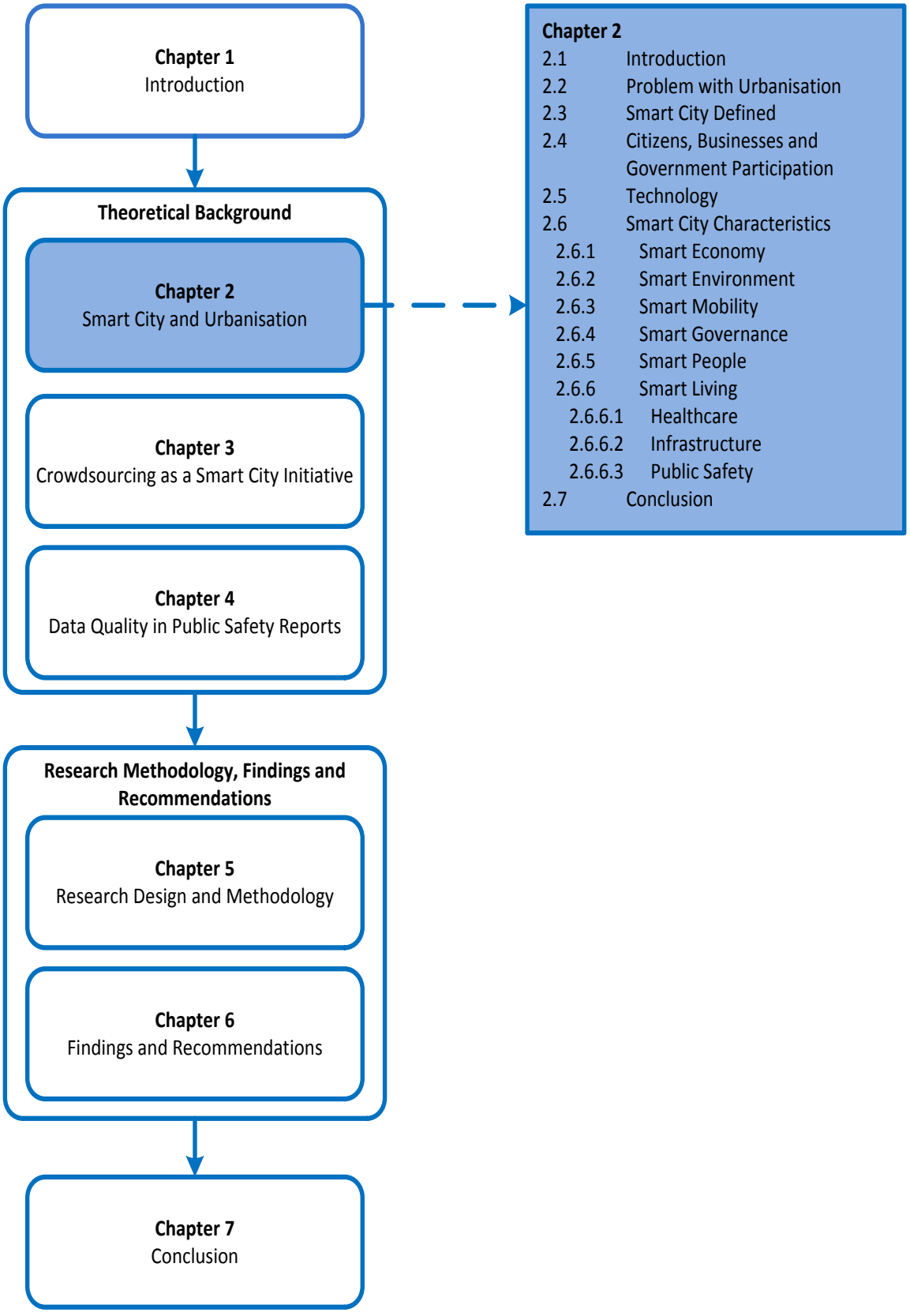
Chapter 1 serves as an introductory chapter and sets out to familiarise the reader with regards to the scope of the study. The chapter includes a background, statement of the problem, objectives of the research, significance of the study, introductory literature review, research methodology, delimitations of the study and ethical considerations. Chapter 2 focuses on the first sub-question mentioned above and explained how the Smart City can contribute to the resolution of problems associated with urbanisation. Chapter 3 addresses the second sub-question which focuses on crowdsourcing and aims at familiarizing the reader of the context in which data quality will apply to this study. Chapter 4 answers the third sub-question by constructing a Public Safety Data Quality criteria to assess the quality of public safety reports. This chapter concludes the secondary research analysis of the study. Chapter 5 discusses this study's research design, methodology, limitations and ethical considerations. Chapter 6 provides the study's findings and recommendations. The final chapter (Chapter 7) provides a conclusion and suggestions for future research. The last section provides a summary of the findings.

1.11 Summary of Findings

This study developed Critical Success Factors to ensure high quality public safety data is collected through the use of participatory crowdsourcing utilising voice technologies (Spoken Web). An expert review was conducted to assess the validity of the proposed Critical Success Factors. In addition, the Critical Success Factors are supported by Wang

and Strong's (1996) Data Quality Framework and the Rational Choice theory. Five Critical Success Factors were developed which include: Relevant Public Safety Data, Public Safety Reporting Instructions, Public Safety Data Interpretation and Presentation Format, Public Safety Data Integrity and Security, and Simple Participatory Crowdsourcing System Setup. Chapter 2 is focused on the Smart City and urbanisation.

Chapter 2 – Smart City and Urbanisation



2.1 Introduction

Worldwide, cities are faced with challenges to improve quality of life while addressing sustainable development (Komninos, Pallot & Schaffers, 2012). These challenges continue to increase and become more complex over time due to increases in urbanisation. Urbanisation means that more people are moving to cities resulting in an increase in city population. The challenges facing all cities around the world is that they cannot keep up with the rate of urbanisation. City resources cannot support all members of the public simply because the ratio of number of citizens to city resources (for example, water and electricity distribution) is low. This is not because city resources are scarce but because they are not managed effectively and efficiently (Exner, Zeile & Streich, 2011). Therefore, improving the management of city resources will assist cities in adapting to increased urbanisation.

The DIKAR (acronym for Data Information Knowledge Action Result) model can be used to explain how one would improve the management of city resources. The DIKAR model, illustrated in Figure 2-1, explains that processed data becomes information and one gains knowledge by interpreting information. Knowledge is then used to decide on a cause of action, which drives a result. Essentially, the DIKAR model explains that data is required to improve decision-making procedures. Therefore, the collection of relevant data is required to improve the management of city resources.

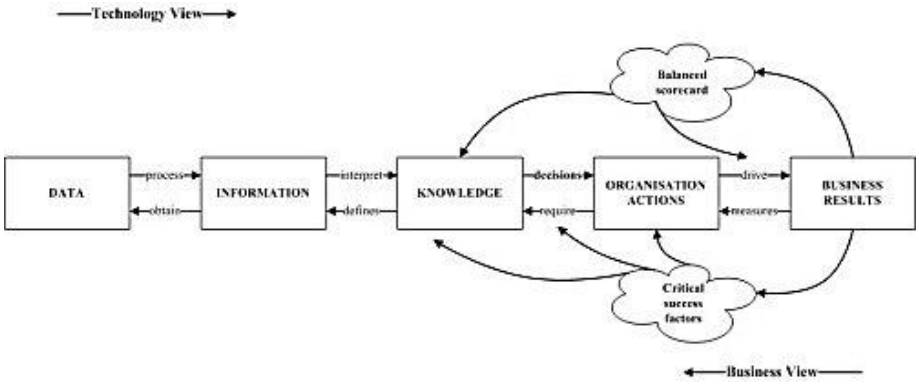


Figure 2-1: DIKAR Model (Venkatraman 1996, in Ward & Peppard, 2002)

This study seeks to determine how a specific project, which focuses on the collection of public safety data, can be used as a Smart City initiative to reduce the problems associated with urbanisation. One needs to firstly determine **how the Smart City concept can be used to reduce problems associated with urbanisation.** Chapter 2 will discuss how the Smart City concept can improve the management of city resources. The various

dimensions of a Smart City and examples of current initiatives contributing to each of these dimensions will be presented. The importance and contribution of the Smart City concept can be realised by firstly understanding the problems resulting from urbanisation.

2.2 Problem with Urbanisation

Cities are constantly competing with each other in an attempt to improve their quality of life while addressing sustainable development. Komninos, Pallot and Schaffers (2012) explain that citizens have certain needs, common to all members of the public, and expect the city to accommodate these needs. These needs include basic services ensuring safety, electricity, healthcare and education. Therefore, one approach to compete with other cities is to improve the quality of such services. Other approaches involve competing by creating employment opportunities and innovation through research (Dodgson & Gann, 2011). Caragliu, Del Bo and Nijkamp (2011) state that urban competition involves investments in physical, social and human capital. Bakıcı, Almirall and Wareham (2012) believe that investments in ICT are a strong approach for competing with other cities. All these suggestions are effective methods for cities to compete with one another. The overall goal of these cities is to create an attractive living environment for citizens. This has resulted in more people moving to urban areas, which continue to increase rapidly over time.

In 1990, 13% of the world's population lived in cities (Dodgson & Gann, 2011); this figure increased to 50% by 2007 – 2010 and is expected to rise by 70% – 75% in 2050 (Bakıcı, Almirall & Wareham, 2012; Nam & Pardo, 2011; Dodgson & Gann, 2011). A timeline presented by Washburn and Sindhu (2010) illustrates the rapid increase in urban population, which is shown in Figure 2-1 below.

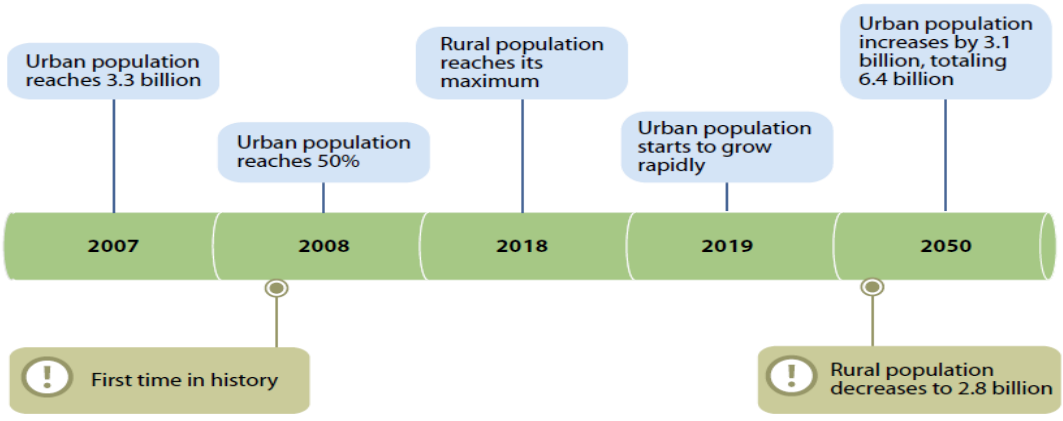


Figure 2-2: Urban Population Increase (Washburn & Sindhu, 2010)

Numerous articles (Bakıcı, Almirall & Wareham, 2012; Nam & Pardo, 2011; Dodgson & Gann, 2011) have different statistical results on urban population increase however; if an approximation is made, Washburn and Sindhu's (2010) diagram can be assumed as credible and can be used to emphasise the continuous increase in urban population over the years. Although a city's growth has numerous benefits for its citizens, it also poses challenges which become more difficult to manage as the city's population increases.

Common challenges include, but not limited to: resource depletion (1), infrastructure deterioration (2), energy shortage (3), environmental issues (4), health concerns (5), unemployment (6) and public safety (7). These challenges will be discussed in greater detail below.

1. Cities inhabit 2% of the earth's surface and consume three quarters of the earth's natural resources (Nam & Pardo, 2011), some of which are non-renewable resources. Natural resources are consumed by cities on a daily basis to maintain certain city operations for citizens, such as ensuring sufficient water and electricity. An increase in population inevitably causes an increase in the consumption of natural resources which poses a problem as many cities are already strained with limited resources. The New York Department of Economic and Social Affairs (2007) predict that consumption of drinking water is expected to increase by 25% by 2030 as a result of urbanisation. Haubensak (2011) suggests that alternative and sustainable methods be found to cope with the demand. Although this may be an effective long term solution, the problem needs to be addressed in this present time. At present, the most appropriate approach to natural resources would be to impose consumption limitations on citizens and improve the management of these resources. Another problem with the population increase is the deterioration of infrastructure.
2. A city's housing, telecommunication and transportation networks require constant, regular maintenance, as a result of constant public use, weather conditions, theft and acts of vandalism. As a city's population increases, the city's infrastructure needs to accommodate this change (Nam & Pardo, 2011). This entails more frequent maintenance activities and the development of additional infrastructure. Infrastructure improvement and enhancements require large investments, time and manpower. Statistics predicting population growth are accessible and should be

considered when planning infrastructure development initiatives. The increasing demand for energy is another concern in cities.

3. In this context, energy consumption refers to electricity and oil use. These two resources are used by all citizens on a daily basis. Energy consumption in South Africa is expected to rise by 44% by 2034 (Debba, et al., 2010). At present, cities are struggling to meet the demand for energy, and as a result this has caused energy prices to inflate in addition to an increase in the importation of these materials. Additionally, large investments in finding alternative solutions for energy generation have been undertaken. Environmental issues resulting from urban population growth is also a problem experienced in all urbanised areas.
4. Environmental concerns in this context relates to air pollution in the city as a result of burning fossil fuels. The burning of fossil fuels has damaging implications to the environment and contributes to global warming, causing droughts, floods and forest fires. Cities are the main cause of global warming, resonating from traffic congestion, import and export activities, manufacturing plants, energy consumption and other minor activities. This is expected to increase as urban population increases (Washburn & Sindhu, 2010). The constant exposure to high levels of air pollution can also cause health problems.
5. Although Haubensak (2011) makes a strong argument that people living in cities have a higher life expectancy rate, it was found that exposure to air pollution (present in all cities) has certain health implications. Air pollution causes damage to lungs, resulting in respiratory problems and lung diseases (Seaton, MacNee, Donaldson, & Godden, 1995). The next problem that will be discussed is the rate of unemployment as population increases.
6. Washburn and Sindhu (2010) explain, as traditional agricultural farming is becoming more automated, jobs are becoming scarcer in this sector as less manpower is required. This has caused many people to seek work in cities as urbanised areas offer employment opportunities and a better quality of life ranging from improved healthcare to education. A city cannot function independently from humans and therefore, employment opportunities will always be present. However, there comes a point when the number of people is higher than the number of job opportunities, causing unemployment. The number of job opportunities is struggling to keep up with the rapid increase in urban population. Unemployment creates additional economic problems such as poverty and an increase in criminal

activity. This is presently a problem in many cities, even in developed countries. The last problem that will be discussed is public safety.

7. Cai and Wang (2008) explain how public safety problems become more serious and complex due to increases in urbanisation. They put emphasis on weak infrastructure construction and increases in safety hazards. Weak infrastructure construction could be a result of three broad possibilities; firstly, substitute resources are used as a result of resource scarcity due to urbanisation. Secondly, rapid urbanisation requires rapid infrastructure construction resulting in decreased focus on quality and more on quantity. Lastly, more people are making use of certain infrastructures which increases the frequency of infrastructure maintenance and repairs. In terms of increases in safety hazards, the majority of safety hazards are created either by people's actions or by people not taking the action required. In addition, the exhaustion of public safety resources will continue to increase from the rise in urbanisation.

The seven challenges discussed above are the major challenges that cities face from an increase in population. Most cities are currently faced with such challenges and as statistics have shown, urban population is expected to rise in the future, making these challenges even more difficult to manage. A concept known as the Smart City is an initiative that aims to exploit the benefits of urbanisation by targeting the common challenges, as discussed above.

2.3 Smart City Defined

A large volume of literature on Smart Cities exists and continues to increase in popularity. Considering the latter, the term Smart City has been used flexibly in previous literature. Nam and Pardo (2011) explain that this idea is a popular concept around the world, and different contexts have created a variety of similar, but different initiatives by simply replacing the word "smart" with an alternative adjective. Examples include: digital city, information city, creative city and knowledge city. It is important to clearly define what is meant by Smart City so to avoid confusion with similar initiatives.

In previous literature the Smart City concept has been variously defined. A common definition used in many articles states that a city is "smart" when "investments in human, social capital, traditional (transportation), modern (ICT-based), infrastructure, fuel, sustainable economic growth and a high quality of life, with a wise management of natural

resources, through participatory government" (Caragliu, Del Bo & Nijkamp, 2011, p. 70). Although Caragliu, Del Bo and Nijkamp's (2011) definition can relate to many Smart City initiatives, many also seek to involve citizens and/or businesses, which are not mentioned in the definition. When defining a Smart City, one should mention the stakeholder groups who participate in the initiatives. Additionally, the definition makes reference to ICT but does not mention the Smart City's focus on Smart Technology such as sensors, cloud computing and other time and special technologies. Caragliu, Del Bo and Nijkamp's (2011) definition also states the intention behind the Smart City concept (sustainable economic growth and improved quality of life) however; the end goal of the Smart City is to make a city more knowledge-intensive by collecting data to improve the management of city services and resources. The last criticism of this widely used definition of a Smart City refers to the areas in which the Smart City is involved. The definition does not specifically mention the contribution areas the Smart City targets. Giffinger *et al*. (2007) identified six characteristics of a Smart City which emphasises the areas influenced by Smart City initiatives. Additionally, a city obtains a Smart City status by running a collection of Smart City initiatives or an individual initiative. Table 2-1 below lists the Smart Cities that have obtained the Smart City status.

Table 2-1: List of Smart Cities (Nam & Pardo, 2011)

Region	Cities
Asia	Bangalore (India); Chongqing (China); Doha (Qatar); Gangnam District, Seoul (Korea); Hong Kong; HwaSeong-DongTan (Korea); Hyderabad (India); Ichikawa (Japan); Jaipur, Rajasthan (India); Jia Ding (China); Kabul (Afghanistan); Mitaka (Japan); Shanghai (China); Seoul (Korea); Singapore; Suwon (Korea); Taipei (Taiwan); Taoyuan County (Taiwan); Tel Aviv (Israel); Tianjin (China); Yokosuka (Japan)
Africa	Cape Town (South Africa); Nelson Mandela Bay (South Africa)
Europe	Besançon (France); Birmingham (UK); Dundee, Scotland (UK); Eindhoven (Netherlands); Glasgow, Scotland (UK); Hammarby Sjostad (Sweden); Issy-les-Moulineaux (France); Karlskrona (Sweden); Malta (Malta); Manchester (UK); Reykjavik (Iceland); Sopron (Hungary); Stockholm (Sweden); Tallinn (Estonia); Sunderland (UK); Trikala (Greece)
North America	US: Albany (New York); Ashland (Oregon); Arlington County (Virginia); Bettendorf (Iowa); Bristol (Virginia); Chattanooga (Tennessee); Cleveland (Ohio); Corpus Christi (Texas); Dakota County (Minnesota); Danville (Virginia); Dublin (Ohio); Florida High Tech Corridor; LaGrange (Georgia); Northeast Ohio; Loma Linda (California); Riverside (California); San Francisco; Spokane (Washington); Westchester County (New York); Winston-Salem (Carolina) Canada: Burlington (Ontario); Calgary (Alberta); Edmonton (Alberta); Fredericton (New Brunswick); Kenora (Ontario); Moncton (New Brunswick); Ottawa (Ontario); Quebec City (Quebec); Stratford (Ontario); Toronto (Ontario); Vancouver (British Columbia); Waterloo (Ontario); Western Valley (Nova Scotia); Windsor-Essex (Ontario); Winnipeg (Manitoba)
Middle/South America	Barceloneta (Puerto Rico); Curitiba, Paraná (Brazil); Pirai (Brazil); Porto Alegre (Brazil)
Oceania	Ballarat (Australia); Gold Coast City (Australia); Ipswich, Queensland (Australia); State of Victoria (Australia); Whittlesea, Victoria (Australia)

One cannot rely on an individual definition to explain the Smart City concept. Nam and Pardo (2011) explain that the most important consideration when defining a Smart City is to view it as one "large organic system". Therefore, when attempting to conceptualize the Smart City concept one must consider the stakeholders involved, Smart Technology, goals and influential areas of the Smart City. Figure 2-2 has incorporated all these components and can be used to explain the Smart City concept. Note that the six boxes (excluding public safety) at the bottom of Figure 2-2 refer to the six characteristics of a Smart City, constructed by Giffinger *et al*. (2007). Public safety is considered under Smart Living, and this is displayed in Figure 2-2 to indicate the focus area of this study. This area will be discussed in further detail in Chapter 3.

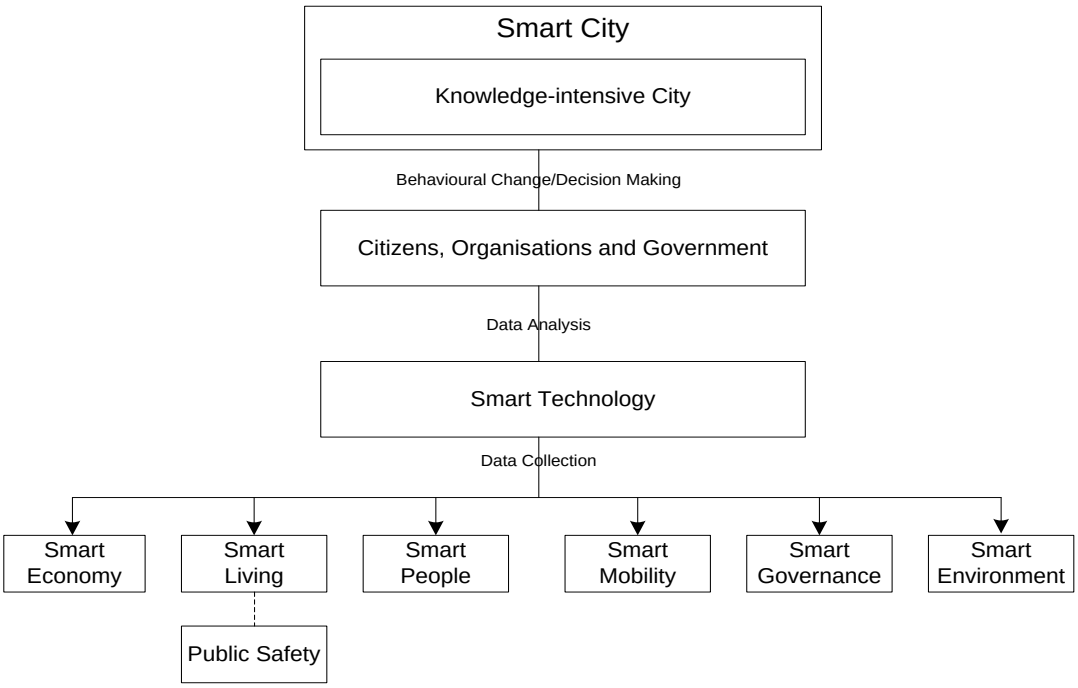


Figure 2-3: Conceptualisation of a Smart City

Smart technologies are used to collect data or provide a direct influence on certain areas of the city (economy, living, people, mobility, environment and governance). In situations where Smart Technologies are used to collect data, the data are used to realise patterns and make decisions. The Smart Technologies are used by citizens, organisations and government, usually in a collaborative effort as limitations exist in individual endeavours. For example, due to the wide dispersion of citizens, they can act as sensors to report potholes around the city, which will assist the government in maintaining roads. These initiatives are used to contribute to making the city smarter and more knowledge-driven. Figure 2-2 incorporates all the elements needed to explain the Smart City concept. This will be used as a discussion guideline for the first part of Chapter 2. The next section will discuss citizens, businesses and local government as key stakeholders of a Smart City.

2.4 Citizens, Businesses and Government Participation

A vast collection of literature on Smart Cities tends to focus on the technological side and fails to emphasise the importance of a collaborative effort from citizens, organisations and local Government. Although technology is an important aspect of the Smart City, additional benefits can be realised with technology and human interaction. Few articles mention that collaborative involvement by these stakeholder groups is necessary to build a Smarter City (Aldama-Nalda *et al*. 2012; Komninos, Pallot & Schaffers, 2012); however,

the questions of how and why are addressed by simply providing examples of Smart City initiatives involving these stakeholder groups. Every person in the city should work to improve the city's overall quality of life, citizens (1), organisations (2) and local Government (3). A further discussion on these stakeholder groups will follow.

1. Kanter and Litow (2009) explain that citizens are rich in information and interconnected to contributing to the Smart City. Citizens are widely dispersed around the city and possess diverse mobility habits as well as other behavioural patterns. Citizens' activities should be exploited to assist in improving certain areas of the city. For example; due to citizens' geographical dispersion, technologies can be used to take advantage of time and spacio behaviour to collect data on activities occurring around the city. Additionally, citizens can modify certain behaviours to reduce urban problems, such as excessive consumption of electricity. If data are presented to individuals on their electricity consumption, this may create a positive behavioural response, resulting in reduced electricity consumption. These are only a few suggestions on how citizens can work collectively with organisations and the government to create a Smarter City. It can be seen that citizens are capable of making much larger contributions than realised. Organisation can also help contribute to the Smart City concept.
2. Gottdiener (2001) explains that cities are being shaped by large businesses located in the city. Although it may be risky to rely on a few large businesses for city growth, it is necessary in many cities. However, risks associated with business-driven urban development are coupled with numerous benefits. Organisations can contribute to economic growth, create new business opportunities and assist small and medium size businesses, encourage international investment and provide job opportunities for citizens. Through these activities organisations assist citizens and the Government to create a Smarter City. Local Government is the last stakeholder group in this discussion.
3. Local government is responsible for managing various resources across the city. Such resources include healthcare, safety, education, water and electricity (Komninos, Pallot & Schaffers, 2012). Since citizens frequently use these services, and many even consider them a necessity, government should aim to improve these service to the citizens. Government has provided citizens with access to certain services online and in addition, have also provided access to information which

may influence citizens' way of life (Nam & Pardo, 2011). In terms of the government's influence on businesses, laws and legislations can be created, modified or removed to assist business growth. Additionally, all businesses conduct recruitment procedures to find talented and educated young employees to continue future business operations. The government's involvement in creating quality education for the youth will create educated employees, eventually contributing to economic growth.

From the discussion above, a joint effort is not only beneficial, but necessary for the city to operate effectively. Each stakeholder group is dependent on the other and therefore, should pool their efforts together to create a Smarter City. Figure 2-3 was developed by incorporating various Smart City initiatives involving the three stakeholder groups to illustrate how these groups can assist each other and the overall growth of the city.

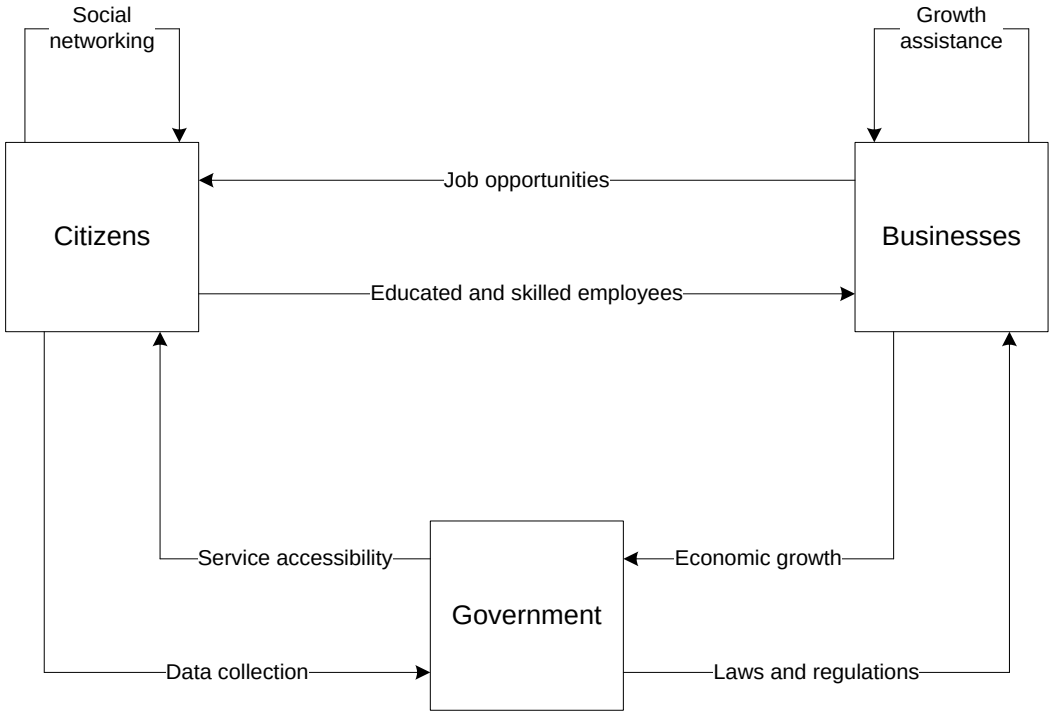


Figure 2-4: Smart City Stakeholder Interactions

Although the need to act collectively is strongly emphasised in the above discussion, no contribution can be made to the Smart City without the use of technology, more specifically, the correct technologies, their application, and their use.

2.5 Technology

Before introducing the way technologies are used in Smart City initiatives, a brief explanation of a city's typical technological infrastructure will be presented. This is necessary as the technological aspect of Smart City initiatives utilise this basic technological infrastructure to run the project. Komninos, Pallot and Schaffers (2012) explain that a typical technological infrastructure in a city is composed of four concentric circles conceptualised in Figure 2-4.

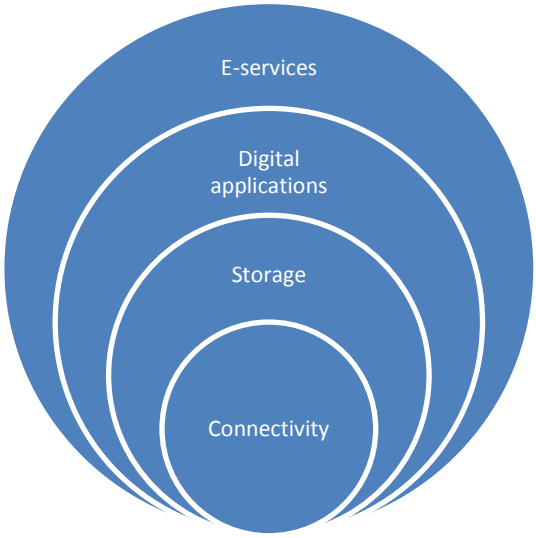


Figure 2-5: Common Technological Infrastructure of a City – adapted from: (Komninos, Pallot & Schaffers, 2012)

The core is comprised of both broadband, which uses telephone cables for connectivity, and wireless communication that utilises mobile phone network towers. This is to facilitate connectivity between devices. The second inner circle refers to the databases used to store data; this is important as all technologies either collect or retrieve data from storage locations. The third circle from the centre is the application domain which includes all applications used to run technology. The outer circle includes all electronic services offered on a regular basis as an alternative to a conventional service, for example; electronic banking. From Figure 2-4, Komninos, Pallot and Schaffers (2012) explain that in a city's technological infrastructure, the outer circles are dependent on the inner circles. However, technological solutions are only fully effective when all four circles are utilised. This is the main concept of "Smart" Technologies.

A vast volume of literature on Smart Cities makes reference to technologies used in Smart City initiatives. However, Smart City contributions in the technological sense are not

limited to a "fixed list" of technologies. Other authors refer to the technologies utilised by Smart City initiatives as "Smart Technologies" or "Smart Computing"; however, they do not explain the difference between these terms to conventional technologies. Washburn and Sindhu (2010) define Smart Computing as "a new generation of integrated hardware, software, and network technologies that provide IT systems with real-time awareness of the real world and advanced analytics to help people make more intelligent decisions about alternatives and actions that will optimize business processes and business balance sheet results" (p. 2). Any technological solution which fulfils the criteria expressed in Washburn and Sindhu's (2010) definition is considered an appropriate technological solution for a Smart City initiative. Since Smart Computing is a collaborative technological innovation, it can make use of new or existing technologies to create a new Smart Computing innovation. Bartels (2009) developed criteria for Smart Computing, referred to as the five A's of Smart Computing. Smart Computing supports these five stages of intelligent activities, illustrated in Figure 2-5 below.

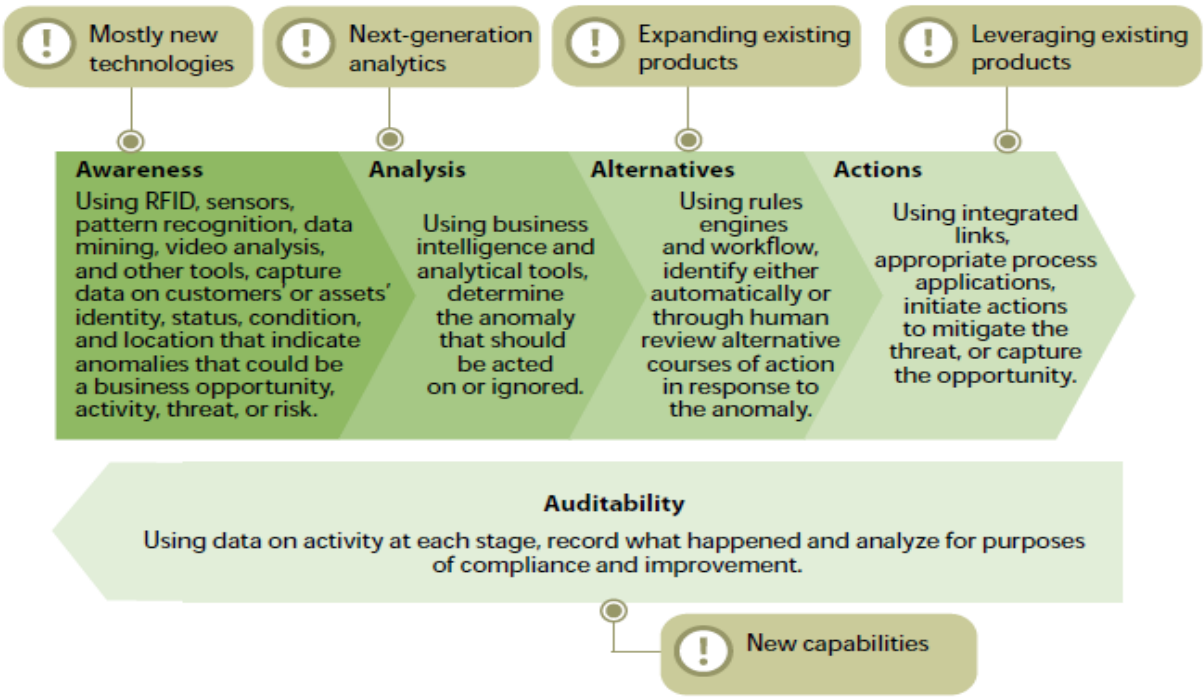


Figure 2-6: The Five A's of Smart Computing (Bartels, 2009)

The five A's of Smart Computing include: awareness (1), analysis (2), alternatives (3), actions (4) and auditability (5) (Bartels, 2009).

1. Awareness – Awareness entails the use of technologies to collect data. The technologies are used to collect specific data on activities occurring in an

environment. It usually includes a combination of hardware and software technologies, for example, CCTV cameras integrated with facial recognition software.

2. Analysis – Analysis refers to hardware and software used to store and analyse the data collected. The software uses complex algorithms to identifying patterns in data, assist decision-making and predict future outcomes. Hardware technology is focused on the flexibility of servers and data storage facilities, such as cloud computing.
3. Alternatives – Alternatives focus on the adaptability of rule engines and workflows. The concept behind *alternatives* is that a system can react to an event automatically, based on predefined rules (if this occurs then do that) or human review. The rule engines and workflows should be able to adapt rapidly to event changes.
4. Actions – Actions refer to the use of integrated links from a process application of a device. The action taken must accomplish at least one of two things, lower a threat or take advantage of an opportunity. For example, a traveller receives an SMS notification that his flight is delayed, or a doctor receives medical information electronically, about a patient's allergic reactions before a diagnosis is complete.
5. Auditability – Auditability entails the use of data from the previous stages to identify any problems or areas of improvement. This stage is also used to determine if the technological solutions conform to specifications, regulations and any legal obligations.

A contribution to the Smart City requires both collaborative stakeholder involvement and a technological solution that needs to be applied to a specific area. Recall the diagram in Figure 2-2 that conceptualised the Smart City and illustrated six broad areas where initiatives can be applied to contribute to the Smart City. These six areas are often referred to as Smart City characteristics, dimensions or axes; for the purpose of this study, the six areas will be referred to as the six characteristics of a Smart City. The six characteristics of a Smart City will be used to direct the following discussion.

2.6 Smart City Characteristics

Giffinger *et al*. (2007) has divided the Smart City concept into six contribution areas, referred to as characteristics of a Smart City. These six characteristics include: Smart

Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment and Smart Living. In the study by Giffinger *et al*. (2007), the characteristics were used to rank 70 European cities according to their "smartness". The characteristics were constructed based on theories relating to traditional neoclassic and regional urban growth and development. This has allowed the characteristics to be used as a theoretical framework by a number of authors such as, Caragliu, Del Bo & Nijkamp (2011), Lombardi, Giordano, Farouh & Yousef (2011), and Desourdis (2010). The six characteristics of a Smart City consist of thirty-one factors and seventy-four indicators. The thirty-one factors were used to rank and describe the six characteristics, and the seventy-four indicators were used to assess the thirty-one factors (1-4 indicators per factor). The list of factors and indicators are available in Appendix B – Characteristics of a Smart City.

Presently, there is no collection of Smart City initiatives that can address all seventy-four indicators. At this point, Smart City initiatives are organised into seven categories: utility, education, transport, healthcare, public safety, infrastructure, and city administration (Washburn & Sindhu, 2010). As mentioned above, a city achieves a Smart City status by running a collection of Smart City initiatives or an individual initiative. Respectively, a single initiative cannot make a city "smart" as initiatives tend to focus on a specific area. However, since the Smart City is a new concept, it is acceptable to keep track of Smart City initiatives by providing the Smart City title to cities that are running one or more Smart City initiatives.

Although targeting these areas will assist a city to achieve a Smart City status, Smart City initiatives should not be limited to these seven categories. For example: none of the seven categories addresses job creation, which is a recurring problem restraining the growth of many cities. In addition, the seven categories of a Smart City (utility, education, transport, healthcare, public safety, infrastructure, and city administration) mentioned above have been used by a number of authors and firms to describe the areas of a Smart City. However, no author states how these seven categories were constructed or even who constructed them. The six characteristics constructed by Giffinger *et al*. (2007) provide a broader scope for Smart City initiatives and are supported by urban development and growth theories. Therefore, the six characteristics constructed by Giffinger *et al*. (2007) are considered to be a more valid set of areas to describe a Smart City's focus areas. Figure 2-6 supports this argument, which shows that Washburn and Sindhu's (2010) seven areas are similar to Giffinger *et al*. (2007); however, Washburn and Sindhu (2010) do not

envision economic growth as a contribution to the Smart City. In addition, Washburn and Sindhu's (2010) seven areas of a Smart City are narrow contribution areas as opposed to Giffinger *et al*. (2007) who have described their six characteristics of a Smart City in a broader sense. For example, Washburn and Sindhu (2010) refer to improving education, whereas Giffinger *et al*.'s (2007) Smart People characteristics include other aspects, including education such as encouraging creativity, job flexibility and citizen participation in government decision-making.

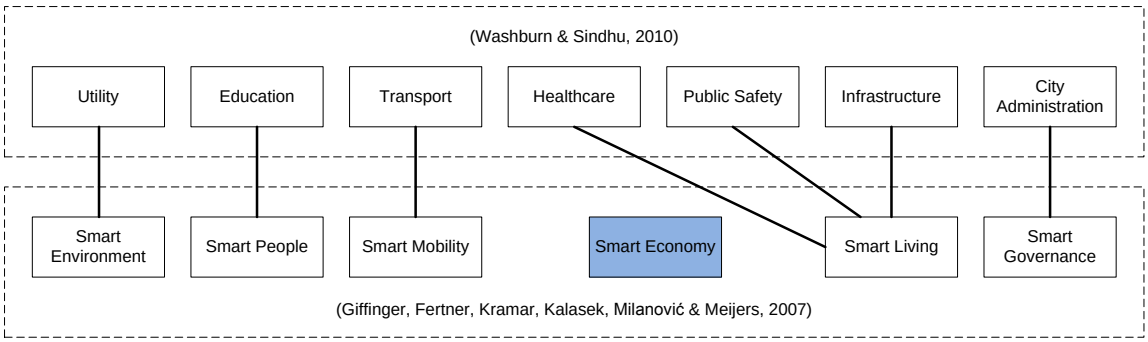


Figure 2-7: Smart City Contribution Areas

The most common Smart City initiatives in commencement will be organised into the six Smart City characteristics constructed by Giffinger *et al*. (2007) to illustrate attempts at creating a Smarter City. Before these initiatives are presented, Table 2-2 illustrates how the six characteristics of a Smart City assist the challenges of urbanisation discussed in section 2.2. Smart Economy is the first characteristic that will be discussed after Table 2-2 is presented.

Table 2-2: Relationship between Urbanisation Challenges and Smart City Characteristic

Urbanisation Challenges	Smart City Characteristic
Resource Depletion	Smart Mobility, Smart Environment and Smart Living
Infrastructure Deterioration	Smart Living
Energy Shortage	Smart Environment
Environmental Issues	Smart Environment and Smart Mobility
Health Concerns	Smart Living and Smart Environment
Unemployment	Smart People and Smart Economy
Public Safety	Smart Living

2.6.1 Smart Economy

Smart economy is dependent on the commercial sector of a city. Some cities rely on large organisations to contribute to city growth and shape the city (Monbiot, 2000). Other cities such as Edmonton (Canada) and San Diego (USA), focus on economic growth by creating

an attractive city for businesses to start-up, relocate or expand (Hollands, 2008). Giffinger *et al*. (2007) add that a Smart Economic city should also support entrepreneurship, business innovation, productivity and international accessibility (imports and exports), as well as create a flexible labour market to reduce unemployment.

Businesses are beginning to realise the benefits cloud computing can have on business processes. Mell and Grance (2011) provide a brief but simple explanation of the entire cloud computing concept and the options it provides to businesses. This explanation has been conceptualised and is illustrated in Figure 2-7 below.

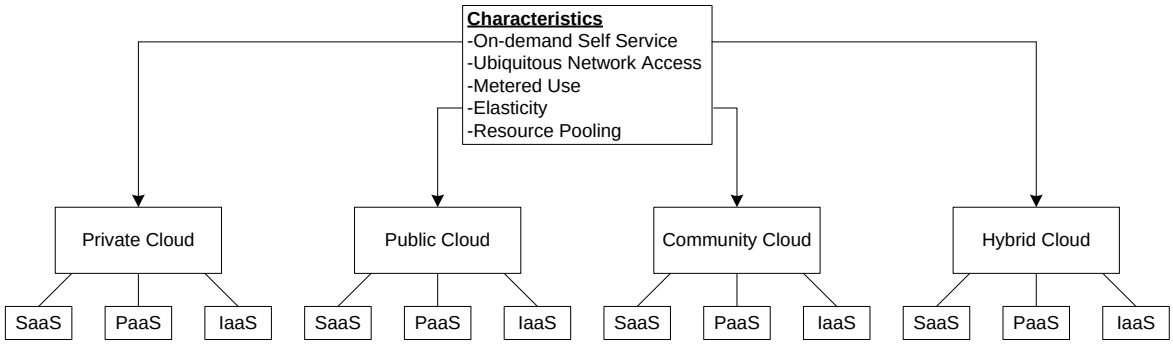


Figure 2-8: Concept of Cloud Computing – adapted from: (Mell & Grance, 2011)

Cloud computing offers numerous applications which can be easily accessed and at an affordable cost. Figure 2-7 shows the characteristics of cloud computing, the types of services offered and how these services can be accessed. The four methods of deployment (private, public, community and hybrid) absorb all five characteristics of cloud computing, and each deployment method can access any three types of services, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Businesses select the type/s of services they are interested in and how they want to access them based on the nature of the business and expected future growth.

2.6.2 Smart Environment

Smart Environment refers to environmental sustainability, ecological protection and efficient use of resources such as water and electricity (Giffinger *et al*., 2007). Although cities are the largest contributors to economic growth, they also create pollution and consume large volumes of natural resources. This is a difficult problem to address because of the contradictions between environmental sustainability and economic growth. A Smart City makes use of technology to create a Smart Environment which consists of advanced technologies. Sample (2004) explains that the rapid innovation of technology has resulted

in frequent upgrades and the disposal of outdated technology which can eventually cause a computer "waste mountain". Sample (2004) adds that it is estimated that the waste of technology from production is five times more than the waste created from the production of vehicles. Therefore, the production and disposal of technology has environmental implications. The use of certain technologies to combat urban environmental problems may create additional problems. This needs to be considered before a Smart Environment initiative is implemented.

A common example of an initiative contributing to the Smart Environment is the Smart Grid initiative. The Smart Grid uses ICT resources to automatically collect and analyse data about the production and consumption of electricity to automatically distribute electricity in a more efficient and effective manner (Massoud, 2005). It is also capable of presenting electricity consumption data to consumers so excessive energy consumption can be realised and may result in consumer behavioural changes. Smart Grid utilises a collection of sensors, digital meters, hot and cold ventilation systems, transformers, distribution systems and supervisory control and data acquisition systems, often known as SCADA systems (Wildeman, Gilpin & Knoll, 2009). Although the long list of technologies emphasises large start-up costs, the Smart Grid solution has proven to be successful in numerous cities across the world. The continuous success of the Smart Grid initiative may encourage the same concept application to the consumption and distribution of clean water.

2.6.3 Smart Mobility

Smart Mobility refers to public transportation in terms of quality, accessibility and environmental sustainability (Giffinger, *et al*., 2007). Numerous transportation methods are used on a daily basis, for the transportation of people and goods. Common transportation methods such as vehicles, trains and airplanes not only consume large amounts of resources such as fuel (oil) but also create air pollution, causing significant damage to the ozone layer. Smart Mobility focuses on public transport as opposed to encouraging private transport that has larger negative implications to the environment. One may argue that Smart Mobility should rather focus its attention on reducing the overall use of transportation. For example, Graham (2002) states that the implementation of ICT technologies in people's homes will allow them to work at home instead of travelling; however, this solution has not proven successful.

A Smart Mobility initiative commonly known as congestion pricing makes use of Smart Technology to determine the most congested roads in the city and charge road users an additional toll for taking the route during peak congestion periods (Washburn & Sindhu, 2010). The intention of this initiative is to raise funds for improvements in public transport and encourage people to make use of current public transport services. With Congestion Pricing, drivers are given the options to reschedule their travel arrangements, carpool or pay the additional toll. This initiative has been proven highly successful in London, England, which implemented the system in 2003. Table 2-3 below displays some positive results from Congestion Pricing in London.

Table 2-3: Statistical Results from Congestion Pricing in London (Washburn & Sindhu, 2010)

Description	Percentage
Congestion	30% Decrease
Traffic speed	37% Increase
Particulate matter and nitrogen oxides	12% Decrease
Fossil fuel consumption and CO2 emissions	20% Decrease

It was mentioned by Washburn and Sindhu (2010) that Congestion Pricing is aimed at encouraging people to make use of public transport services. However, charging drivers a high toll to use certain roads may also cause drivers to rather find alternative routes which are longer, resulting in additional consumption of fossil fuel. Additionally, it was not mentioned if people who decide to use public transport, a bus for example, will pay a slightly higher fee if the bus uses the congested road during peak congestion. Public transport by road such as buses should not be charged for using the congested roads, unless the goal is to encourage people to make use of other public transport services, such as rail.

2.6.4 Smart Governance

Smart Governance refers to the efficient and effective delivery of government services, including attempts to reduce corruption (Giffinger *et al*., 2007). Giffinger *et al*. (2007) have used the term Smart Governance to describe Government activities. Authors (International Bank for Reconstruction and Development (2006); and The World Bank (1991) refer to government responsibilities as governance, Griffith (2000) explains that Smart Governance involves a collaborative effort from all local Governments in a region. Although governance has been used to refer to government activities, governance is not limited to government responsibilities, and can be exercised in any entity whether a household, small business or large organisation (Hufty, 2011; Bell, 2002). Therefore, it is

suggested that the characteristic, Smart Governance, be renamed Smart Government as this characteristic only relates to government activities. From this point onwards the term Smart Government will be used when referring to this characteristic.

This is an important characteristic of a Smart City, as many city operations rely on government support; as previously illustrated in Figure 2-3 citizens and businesses depend on government services. Additionally, as urbanisation increases, the government is becoming more complex and growing larger (Washburn & Sindhu, 2010). Local governments should be aware of their city's condition and be able to assist citizens where necessary.

A Smart City initiative in Southampton, England, has implemented a Smart Card system which provides citizens with access to numerous government services (Hollands, 2008). The project aims to mitigate issues of social inclusion and create stronger connections between government, businesses and citizens. Two sets of Smart Cards were issued to the public; one for students and the other set were for ordinary citizens. At the moment Smart Cards provide access to public transportation, libraries and other government services. However, the initiative is aimed at linking additional services and eventually for the Smart Card to be used for electronic voting.

Another Smart City initiative, improving service delivery, is the e-government initiative which provides citizens with access to certain services online. The initiative creates a centralised point of entry for citizens to access various government services (Aldama-Nalda *et al.*, 2012). This will minimise over-crowding (long cues) at government institutions, provide 24/7 access to certain services and increase the turn around time of services. The e-government initiative also reduces counter service corruption by cutting out the "middle man". Grönlund, Heacock, Sasaki, Hellström and Al-Saqaf (2010) state that automating tasks originally carried out manually will remove the risk of the public interacting with corrupt government workers.

2.6.5 Smart People

Smart People refers to the knowledge base of all citizens, businesses and government. Smart City initiatives focus on assisting people to learn, adapt and innovate, through the use of Smart Computing technologies (Coe, Paquet & Roy, 2000). This is accomplished through education, social learning, as well as investments in social capital (Eger, 2003). It is clear that focus on education is important to create Smarter People. However, Hollands

(2008) emphasizes that some authors do not agree with social learning initiatives. Social sustainability creates a sense of belonging and social cohesion amongst a community (Carley, Jenkins & Small, 2001). It can be argued that this approach to creating Smarter People is ineffective. Although some social approaches may not form part of enhancing people's ability to gain knowledge, it is necessary as an overall contribution to the Smart City as it forms part of improving the quality of life. Additionally, Washburn and Sindhu (2010) point out that collaboration technology improves the quality of education. Social sustainability also allows people to learn from others. In this manner, social approaches can contribute directly to Smart People, and other approaches will contribute to the Smart City as a whole.

An educational related Smart People initiative, run by the Jack Welch Management Institute, has created an independent online Master of Business Administration (MBA) program (Gloeckler, 2009). The initiative offers the opportunity to obtain an MBA at a low cost and supports part-time students. Jack Welch Management Institute (Gloeckler, 2009) boasts that their online MBA program is offered at a cost of approximately five times cheaper than most MBA degrees. This is an example of an initiative that increases access to education, amongst other things.

2.6.6 Smart Living

This characteristic of the Smart City relates to the lifestyle of the citizens by addressing basic living needs. Smart Living includes the city's healthcare, infrastructure such as entertainment facilities and housing, and safety (Giffinger *et al*., 2007). These will be discussed in further detail below. Note that public safety forms part of Smart Living and therefore, is the focus area of the project for this study and will be discussed in Chapter 3.

2.6.6.1 Healthcare

Healthcare refers to the quality of health services provided in the city, such as availability of medical staff, hospital beds and ambulance services. Contribution to this aspect of Smart Living can be achieved by improving the quality of health services. Health services can be described as any service which assists a person with illness (Rosenstock, 2005). Washburn and Sindhu (2010) explain that a Smart Healthcare system is comprised of storage and communication technologies. The initiative provides numerous advantages to the health sector. For example, patients' medical records can be stored, backed-up, shared and made easily accessible almost immediately, as and when required. A city in Spain has

utilized a similar initiative where they have centralised all patient information and provided access to all health institutions in the region. The initiative has connected 13 000 medical professionals and 9 million outpatient visits annually (Washburn & Sindhu, 2010). The second segment of Smart Living relates to the city's infrastructure.

2.6.6.2 Infrastructure

In this context, infrastructure refers to the development of buildings ranging from residential houses to office buildings to entertainment facilities such as casinos and malls. The initiatives focus on financial benefits, quality and environmentally friendly construction and development methods by utilising Smart Computing technologies (Washburn & Sindhu, 2010). For example, in the USA, Smart Computing was incorporated into building management systems. This provided numerous benefits, such as automated cooling and heating systems. Sensors were also installed in rooms to automatically switch off lights when the room was empty. Haubensak (2011) states that automated data collection techniques allow data to be collected on household behaviours. The data is then presented to home owners in the expectation that it will result in a change of negative behaviours, such as excessive water and electricity consumption. Alternatively, the same method can be applied to building managers to influence maintenance decisions. Table 2-4 below shows the results of such building management systems that integrated Smart Computing technologies.

Table 2-4: Statistical Results of Smart Computing Building Management Systems (Washburn & Sindhu, 2010)

Description	Percentage
Operating costs	8% - 9% Decrease
Building value	7.5% Increase
Return on investment	6.6% Increase
Occupancy ratio	3.5% Increase
Rent ratio	3% Increase
Energy consumption	26% Decrease
Maintenance costs	13% Decrease
Occupant satisfaction	27% Increase
Greenhouse gas emissions	33% Decrease

Smart Computing technologies are capable of numerous financial and environmental benefits for infrastructure development. The last section of Smart Living will discuss public safety.

2.6.6.3 Public Safety

Public safety can be defined as any incident which creates or has the potential to create harm to citizens. Therefore, public safety does not only involve criminal activity, as many people assume, but also forest fires, floods, as well as road and infrastructure damage. With the increase in urbanisation, emergency units such as police officers and fire fighters need to respond to public safety issues more quickly and simultaneously maintain a safe living environment for everyone. The project on which this research is based focuses on contributing to this area (public safety as part of Smart Living) of a Smart City.

The Smart Computing technologies used for public safety initiatives tend to focus on communication technologies. For example, the New York Police Department (NYPD), USA, collects criminal data at a centralised location for analysis. Once the data have been analysed, it is available instantaneously to any officers who require such information. This initiative has resulted in a 27% decrease in crime (Washburn & Sindhu, 2010). In addition, a real-time dashboard is also available to the police department, which consolidates all reported incidents that need to be addressed as well as the resources available, in a single view (Washburn & Sindhu, 2010). Police officers can also access the portal from any location through e-mail, public computers and mobile phones. This portal has allowed the police department to manage their resources more effectively and efficiently. Other public safety initiatives that do not focus on communication technologies also contribute to creating a safer city.

A collaborative system of video analytics and closed-circuit television (CCTV) has created an opportunity for technology to detect criminal activity or reckless driving on the road (Washburn & Sindhu, 2010). The CCTV cameras are installed with sensor technologies that use motion to pick up any abnormalities. For example, CCTVs used to monitor a building, can detect break-in attempts based on a sudden motion increase in the captured area. In addition, the mere presence of CCTV cameras has known to reduce criminal activity.

Washburn and Sindhu (2010) state that Smart City initiatives relating to public safety essentially focus on developing a state of the art public safety network. The public safety network described by Washburn and Sindhu (2010) is conceptualised in Figure 2-8. This study will focus on a specific Smart City public safety initiative, which will focus on improvements in *emergency dispatch and coordination* through participatory

crowdsourcing. Before the specific participatory crowdsourcing project is introduced, a basic understanding of crowdsourcing is necessary. This will be discussed in the next chapter. Concluding remarks about this chapter will now be provided after Figure 2-8.

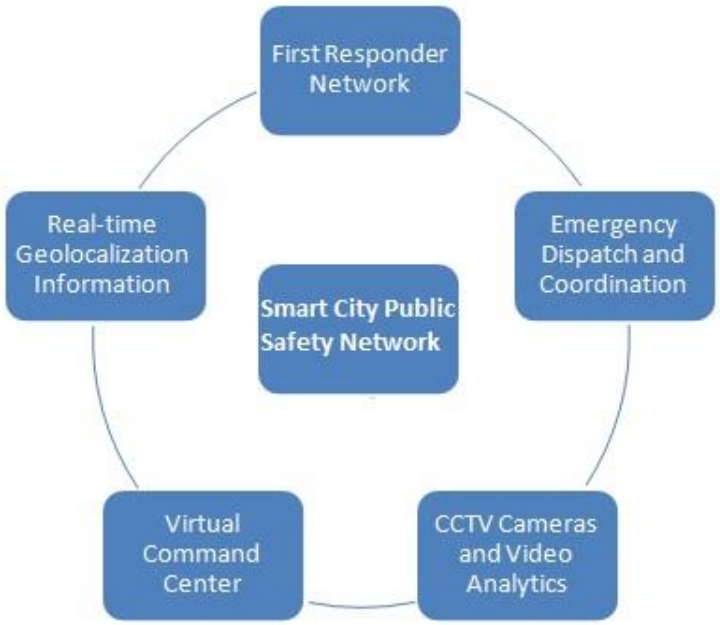
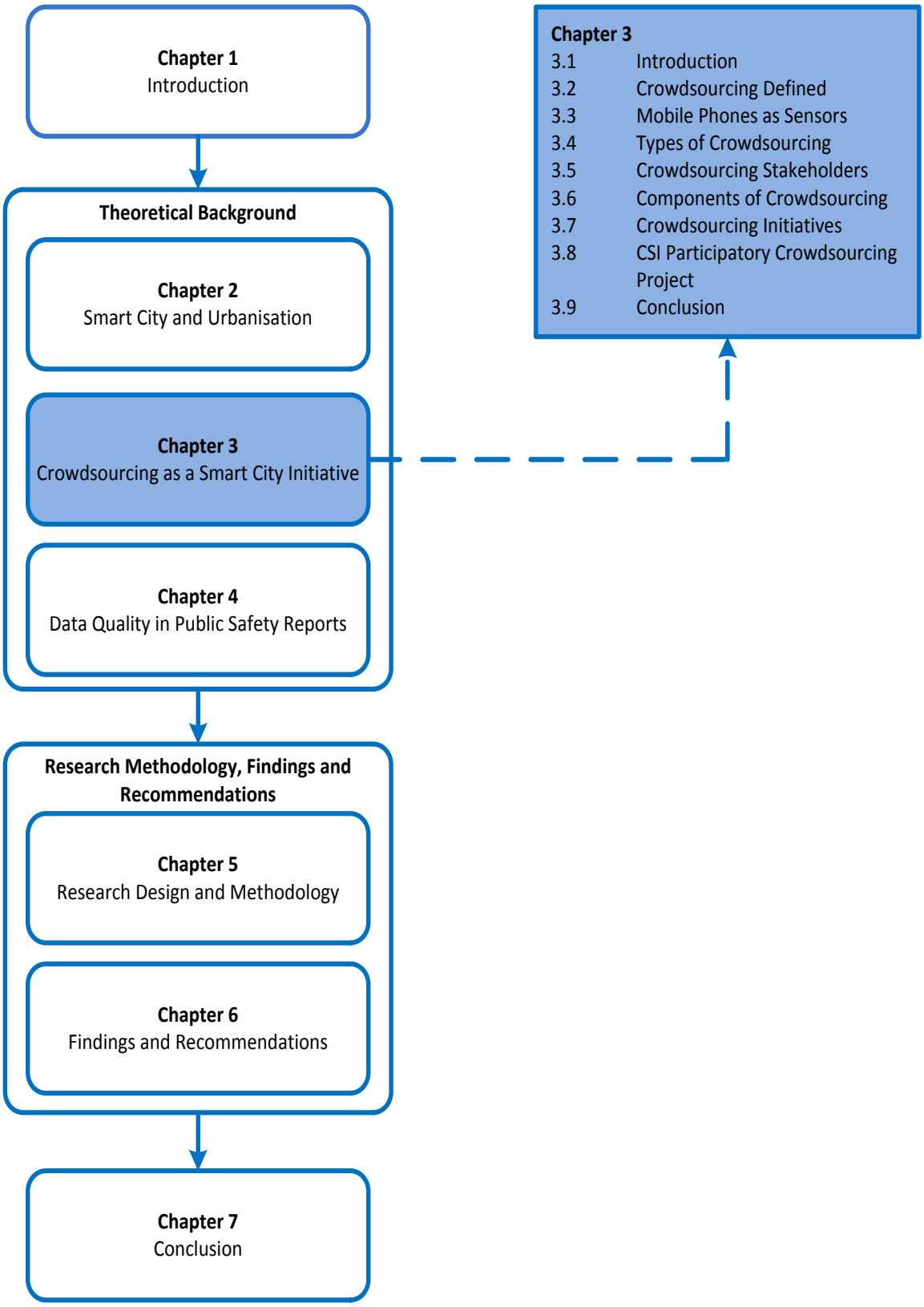


Figure 2-9: Conceptual Representation of a Smart City Public Safety Network – adapted from: (Washburn & Sindhu, 2010)

2.7 Conclusion

The Smart City concept can be used to assist cities to cope with urbanization by making contributions to six broad areas. These areas include: Smart Economy, Smart People, Smart Government, Smart Mobility, Smart Environment and Smart Living. All Smart City contributions are made through the adoption of Smart Technologies. In addition, a collaborative effort from citizens, local Government and businesses is required in order to make an effective contribution. Public safety forms part of the Smart Living characteristic of a Smart City. The project on which this study is based, contributes to this area of a Smart City, more specifically improving *emergency dispatch and coordination*. Chapter 3 will provide a more in-depth discussion on the project and how it will be used as a Smart City initiative to reduce problems with urbanisation.

Chapter 3 – Crowdsourcing as a Smart City Initiative



3.1 Introduction

Chapter 3 will determine **how crowdsourcing can be used as a Smart City initiative to reduce public safety issues.** The chapter will develop a crowdsourcing map which illustrates what one must consider before ensuring the successful implementation of a crowdsourcing initiative. Various crowdsourcing projects which are in progress, will follow the crowdsourcing map. These crowdsourcing projects will be presented according to the 6 characteristics of a Smart City (Smart Economy, Smart People, Smart Government, Smart Mobility, Smart Environment and Smart Living) discussed in Chapter 2. This will allow one to determine how crowdsourcing can be used as a Smart City initiative and assist a city to cope with urbanisation. The last section of Chapter 3 will discuss the specific crowdsourcing project on which this study will focus, the CSI participatory crowdsourcing project. This chapter will begin by defining the term crowdsourcing.

3.2 Crowdsourcing Defined

Crowdsourcing is a term which refers to a group of individuals who collect data, or report on certain events of similar nature and pool all the data collected (Väätäj, Vainio, Sirkkunen, & Salo, 2011). Originally, crowdsourcing related to human observations while crowdsensing referred to the use of hardware and software to act as sensors. Numerous authors use the word crowdsourcing and crowdsensing interchangeably, by explaining that humans are used as sensors when using the term crowdsensing. This has been accepted by many publishers but to avoid any confusion, this study will use the term crowdsourcing.

Alt, Shirazi, Schmidt, Kramer and Nawaz (2010) parameterised the definition of crowdsourcing by explaining that the participants are the undefined public rather than a specific group. This is the underlying concept of Crowdsourcing, where tasks can be solved or data generated by taking advantage of the wide geographical dispersion of citizens. The dynamic mobility behaviour of citizens and their observations and knowledge of their surroundings makes them perfect candidates for collecting data on activities occurring in their environment. This data can be analysed and shared with the public and/or be used to improve decision-making.

Crowdsourcing has received a great deal of attention in recent years due to its innovative method of distributed data collection (Hester, Shaw & Biewald, 2010). Instead of outsourcing a problem or task to an individual or a group, crowdsourcing provides access

to additional data resources. For example, the problem can be presented to the public, allowing anyone to contribute to solving the problem. In this manner, large volumes of data can be collected in a short period of time, at a low cost. The data collected can be used to make decisions, identify patterns (Yang, Zhang & Roe, 2011), solve problems (Alt, Shirazi, Schmidt, Kramer & Nawaz, 2010) and even influence behavioural change (Christin, Reinhardt, Kanhere & Hollick, 2011). Before a crowdsourcing initiative can be initiated, one must first determine exactly what data will be collected from crowds. For example, data on weather conditions will entail temperature, wind speed and a geographical location to identify the weather conditions experienced in a specific position. The data expected will influence the method of collection. Methods can range from social networks to sophisticated sensor devices. Note that any method utilising the advantages of a crowd is considered an appropriate method for a crowdsourcing initiative. Due to the vast variety of methods, the most common crowdsourcing method will be discussed, mobile phones.

3.3 Mobile Phones as Sensors

Over the last decade, there has been a large proliferation of mobile phone users, that continues to growth over time. As of 2011, it was found that over 5 billion people worldwide have access to a mobile phone (Christin, Reinhardt, Kanhere & Hollick, 2011). As more sophisticated mobile phones are developed, the prices of standard phones (calls and SMS compatible) become cheaper and more available for low income groups. This has led to increases in the number of mobile phone users around the world. Additionally, this emphasises that Smart Phones (sophisticated mobile phones) are more ubiquitous in developed countries than developing ones.

Smart phones are enhanced with numerous sensor capabilities that utilise typical functions, such as GPS and microphones (Alt, Shirazi, Schmidt, Kramer & Nawaz, 2010). These sensors allow large amounts of data to be collected over a large geographical area at little cost. However, this does not mean that crowdsourcing is not possible in developing countries due to its low Smart Phone penetration. Although it is still possible to collect data on Smart Phone users in developing countries, it is not feasible as the accuracy of the data collected is based on the number of sensors operational for data collection (more sensor operators will increase data accuracy). Therefore, it is suggested that standard phones be used for crowdsourcing in countries with low Smart Phone penetration. Crowds

will act as data sources, by reporting observations of their environment through text, camera (photos or videos) or phone call.

Using mobile phones as a method for crowdsourcing, still requires decision-making on how the mobile phones will be used to collect data. This is usually the case with other methods as well. A common decision that must be made, regardless of the method chosen, is the type of crowdsourcing initiative to be undertaken.

3.4 Types of Crowdsourcing

Crowdsourcing initiatives can be organised into three categories based on the role and extent of participation (Srivastava, Abdelzaher & Szymanski, 2012). Srivastava, Abdelzaher and Szymanski (2012) strongly emphasise that these three categories are not mutually exclusive; however, they can be used to understand the broad range of crowdsourcing initiatives. Therefore, a crowdsourcing initiative can incorporate a combination of characteristics from more than one category.

Three types of crowdsourcing exist namely, opportunistic sourcing and two kinds of participatory sourcing. Opportunistic sourcing is described as crowds being the targets of sourcing (Srivastava, Abdelzaher & Szymanski, 2012). This type of crowdsourcing is common in the security domain and to indicate user interactions with websites or applications. Sensing technologies are deployed to monitor individual or group behaviours, activities and trends. Opportunistic sourcing can be conducted with or without the permission of the device users. Opportunistic sourcing without the knowledge of the user is acceptable in some cases, as long as no private information is captured. For example, a mobile phone user downloads an application from a website. The website will discover automatically that a mobile user has downloaded the application and automatically increment (record) the amount of users who have downloaded the application to indicate its popularity. This can be used to determine which applications should release an improved version. Another popular example is YouTube, which indicates how many people have watched a video on YouTube by displaying the number of users who opened the video link.

Participatory sourcing has a high level of human involvement in the task of collecting data, as opposed to opportunistic sourcing which focuses more on the data collection device, securing minimum user involvement throughout the collection process (Ganti, Ye & Lei,

2011). Two types of participatory sourcing exist: *crowds as sensor operators* (1) and *crowds as data sources* (2). This will be further discussed below.

1. Crowds as sensor operators – Using crowds as sensor operators entails the use of a sensor device, usually mobile phone due to its ubiquity. In this case the participants (crowd) use the sensor device to collect data on their surroundings. Since sensor devices are used, the type of data collected is limited. For example, certain activities occurring in an environment that cannot be measured by a sensor device cannot be collected as some activities require human interpretation. However, this does ensure the legitimacy of the data, as interpretation may lower the accuracy of the data. Therefore, these types of crowdsourcing initiatives usually focus on data that cannot be accurately interpreted by humans.
2. Crowds as data sources – This type of crowdsourcing relates to humans collecting and disseminating data without the use of any sensor devices. The data collected is usually disseminated to a central location which is accessible to others, usually social networks like Facebook or Twitter. Srivastava, Abdelzaher and Szymanski (2012) mention that the data collected from this type of sourcing can also be uploaded on YouTube. Consequently, electronic devices such as cameras and mobile phones can be used to capture activities identified by individuals. Therefore, unlike the other two types of sourcing, crowds as data sources require a high level of human involvement. Thus, motivational strategies to participate are vital for the success of this type of sourcing. This option has been chosen for the project. Further explanation will be provided in Chapter 3.

The majority of crowdsourcing initiatives can be organised easily into one of these three categories. A multi-categorical crowdsourcing initiative should clearly define the areas using different categories for sourcing data and whether the data will be analysed separately or collectively. When the crowdsourcing initiative has been clearly defined, the stakeholders involved must be selected.

3.5 Crowdsourcing Stakeholders

A typical crowdsourcing project has four stakeholder groups: evaluators (1), initiators (2), gatherers (3) and analysts (4) (Yang, Zhang & Roe, 2011). Figure 3-1 graphically illustrates the four stakeholder groups and their general roles and responsibilities in a crowdsourcing project.

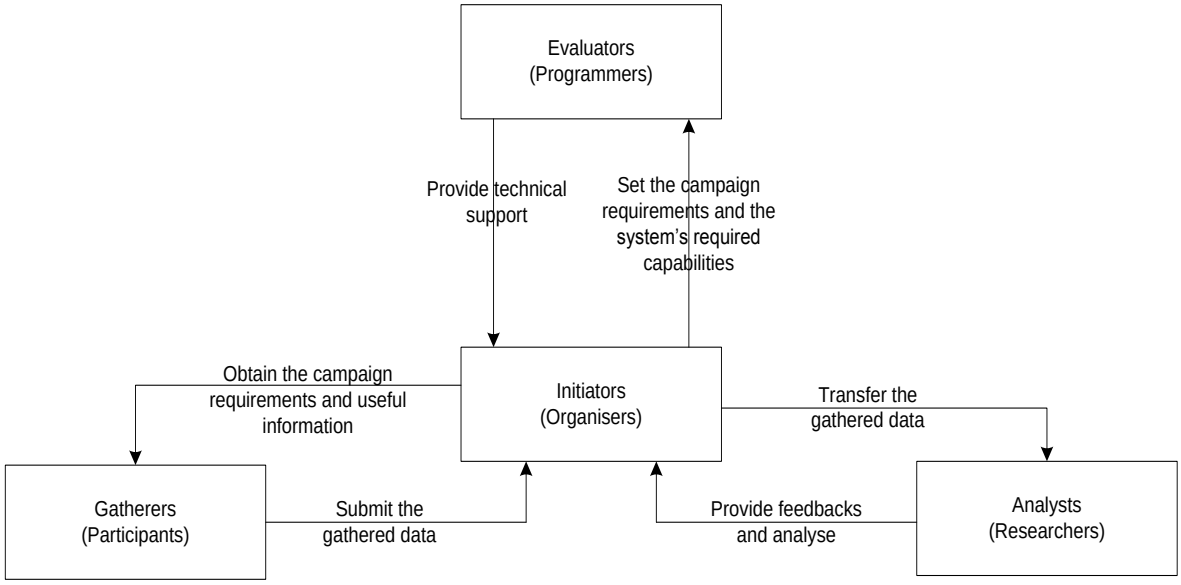


Figure 3-1: Crowdsourcing Stakeholders (Yang, Zhang & Roe, 2011)

Note that an entity can play the role of more than one stakeholder group, for example initiators can also be analysts. Additionally, a stakeholder group can occupy more than one entity. This concludes the last segment that should be considered in a crowdsourcing initiative. The next section will discuss how all the considerations of crowdsourcing are combined to create an effective crowdsourcing initiative.

3.6 Components of Crowdsourcing

The discussion above shows that there are four components of a typical crowdsourcing initiative. These four components include: the type of data collected (1), the method used to collect the data (2), the type of crowdsourcing initiative (3) and the stakeholders groups involved in the crowdsourcing initiative (4). Figure 3-2 incorporates all four components to illustrate what needs to be considered for a typical crowdsourcing initiative to successfully facilitate its data collection purpose. Note that the bottom box was constructed by Yang, Zhang and Roe (2011), as illustrated in Figure 3-1. The greyed boxes in Figure 3-1 indicate the direction adopted by the CSI participatory crowdsourcing project. The next section after Figure 3-2, will present examples of crowdsourcing initiatives currently in progress.

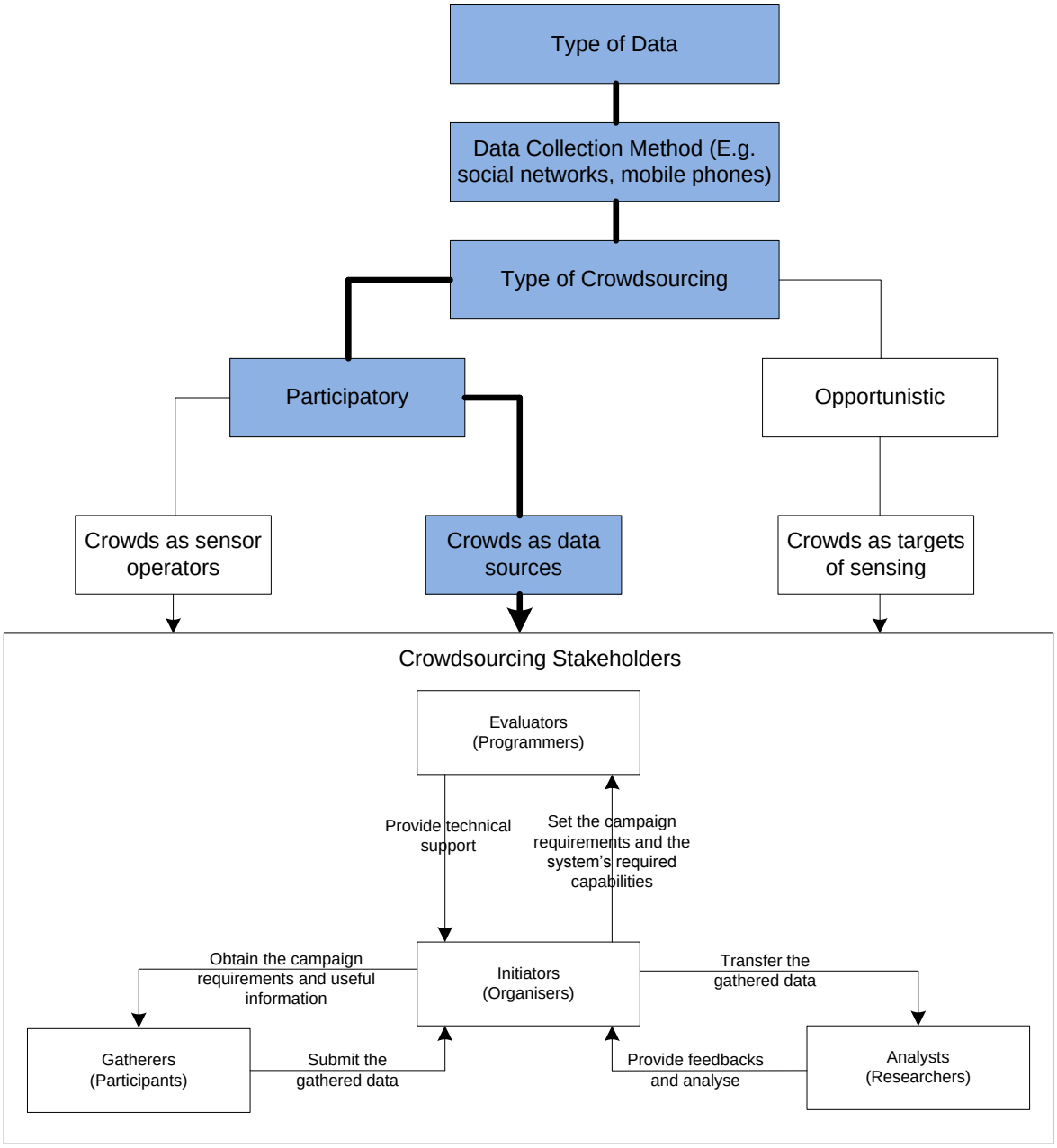


Figure 3-2: Crowdsourcing Components

3.7 Crowdsourcing Initiatives

A number of crowdsourcing initiatives are in progress across the world, and not only in developed countries. Developing countries are beginning to realise the benefits and advantages of these initiatives irrespective of certain limitations (low Smart Phone ubiquity level) as previously mentioned. Examples of crowdsourcing initiatives in developed and developing countries will follow. Since the focus of this study is the use of crowdsourcing to contribute to the Smart City, the initiatives will be discussed according to the six characteristics of a Smart City, as discussed previously.

- Smart Economy
    - A crowdsourcing project known as *txteagle* aims to utilise crowd knowledge to complete tasks. Small tasks, such as transcription or surveys are uploaded on a website which can be accessed through a direct link or an application from a desktop computer or mobile phone. Open participation is provided, in addition to a small financial incentive to encourage involvement. A person can access *txteagle* and browse through a list of tasks until they find one that they can complete. Once the task is completed, the contribution is verified and the allocated money is transferred to the participant's bank account.
    - A similar initiative is run by CNN called *iReporters*, a crowdsourcing project allowing the public to notify CNN about any news worthy material. *iReporters* presents generic areas of news which CNN are interested in. The initiative also has a selective participant arrangement for investigative journalism, where journalists work collectively with the newsroom.
- Smart Mobility
    - Numerous crowdsourcing initiatives based on traffic conditions are presently active. Any Smart Phone with the Google Maps 10 application sends anonymous data to Google to measure traffic conditions (Alt, Shirazi, Schmidt, Kramer & Nawaz, 2010). The application sends data on the phone user's geographical position and the speed the user is travelling at by using the phones GPS. *CarTel* uses an integration of accelerometers and GPS to collect data on road conditions such as potholes and speed bumps (Christin, Reinhardt, Kanhere & Hollick, 2011). *PetrolWatch* is a more specific and simpler initiative that captures data on the changes of petrol prices in different locations (Dong, Kanhere, Chou & Bulusu, 2008). Participants contribute to *PetrolWatch* by sending a photo of the petrol price. One may argue that restricting input to photos alone may lower the level of participation (as a camera phone is required); however, it ensures a certain level of validity in the participant's contribution as one cannot simply make-up an amount. Crowdsourcing has not only been used to assists drivers but also cyclists.
    - *Biketastic* and *BikeNet* use a cyclist's mobile phone to measure their speed, position and the number of calories burnt (Christin, Reinhardt, Kanhere &

Hollick, 2011). These initiatives not only improve cyclists' experiences but may even encourage others to undertake cycling as an alternative means of transport or simply for health benefits.

- Smart People
    - *LiveCompare* is a crowdsourcing initiative that compares the prices of grocery products in different shopping outlets (Deng & Cox, 2009). Participants contribute to *LiveCompare* by accessing an application on their mobile phones and taking a photo of the product's price and barcode. When a picture is taken the application automatically adds a time stamp and captures the geographical location of the phone when the photo was taken. The barcode is tagged to text when *LiveCompare* users search for products. When a user searches for a product the application displays the price and the closest locations of purchase.
    - Another initiative that also seeks to improve a shoppers experience is *Fashism*. *Fashism* is a crowdsourcing initiative that acts as a social network for fashion advice (Alt, Shirazi, Schmidt, Kramer & Nawaz, 2010). Users take photos of outfits in a store and upload them on *Fashism*, where anyone is free to comment on the picture.
- Smart Environment
    - Crowdsourcing has many capabilities in assessing aspects of the environment. *NoiseTube* and *NoiseSpy* use a mobile phone's microphone and GPS to capture noise pollution in a location. These initiatives are aimed at identifying noisy areas that could cause changes in one's behaviour or hearing loss (Maisonneuve, Stevens, Niessen & Steels, 2009; Kanjo, Bacon, Roberts & Landshoff, 2009).
    - *PollutionSpy* is used to measure weather conditions such as wind speed and temperature (Kanjo, Bacon, Roberts & Landshoff, 2009). The data collected from *PollutionSpy* is more specific than an average weather report, as the readings are more centralized (time stamped and geo-tagged). Note that although the data collected is more specific, the quality of the data collected is lower than the sensors used at weather stations.
    - *Haze Watch* is an air pollution-specific crowdsourcing project that captures levels of sulphur, carbon and nitrogen in the air (Hollands, 2008). Although the data collected is not as accurate as the air pollution sensors used at the

meteorological station, *Haze Watch* can monitor sudden changes in air pollution which, at times, cannot be detected by the meteorological station.

- Another similar crowdsourcing initiative relating to air pollution is *PEIR* (Personal Environmental Impact Report), which presents the phone users with exposure to air pollution (Mun, *et al*., 2009). *PEIR* captures one's geographical location and based on the speed of movement the application can determine the mode of transport used by the phone user. Data is then transferred to a central location and used to calculate one's exposure to air pollution.

- Smart Living
  - A number of crowdsourcing initiatives are focused on monitoring health issues. A crowdsourcing project known as *DietSense* is used to help participants seeking to lose weight by recording their eating habits (Reddy, Parker, Hyman, Burke, Estrin & Hansen, 2007). The participants wear their mobile phones around their necks as a necklace which captures visual and audio data. *DietSense* automatically captures pictures of food presented in front of the participant to approximate the volume of food and how much was consumed. *DietSense* also captures the time taken to consume the meal, as well as location and sound. This data is used to decipher whether the participant is at a restaurant, food court or at home sitting on the sofa. The data collected is sent to a central location which can only be accessed by the participant. The participant can then select that data he/she wishes to share with the doctor by removing any unwanted data.
  - *Ushahidi* is a crowdsourcing project used to assist the police in the fight against violence (Alt, Shirazi, Schmidt, Kramer & Nawaz, 2010). The public can submit data on violent acts to the police through SMS, e-mail or web link. This initiative is currently running in Kenya and has proven to be successful in assisting the police to reduce violent behaviour.
  - *CSI (Crowdsourcing Safety Initiative)* is another crowdsourcing project currently running in East London, South Africa. Where *Ushahidi* only focuses on violent crimes, *CSI* has a much broader scope. The public provide data on any incident which is currently or has the potential to become a public safety issue. The public report any public safety incidents

through a phone call. This research will focus on this particular project which will be discussed in greater detail in section 3-3.

- Smart Government
    - An initiative known as *Government 2.0* utilises the knowledge of crowds to gain their insight on various government processes (Nam & Sayogo, 2011). This allows the government to better understand what the public expect from them and also identify problems they were unaware of. The intention of *Government 2.0* is to improve decision making and policy making, and resource allocation by considering information gathered from crowds. *Government 2.0* gathers data from crowds in four areas: competitions, wikis', social networks and social voting. The public are free to voice their opinions and knowledge in any four areas.
    - Similar crowdsourcing initiatives to *Government 2.0* are used to assist Government but are too many to mention. Two separate ideas have been discovered which can be used together to assist the Government but no such initiatives are in progress at this point. These ideas make use of "cheering meters" (sound level meters) (Barkhuus & Jørgensen, 2008) and interactive displays (Schneider, Moraes, de Souza & Esteves, 2012) to collect data on audience thoughts when Government speeches are given to the public. The cheer meters can be used to discover the amount of support the audience gives to the speaker by measuring the level of noise from clapping and cheering. Sophisticated interactive displays incorporate technologies that can measure audience body language and facial expressions. Incorporating these ideas when Government speeches are delivered will allow the Government to discover how their speeches and ideas influence the audience.

It is clear to see from the discussion above that crowdsourcing is capable of contributing to the Smart City. It can also be concluded that any components of crowdsourcing can be used effectively; however, it is suggested that components should be chosen based on the environment in which the crowdsourcing initiative will be applied. The next section will discuss the details of the CSI participatory crowdsourcing project according to the components of crowdsourcing.

3.8 CSI Participatory Crowdsourcing Project

The University of Fort Hare and IBM are pooling their resources together to run a public safety crowdsourcing pilot study in East London, South Africa. East London is a small district situated in Buffalo City, located in the developing country of South Africa. Figure 3-3 indicates the location of East London with the letter "A".

Figure 3-3: East London (Google Maps)

East London's coastal location and deep water river harbor has attracted businesses and created economic growth for this city. For example, East London harbor accommodates the Mercedes Benz production plant with the import of production parts and export of finished goods. According to Statistics South Africa, the city is gradually growing in population. In 2000 the population was 715 800, rising to 734 980 in 2005, and is now standing at 761 996 as of 2010 (Eastern Cape Socio Economic Consultative Council, 2012). Although the city's population continues to increase, Buffalo City remains a small developing city in comparison to other cities in South Africa.

It was informally found that public safety issues go unreported because people are not aware of whose responsibility it is to resolve the problem. It was also found that East London has a public safety department (http://www.buffalocity.gov.za/municipality/dep_publicsafety.stm) but the majority of citizens are unaware of such a department. This is possibly due to lack of promoting of the department to increase awareness as the public assume that all emergency and non-emergency units are separate entities. The Buffalo City public safety department claims responsibility for the police force, fire department and disaster management. These factors make East London an interesting city to pilot a participatory crowdsourcing Smart City initiative.

The CSI participatory crowdsourcing project is best explained using the components of crowdsourcing integrated in Figure 3-2. Figure 3-4 will be used to explain how the CSI participatory crowdsourcing project was deployed.

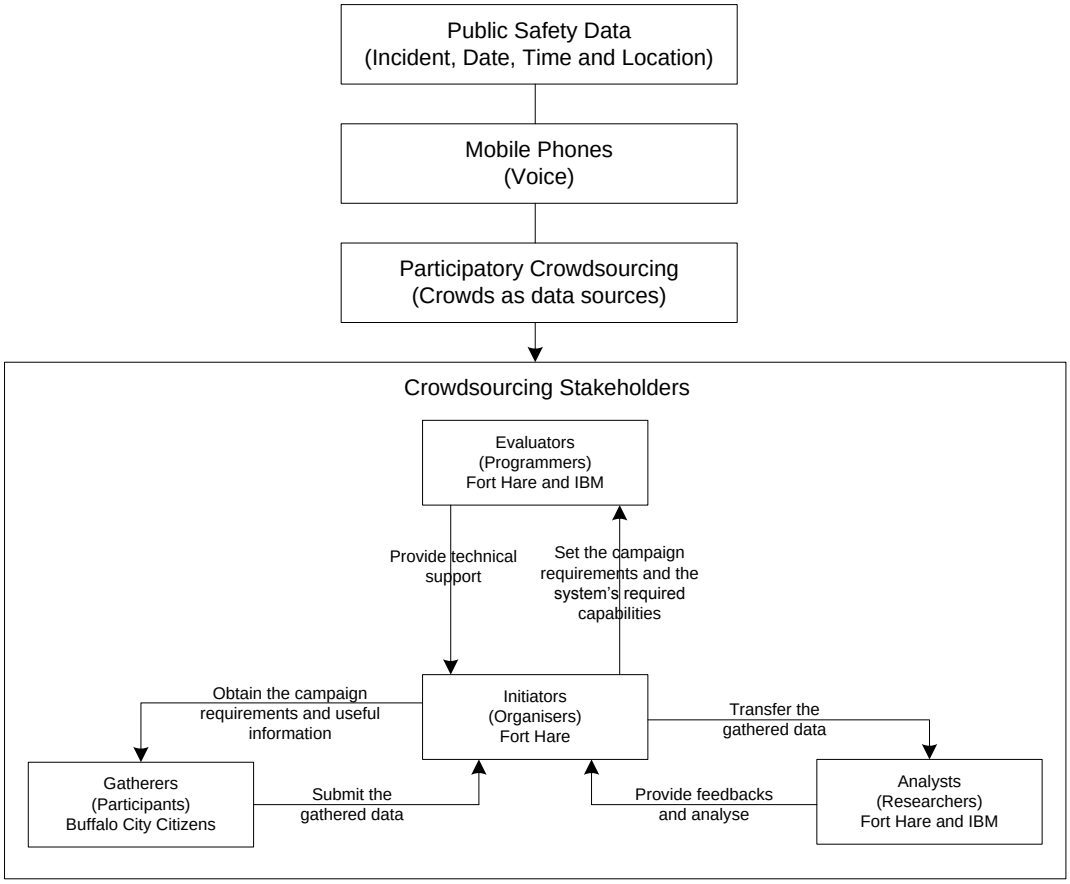


Figure 3-4: CSI Participatory Crowdsourcing Project

The first thing that needs to be decided is exactly what data will be collected from the crowds. In the case of the CSI participatory crowdsourcing project, public safety data will be gathered. Therefore, the initiative will contribute towards the Smart Living area of a Smart City. The crowdsourcing initiative is required to accept any public safety incidents and because of the vast array of public safety issues, general data common to all safety incidents should be collected to avoid data overload. Simultaneously, the data must be sufficient enough to facilitate an effective and efficient response. Through conversational analysis between IBM and the University of Fort Hare, it was found that the data should include: type of incident, date and time of the incident and geographical location. This simple and timeless request also ensures that participants can report an incident with little effort and therefore, facilitating and encouraging participation, as crowdsourcing is easier than other methods of reporting by the public. The next step was to choose the data collection method for example, social networks, mobile phones or blogs.

The CSI participatory crowdsourcing project will utilise mobile and landline phones for data collection. The next decision was how the chosen data collection method would be used to collect the data; this is influenced by the type of crowdsourcing used. The type of crowdsourcing utilised by the CSI participatory crowdsourcing project will be *crowds as data sources* (participatory sourcing). This method was chosen as all other methods were found to be impractical. Therefore, the CSI participatory crowdsourcing project is limited to reporting public safety incidents through text messaging (SMS) or email. Because of South Africa's low literacy rate, text messaging may restrain some participants from reporting and therefore, making a phone call was decided as an appropriate method for reporting for the CSI participatory crowdsourcing project. Additionally, reporting through speech is faster and less effort than typing a message. It is safe to assume that the majority of the public are capable of making a simple phone call from a mobile or landline phone. When phone calls are used as a data collection method one must decide if calls will be managed by computer interaction (message prompts), human interaction or a combination of both. In the case of the CSI participatory crowdsourcing project, participants will be directed by message prompts (computer interaction).

The CSI participatory crowdsourcing project will make use of an advanced IVR system known as the Spoken Web. IVR systems are an interactive telephonic interface (Greeff, Coetzee & Pistorius, 2008) in which pre-coded messages are provided to the user, who supplies an audio input to the system (Dawes & Rowley, 1998). The Spoken Web serves as a substitute to the World Wide Web by sending and receiving data to people through their telephones (Agarwal, Kumar, Nanavati & Rajput, 2011). The Spoken Web is capable of allowing users to create "audio websites" commonly known as VoiceSites. The simple development interface and audio embedded hyperlinks allow novice users with no web development background to develop VoiceSites and makes for a seamless navigational experience for users. However, the CSI participatory crowdsourcing project will not make use of this function since callers are only required to send data as opposed to receiving data. Instead, the CSI participatory crowdsourcing project will take advantage of the Spoken Web's advanced voice recognition system when developing the message prompts that will direct the caller to effectively respond. Therefore, the advanced IVR system can be seen as the intermediary between the different crowdsourcing stakeholders for the project.

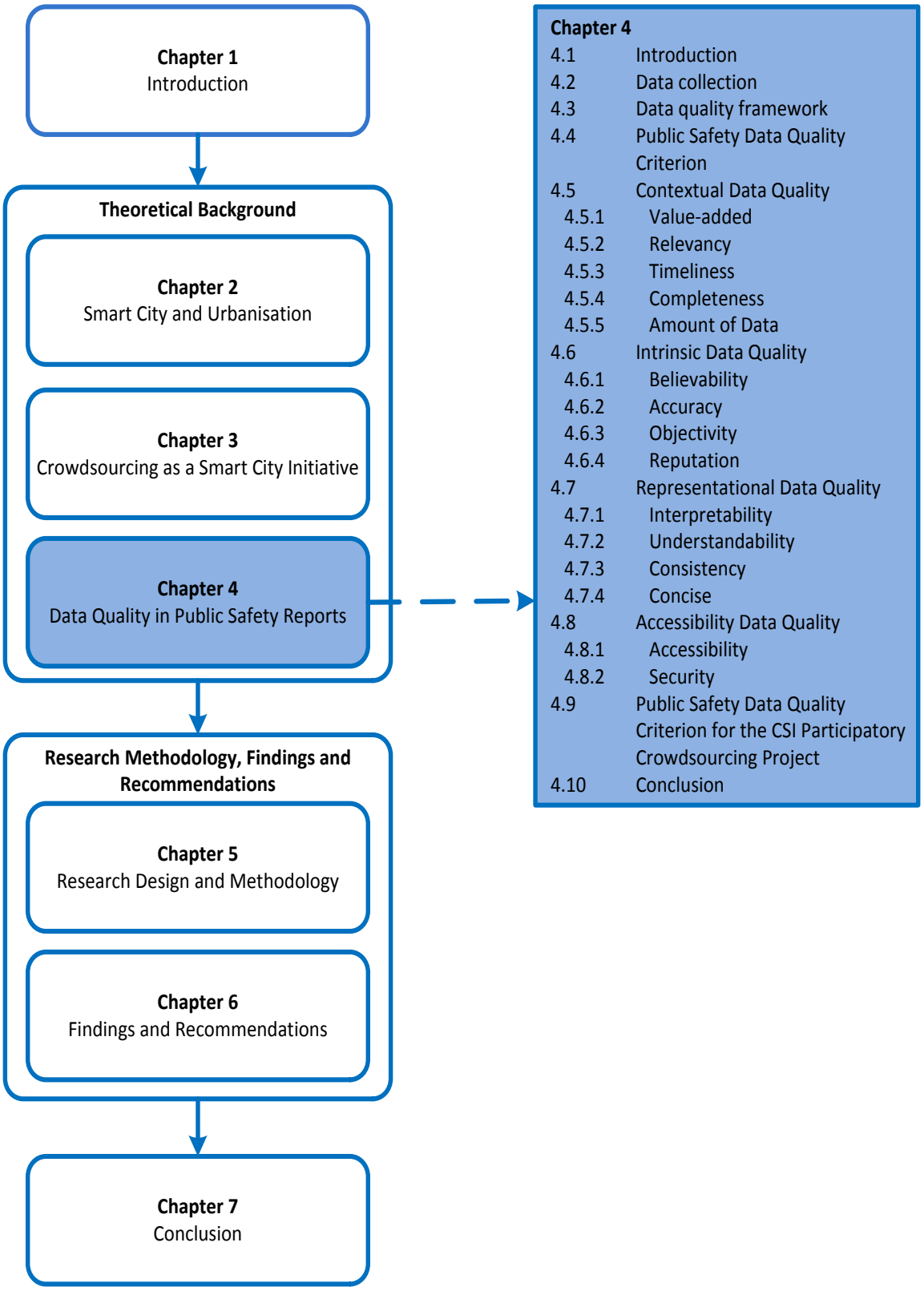
Figure 3-2 illustrates that there are four types of stakeholder groups in a typical crowdsourcing initiative. As mentioned above, an entity can occupy the role of more than one stakeholder group, which is usually the case in many crowdsourcing initiatives. The CSI participatory crowdsourcing project also follows this stakeholder map with the University of Fort Hare acting as the *initiators*, while the University of Fort Hare and IBM play the joint role as *analysts* and *evaluators* (strictly in terms of setting up the Spoken Web), and the *participants* include the citizens of East London. The Fort Hare stakeholder team includes one Professor, one Doctor, two information system lecturers, one PhD student and one Master's student. The IBM stakeholder team is composed of one researcher, one analyst and one programmer.

The objective of this CSI participatory crowdsourcing project is to utilise the citizens' geographical dispersion and mobility to act as human sensors that notify emergency and non-emergency units of any public safety issues in the city. The public safety data collected will be used to resolve public safety issues, identify patterns and appropriately distribute emergency and non-emergency units. This will result in effectively and efficiently managing a city's emergency and non-emergency response resources.

3.9 Conclusion

Crowdsourcing as a Smart City initiative has the potential to improve the management of city resources; however, it is dependent on data collection. Data captured on various activities occurring in the city can be used to improve decision making on resource generation and allocation. Therefore, the review of low quality data will result in ineffective decisions on managing city resources. Crowdsourcing projects must be managed effectively for quality data to be collected. This requires a collaborative effort from businesses, citizens and the Government, as well as a correct method of data collection and type of crowdsourcing initiative. The crowdsourcing map which illustrates all the components that a crowdsourcing project should consider will ensure the success of a crowdsourcing project's ability to collect data. However, data collected may not be of quality. This will be the focus area in the next chapter.

Chapter 4 – Data Quality in Public Safety Reports



4.1 Introduction

Data consumers use data to gain knowledge of a certain context so that an appropriate action can be taken (Burke, *et al*., 2006). Chapman (2005) states that data is often used uncritically without considering its quality which results in inaccurate information, bad decisions, erroneous results and increased costs. Conventionally, data are considered to be of quality if it satisfies the needs for its use (Wang & Strong, 1996). This statement is consistent with other authors' definitions of data quality, using the phrase "fit for use". Note that data quality is described according to a collection of attributes that the data possesses (for example, accuracy, timeliness, relevance and reliability). Therefore, data quality assessment methods are focused on assessing the presence of relevant data quality attributes within the data. Data quality assessments are not only conducted to determine the quality of the data, but also to improve its quality or the quality of future data. Improvements in data quality either relate to error prevention or data correction (Chapman, 2005).

Shilton *et al*. (2008) explain that the difficulty with crowdsourcing data is not the collection process, but influencing the data consumers that the data is of quality. Dalvi and Suciu (2007) state that crowdsourcing data usually lacks quality, creating new challenges to meet data consumers' expectations. Conducting a data quality assessment on the data will influence the perception of data consumers that the data is of quality, provided that the assessment verifies this. McGilvray (2008) explains a logical approach to assessing data quality: identify the relevant data quality attributes, construct assessments for each data quality attribute and then combine all individual attribute results from the assessment. Therefore, firstly one must identify the relevant data quality attributes as it is not feasible and/or not possible to assess all certain attributes given the context and type of data. Secondly, each attribute must be assessed individually to easily identify problem areas with the data. And lastly, all results from the assessment must be combined so that an effective decision can be made to improve the quality of data.

This chapter is focused on developing a Public Safety Data Quality criteria for the CSI participatory crowdsourcing project that collects public safety reports from citizens in an audio format. The Public Safety Data Quality criteria is constructed based on Wang and Strong's (1996) Data Quality Framework, which lists 15 data quality attributes that are important to a data consumer. Before, the Public Safety Data Quality criteria is developed it is necessary to explain the CSI participatory crowdsourcing project's data collection

process. A discussion on how Wang and Strong's (1996) framework will be used to construct the Public Safety Data Quality criteria will follow the data collection process. Thereafter, the parameters of the Public Safety Data Quality criteria will be presented before discussions focussing on the construction of the Public Safety Data Quality criteria are presented.

4.2 Data collection

This section will discuss the method used by the CSI participatory crowdsourcing project to collect public safety data. A large percentage of the population in developing countries have low literacy levels and do not have access to the internet; however, there is an ubiquity of phone usage, which continues to proliferate (Agarwal, Jain, Kumar & Rajput, 2010). This has created an opportunity to assist people deprived from accessing the internet by providing them with an alternative to the World Wide Web (WWW). Although Smart Phones and other WAP (Wireless Application Protocol) enabled mobile phones have access to the internet, these devices have not reached all members of the public, especially in developing countries the where majority of the population uses standard mobile phones. The Spoken Web platform is not restricted to Smart Phones and other WAP enabled phones.

The vision of the Spoken Web is to create an internet environment that is accessible from an ordinary phone (Agarwal, Kumar, Nanavati & Rajput, 2011), mobile or landline. An advanced IVR system, known as the Spoken Web, will be utilised to collect data from the public. An IVR system is an interactive telephonic interface that provides the user with pre-coded messages and accepts audio input or dial tone as a response from the user (Dawes & Rowley, 1998). The Spoken Web is a sophisticated IVR system with two distinct differences; firstly, the Spoken Web can connect numerous Voice Sites through hyperlinks and secondly, the Spoken Web provides the option for a two-way interaction by allowing users to be data generators as well as data consumers. Concerning the CSI participatory crowdsourcing project, the public will only use the Spoken Web to generate data on public safety issues. Therefore, the public will not interact with the Spoken Web as data consumers. In other words, the Spoken Web will not produce any output besides the pre-recorded messages used to instruct the user on how to provide an input. Figure 4-1 conceptualises how the sophisticated IVR system will be used for the CSI participatory crowdsourcing project.

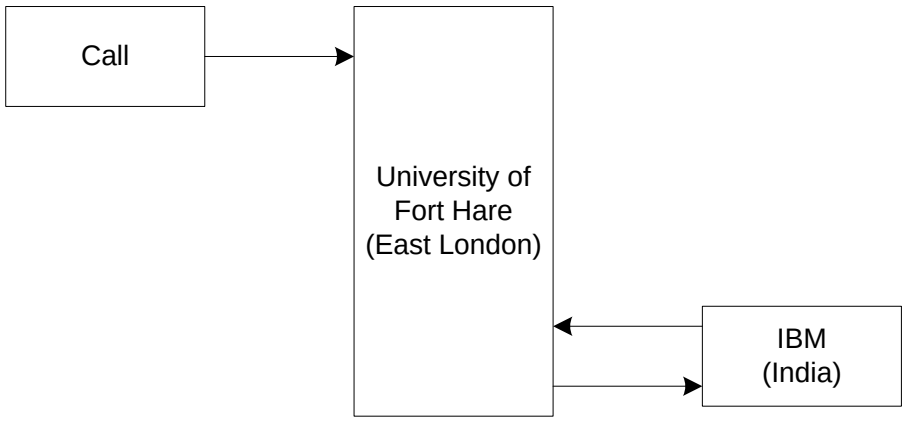


Figure 4-1: Data Collection Process

The data collected will be in audio format (.wma). Once a citizen ends a call the data is passed through a portal hosted by the University of Fort Hare (in East London, South Africa) into a gateway to IBM, New Delhi, India. As soon as a call ends, an audio file will automatically be generated. Therefore, based on the audio file's time stamp, one will be able to determine the time the call was made. An employee at IBM will then upload the audio files to the University of Fort Hare every 72 hours for analysis. Weekly back-ups of the audio files will be undertaken at IBM and at the University of Fort Hare. Due to technical complexities found through conversational analysis, this process was considered the most simple, efficient and effective method for data collection.

The audio interface is vital for data collection, as this will direct the user to provide the required data about the public safety issue. Greeff, Coetzee and Pistorius (2008) have found that users have difficulty understanding the message prompts at times. The presentation of message prompts to the user must be clear, ensuring that no ambiguity is present. This is not a simple task as lengthy message prompts and countless options usually create user dissatisfaction (Lobo, Doke & Kimbahune, 2010). Therefore, the whole instruction process must be clear and short. Numerous articles (Greeff, Coetzee & Pistorius, 2008; Farrell, Rajput, Das, Danis & Dhanesha, 2010; Abu-Hasaballah, James & Aseltine Jr., 2007) have expressed methods of developing an audio interface that will ensure message prompts do not negatively affect the satisfaction of the user. These methods need to be considered in conjunction with other methods that ensure quality data is captured by the system. This will ensure that quality data is collected without compromising user satisfaction.

4.3 Data quality framework

Wang and Strong's (1996) findings have appeared in more articles on data and information quality than any other work, suggesting that they are pioneers in the field of data and information quality. The consequences of poor data quality and the issue that firms tend to focus solely on accuracy when addressing data quality have created a research opportunity. Kahn and Strong (1998) explain that data quality is usually assessed by IS (information systems) professionals and data consumers. They found that data consumers have a broader perception of data quality than IS professionals. In addition, Katerattanakul and Siau (1999) claim that data providers have a different perception of data quality from data consumers. They continue to explain that data consumers should be the primary focus, as they will ultimately decide whether the data is of quality. Strong, Lee and Wang (1997) support this proclamation by explaining that it is not possible to assess data quality without considering the data consumer. This has provided them with the opportunity to create a framework for data quality.

Wang and Strong (1996) developed a data quality framework that depicts the attributes of data quality important to a data consumer. Although Wang and Strong's (1996) framework was originally developed for data quality, numerous articles (Gendron & Shanks, 2003; Peng, 2002; Najjar & Schniederjans, 2006) use the framework when explaining information quality attributes suggesting that the attributes of data quality important to data consumers are consistent with those of information quality. However, this study will utilise Wang and Strong's (1996) framework to assess the quality of public safety data.

The data quality attributes were assessed and organised into four distinct categories. The derivation of the four categories relates to previous literature on data quality and data consumers' experiences (Wang & Strong, 1996). The four categories in which the attributes of data quality were organised include: contextual data quality (1), intrinsic data quality (2), representational data quality (3) and accessibility data quality (4). These data quality categories will be discussed in detail below:

1. Contextual data quality
    - Data consumers often protest that the data does not align with its expected use (Katerattanakul & Siau, 1999). This category of data quality denotes that data must apply to the task, depending on the context in which it will be

in use. Concerning the CSI participatory crowdsourcing project, the context in which data applies will be public safety issues. The main attribute supporting the construction of this category is the relevancy of data. Wang and Strong (1996) explain that tasks in a certain context are unique, and at times the context itself differs from others. This category refers to providing data that is strictly required to perform the task that needs completing. They emphasise that achieving data quality in this category is a challenging task as each task has different data requirements.

2. Intrinsic data quality
    - Accuracy is the main attribute of intrinsic data quality, as this forms part of the literature review used to construct this category of data quality. Katerattanakul and Siau (1999) explain that accuracy is vital as inaccurate data leads data consumers to question the data source. Therefore, intrinsic data quality refers to the validity of origin of the data. In terms of the CSI participatory crowdsourcing project, the data source will be the citizens, as they will be reporting on public safety issues. Wang and Strong (1996) found that intrinsic data quality goes beyond the traditional view of quality, that objectivity and accuracy are the only attributes that belong in this category. They explain that objectivity and accuracy are not enough to ensure high quality data and that believability and reputation are important to data consumers in terms of intrinsic data quality.
3. Representational data quality
    - Representational data quality refers to the presentation format and meaning of the data to the data consumer. The format of the data refers to the conciseness and consistency of the data, while the meaning of the data relates to the ease of understanding and interpretability of the data. Wang and Strong (1996) state that the meaning of data involves semantic reconciliation, and the format relates to syntax. This suggests that the system and the databases used by the system influence representational data quality. The CSI participatory crowdsourcing project will be using an IVR system and a single database to store all public reporting data.
4. Accessibility data quality
    - Accessibility data quality assumes the systems and the data located in the databases is accessible but also secure. This category of data quality

considers the action of accessing the data and the methods used to restrict user accessibility. Katerattanakul and Siau (1999) state that privacy and confidentiality are the main reasons why access restrictions are imposed on users. In terms of this category, the users refer to the data consumers. Accessibility data quality relates to the data consumer's accessibility of the database through the system. Restricting user access to certain data will ensure data integrity, and by doing so maintain quality data. In connection with the CSI participatory crowdsourcing project, the users of the IVR system will be the emergency and non-emergency units.

The attributes of data quality that are important to a data consumer are organised into the four categories (intrinsic, contextual, representational and accessibility) discussed above, and shown in Table 4-1.

Table 4-1: Data Quality Framework (Wang & Strong, 1996)

Category	Data Quality Dimensions
Contextual Data Quality	Value-added
	Relevancy
	Timeliness
	Completeness
	Appropriate Amount of Data
Intrinsic Data Quality	Believability
	Accuracy
	Objectivity
	Reputation
Representational Data Quality	Interpretability
	Ease of Understanding
	Consistency
	Concise
Accessibility Data Quality	Accessibility
	Access security

The Data Quality Framework in Table 4-1 only lists the attributes of data quality without any formal definition, because some attributes have an ambiguous meaning and therefore interpretations may vary. Table 4-2 provides a definition of the data quality attributes listed in the Data Quality Framework.

Table 4-2: Data Quality Attributes Defined (Pipino, Lee & Wang, 2002)

Category	Attributes	Definition
Contextual Data Quality	**Value-Added**	extent to which data is beneficial, provides advantages from its use
	Relevancy	extent to which data is applicable and helpful for the task at hand
	Timeliness	extent to which the data is sufficiently up-to-date for the task at hand
	Completeness	extent to which data is not missing and is of sufficient breadth and depth for the task at hand
	Amount of data	extent to which the quantity or volume of available data is appropriate
Intrinsic Data Quality	**Believability**	extent to which data is regarded as true and credible
	Accuracy	extent to which data is correct, reliable and certified free of error
	Objectivity	extent to which data is unbiased, unprejudiced and impartial
	Reputation	extent to which data is highly regarded in terms of source or content
Representational Data Quality	**Interpretability**	extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
	Understandability	extent to which data is clear without ambiguity and easily comprehended
	Consistency	extent to which data is presented in the same format and compatible with previous data
	Concise	extent to which data is compactly represented without being overwhelming (i.e. brief in presentation, yet complete and to the point)
Accessibility Data Quality	**Accessibility**	extent to which data is available, or easily and quickly retrievable
	Security	extent to which access to data is restricted appropriately to maintain its security

The definition of the data quality attributes listed in Table 4-2 provides clarity on what makes up the four categories of the Data Quality Framework. The above discussion explains the way in which the CSI participatory crowdsourcing project can ensure data quality. This is summarised in Table 4-3 below.

Table 4-3: Research Application

Category	Focus	CSI Project
Contextual Data Quality	Task at hand	Public safety issues
Intrinsic Data Quality	Data source	Citizens
Representational Data Quality	Format of the data	Voice to text
Accessibility Data Quality	Accessible but secure	System accessibility and security

Table 4-2 will guide the discussion on data quality in this chapter by using the data attributes to develop a criteria to assess the quality of public safety data collected by the CSI participatory crowdsourcing project. The next section will discuss the rules for the development and use of the Public Safety Data Quality criteria.

4.4 Public Safety Data Quality Criteria

Wang and Strong's (1996) data quality framework will be used to guide the development of the criteria to manually assess the data provided in public safety reports from the Buffalo City citizens. A discussion on each data quality attribute in Wang and Strong's (1996) framework will be provided to develop a criteria to assess the existence of each attribute in the public safety reports. Note that the criteria is developed to assess the data quality of individual public safety reports and therefore, the criteria is based on identifying data quality problems by assessing individual data units (single reports) and not collectively as a dataset. The assessment will be developed into "yes/no" questions; "yes" indicating that the data attribute exists in the report. Therefore, if all questions result in "yes" after a public safety report is assessed, then the data provided in the report is considered to be of high quality. This method was chosen over a conventional Likert scale as assigning numerical high quality and low quality indicators are not possible for public safety data. In addition, no weights will be allocated to questions in the Public Safety Data Quality criteria as all attributes are considered equally important. Consequently any "no" result will indicate that data quality is compromised and based on the data quality attribute one would be able to identify the problem area. For example, a "no" is scored for the believability attribute of data and since it forms part of the intrinsic data category one will be able to deduce that the data quality problem relates to the data source (focus area of intrinsic data quality). The entire Public Safety Data Quality criteria will be used to manually assess data quality as certain data quality attributes require a certain level of judgement when answering the questions. One may argue that data quality assessments requiring intuition may affect the consistency of assessment results; however, the questions will be developed in an obvious manner where the person assessing the data can clearly answer all questions with confidence and consistency.

In summary, the above discussions firstly focused on emphasising the importance of the data quality in this study's context. The next section explained the data collection process of the CSI participatory crowdsourcing project. The presentation of Wang & Strong's (1996) Data Quality Framework followed, discussing the four categories (intrinsic, contextual, representational and accessiblity) of the framework and how they relate to the CSI participatory crowdsourcing project before presenting the data quality attributes and illustrating how the framework will be used in this chapter. The rest of Chapter 4 will focus on the construction of the Public Safety Data Quality criteria. The first area of

discussion will be *contextual data quality* as it is necessary to define context before assessing other categories.

4.5 Contextual Data Quality

Contextual data quality denotes that data must be relevant to the task, depending on the context in which the data will be used. For example, one cannot collect and review data on mobile phones when deciding on what television to purchase. In terms of the CSI participatory crowdsourcing project, public safety data will be collected to improve the service delivery of emergency units such as the police and fire brigade, and non-emergency units such as road maintenance teams. The data quality attributes of contextual data quality include: value-added, relevancy, timeliness, completeness and the amount of data (Wang & Strong, 1996). These attributes should be present to ensure that data is relevant for its use. This section will continue by determining how data can be accessed to determine the presence of these data attributes. This will not only reassure that data are not only relevant (contextual), but one will also be able to identify the problem area. The first attribute of data quality that will be discussed is *value-added*.

4.5.1 Value-added

This attribute of contextual data quality refers to the degree of positive outcomes that can be created from analysing the data collected (Pipino, Lee & Wang, 2002). These outcomes usually improve decision-making, realising an opportunity or identifying problems, trends or patterns. Therefore, to determine if data adds value, one must firstly identify how the data will be used and the intended decisions produced from reviewing the data. In terms of the CSI participatory crowdsourcing project, the intention of the public safety data collected is to identify public safety issues occurring in the city and discover trends and patterns to improve the distribution of Government services such as the police force and fire brigade. Therefore, to assess if data adds value to decision-making, one must predict potential outcomes that may influence decision-making. However, because of the vast possibilities (predictable and unpredictable) that data could have on decision-making, one should use a common identification as predicting specific potential patterns that could arise is not possible. For instance, a pattern identified in public safety data would be a common public safety issue frequently occurring in a specific area and time period (mornings, afternoons or at night); for example, motor vehicle break-ins in a specific street or suburb at night.

Decisions, in term of the CSI participatory crowdsourcing project, are influenced by two aspects; responding to public safety issues and identifying patterns from the reports. The common decision that must be made when reviewing a public safety report is to determine **the emergency or non-emergency units responsible for responding to the issue** and **the number of emergency or non-emergency teams that will be required to resolve the issue** (for example, three squad cars and one fire truck). The identification of patterns in data to improve decisions on how to allocate and distribute emergency and non-emergency units requires one to assess a collection of reports. However, the Public Safety Data Quality criteria will be used to assess individual public safety reports and therefore, a method to identify whether patterns can be identified from a report must be found. Numerous methods can be used to identify patterns; the methods which can be used for the public safety data include: Classification Method (1), Associations Method (2) and Clusters Method (3) (Wasan, Bhatnagar & Kaur, 2006). These methods will be discussed below. Note that authors have different naming conventions for methods.

1. The Classification Method organises data collected into pre-defined categories (Wasan, Bhatnagar & Kaur, 2006). This method of identifying patterns is restricted to specific patterns which are known before the data is categorised. Since there are numerous possibilities and types of public safety issues, it is difficult to create a list of public safety categories. However, it is possible to create vague public safety categories if one is not interested in identifying specific patterns. For example, Graves (2010) created four broad categories of public safety (non-offensive events, offensive events, false events and other events), which were further organised into more specific public safety issues. Therefore, if one is searching for specific patterns then this would be an appropriate method for identifying patterns in data. Thus, this classification method can be used to identify public safety patterns regarding time and location data.
2. The Association Method identifies reaction patterns in data (Wasan, Bhatnagar & Kaur, 2006). For example, if events A and B occur, event C usually transpires afterwards and therefore, it is assumed that events A and B usually cause event C. A public safety example of an Association Method could be an increase in infrastructure and road damages after heavy rainfall (flood damage). Therefore, this method can be used to predict increases in specific public safety reports based

on the occurrence of an external factor, such as weather or public safety issues that have the potential to create additional public safety issues.

3. The Cluster Method is a knowledge discovery process where data is sorted into categories, which are discovered by reviewing data (Wasan, Bhatnagar & Kaur, 2006). This differs from the classification method (pre-defined categories) as the categories are defined after data is reviewed based on common attributes between data segments. The cluster method is the most appropriate method for identifying patterns as public safety issues have a vast variety of different categories. This will allow one to determine the level of specification when developing the categories of public safety issues.

Note that the analytics conducted on the public safety data will identify the patterns. However, this section intends to identify public safety data segments required to allow for patterns identification. Based on the discussion above it is clear that one can find predicable and unpredictable patterns in public safety data. However, identifying patterns requires the discovery of common data attributes and therefore, specific data attributes need to be present in all reports to find patterns in data. It is difficult to identify and present (usually in the form of graphs) patterns from detailed data and therefore, the common data attributes should be simple. In terms of the public safety data, the best method to assess if patterns can be identified is to ensure that common data is found in all public safety reports. This data includes **type of incident, date, time and location**. After each data quality attribute in Wang and Strong's (1996) framework has been discussed, a table will be presented to indicate how a data quality attribute will be assessed. All these tables have been consolidated in the next chapter to create a completed version of the Public Safety Data Quality criteria. Table 4-4 summarises the findings from the discussion on value added data in the form of questions that will be used to assess if the public safety data adds value to emergency and non-emergency services. The next section of contextual data quality will discuss the relevancy of data. This will follow the presentation of Table 4-4.

Table 4-4: Value-added Assessment Criteria

Value-added	Assessment
Public safety issues	Is it possible to determine the team accountable for responding to the report?
	Is the report clear enough to anticipate the number of response teams required to resolve the issue?
Patterns	Can the type of incident be determined?
	Can the date the incident occurred/observed be determined?
	Can the time the incident occurred/observed be determined?
	Does the report mention the suburb/highway at which the incident occurred/observed?
	Does the report mention the street/landmark at which the incident occurred/observed?

4.5.2 Relevancy

Relevancy in terms of data means that the data must be applicable and helpful for its intended use. To determine if data are relevant, one must firstly know what the data will be used for, and then find a correlation between the data and its use (Burke, *et al.*, 2006). The CSI participatory crowdsourcing project will collect data on public safety issues to improve the service delivery of emergency and non-emergency units in East London. Therefore, one must discover what is considered a public safety issue in order to determine if the data collected is relevant.

The broad spectrum of public safety possibilities makes it difficult to develop a list of all public safety incidents. Therefore, the most appropriate approach to determine if a phenomenon is regarded as a public safety issue is to create a criteria, and any event that fulfils the criteria is considered a public safety issue. However, this becomes problematic as citizens can report potential public safety issues in addition to those that are currently a problem. For example, a large pothole is reported as a public safety issue as it compels vehicles to dodge the object by swerving into the oncoming traffic lane and therefore, all pothole reports are considered a public safety issue. Potential public safety issues refer to an event that is currently not harming anyone but may be hazardous in the future. Therefore, potential public safety issues should be measured on the **level of possibility that the issue can escalate into a problem**. Thus, a certain level of judgement is required when assessing the data in order to determine its relevance. Table 4-5 illustrates the assessment questions to determine if public safety data are relevant for its intended use. The development of an assessment criteria for the timeliness of data will follow Table 4-5.

Table 4-5: Relevancy Assessment Criteria

Relevancy	Assessment

Public Safety Event	Does the reported incident pose a threat to citizens? *If Yes, skip next question (Potential Public Safety Event).* *If No, answer next question (Potential Public Safety Event).*
Potential Public Safety Event	Is there a high possibility that this incident could escalate to a point of endangering citizens?

4.5.3 Timeliness

Timeliness as an attribute of contextual data quality refers to data being up-to-date for its intended use. Timeliness represents the time when the phenomenon occurred and the time the data are ready for processing or analysis (Reddy, *et al*., 2008). Timeliness of data is an important attribute for the CSI participatory crowdsourcing project as public safety incidents will not always be reported immediately after occurrence or identification. For example, the participant may not have access to a mobile phone (low battery, no signal or no airtime) or they may simply be in a fast moving vehicle. Based on the time gap between incident identification and the time of reporting, is important to determine if the nature of the incident has changed.

Data are considered timely when velocity is lower than currency (Cappiello, Francalanci & Pernici, 2002). Velocity in terms of data timeliness is the time data remains valid, while currency refers to the capture time of the data. Therefore, the velocity of public safety data would be the time during which the reported incident has remained unchanged (has not escalated further than described in the report). Currency constitutes three time events, the age of the data before it was collected, the time the data consumer receives the data when requested, and the time taken to input the data (Ballou, Wang, Pazer & Tayi, 1998). In other words, currency = age before capture + delivery time + input time. Therefore, the timeliness attribute of data quality considers the period from data collection to the time it is available for use. Since data is timely while currency is lower than velocity, the criteria for timeliness in data will focus on assessing if the currency is lower than the velocity of public safety reports. Therefore, the criteria will relate to three aspects of currency: age before capture (1), the input time (2) and the delivery time (3), which will be developed in the following discussion.

1. The age of the data refers to how old the data are before collection (Ballou, Wang, Pazer & Tayi, 1998). The age of the public safety data will be the time the citizen takes to report the public safety issue from the time the incident was identified. For example, a citizen witnessed a public safety issue on Monday but only reported the incident on Friday. If the incident required immediate attention, the data's age would have been out of date. However, this does not mean that the data is not

useful; the data can still be used to identify patterns in public safety incidents. Although it is possible for reporting delays to occur, the CSI participatory crowdsourcing project allows citizens to report incidents from their mobile and landline phones. This is intended to encourage citizens to report a public safety issue as soon as it is identified. Therefore, this measurement will also determine the effectiveness of mobile phones as a method of reporting public safety incidents. To determine the age of public safety data one will need to know when the citizen saw the incident (time mentioned in the citizens report) and when it was actually reported (time the call was made). The age is calculated by subtracting the time the incident occurred from the time the report was made **(age of public safety data = time mentioned in the report – report time stamp)**. Based on the type of incident reported, a judgement call will have to be made on whether the age of the data is timely for an appropriate response. The age is deemed appropriate if it is anticipated that the nature of the public safety issue has not escalated.

2. The input time relates to the length of time taken to transfer the data set into the system (Ballou, Wang, Pazer & Tayi, 1998). The system is expected to automatically create an audio file once a caller has completed his/her call. Therefore, the input time of the audio file is immediate. However, input time also focuses on human actions that may cause delays, such as the capturing of data into the system. Data capturers need to ensure that the data is in the system before it becomes outdated. In terms of the CSI participatory crowdsourcing project, the data capturers will be the people responsible for transcribing the data from audio to text (into a spreadsheet). This requires time as the speed of capture on individual audio files cannot be faster than the length of the audio file. The public safety reports that require immediate or timely response are most likely at risk of becoming outdated. One can calculate the input time by subtracting the time stamp on the audio files from the time when the data was captured in the system **(input time = data capture time – report time stamp)**. The same judgement used for the age of data will now consider the age of the data and the input time to deduce if the nature of the public safety issue remains unchanged.
3. Delivery time (output time) is the period of time the system takes to provide the user with the requested data (Ballou, Wang, Pazer & Tayi, 1998). Any system delays will affect the delivery time of data. As illustrated in Figure 4-1 above, the audio files will be saved on a secure database in IBM India. These files will then

be manually uploaded on a designated University of Fort Hare server. Therefore, the delivery time will refer to the amount of time taken to send the audio files from IBM India to the University of Fort Hare. It was agreed through conversational analysis that the audio files will be sent as a batch every 72 hours to the University of Fort Hare. It was decided that the transcription of the audio file to text and the analytics be performed on the batch before the next batch is sent to the University of Fort Hare (i.e. within 72 hours). Therefore, a new set of audio files should be **accessible every Monday**. This will be used to test the system's capability to provide the data in a timely manner. Since the audio files will be transcribed by the University of Fort Hare, the transcribed files will be readily accessible once the transcribers have completed the transcription of the weekly audio batch.

Although it would be more convenient to assess the timeliness of data collectively (consider age, input time and delivery time simultaneously) as the judgement is the same for all three cases, this would make it difficult to discover the specific problem area affecting the timeliness of data. Therefore, it is best if the three phases are assessed individually. The next section will discuss the completeness attribute of data quality. Discussion on completeness will follow the presentation of Table 4-6, which illustrates the criteria to assess the timeless of data.

Table 4-6: Timeliness Assessment Criteria

Timeliness	Assessment
Age	Is the time mentioned in the report less than the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?
Input Time	Is the time the data was captured less than the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?
Delivery Time	Are all audio files for the week available for transcription every 72 hours?

4.5.4 Completeness

Data is considered complete if no data is missing and it is appropriate for its expected use. Data completeness in sensor data exists when all the segments of data (geographical location, time and sensor reading) are integrated accordingly (Klein & Lehner, 2009). Since humans act as sensors, the mobile phone could cause failures or malfunctions. For example, the battery of the mobile phone may die during the report. In this case it is expected that the caller will call back once the mobile phone is functional again. However, there are situations when one experiences signal problems that only affect the call receiver. This most likely will affect the completeness of data in terms of device malfunctions. In addition, human error could jeopardise the completeness of data. For example, certain data

need to be provided clearly by the caller in order for an effective response to occur. Therefore, before one can assess the completeness of data, one must determine what is considered to be complete data.

Ząbkowski (2010) states that data completeness can be considered in two dimensions, quantitatively and qualitatively. Quantitative completeness refers to the completeness of data sets, while qualitative completeness relates to the completeness of individual data sets. For the purpose of the public safety data, both quantitative and qualitative completeness will be assessed to ensure precise completeness as well as collective completeness. Quantitative completeness (1) and qualitative completeness (2) in public safety data will follow.

1. Quantitative completeness refers to all the data being present in a data set. For example, a full list of customer names should be found in a data set containing customer names. Since it is not possible to determine if the system has captured all calls made by the public, the quantitative completeness test will focus on the transcribed reports. The **amount of calls** collected should be equal to the **amount of transcribed reports**, as the transcribed reports are a textual replication of the audio calls.
2. Qualitative completeness can be viewed as a more detailed assessment of completeness as it measures data completeness in each data segment (type of incident (a), time (b) and location (c)).
    a. The type of incident requires a certain level of judgement when assessing its completeness. One should be able to **understand the extent of the incident** for qualitative completeness to exist in the data segment. For example, an active electric cable causes the cable to dangle around on the side walk.
    b. The time the incident occurred/identified includes the **date and the time**, since callers may only report the public safety issue two or three days later. If the report mentions the day (example; Monday or Tuesday) as the date, it can be assumed that the incident was seen in the current week. Likewise, if a date is not provided, it will also be assumed that the caller identified the public safety issue on the current day that the call was made. Although these deductions may be reasonable to presume as true, that assumption will affect the completeness of data. Therefore, one should be able to clearly

determine the specific day of the week that the incident occurred. This is the same case for the expression of time; the specific time of day is required in certain situations where the seriousness of the incident may change rapidly. Since the caller is not expected to be aware of the dynamics of public safety issues, the caller should provide the time of day in all reports. The time should specifically mention or emphasise that the issue occurred before-midnight (a.m.) or after-midnight (p.m.). For example; the public safety incident occurred in the morning, afternoon or night.

c. Location should be described in such a way that the average person should be able to find the location on a map. This means that the report must clearly indicate the position of the public safety incident. For example, if a report states that there is an uncovered manhole on the sidewalk in Vincent (suburb), the location of the incident is considered incomplete as numerous manholes are dispersed around Vincent. The reporter must be more specific and state the **street** name as well as the **suburb**. Another appropriate method for stating the location of the incident would be to mention a unique **landmark** and describe the distance and direction from the landmark; for example, a shop name or gas station in the area. This method will most likely be used in public safety cases that occur on highways where the closest turn-off will be used as a landmark to express the position of the incident. The purpose of stating the location is to find the public safety issue with as little effort as possible.

The discussion of data completeness has allowed for a criteria to be developed to assess the completeness of public safety data. The criteria is shown in Table 4-7, before the last attribute of contextual data quality (amount of data) is discussed.

Table 4-7: Completeness Assessment Criteria

Completeness	Assessment
Quantitative Completeness	Is there an audio version of the report?
	Is there a textual version of the report?
Qualitative Completeness	Is it possible to understand the extent of the public safety incident?
	Can the specific date (example: 3 January) be determined?
	Can the specific time period (am or pm) be determined?
	Does the report include an area (example: suburb or highway) name?
	Does the report include a street name or landmark?

4.5.5 Amount of Data

Amount of data refers to the quantity of data and the extent to which the data is of an appropriate volume. Amount of data can be measured in size (bytes, kilobytes or megabytes) of the data unit (Russell, 2010). However, in the case of audio data, the size of the data unit may differ based on the quality of the audio. For example, the size of a high quality audio file may be twice the size of a low quality version of the same audio file. Note that if the level of quality is consistent in all audio data units, measuring the amount of data in bytes would be acceptable. Measuring the length (time) of the data unit may be a more appropriate method for measuring the amount of data when assessing audio data. The length of the data unit usually correlates with its size (example: 3 megabyte audio file is equivalent to 5 minutes of audio), assuming the level of audio quality is consistent in all files. The next step is to determine the appropriate length parameter for the amount of data in a data unit.

The appropriate length parameter refers to the amount of detail the caller provides in the public safety report. The report should be clear and ensure no data are excluded while avoiding a lengthy descriptive report. For example, when stating the type of incident the caller explains: "There was a robbery at a shop that involved two robbers and a teller. The robbers were in possession of firearms, fortunately no citizens were harmed. The robbers managed to take all the cash from the register and fled in a white car. The car had no number plate however; I did manage to see a dent on the back bumper as well as a left broken tail light." Although this is an extremely detailed and helpful report, the caller is only expected to provide additional data that could affect the dispatch decision. Using the same example, the caller only needs to mention that there are two armed robbers, who used a white getaway car with a visibly damaged back bumper and a broken left tail light. This will allow one to identify that the issue is a police matter (robbery), the response team must approach with caution (armed robbers) and the unit should be on the lookout for a white car with specific identifiers. Therefore, any data that resembles "waffling" would constitute an inappropriate amount of data. In addition, other unpredictable factors could affect the length parameter of a data set.

Some individuals are capable of reporting an incident at a fast pace, while others speak in a moderate to slow speed. This may be because English is their second language (the Spoken Web for this project will only accept English speaking candidates) or they simply need time to gather their facts together to ensure that the correct data is provided. An

appropriate amount of data will also be affected by the type and nature of the public safety incident. Additionally, traumatizing events could make it difficult for the caller to provide the data in a timely manner.

When considering these possibilities it is difficult to determine an appropriate length of time to assess the amount of data. The best method to assess the amount of data is to decipher the useful data from the unnecessary data. Note that data are not considered useful if it simply relates to the public safety issue; more specifically the minimum data needed to understand the public safety issue and what could **affect the dispatch decision** is considered useful. This requires a certain amount of judgement as useful data may differ based on the type and nature of the public safety issue. Table 4-8 presents the criteria used to assess the amount of public safety data. This concludes the section on the contextual data quality attributes. The data attributes that form the intrinsic data quality category will be introduced after Table 4-8.

Table 4-8: Amount of Data Assessment Criteria

Amount of Data	Assessment
Data Volume Parameter	Does all the additional data provided influence the dispatch decision?

4.6 Intrinsic Data Quality

Intrinsic data quality refers to the validity of the data provided by explaining that data must have accuracy (Wang & Strong, 1996). If data is inaccurate one begins to question the validity of the data source. In terms of the CSI participatory crowdsourcing project, the data sources (data providers) are the citizens of East London as they will be reporting any public safety issues. The attributes of intrinsic data quality include, believability, accuracy, objectivity and reputation (Wang & Strong, 1996). The existence of these attributes will gain the data consumers' confidence about the data source's capabilities to provide high quality data. Therefore, the assessment criteria developed for intrinsic data will ensure that citizens provide valid public safety data. Believability is the first attribute that will be discussed, followed by accuracy, objectivity and reputation.

4.6.1 Believability

Believability refers to the degree to which the data is accepted as true. Prat and Madnick (2008) explain that believability in terms of data is composed of three parts, trustworthiness, reasonableness and temporality. Trustworthiness (1) refers to the data

consumer's faith in the data source to provide high quality data. Reasonableness (2) relates to the degree to which data is considered realistic. And lastly, temporality (3) refers to the time believability of the data. These three aspects will be used as a guide when developing the data quality criteria for believability.

1. Trustworthiness is assessed according to the knowledge domain of the data source. For example, IT related data is provided by an IT professional (data source). In terms of the CSI participatory crowdsourcing project, it is difficult to determine the knowledge domain of the data provider as any citizen can report a public safety issue. The message prompts of the CSI participatory crowdsourcing project are expected to assist (by directing) the caller to provide the necessary data required to respond to the problem. It is assumed that all members of the public are capable of reporting any public safety issue if they follow the message prompt request. Therefore, the only method of testing trustworthiness in the CSI participatory crowdsourcing project is to determine if the report includes all the data requested by the message prompts. The caller's capabilities to follow instructions (message prompts) correctly will reflect in their capacity to report a public safety issue. The message prompts will ask the caller to provide data on the **type of incident**, **date and time** of occurrence/discovery and the **location** of the issue.
2. Reasonableness is assessed through one's intuition that the data collected is realistic. Assessing reasonableness accurately is difficult as reasonableness is usually influenced by the data consumer's knowledge base and experiences. For example, a public safety report which sounds unreasonable to one person may be realistic to another based on related past experiences or knowledge of a similar event. In addition, the caller may exaggerate a public safety report to ensure a timely response and in doing so, create the illusion that the report is unrealistic. The likelihood that callers may exaggerate a report is the reason why reasonableness needs to be assessed. In addition, the reasonableness assessment of data will also identify any prank calls from legitimate calls. To ensure consistency in the assessment of reasonableness, any report which is found to be **extremely unrealistic or impossible** will be considered unreasonable data. For example, a shark attack in a shopping mall (located 10 km away from the sea).
3. Temporality relates to the time believability of the data based on the occurrence of a popular event/s. For example, the price of baking flour has suddenly increased as

a result of a shortage in wheat. The increase in the price of baking flour may be considered as "unbelievable" if the shortage of wheat is absent. A public safety example of such a situation can be explained by an increase in flood reports after a heavy rainfall. Note that in terms of public safety reports, temporality cannot always be explained as it is highly possible that certain events occur at random times without the reaction of a certain event. Although temporality is an effective method to assess believability, it cannot be applied to all situations. Therefore, temporality will only be used to verify the believability of data in certain cases. In this case, temporality will merely be used to **support** any unreasonable (extremely unrealistic or impossible) reports. Therefore, temporality assessment will only be conducted if a "no" is answered for the question constructed for *reasonableness*.

The findings from the discussion above has been summarised in Table 4-9. Note that consistency needs to be applied when assessing this data quality attribute. A discussion on the accuracy of data will follow the presentation of Table 4-9.

Table 4-9: Believability Assessment Criteria

Believability	Assessment
Trustworthiness	Can the type of incident be determined?
	Can the date the incident occurred/observed be determined?
	Can the time the incident occurred/observed be determined?
	Does the report mention the area (suburb or highway) at which the incident occurred/observed?
	Does the report mention the street at which the incident occurred/observed?
Reasonableness	Is the report realistic and likely to occur given its expressed location and time? *If Yes, skip next question (temporality).* *If No, answer next question (temporality).*
Temporality	Does the occurrence of a certain unrelated incident (example: heavy rain caused floods) still make reasonableness in the report "false"?

4.6.2 Accuracy

Accuracy refers to the degree of error-free data and correctness. Accuracy assumes that the data is complete; however, the data may have errors or may be incorrect (Pipino, Lee & Wang, 2002). Accuracy is difficult to assess without verification; for example, the accuracy of the public safety reports would be assessed by physically sending a response team to the site to verify that the data provided in the report is accurate. Therefore, in the context of the CSI participatory crowdsourcing project, a public safety report is considered accurate if the data provided is capable of eliciting an accurate dispatch decision by the necessary authorities. The best method to assess accuracy from the report is to identify obvious errors. Note that the analytics conducted on the public safety data will identify

unwanted patterns in the data as well as anomalies. However, certain factors negatively affecting accuracy cannot be discovered through analytics, but may be found after considering the following discussion on data accuracy. Accuracy will be assessed by dividing each report into the data segments requested from the caller: type of incident (1), location (2) and time (3). These three groups will be discussed further below.

1. The data provided for the type of incident is considered to be accurate if errors do not affect the understanding of the type of public safety issue. The purpose of understanding the type of public safety incident is so that the **correct response team** (police or fire fighters) can be dispatched as well as an appropriate **number of emergency or non-emergency units**. One should also be able to determine the **seriousness** of the public safety issue so that priority can be given and the problem can be addressed before the nature of the incident changes. This requires a certain level of intuition from the data consumer who will assess the reports for accuracy.
2. An emergency and non-emergency unit responding to a public safety issue should be able to find the location with little effort. Mispronunciations or unmatched locations (for example; correct street name but incorrect suburb, or the landmark does not exist in the stated suburb) are errors that affect the accuracy. To assess the accuracy of the location data provided, one should easily find the place of the incident on a **standard street marked map** (Google Maps for example).
3. The time segment in the report should express the date and time of the public safety incident. The accuracy of time data can only be measured to a certain extent from assessing the report. The assessment is based on the reasonable assumption that citizens will not delay reporting a public safety issue more than a week after they have witnessed it. Although citizens may report an incident a week later, a delay in reporting may change the nature of the public safety issue, which will affect the accuracy of the type of incident segment of the report. Therefore, the accuracy assessment for the time section in the report will seek to determine if the incident described in the report is likely to change from the time the incident occurred/spotted to date. One can make such an assessment by comparing the time stamp of the report and the time provided in the report **(time accuracy = time stamp – report time)** to determine if the reporting window is an appropriate length of time.

Based on the discussion above, a certain level of judgement is required to assess the data accuracy of the type of incident reported, while the accuracy of location and time will be a more objective assessment. Table 4-10 illustrates the assessment criteria for accuracy in public safety data. Objectivity as a data quality attribute will be discussed after Table 4-10.

Table 4-10: Accuracy Assessment Criteria

Accuracy	Assessment
Type of Incident	Can the correct response team accountable for addressing the issue be identified?
	Can the number of response teams that need to be dispatched be identified?
	Can the nature (seriousness) of the public safety issue be determined?
Location	Can the location of the incident be found on a map without difficulty?
Time	Is the length of time of the time stamp less the report time appropriate in terms of the likelihood that the nature of the issue has changed?

4.6.3 Objectivity

Objectivity relates to the data source providing unbiased, unprejudiced and impartial data. Sackmann (1991) used numerous people to assess the same data to find any unbiased, unprejudiced and impartial feelings that could affect results. This method is also applicable to data sources by collecting the same data from numerous sources. In term of the CSI participatory crowdsourcing project, objectivity can be determined by finding repetitive reports on the same public safety incident. However, due to the system ensuring anonymity, there is no way to determine if the same person reported the issue more than once or if it was reported by different individuals. One could discover repetitive reporting on the same incident by matching the voice characteristics of the reports. However, this method remains ineffective as this only ensures data objectivity of collective reports on the same event. Another successful method to assess objectivity which has been applied in numerous situations such as law and medicine, is the assessment of non-verbal communication.

Non-verbal communication such as body language, hand gestures, eye contact and voice tone can be used to assess one's unbiased (Fiske, 2002), unprejudiced (Benkert & Peters, 2005) and impartial (Grabe, Samson, Yegiyan & Zelenkauskaite, 2009) behaviour. Since the data source will report public safety issues in an audio format, visual assessments (of non-verbal communication) for objectivity cannot be conducted. Therefore, the voice tone of the data source is the only non-verbal communication method that can be assessed. According to Paddock, Terranova and Giles (2001), objectivity is compromised if one's **voice tone is sharp or loud**. However, a caller reporting in a noisy area is likely to raise

their voice to ensure the report can be clearly heard by the person receiving the call. The sharpness and loudness in one's tone as a result of emotional anger or frustration, is capable of affecting data objectivity. Therefore, the objectivity of data will only be compromised if the caller's voice tone is sharp or loud with the **absence of any noisy background**. If noisy background is present, the data consumer assessing the report for objectivity will need to identify any **anger or frustration** in the caller's tone, as these emotions are likely to cause unbiased, unprejudiced and impartial reporting. The last attribute of intrinsic data quality refers to the reputation of the data source. This attribute will be discussed after the presentation of Table 4-11.

Table 4-11: Objectivity Assessment Criteria

Objectivity	Assessment
Voice Tone	Is background noise absent? *If yes, skip next voice tone question.* *If no, answer next voice tone question.*
	Is the caller's voice tone absent of aggression?
	Is the caller's voice consistently calm and soft throughout the report?

4.6.4 Reputation

Reputation refers to the trustworthiness of the data provided by the data source (Angeles & García-Ugalde, 2012). Ishida (2011) states that datasets which are willingly shared with others, are considered to be more reputable than datasets provided from requests. Nonetheless, one cannot assume that because citizens are providing the data freely that the data source is highly reputable. For example, if this notion is used to assess the reputation of public safety data, prank calls will be considered as highly reputable as the "data" is provided willingly. Yang, Zhang and Roe (2011) state a reputation management system/model can be used to indicate high and low quality data from a crowdsourcing initiative based on the data source's reputation. The model works by assigning ratings to a data source, based on the quality of data it provides. The data sources are provided with specific data quality rules, the data is then analysed after the data consumer collects the data, and feedback of the quality of the data is provided back to the data source. The data source is then assigned a reputation ranking based on a logical average formula that considers all data collected from that person or group, which is used to indicate their overall reputation. The average ranking encourages the data source to improve their reputation ranking. This has proven to be a successful method to assess data quality in terms of reputation; however, the method does have its limitations. If a data source wishes to remain private, this method cannot be used to assess the reputation of the data source.

Therefore, this method cannot be used in the CSI participatory crowdsourcing project to assess the reputation of the data source.

Gackowski (2006) explains that there are three alternative methods to assess the reputation of the data source: replicate, warrant or verify the data. Replication is a future reputation assessment method where the data is stored for future tests when errors in the data are discovered. This assessment method is not feasible as the reputation is required to be assessed in the current course of time. Warrant is the level of risk (financial or criminal for example) the data source is exposed to if one finds the data to be of low quality. The level of risk can be used to indicate the reputation of the data source. Warrant cannot be used in the CSI participatory crowdsourcing project to assess reputation as the data source will not have anything to lose by providing low quality data. Verification assumes that data is highly reputable if data sources are professional in the relevant field to which the data relates. For example, accounting data is assumed to be highly reputable if it is provided by an accountant. In terms of the public safety data, any member of the emergency and non-emergency units is regarded as a professional. Although Shilton *et al*. (2008) argue that participants in a participatory crowdsourcing project are experts in their environment, they do not know what data the emergency or non-emergency unit will require to effectively act on the report. Participants may know more about the incident than any other person, but they may lack the ability to express this knowledge to the emergency and non-emergency unit. Citizens are expected to report public safety issues by providing the required data requested by the message prompts. Although the message prompts are constructed in a manner that allows any caller to provide the data to resolve the public safety issue, one cannot assume that the data source is capable of providing quality data. Therefore, it is not possible to accurately assess the reputation of the data source due to the anonymity of callers. This concludes the discussion of the attributes included in the intrinsic data quality category. The next data quality category refers to representational data quality.

4.7 Representational Data Quality

Representational data quality refers to the presentation of the data collected. This relates to the format and the meaning of the data. The CSI participatory crowdsourcing project intends to collect the data telephonically through an automated interface. Therefore, the data will initially be saved in an audio format (.wma files) and then transcribed into text in a tabular format. The data must be flexible enough to be transcribed into tabular format for data to possess reputational quality. Therefore, the data must be capable of being

represented appropriately and as required by the data consumer. Representational data quality attributes include, interpretability, understandability, consistency and conciseness (Wang & Strong, 1996). The discussion on reputational data quality will continue with the interpretability attribute.

4.7.1 Interpretability

Data is considered to be interpretable if the data consumer can understand the language, symbols and definitions used in the data. Interpretability usually becomes a problem when various forms of technical data need to be analysed (Strong, Lee & Wang, 1997). Technical data refers to any data that requires interpretation from a professional in the respective field. For example, patient data uses medical terminologies and diagnoses which only a person with a medical background (doctor or nurse) can understand. In some cases, data requires interpretation by multiple professionals in different fields. If patient data needs to be interpreted by a person with insufficient medical background, the data will need to be presented in an easily understandable manner so that this person will be capable of interpreting the data. An example of public safety data is reporting codes which police assign to the occurrence of common events (for example; "possible code 10 in progress" is code for a robbery). In addition to coded and technical data, language and cultural barriers can make interpretation difficult. Certain data requires local interpretation based on their language or cultural background. Language is not an issue with respect to crowdsourcing data on public safety as the system is restricted to English speaking candidates. However, if a caller provides a report in any language besides English, the interpretation of the data will be difficult. The public safety data may require experienced emergency or non-emergency personel to interpret the data so an effective dispatch decision can be made; however, it is assumed that an experienced emergency or non-emergency personel will be appointed to such a task. Under this assumption, the only difficulties that emergency or non-emergency personel will experience with interpreting the data would be if the public safety report is in **another language other than English**. Although the caller will be notified that the report must be provided in English, some callers may, directly or indirectly, provide **portions of the report in another language** (for example a mixture of English and Xhosa words). This is most likely to occur by callers whose first language is not English. The next section will focus on the understandability attribute of data quality. This will follow Table 4-12 which illustrates the assessment criteria to determine if data is interpretable.

Table 4-12: Interpretability Assessment Criteria

Interpretability	Assessment
Full Language	Is the majority of the report in English?
Partial Language	Is the report absent of any words or phrases from a language other than English?

4.7.2 Understandability

Data is considered easy to understand if the data is clear (1), unambiguous (2) and comprehendible (3) (Pipino, Lee & Wang, 2002). The criteria for the understandability of public safety reports will focus on these three areas. The three factors of understandability will be discussed further below.

1. Since the public safety data will be provided in audio format, any form of noise could affect the clarity of the data. Noise refers to any unwanted effects that create disruptions in the original intention of the message. Therefore, noise does not only include **background noise** such as car engines, radio (music) or people talking. **Radio interferences** are common when phones are used. Radio interference occurs when a nearby mobile device (for example, laptop, tablet or mobile phone) exchanges data (connect to the internet) while a call is in progress from a mobile phone (such as calling to report a public safety issue). Many mobile and landline phone users have experienced this type of noise at some point in time. The nearby mobile device emits a signal interruption that creates a disturbing tone during the call. Any calls made outdoors in **windy conditions** may also disrupt the clarity of the report. Strong winds that blow into the talking piece of the phone at times can be loud enough to drown the intended message of the call. Note that it is difficult to decipher the difference between wind and a caller's **heavy breathing** and therefore, both will be considered as **wind noise**. Other types of "noise" which could be experienced in reports include the caller's **accent and the use of slang.** A strong accent from a caller may make it difficult to understand some sections or the entire report. Although popular slang may be understood in the context of the report, it was found that certain slang is ambiguous and vague which could make the report difficult to understand. In addition to human and external noise, the system itself must also be assessed for the absence of noise. **Lag time** is common in many technologies and in this case it could affect the quality of data by distorting the public safety report. To ensure that the data in the report is clear, the report must be free from noise. In terms of audio data, noise includes radio interferences, wind, strong accents and slang.

2. Ambiguity exists when any data provided can be interpreted in more than one way and it is difficult to determine the intended meaning. Ambiguity can be found in three parts of the report, the type of public safety issue, the time and the location. Any person assessing the report should be able to understand the type of public safety issue from a **single viewpoint**. This is an important factor in the CSI participatory crowdsourcing project as it is not possible to verify the report from the data source due to anonymity. This factor will reduce participation, if the caller believes that the issue was reported but the necessary authorities took no interest in addressing the problem. Ambiguity in terms of the time that the incident was identified is present if the report does not emphasise if the issue occurred **in the morning (AM) or in the evening (PM)**. Location ambiguity exists when the street name or landmark is provided **without a suburb name**. For example, Pick n Pay (grocery chain store) was used as a landmark to describe the location and since there are a number of Pick n Pay outlets, it is difficult to determine which store was mentioned in the report. In addition, it is common for street names to repeat in different suburbs. Note that the absence of a street name or landmark while the suburb is expressed is considered vague or incomplete data and therefore, it is not considered in the assessment of ambiguity.
3. The caller's use of words and descriptive capabilities will affect the comprehensibility of the public safety data. This includes the use of **acronyms** or a short version of a word or phrase that is not recognisable. In addition to acronyms, any speech impediments that interrupt the flow of words will make it difficult to grasp the public safety report. This includes stuttering, lisps, mumbling, or lengthy pauses during the report. These speaking disorders are likely to occur in people whose first language is not English. This problem must be considered, as the CSI participatory crowdsourcing project is restricted to English speaking candidates. In addition to **speech impediments** that make data difficult to comprehend are the caller's **vocabulary and use of grammar**, which could also pose a problem in candidates whose English is not fluent. The assessment of speech impediment and vocabulary and grammar usage will not only test if data are easy to comprehend, but will also illustrate the effects of restricting the reporting to English speaking callers.

Table 4-13 is a summary of the above discussion in the form of an assessment criteria to determine if public safety data collected is understandable. The next section will discuss the consistency of data. This will be discussed after the presentation of Table 4-13.

Table 4-13: Understandability Assessment Criteria

Understandability	Assessment
Clarity	Is background noise absent?
	Is radio interference absent?
	Is wind created noise absent? (including caller breathing distortion)
	Does the caller have a clear accent?
	Is the use of slang absent?
	Are system glitches absent? (example: lagging)
Ambiguity	Can the type of public safety issue be interpreted in a single point of view?
	Does the time express where the incident occurred in AM or PM?
	Does the location state the suburb or highway and street name or landmark?
Comprehensiveness	Is the caller's use of acronyms appropriate to comprehend the report?
	Are speech impediments absent?
	Is the caller's vocabulary and use of grammar appropriate to comprehend the report?

4.7.3 Consistency

Consistency states that all data units in a dataset should be in the same format and compatible with other datasets already collected (Pipino, Lee & Wang, 2002). Note that the majority of authors focus on the database and network capacities of maintaining consistent datasets as the data passes through different networks, applications and users. For example, if changes are made to one file, the changes should be visible in copies of that file in the system. Therefore, assessments for data consistency relate to technical tests on the database and networks (transferral and storage of data). Katerattanakul & Siau (1999) measured data consistency on websites as the constant use of the same font size, font type, layout and use of graphics on all web pages. The message prompts used for this CSI participatory crowdsourcing project assist the collection of consistent data by asking the caller to provide specific data such as the type of incident, date, time and location. Therefore, consistency in public safety reports requires all individual reports to state these segments of data. Consistency tests can be made more specific by assessing the order of the data segments. For instance, the message prompts ask the caller to provide the type of incident (first data segment), the time (second data segment) and the location (third data segment). However, the ordering of data segments is not required at this stage of the CSI participatory crowdsourcing project. For instance, ordering would be considered important if the reports were automatically categorised by the system. Therefore, the assessment will not consider the ordering of data segments as inconsistent data. The last attribute under the representational data quality attribute is the conciseness of data. This will follow the

presentation of Table 4-14, which illustrates the assessment criteria for the consistency of data.

Table 4-14: Consistency Assessment Criteria

Consistency	Assessment
Type of incident	Can the type of incident be determined?
Time	Can the date the incident occurred/observed be determined?
	Can the time the incident occurred/observed be determined?
Location	Does the report mention the area (suburb or highway) at which the incident occurred/observed?
	Does the report mention the street/landmark at which the incident occurred/observed?

4.7.4 Concise

Conciseness refers to the capability to summarise data while maintaining the ability to understand and interpret the data (Pipino, Lee, & Wang, 2002). To assess the conciseness in data, one must be able to provide the minimum data possible by maintaining an understanding of the data after summarisation. Conciseness usually affects data quality when data of different formats need to be merged into a single format (Niermann, 2005). In terms of the CSI participatory crowdsourcing project, the reports must be presented in a visual form of some sort. This will allow for an easier assessment of reports as one will not need to re-listen to an audio file every before reviewing the report. The public safety data in reports should mention the three data segments but do so in a summarised format. Niermann (2005) states that conciseness can be tested by assessing the data's capability to be presented visually through a graph or table using keys to further summarise data. Regardless of the method chosen to present data, it must be displayed in a brief format, requiring the minimum amount of data. An example of a table format can be seen in Table 4-15.

Table 4-15: Presentation of Public Safety Reports

Report Number	Type of Incident	Time	Location
xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx

Each column presented in Table 4-15 must present the data in a consistent manner. Therefore, the criteria will be based on summarising similar data segments on a consistent basis.

- The report number column in Table 4-15 is used as an index to the audio and transcribed files. Therefore, a report number should be unique to each report. Each audio file will have a unique name; therefore, the name of the audio file will be used as the report number.
- Emergency and non-emergency units usually have codes for specific public safety incidents; however, these codes do not cover all types of public safety incidents and some situations require more than one code to describe the situation. Based on the type of incident and location, one should be able to understand how the issue is endangering the public. Therefore, the **type of incident** should only state the issue endangering the public and not explain how it is creating a public safety issue. Note that any description of how an incident is creating a public safety issue may be helpful in understanding the situation but it may also limit the magnitude of the incident from the caller's experience.
- The time of the incident can be presented as the date (for example; 10 January 2012 or 10/01/12). The specific time during the day complicates presentation of time since a caller either will state a specific time, emphasise the time by describing the time of day or may not state the time at all. If the caller does not state the time, a symbol should be used to indicate its absence. Since it is not possible to determine the exact time from a caller emphasising the time in a report, the best method to present time is through **periods of the day**. For example, the incident occurred in the morning, day or night.
- The presentation of location is straightforward if the street name and suburb is provided. Any landmarks used to describe the location of the incident must be able to identify the street name from the landmark or where the incident occurred. For example; a caller reports that there is a broken concrete slab in the middle of the third road to the left of Edgars on Oxford Street. The street name of the third road should be easily identified by simply finding Edgars on a map and moving three streets left. Therefore, all locations can be presented by the street name and suburb. In terms of highway public safety issues, the street name will represent the closest off-ramp and the name of the highway can be used as the substitute for the suburb.

The discussion on conciseness concludes the representational category of data quality. Table 4-16 presents a summary of the discussion above in the form of a criteria to assess

the conciseness of public safety data. The last category is the accessibility of data, which will be discussed after the presentation of Table 4-16.

Table 4-16: Concise Assessment Criteria

Concise	Assessment
Type of incident	Can the type of public safety incident be identified?
Time	Is it possible to identify the time period in which the incident occurred/observed? *For example; morning, afternoon or night.*
Location	Was the suburb/highway stated?
	Was the street name or landmark mentioned or can it be identified?

4.8 Accessibility Data Quality

Accessibility category of data quality is based on the notion that data must be secure but accessible. This emphasises that this category is focused on the integrity of the data. Accessibility data quality is made up of two attributes, accessibility and security. Security refers to the security procedures of restricting access of data to any unauthorised users. Likewise, accessibility relates to ensuring that the data available to authorised users are accessible when required. Addressing these two attributes of data quality will ensure accessibility data quality. Accessibility data quality is the first attribute that will be discussed.

4.8.1 Accessibility

The data collected must be accessible to users when required (Pipino, Lee & Wang, 2002). Therefore, data should be rapidly retrievable with little effort from the user. This means that the data must be available at all times, unless other external rules state otherwise. For example, users are authorised to access data for a certain time period (for example, 08:00 to 17:00). Although a user may have access to data, an external rule dictates that data will only be available during certain times of the day. Restrictions can also be imposed based on the occurrence of a specific event/s for example, a fraud or corruption investigation.

Strong, Lee and Wang (1997) state that data accessibility problems are due to a lack of computing resources. Lack of computing resources refers to insufficient system resources that result in users being unable to access any authorised data; for example, unreliable networks. The lack of computing resources not only relates to the system resources responsible for the output of data to users, but also input capabilities. Insufficient system resources can also cause delays in data updates, resulting in access to outdated data. Katerattanakul and Siau (1999) found that the assessment of this attribute of data quality

could be completed by assessing the user's capability to navigate through a system interface. Although Katerattanakul and Siau's (1999) study focused on the accessibility of website content, this assessment method can be used for the CSI participatory crowdsourcing project. Therefore, to test whether data is accessible, one should expect to find the required data in the **correct location** (for example, audio files in the audio folder). Data accessibility will be compromised if one struggles to **locate the files or folders**, or has trouble **opening any the location paths**. Note that this assessment does not consider the system used by emergency and non-emergency units that provide them with access to this data. The last attribute that will be discussed refers to ensuring that data are secure. This will follow the presentation of Table 4-17 which shows the criteria that will be used to assess the accessibility of data.

Table 4-17: Accessibility Assessment Criteria

Accessibility	Assessment
File and Folder Logic	Are the naming conventions of files and folders logical? *For example; a folder named "January Week 1- Audio" includes audio files from the 1 January to 7 January.*
File Accessibility	Can all authorised files be accessed?
Folder Accessibility	Can all authorised folders be accessed?

4.8.2 Security

Data security procedures are put in place to protect the privacy and confidentiality of the data (Pipino, Lee & Wang, 2002). Privacy and confidentiality restrictions are imposed to ensure that sensitive data is not used illegally or in any harmful manner. Protecting the integrity of data reduces the risk of any unwanted changes, which could corrupt and reduce its quality (Wang & Strong, 1996). The CSI participatory crowdsourcing project ensures anonymity of callers reporting public safety issues. This means that callers are not obligated to provide any personal information. However, this does not ensure that the caller will not provide personal information either intentionally or unintentionally. Therefore, to ensure that no personal information is captured, any data that could be used to identify the caller will be removed during the transcription process. Through conversational analysis, it was found that this is referred to as the sanitisation of data. Thus, there should be no data in the transcribed reports that could be used to identify the caller. Examples of data that could be used to discover the identity of the caller include, ID number, name, contact number and residential address. All **personal data** from the caller must be removed before the data is made available to users. The data must be removed in such a manner that **data integrity** is still maintained. Data integrity is

maintained if the report can still be understood after the personal data of the caller is removed. For example, a caller reports that a person was mugged outside his shop, on Fifth Avenue yesterday. The caller mentions in the report that the incident occurred outside his shop; this is considered personal data as the identity of the caller can be traced back to the shop. Table 4-18 summarises the findings of the security discussion in the form of a criteria to assess if public safety data is secure.

Table 4-18: Security Assessment Criteria

Security	Assessment
Personal Data	Has any personal data been provided in the report? *If yes, answer next question.* *If no, skip next question.*
Data Integrity	Is the report still readable after the removal of all personal data?

This concludes the discussion on data quality attributes. The next section presents a brief summary of the data quality criteria of the segments developed in the above discussions of data quality attributes.

4.9 Public Safety Data Quality Criteria for the CSI Participatory Crowdsourcing Project

This Public Safety Data Quality criteria for each attribute developed from the above discussion is consolidated. The consolidation can be found on in Chapter 4, Table 4-4. This is the data quality criteria that will be used to assess the quality of the crowdsourcing public safety reports. The only data quality attribute excluded in the Public Safety Data Quality criteria is the reputation of the data as it was found that this could not be assessed in this context due to this project's requirements that ensure anonymity of data sources. A summary of what was covered in this chapter will be discussed in the next section.

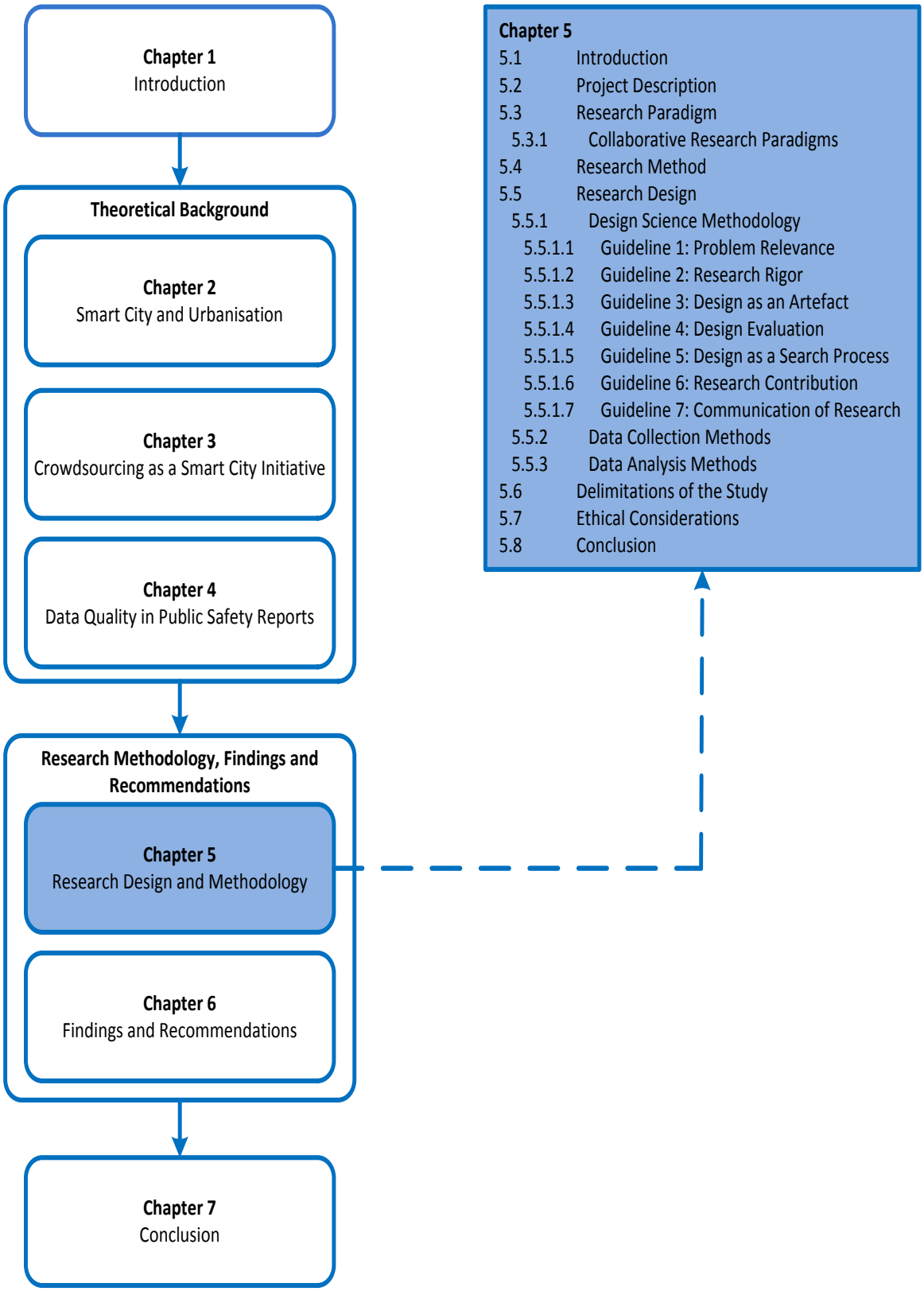
4.10 Conclusion

The majority of authors have focused on automatic data quality assessments methods (Borek, Woodall, Oberhofer, & Parlikad, 2011). However, most of these automatic assessment methods cannot be applied when dealing with audio data. Additionally, these automatic assessment methods tend to focus more on the system and network capabilities to ensure data quality rather than the actual content of the data. It was also found that some data quality attributes cannot be accurately assessed in an automatic manner as a certain level of judgment and intuition is required. Some data quality attributes cannot be assessed as a result of restrictions imposed in the form of project requirements (for

example, ensuring the anonymity of data). One should also note that although data quality attributes can be assessed there are some attributes that do not relate to the project requirements. It was found that questions developed for attribute assessment are the same or similar questions to assess other attributes. This emphasises that certain attributes are dependent on each other and one cannot exist without compromising the other (for example, the quality attributes: value added, believability, consistency and conciseness).

The Public Safety Data Quality criteria is used to assess the quality of public safety reports to identify any problem areas with the quality of data. The quality assessment assists the development of this study's research artefact (Critical Success Factors). This chapter has found that one must consider various areas when assessing the quality of data, such as the data consumer's capabilities, the context in which data applies, the presentation format of the data and the accessibility of the data. This chapter has also determined how to assess the existence of data quality attributes. Any quality problems discovered after the quality assessment will allow one to determine the factors which need to be considered to successfully collect quality data from a participatory crowdsourcing initiative. The next chapter (Chapter 5) will explain how this study will be conducted through the presentation of the research design and methodology.

Chapter 5 – Research Design and Methodology



5.1 Introduction

This chapter is focused on the research design and methodology used for this study. The specific research design and methodology chosen for this study was based on the study's problem statement and the accessibility of primary data and secondary data. Hofstee (2006) explains that there are numerous methods of investigating a study and reaching a conclusion. The research design and methodology states the specific method chosen and explains why this is the best method to approach the study, considering limitations (for example, data accessibility or time). He continues to explain that the most appropriate research design and methodology can be found by considering the problem statement and the data available. In a more descriptive layout of this chapter, a description of the CSI participatory crowdsourcing project will be provided as this is necessary to understand why specific components were chosen to conduct this research. The research paradigm adopted for this study will then be discussed, followed by the research design chosen. The next section will discuss the selection of the research design which includes the Design Science methodology and the data collection and analysis methods. The last section will list the ethical considerations before concluding the chapter.

5.2 Project Description

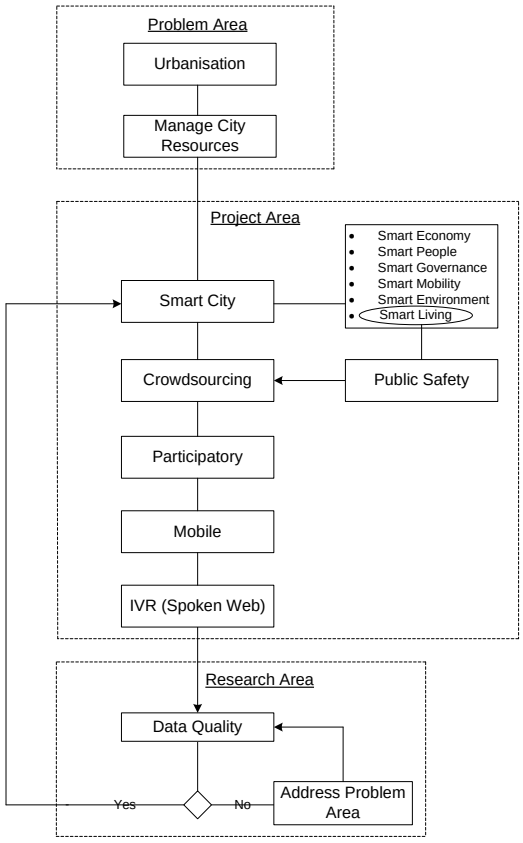


Figure 5-1: Project Description

Figure 5-1 illustrates the connection between the problem area, CSI participatory crowdsourcing project and this study. This study was aligned with a project run by the University of Fort Hare and IBM which is currently in its pilot phase. The project seeks to contribute to the Smart City initiative through the use of a specific crowdsourcing project (CSI participatory crowdsourcing project). The significance of this project can be identified by the problem area, which relates to the problems created due to increased urbanisation. As urbanisation continues to increase, and is expected to increase in the future, cities are struggling to accommodate the influx. The improved management of city resources will assist cities to cope with increases in urbanisation. The project aims to use the Smart City concept to improve the management of city resources.

A Smart City is a concept initiated by IBM which aims to improve urban lifestyle through the use of Smart Technologies. All Smart City constructs can be arranged into one of six broad, but not concrete, categories: Smart Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment and Smart Living. Since it is not viable for a single project to contribute directly to all six categories, this project is focused on public safety which forms part of the Smart Living characteristic of the Smart City. The project plans to use crowdsourcing to improve public safety services. Once this crowdsourcing model is refined, one could use the lessons learnt and work on another Smart City construct.

Crowdsourcing is a concept which takes advantage of the wide dispersion of the public by utilising them to collect data on certain events, usually occurrences in their environment. Two types of crowdsourcing exist, participatory crowdsourcing and opportunistic crowdsourcing. Due to the technological accessibility constraints of South Africa and the project's requirements, participatory crowdsourcing was considered to be the most appropriate type of crowdsourcing option. Numerous methods can be used to collect data through participatory crowdsourcing; for example, social networks, blogs and mobile application. Based on the low levels of internet access and Smart Phone possession in South Africa, simple mobile phones or landline phones were used as a method for collecting data on public safety issues. The public used their mobile or landline phones to provide data through an advanced IVR (Interactive Voice Response) system.

The advanced IVR system used for this project is known as the Spoken Web. The Spoken Web is best explained as an audio version of the World Wide Web. The system allows users to provide and access data and/or information telephonically, through mobile or

landline phones by dialling a designated number. The Spoken Web's high-level voice recognition and hyperlink capabilities make it a suitable method for data collection through participatory crowdsourcing. Although this is an appropriate method of collecting public safety data from citizens, there is a risk that the data may lack quality.

This study focused on ensuring that the correct methods are taken to ensure quality data is collected from citizens. This is an important aspect of the project, as low quality data will lead to ineffective decisions on improving the service delivery of emergency and non-emergency units. Therefore, when high quality public safety data is collected a positive contribution can be made to the Smart City. This will encourage participatory crowdsourcing initiatives to be integrated into other areas of the Smart City initiative such as Smart Economy or Smart Mobility. This will ultimately improve the management of numerous city resources and assist cities to cope with increases in urbanisation.

5.3 Research Paradigm

Research can be conducted in numerous ways based on the manner of interpretation (view) (Hofstee, 2006). Therefore, all research should be conducted according to a specific research paradigm. The research paradigm chosen will have an influence on both the nature of the research questions and how the questions will be answered (Blanche, Durrheim & Painter, 1999). Three broad research paradigms exist namely, positivist, interpretivist and critical theory. These research paradigms enforce the researcher to use specific data collection methods, perform observations in a certain manner and interpret both primary and secondary data in a specific way. Further discussions of the three research paradigms follow.

- Positivist
    - Positivist research focuses on the facts associated with the occurrences of social phenomena (Collis & Hussey, 2009). Therefore, logical reasoning is used to support all assumptions made by the researcher. This means that reality can be analysed through measurable properties, independent from the researcher (Myers, 2009). This ensures that research is conducted objectively and can be replicated.
- Interpretivist
    - Interpretivist research focuses on other's views and activities as they understand their own reality (Collis & Hussey, 2009). Therefore,

interpretivist research attempts to understand human behaviour in a specific context. Interpretivist research is conducted in a subjective manner by focusing on the meaning of the social phenomenon rather than the measurement.

- Critical Theory
    - Critical theory research focuses on the oppositions, conflicts and contradictions in social reality (Myers, 2009). It asserts that social reality is historically created and re-created by people in a consistent manner (Oates, 2006). This paradigm asserts that social reality has certain objective qualities which change experiences and the way of seeing the world; for example, political and economic systems.

Blanche, Durrheim and Painter (1999) illustrate the differences between these three research paradigms, ontology, epistemology and methodology. Ontology refers to what is known about the nature of reality to be studied. Epistemology relates to the relationship between the researcher and what can be researched. Methodology states how the research will be conducted. These three dimensions and the three research paradigms are illustrated in Table 5-1 below.

Table 5-1: Research Paradigms (Blanche, Durrheim & Painter, 1999)

	Ontology	Epistemology	Methodology
Positivist	-Stable external reality -Law-like	-Objective -Detached researcher	-Experimental -Quantitative -Hypothesis
Interpretivist	-Internal reality of subjective experience	-Empathetic -Interactively subjective researcher	-Interactional -Interpretative -Qualitative
Critical	-Socially constructed reality	-Suspicious -Political -Researcher constructs versions of reality	-Deconstruction -Textual analysis -Discourse analysis

Although there are distinct differences between the three paradigms, researchers can combine certain characteristics from these research paradigms. However, this does not mean that a single study can have two paradigms.

5.3.1 Collaborative Research Paradigms

Oates (2006) describes the clear difference between two research paradigms, positivism and interpretivism. Taking into consideration the separated characteristics of both

positivism and interpretivism, conducting research under the strict characteristics of positivism or interpretivism is extremely unlikely (Collis & Hussey, 2009). Collis and Hussey (2009) express a more realistic alternative by providing research paradigms which combine characteristics of both the positivism and the interpretivism approach. This allows one to take a more extensive view of the research context. Figure 5-2 illustrates the research paradigms mentioned by Collis and Hussey (2009).

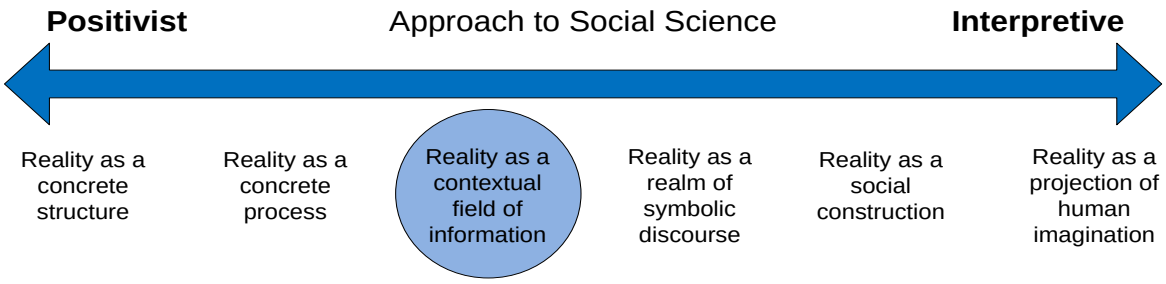


Figure 5-2: Continuum of Core Ontological Assumptions (Morgan & Smircich, 1980, p.492 in Collis & Hussey, 2009, p.51)

This research study attempted to determine the research subject's cognitive interpretation of situations and information regarding those situations. Theories and experts in the relevant fields were used to verify all logical assumptions. The reflection of cognitive interpretations and verification from theories and experts emphasise that the *reality as a contextual field of information* research paradigm was most suitable for this study. The next section will discuss the research methods commonly used with research paradigms.

5.4 Research Method

The approach in which research is conducted is commonly divided into one of three categories namely: qualitative methods, quantitative methods and both qualitative and quantitative (commonly known as mixed methods). Although many authors refer to positivist research as quantitative research and interpretivist research as qualitative research, it was explained in the above discussion that research may include characteristics of both positivist and interpretivist research. Therefore, it is necessary to indicate which research method was adopted. A discussion on qualitative methods (1), quantitative methods (2) and mixed methods (3) will follow.

1. Qualitative research is commonly used in the social sciences context to understand human behaviour and what causes such behaviour. This research method focuses on a collection of qualitative data. Qualitative data can be characterised as low in

volume and high in detail (in-depth). Open-ended questions are used that result in data being presented in a textual format. The most common methods used in qualitative research are case studies, observations, in-depth interviews, experiments, relevant literature, action research, physical experiences, theories, social interaction, surveys, Delphi Technique and expert review.

2. Quantitative research is used for natural sciences research to observe and learn about occurrences of natural phenomena. The research conducted quantitatively can usually be measured numerically by using closed-ended questions. Quantitative data can be characterised as high in volume and low in detail (in-depth). Research methods include written surveys, questionnaires, experiments, facts, methods and theories. Presentation of data is usually in a statistical (for example graphs and tables) format for statistical analysis (trends, average and totals, for example).
3. Mixed methods combine both qualitative research and quantitative research. Qualitative research and quantitative research have limitations, which can be avoided through the use of mixed methods. Mixed methods are usually adopted when assessing a broad range of research questions. The data are collected and analysed concurrently or sequentially to collectively solve research problems. Therefore, it is important to select the appropriate qualitative and quantitative research where the findings can be merged together in a viable manner.

This study adopted a mixed methods approach. This was found to be the most effective approach when considering the process required when developing the Critical Success Factors for public safety data quality. The qualitative research approaches used in this study included case studies on the Smart City and crowdsourcing initiatives, theories on data quality, and an expert review (see Design Science methodology) to assess the Public Safety Data Quality criteria (see Table 5-4) and the proposed Critical Success Factors. The quantitative research approach involved the assessment of the public safety reports collected from the CSI participatory crowdsourcing project. The Public Safety Data Quality criteria developed from relevant theories and literature on data quality were used to assess 100 public safety reports in order to identify the areas jeopardising the quality of data. The research design used directed the use of the research method.

5.5 Research Design

Hevner, March, Park and Ram (2004) present a conceptual framework which assists researchers in understanding, executing, and evaluating research in the information systems discipline. The framework is illustrated in Figure 5-3 below. All information system research incorporates people, businesses (organisations) and technology; this is shown in the far left box in Figure 5-3, which includes the characteristics of people, businesses and technology. This box is intended to ensure research relevance as these three combined groups allow the researcher to define the problem or opportunity (Hevner, March, Park & Ram, 2004). The knowledge base (box on the far right) ensures the credibility of the research by using previous information system research relevant to the researcher's study and methodology used to guide the research process. All information system research is conducted in two phases, the development or criticism of an artefact or theory, and the assessment of this discovery through appropriate evaluation methods.

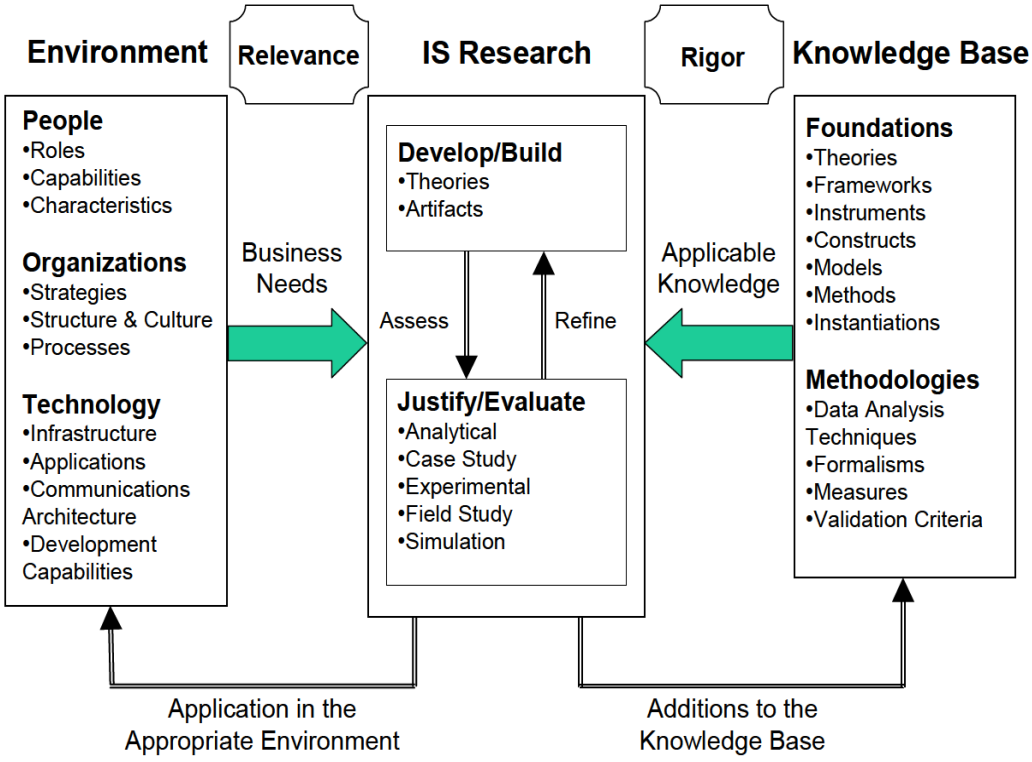


Figure 5-3: Information Systems Research Framework (Hevner, March, Park & Ram, 2004)

Hevner, March, Park and Ram (2004) refined this framework by combining the Behavioral Sciences paradigm with the Design Sciences paradigm. Hevner, March, Park and Ram (2004) explains the differences between these two paradigms which are tabulated in Table 5-2 below.

Table 5-2: Differences in Behavioral Sciences and Design Sciences (Hevner, March, Park & Ram, 2004)

Design Sciences	Behavioral Sciences
Develop artefact to solve problem/s	Develop theories to explain problem/s
Improve reality	Understand reality
Truth – Informs design	Utility – Informs theory

Note that truth and utility (last row in Table 5-2) cannot be separated, as utility may exist in an artefact based on the possibility of yet to be discovered truths. In addition, the Design Sciences paradigm addresses either unsolved problems or alternative (better) solutions for previously solved problems. When considering the differences between these paradigms, the Design Sciences paradigm was best suited for this study, the public safety data quality problem ensuring that the resources of emergency and non-emergency units are managed more effectively and efficiently.

5.5.1 Design Science Methodology

Design Science is a problem solving paradigm which "...seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished" (Hevner, March, Park & Ram, 2004, p. 76). The Design Science paradigm consists of seven guidelines which must be considered to effectively utilise this research methodology (Hevner, March, Park & Ram, 2004). The seven guidelines of Design Science are displayed in Table 5-3, which includes a description of each guideline. Figure 5-4 follows Table 5-3, illustrating the application of the Design Science methodology to this study. A detailed description of each guideline will then be presented.

Table 5-3: Design Science Research Guidelines (Hevner, March, Park & Ram, 2004)

Design Science Guidelines	
Guidelines	**Description**
Guideline 1: Design as an Artefact	Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.
Guideline 6: Design as a Search Process	The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Note that the Design Science guidelines presented by Hevner, March, Park and Ram (2004) have been re-arranged for the purpose of this study.

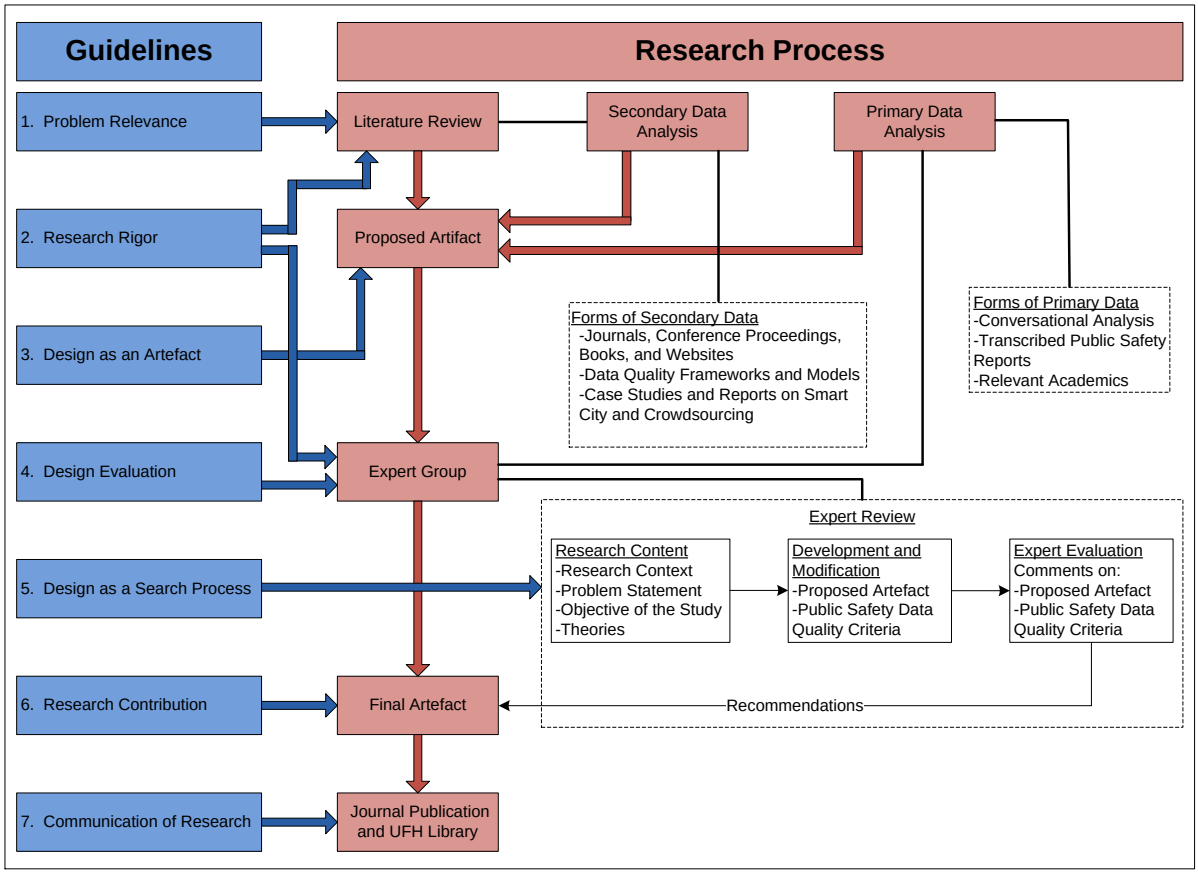


Figure 5-4: Applied Design Science Guidelines

5.5.1.1 Guideline 1: Problem Relevance

Information systems research is aimed at understanding and developing technological solutions that solve problems. Behavioural Sciences research is focused on the development of theories to explain what would most likely occur in a specific situation. Design Sciences research seeks to develop an artefact that focuses on changing the occurrence. For example, the Technology Acceptance Model is used as a theory to explain what influences businesses to accept technology in the work place. Design Sciences research is then used to develop appropriate artefacts that assist businesses to overcome acceptance problems.

The problem identified in this study refers to the scarcity of city resources because of an increase in city population. The improved management of these resources will solve this problem by analysing data collected on specific occurrences in the city environment. An optimal method of collecting data on the city's environment is by using the public to act as sensors. Nam and Pardo (2011) explain that the public are experts on their surrounding environment. In addition, the wide geographical dispersion of the public enables data collection across an extensive area (entire city). This method of utilising the public to collect data on certain occurrences in their environment is commonly known as crowdsourcing/crowdsensing. However, the data collected must be of quality to ensure that effective and efficient decisions are made. In terms of the research context, this study focused on the collection of public safety data to ensure that quality data are collected from the public and analysed to improve the management of city resources (in this case, emergency and non-emergency units).

5.5.1.2 Guideline 2: Research Rigor

Rigorous methods must be applied in the construction and evaluation of the artefact. This ensures that valid research is conducted thoroughly and the artefact designed is convincingly correct. As mentioned, research rigor must be proven during the construction of the artefact (1 to 6) and the artefact's evaluation (7 to 8). Note that two expert reviews were conducted; the first assessed the Public Safety Data Quality criteria and the second assessed the proposed artefact.

1. Theories were used to guide the research and validate any assumptions.
2. Related research also served as a guideline for the layout and direction of this study to ensure all views relevant to the research area were considered.

3. The research design adopted for this study conformed to the most appropriate philosophical view.
4. Views from academic colleagues familiar with aspects of this study were used throughout this study.
5. Conversational analysis with the CSI participatory crowdsourcing project stakeholders' (IBM team and Fort Hare University team) provided relevant information throughout this study.
6. The Public Safety Data Quality criteria was assessed by six experts, carefully selected based on their relevant educational backgrounds and experience. This ensured quality feedback from these experts (expert review 1).
7. The designed artefact was assessed by presenting the Critical Success Factors to an expert group (six experts). The experts have sufficient and relevant educational backgrounds and experience (expert review 2).
8. Recommendations and comments from relevant academics were considered before the expert review was conducted.

In summary, the use of theories and expert groups ensured rigorous research for this study.

5.5.1.3 Guideline 3: Design as an Artefact

The Design Science methodology states that the research must produce a viable artefact in the form of a construct (1), a model (2), a method (3), or an instantiation (4) (Hevner, March, Park & Ram, 2004). The different types of artefacts are explained below.

1. A construct is the vocabulary and symbols used to define problems and solutions. An information system example of a construct would be a Data Flow Diagram used to illustrate the distribution capabilities of data. This diagram uses specific shapes to illustrate external entities, processes, data stores and data flows.
2. A model is explained as abstractions and representations of a problem or opportunity. Its visualisation summarises findings and conclusions from the research.
3. Methods are detailed algorithms and practices, which can be applied in respective fields.
4. An instantiation is focused on systems that have been implemented or prototype systems that have been developed.

After a thorough evaluation of all four types of artefacts in the context of this research, the most appropriate artefact for this study was found to be the design of a model in the form of Critical Success Factors. More specifically, this study produced the Critical Success Factors necessary for the collection of high quality public safety data through Participatory crowdsourcing, utilising voice technologies.

5.5.1.4 Guideline 4: Design Evaluation

Design evaluation explains that the designed artefact must be rigorously assessed to ensure utility, quality, and efficacy. The evaluation chosen will be based on the specific design artefact constructed by the researcher (Hevner, March, Park & Ram, 2004). After assessing a number of evaluation methods including those suggested by Hevner, March, Park and Ram (2004), the expert review was found to be the most appropriate method to evaluate the artefact.

The expert review is an evaluation approach which uses an expert group/s to criticize the design artefact (Molich & Jeffries, 2003). The expert group provides comments on the presented material, which is then used to refine the material based on these comments. Two expert reviews were conducted; the first assessed the Public Safety Data Quality criteria and the second assessed the proposed artefact. Relevant experts must be chosen to ensure that the presented material is assessed effectively and valuable comments are received from the experts. In terms of this study, the experts (six experts) were selected based on their educational background and experience relevant to this study. These experts commented on the Public Safety Data Quality criteria and proposed artefact and provided feedback in the form of recommendations. The recommendations were considered and both the Public Safety Data Quality criteria and the proposed artefact were appropriately modified. The expert review ensured that the Public Safety Data Quality criteria and the proposed artefact were appropriately assessed and confidently presented a valid Public Safety Data Quality criteria (assessing voice data collection) and accurate and relevant Critical Success Factors that would ensure high quality public safety data in a Participatory crowdsourcing utilising voice technologies.

5.5.1.5 Guideline 5: Design as a Search Process

This guideline entails the description of the research process leading to the development of an effective solution to a problem. To ensure applicability to the problem domain, the Public Safety Data Quality criteria and proposed artefact will be constructed from the

assessment of related primary data and secondary data. An assessment by an expert group followed the development of both the Public Safety Data Quality criteria and the proposed artefact. Relevant comments and recommended modifications to the Public Safety Data Quality criteria and the proposed artefact were taken into consideration.

5.5.1.6 Guideline 6: Research Contribution

All Design Sciences research must contribute to the specified area of the designed artefact. In Design Sciences research at least one or more of three contributions must be provided. These contributions include, the design artefact (1), foundations (2) and methodology (3) (Hevner, March, Park & Ram, 2004). Each contribution is based on the design artefact's novelty, generality and significance. These three contributions are explained further below.

1. This is the most common contribution in Design Sciences research. The designed artefact must either provide a solution to an unsolved problem or provide an alternative solution to a problem which has been previously solved.
2. Foundation refers to the contribution to the Design Sciences knowledge base in the form of constructs, models, methods, or instantiations.
3. Methodology refers to the development and expressed use of evaluation methods. These methods include: experiments, analysis, tests, informed arguments, and case studies or scenarios.

This research focused on developing a designed artefact. The design artefact was in the form of Critical Success Factors, which addressed the issue of receiving poor quality public safety data; acting effectively on poor data is not a viable option.

5.5.1.7 Guideline 7: Communication of Research

The final guideline refers to the manner in which the researcher presents (communicates) the research to other entities. Hevner, March, Park and Ram (2004) explain that the research must be communicated to technical and management audiences. Technical audiences require a detailed description of the artefact in order to replicate and effectively implement it in relevant contexts. This requires a detailed explanation of the construction and evaluation process of the artefact. This will allow others to adopt the artefact and encourage researchers to extend the knowledge base relevant to the artefact. The communication of the artefact for managerial purposes is needed to ensure that the artefact is feasible for adoption. Management audiences are concerned with whether the required

resources necessary for implementation are available and acceptable. Therefore, in order to target managerial audiences, the research must provide the necessary knowledge to apply the artefact to other relevant contexts.

This research ensured effective communication in the following manner:

- This study provided the management audience with an awareness of the importance of data quality, as well as the way in which data can be assessed to discover factors which produce low public safety data quality.
- From a technological perspective, an illustration of how high quality public safety data can be achieved through the use of technology was realised.
- The findings of this research were presented to IBM for future research purposes.
- The findings will be published in academic journals and conferences, and made available to the public for future research.
- The completed dissertation will be sent to the library of Fort Hare University and made available to all students and staff members.

The assurance that the seven Design Science guidelines are accounted for confidently reinforced that valid conclusions have been reached. The next section of the research methodology refers to data collection methods.

5.5.2 Data Collection Methods

The primary data collection included, conversational analysis, transcribed reports, recommendations from the expert review and comments from relevant academic experts.

- Primary Data Collection Methods
    - Conversational analysis is a Social Sciences method of data collection through social interactions, usually including verbal and non-verbal cues (Goldkuhl, 2003). Data was gathered from continuous meetings with the IBM team regarding the Smart City and the CSI participatory crowdsourcing project matters. This form of primary data is commonly referred to as conversational analysis.
    - The first phase of public safety incident reports from citizens was a free flowing response (unstructured, as the message prompts will not be used to extract specific data from callers), which were transcribed into text. These

transcribed responses were used to identify what data is missing from the reports.

- The expert review was used to assess the Critical Success Factors by collecting recommendations from 6 experts based on their sufficient and relevant educational background and experience.
- All published material and this dissertation were reviewed by academic experts in the respected field. The comments provided by them were taken into consideration.

- Secondary Data Collection Methods
  - Literature relating to this study was collected from journals, conference proceedings, books, and websites, the theories, methodologies, and previous studies.
  - Frameworks and models, related to data quality and case studies and reports referred to Smart City and various crowdsourcing initiatives.

As previously mentioned, the message prompts were used to direct the caller to provide the necessary data on the public safety incident. Through numerous iterations, the selected and tested message prompts, illustrated in Figure 5-5, were considered most suitable. When constructing the message prompts, the Spoken Web limitations, quality of data and user experience were taken into consideration.

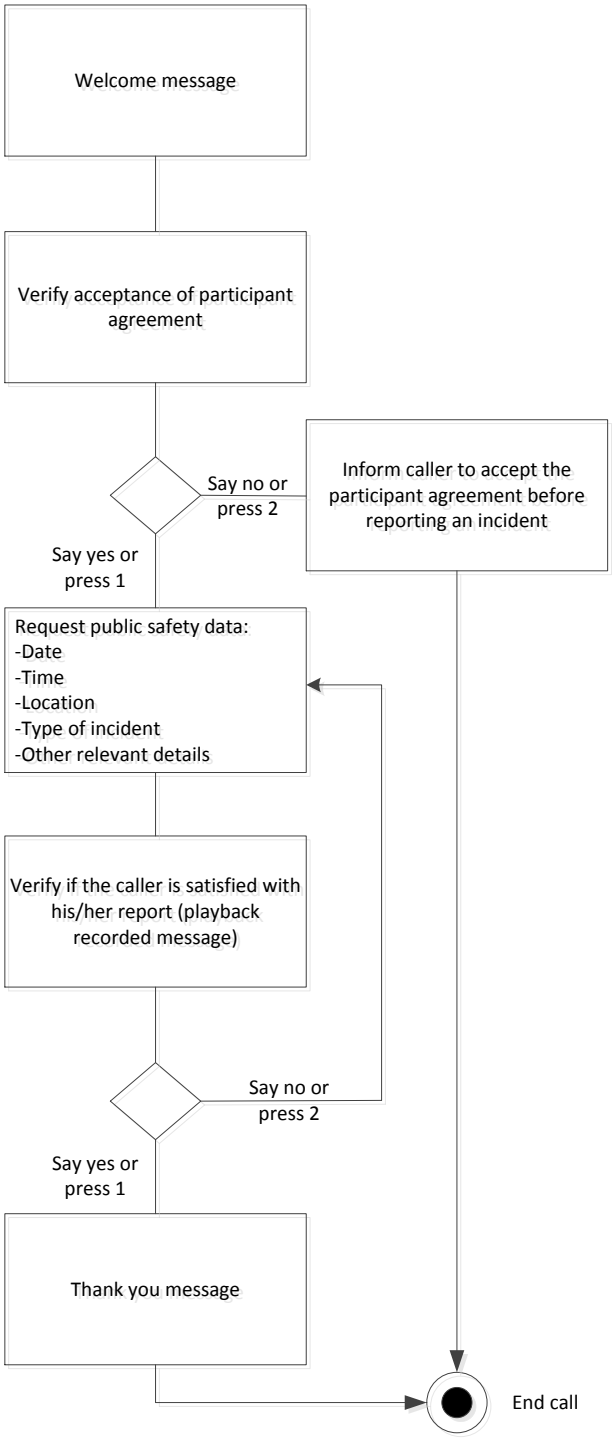


Figure 5-5: Crowdsourcing Message Prompts

The caller can use audio inputs or dial tone inputs to manoeuvre through the system. If audio inputs fail (the voice recognition software struggles to recognise the callers input), the system will request a dial tone input. Although a simple dial tone input is more reliable, audio inputs were integrated to improve the user experience. In addition, it was found (through conversational analysis) that using dial tone input on touch screen mobile

phones is tedious, as the phone's dialling interface needs to be called up before use. This is common on the majority of touch screen phones.

The participatory agreement was developed to notify the caller about the CSI participatory crowdsourcing project and how the data will be used (including that the data will only be used for research purposes at this stage). Callers are unable to report a public safety issue until they have accepted the participatory agreement. A copy of the agreement can be found in Appendix A or alternatively on http://csi.ufh.ac.za/ (see Appendix C). The next section will state how this data and other data collected was analysed for the purpose of this study.

5.5.3 Data Analysis Methods

All data collected will be analysed appropriately. This includes all relevant data collected through conversational analysis, transcribed reports, expert review and recommendations from relevant academic experts.

- Primary Data Analysis
    - The conversational analysis based on meetings with the IBM team was taken into consideration when developing the Critical Success Factors.
    - The transcribed reports were assessed using the Public Safety Data Quality criteria constructed in Chapter 3. A sample of 100 reports was deemed appropriate for analysis.
    - The data collected from the expert review either supported or opposed the Critical Success Factors; the opposed responses were considered, analysed and addressed appropriately.
    - All comments received from academic experts were considered before publication and before the final dissertation is submitted.
- Secondary Data Analysis
    - Secondary data was analysed in an inductive manner to determine whether the logical conclusions agree or disagree with supported theories.

Figure 5-5 illustrated the message prompts used to assist the caller to provide the necessary public safety data. This study focused on ensuring the data collected from the callers is of quality. Chapter 3 was based around developing a Public Safety Data Quality criteria to identify the areas which lack quality and subsequently mitigate these problems by addressing the weak areas and maintaining the strong areas of quality. The criteria was

based on data quality attributes which are considered important to a data consumer. The criteria used to assess each caller's report are shown in Table 5-4.

Table 5-4: Public Safety Data Quality Criteria

Public Safety Data Quality Criteria		
Contextual Data Quality		**Yes/ No**
Value-added	**Assessment**	
Public safety issues	Is it possible to determine the team accountable for responding to the report?	
	Is the report clear enough to anticipate the number of response teams required to resolve the issue?	
Patterns	Can the type of incident be determined?	
	Can the date the incident occurred/observed be determined?	
	Can the time the incident occurred/observed be determined?	
	Does the report mention the suburb/highway at which the incident occurred/observed?	
	Does the report mention the street/landmark at which the incident occurred/observed?	
Relevancy	**Assessment**	
Public Safety Event	Does the reported incident pose a threat to citizens? *If Yes, skip next question (Potential Public Safety Event).* *If No, answer next question (Potential Public Safety Event).*	
Potential Public Safety Event	Is there a high possibility that this incident could escalate to a point of endangering citizens?	
Timeliness	**Assessment**	
Age	Is the time mentioned in the report less than the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	
Input Time	Is the time the data was captured less than the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	
Delivery Time	Are all audio files for the week available for transcription every 72 hours?	
Completeness	**Assessment**	
Quantitative Completeness	Is there an audio version of the report?	
	Is there a textual version of the report?	
Qualitative Completeness	Is it possible to understand the extent of the public safety incident?	
	Can the specific date (example: 3 January) be determined?	
	Can the specific time period (am or pm) be determined?	
	Does the report include an area (example: suburb or highway) name?	
	Does the report include a street name or landmark?	
Amount of Data	**Assessment**	
Data Volume Parameter	Does all the additional data provided influence the dispatch decision?	
Intrinsic Data Quality		**Yes/ No**
Believability	**Assessment**	
Trustworthiness	Can the type of incident be determined?	
	Can the date the incident occurred/observed be determined?	
	Can the time the incident occurred/observed be determined?	
	Does the report mention the area (suburb or highway) at which the incident occurred/observed?	
	Does the report mention the street at which the incident occurred/observed?	
Reasonableness	Is the report realistic and likely to occur given its expressed location and time? *If Yes, skip next question (temporality).* *If No, answer next question (temporality).*	

Temporality	Does the occurrence of a certain unrelated incident (example: heavy rain caused floods) still make reasonableness in the report "false"?	
Accuracy	**Assessment**	
Type of Incident	Can the correct response team accountable for addressing the issue be identified?	
	Can the number of response teams that need to be dispatched be indentified?	
	Can the nature (seriousness) of the public safety issue be determined?	
Location	Can the location of the incident be found on a map without difficulty?	
Time	Is the length of time of the time stamp less the report time appropriate in terms of the likelihood that the nature of the issue has changed?	
Objectivity	**Assessment**	
Voice Tone	Is background noise absent? *If yes, skip next voice tone question.* *If no, answer next voice tone question.*	
	Is the caller's voice tone absent of aggression?	
	Is the caller's voice consistently calm and soft throughout the report?	
Reputation	**Assessment**	
N/A	N/A	
Representational Data Quality		**Yes/ No**
Interpretability	**Assessment**	
Full Language	Is the majority of the report in English?	
Partial Language	Is the report absent of any words or phrases from a language other than English?	
Understandability	**Assessment**	
Clarity	Is background noise absent?	
	Is radio interference absent?	
	Is wind created noise absent? (including caller breathing distortion)	
	Does the caller have a clear accent?	
	Is the use of slang absent?	
	Are system glitches absent? (example: lagging)	
Ambiguity	Can the type of public safety issue be interpreted in a single point of view?	
	Does the time express where the incident occurred in AM or PM?	
	Does the location state the suburb or highway and street name or landmark?	
Comprehensiveness	Is the caller's use of acronyms appropriate to comprehend the report?	
	Are speech impediments absent?	
	Is the caller's vocabulary and use of grammar appropriate to comprehend the report?	
Consistency	**Assessment**	
Type of incident	Can the type of incident be determined?	
Time	Can the date the incident occurred/observed be determined?	
	Can the time the incident occurred/observed be determined?	
Location	Does the report mention the area (suburb or highway) at which the incident occurred/observed?	
	Does the report mention the street/landmark at which the incident occurred/observed?	
Concise	**Assessment**	
Type of incident	Can the type of public safety incident be identified?	
Time	Is it possible to identify the time period in which the incident occurred/observed? *For example; morning, afternoon or night.*	
Location	Was the suburb/highway stated?	
	Was the street name or landmark mentioned or can it be identified?	
Accessibility Data Quality		**Yes/ No**
Accessibility	**Assessment**	
File and Folder Logic	Are the naming conventions of files and folders logical? *For example; a folder named "January Week 1- Audio" includes audio files from the 1 January to 7 January.*	

File Accessibility	Can all authorised files be accessed?	
Folder Accessibility	Can all authorised folders be accessed?	
Security	**Assessment**	
Personal Data	Has any personal data been provided in the report? *If yes, answer next question.* *If no, skip next question.*	
Data Integrity	Is the report still readable after the removal of all personal data?	

This section has discussed the research paradigm, research design and the data collection and analysis methods of the study. The next section will provide delimitations of the study.

5.6 Delimitations of the Study

Delimitations of a study allow the researcher to create a parameter around the study by clearly stating what was included and excluded in the research (Hofstee, 2006). The research parameter for this study will follow.

- The data, collected from the callers, was only related to public safety issues occurring in Buffalo City.
- The reporting of public safety issues was reported through mobile or landline phones; therefore this study specifically focused on citizens of Buffalo City, who possess or have access to a mobile or landline phone.
- The Spoken Web was restricted to English speaking candidates, as the system will not recognize any other language.
- The data collected from the callers was in an audio format, which was transcribed manually into a textual format. This was done for an easier method of analysing the reports from callers.
- At this stage, all the information collected from callers' reports was used strictly for research purposes; therefore the callers still had to report the issues to the authorities responsible for acting on the information.
- Although it was mentioned throughout this study that analytics will be performed on the public safety data collected, it was not considered for this study.

The boundaries of this study have been emphasised through delimitations, which will lead to the ethical considerations of the study.

5.7 Ethical Considerations

Ethical considerations are vital in this study as all data and findings was used by researchers at the University of Fort Hare, IBM India and IBM Israel. In addition, all published papers will be accessible to the public. Punch (2006) states that it is the researcher's responsibility to comply with academic integrity and honesty, as well as to respect other people. Punch (2006) continues by listing categories of ethical issues which should be considered by researchers. These ethical considerations are listed below.

- Informed consent
    - Informed consent was obtained by sharing all necessary information about the research to all participants.
    - The data collected from participants was used for research purposes only, as instructed by IBM's terms and conditions (see Appendix A). This was explained to all participants before any data was collected from them.
    - All callers were notified to accept a participatory agreement before providing any data. The participatory agreement is accessible online and explains all necessary information about the CSI participatory crowdsourcing project and how the data will be used. The length of the participatory agreement is reasonably short ensuring that the intention of the agreement is not obscured by its length. As mentioned above, a copy of this agreement can be found in Appendix A or alternatively on http://csi.ufh.ac.za/ (Appendix C).
- Confidentiality and anonymity
    - The expert group used for the expert review was guaranteed anonymity by constructing the questions in a manner which ensured no personal information was captured.
    - Callers providing a public safety report were instructed to avoid any personal information. In the possibility that callers did provide any personal information, a person was appointed to remove all personal information when transcribing the audio reports to text before the data is distributed and used by researchers.
- Ownership of data and conclusions
    - As mentioned above all data and findings were distributed to the University of Fort Hare, IBM India and IBM Israel for future research intentions.

    - All published material relating to the study will acknowledge the relevant stakeholders involved in the project.
- Use and misuse of results
    - The misuse of results was mitigated by ensuring information gathered from participants were only used for the indented and expressed purpose of the research.
- Honesty and trust
    - Participants were also allowed to refuse participation in the research at any time.
    - All participants were aware of the use of the information and were prompted to report any issues to the relevant authorities first and then report the issue on the IVR system.
- Reciprocity
    - Citizens were encouraged to participate in the CSI participatory crowdsourcing project by realising that their contribution will assist in creating a safer living environment.
- Intervention and advocacy
    - All public safety incidents reported was not forwarded to the necessary authorities as the data collected was only used for research purposes at this point.
- Harm and risk
    - It is unlikely that harm could come to the participants if any personal information is provided. Nonetheless, all personal data was removed from reports, as mentioned above.

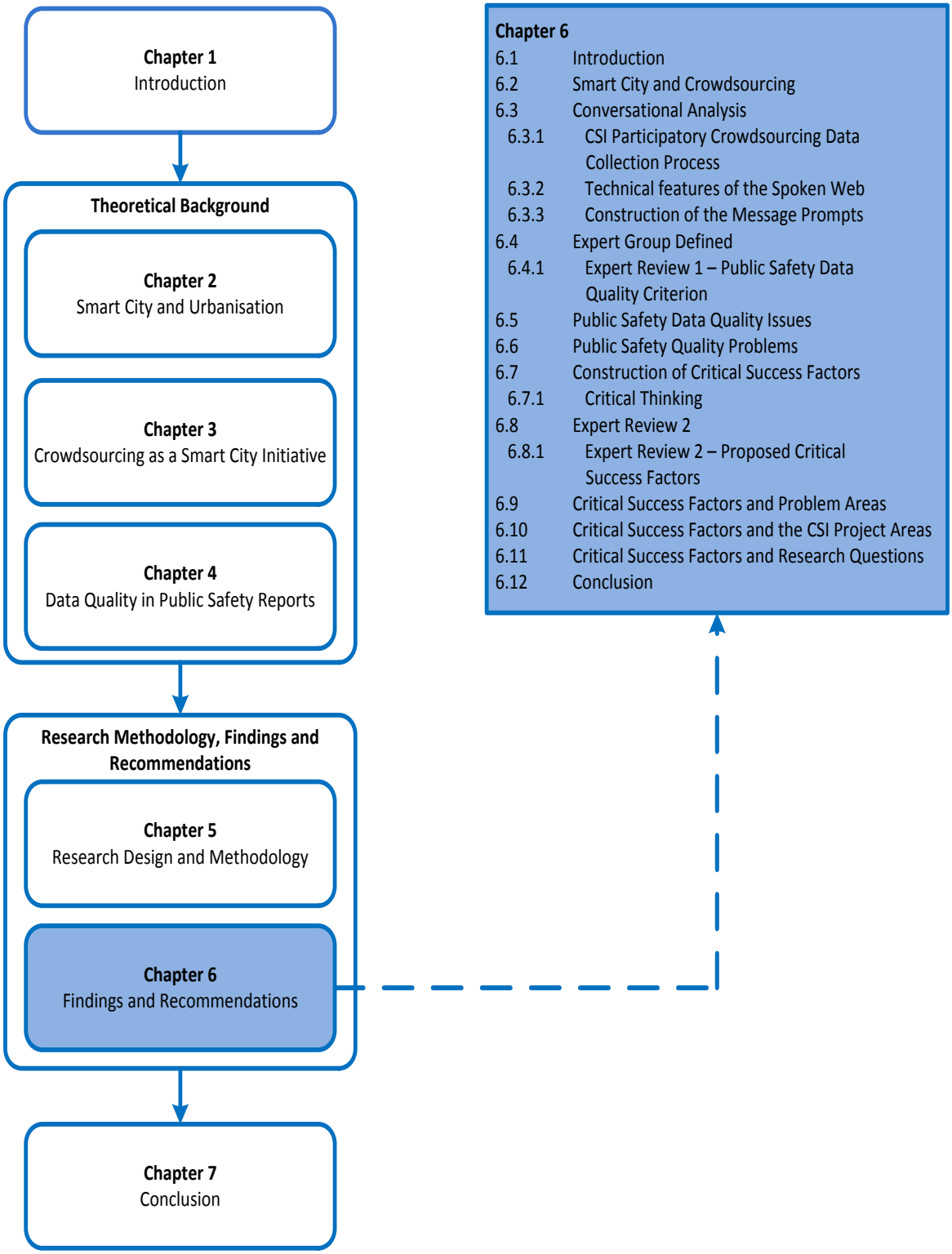
All these ethical issues have been considered and the research undertaken complied with these considerations.

5.8 Conclusion

This chapter provided a detailed description of the manner in which this research was conducted. The chapter firstly provided the context in which the research was applied by describing the CSI participatory crowdsourcing project. A discussion on the research paradigms, positivistic, interpretive and critical theory followed by a supportive argument on the most appropriate research paradigm for this study, that is *reality as a contextual field of information*. The research methods, qualitative, quantitative and mixed were

explained before stating that the qualitative research method was used for this study. The Information Systems Research Framework was then presented before a discussion supporting the selection of the Design Science Methodology as the most suitable methodology for this study. The next section discussed the data collection and analysis methods that included the CSI participatory crowdsourcing project's message prompts and the Public Safety Data Quality criteria. The delimitations of the study followed the data analysis section to clarify the research area. The last section of this chapter listed the ethical issues considered for this study. All discussions above allowed this study to be conducted successfully, resulting in the generation of viable Critical Success Factors for the quality of public safety data within a participatory crowdsourcing project. The next chapter focuses on the Findings and Discussion of this study.

Chapter 6 – Findings and Recommendations



6.1 Introduction

Chapter 6 discusses the findings and illustrates how the previous sub-conclusions are integrated into the study (Hofstee, 2006). A discussion on the relationship between Smart City, crowdsourcing and Data Quality will be introduced. The next section provides a discussion on the data collected through conversational analysis. The expert review process will then be discussed, which is followed by a discussion on the results from the evaluation of the public safety reports. The public safety reports were evaluated to identify data quality issues. After these issues were listed, the next section illustrates the thought process used to construct the proposed Critical Success Factors to collect high quality public safety data through participatory crowdsourcing. The chapter then concludes with an explanation on each proposed Critical Success Factor.

6.2 Smart City and Crowdsourcing

The literature in the previous chapters assisted in reaching the findings for this study. It is important that the findings be linked back to the literature to emphasise this study's significance and contribution to the body of knowledge. This will also improve one's understanding of how the Critical Success Factors were decided upon and why they were deemed important (critical) to the research area of this study. Figure 6-1 illustrates the relationship between Smart City and crowdsourcing, but more specifically, how participatory crowdsourcing can contribute to the Smart City. The two diagrams in Figure 6-1 were developed from relevant literature in Chapters 2 and 3.

The link between "Type of Data" and "Smart City" illustrates that high quality data will create a robust knowledge intensive city. The relationship between "Citizens, Businesses and Government" and the "Crowdsourcing Stakeholders" emphasises that crowdsourcing stakeholders should be made up of citizens, businesses and local Government. The last link between "Smart Technology" (and the Smart City characteristics) and the "Data Collection Method" (and Type of Crowdsourcing) emphasises that different types of Crowdsourcing can be used as Smart Technology to contribute to six areas (Smart Economy, Smart Living, Smart People, Smart Mobility, Smart Governance and Smart Environment) of a city. The next section will provide further detail into the CSI participatory crowdsourcing project findings collected from conversational analysis.

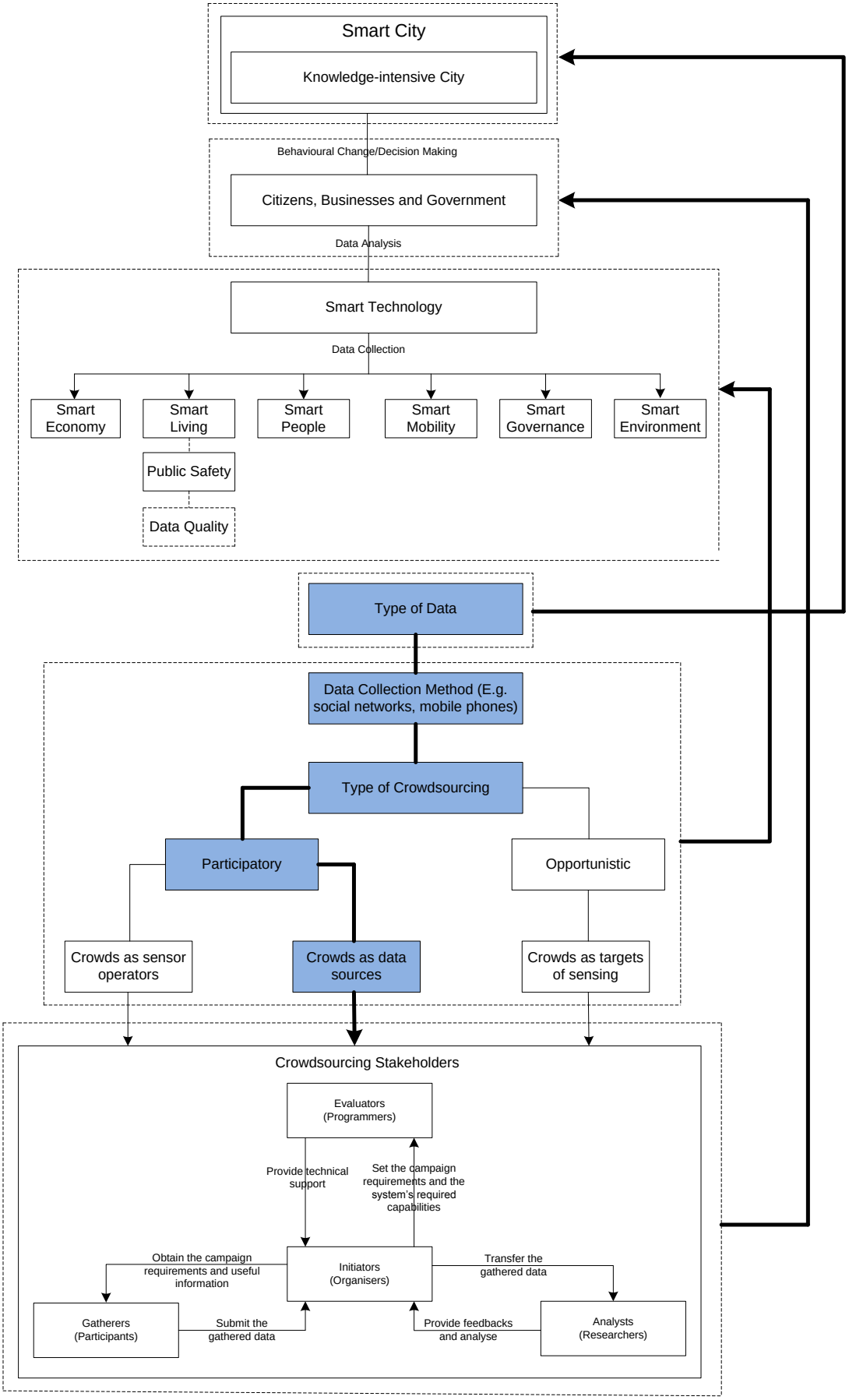


Figure 6-1: Smart City and Crowdsourcing Relationship

6.3 Conversational Analysis

Conversational analysis is a Social Sciences method of data collection through social interactions, usually including verbal and non-verbal cues (Goldkuhl, 2003). Data relating to the CSI participatory crowdsourcing project's scope was collected through continuous meetings with IBM and the University of Fort Hare CSI participatory crowdsourcing team. These meetings were conducted on a bi-weekly basis throughout the project's progress. The following data were collected through conversational analysis with IBM and Fort Hare University, and in discussions between the Fort Hare University project team.

6.3.1 CSI Participatory Crowdsourcing Data Collection Process (see section 4.2 for further detail)

The data collection process of saving audio files on a secure IBM database in India, which are then sent manually by an IBM employee to the University of Fort Hare (on a secure sever) was agreed upon. It was agreed that the audio files would be sent every 72 hours. The University of Fort Hare would then be responsible for transcribing the audio files to text format and removing all personal data before analysis could be conducted. This ensured the adherence to all ethical considerations. This data collection process was considered to be the most simple, efficient and effective method based on the technical complexities.

6.3.2 Technical Features of the Spoken Web (advanced IVR system)

Due to the low computer literacy rate in South Africa, the Spoke Web needed to be utilised in such a manner that any user interacting with the system could use it without difficulty. Therefore, simplicity was key when selecting the Spoken Web features. It was found that using dial tone inputs were tedious, especially when touch screen mobile phones were used, as the phone's dial tone interface must be called up before use. In addition, it was found that if the caller constantly needed to move the phone away from their ear to provide an input, it negatively affected the user experience. Therefore, voice inputs were decided on as the Spoken Web has advanced voice recognition software. Although the Spoken Web has advanced voice recognition software, at times the system still struggled to recognise the caller's input. Therefore, it was decided that in a situation where voice input fails during a call (voice recognition system fails to recognise the voice input), the system would immediately ask the caller to provide a keypad input (for example; press 1 for yes or press 2 for no).

The CSI participatory crowdsourcing project utilised the Spoken Web's voice recognition feature to allow the caller to manoeuvre through the audio interface. The caller does this by simply providing a "yes" or "no" audio input. In order to further improve the Spoken Web's voice recognition software, a number of "yes" and "no" audio responses from the local citizens were collected. These audio responses were collected because of the large number of languages spoken in East London. Recall that at this stage of the CSI participatory crowdsourcing project, reporting will be restricted to English speaking candidates. Therefore, "yes" and "no" audio inputs in different accents will make it difficult for the Spoken Web to recognise the input.

In addition to further improving the user experience, a barge-in feature was considered. This allows callers who have previously called and are familiar with the system, to save time and cut the instructions provided by the system. However, due to the first set of message prompts verifying if the caller is aware of the project's intention (data will only be used for research purposes at this stage), the barge-in feature was excluded to ensure that the project conforms to all ethical considerations. In other words, the caller is obligated to understand the project's intentions by restricting them from skipping the message prompts explaining the project's intentions.

6.3.3 Construction of the Message Prompts (see Figure 5-4 for graphical illustration of agreed message prompts)

The message prompts are intended to instruct the caller to provide the correct data in order to appropriately use the public safety data collected. It was originally planned to develop distinctive public safety categories and present these options to the caller. For example, the message prompts would read: "if the public safety issue relates to crime press 1, if the public safety issue relates to infrastructure press 2, if the public safety issue relates to fires press 3, and if the public safety issue relates to anything else press 4." This would allow for an automatic organisation of public safety reports based on the emergency unit or non-emergency unit responsible for attending to the problem. However, this method was found to be ineffective due to the lengthy list of grouped public safety issues which could occur. For example, the longer the list of message prompts, the high the risk that the caller will lose patience and end the call as this makes the reporting process more tedious. In addition, this method added additional responsibility on the caller to determine the type of public safety incident before providing the report.

Alternatively, a simpler approach was discussed where the message prompts would read; "if the public safety issue requires immediate response press 1 and if the public safety issue is a non-emergency incident press 2." This method separates important issues which require immediate attention from those that could be solved after a couple of days (incidents that are not life threatening and will remain the same for a reasonable time after reporting, for example, potholes). However, this prioritisation method of message prompts was also found to be ineffective as callers were tempted to express every incident as an emergency so the problem can be resolved in a timelier manner. Additionally, callers would be required to judge whether the incident reported is an emergency or non-emergency issue; however, callers by and large do not have that kind of discernment. For example, a public safety incident may not seem life threatening to one caller but seen as a high priority incident by another person with emergency response experience or knowledge.

After a number of discussions and iterations on potential message prompts, a free flowing data reporting method (see Figure 5-4) was agreed upon. These message prompts will be used during the first phase of the project and may be modified based on the manner of reporting provided by the callers. This will be identified by assessing the public safety reports provided by callers. Therefore, the next phase of the CSI participatory crowdsourcing project may require change management practices to ensure participation is maintained after changes are made to the message prompts. An explanation of how and when the expert review process was conducted will follow.

6.4 Expert Group Defined

An expert group was used to analyse and receive feedback on two segments of findings in the study. Expert review is an ad hoc method used to evaluate research findings by using a group of relevant professionals (Molich & Jeffries, 2003). The number of experts in a group is not limited; however, the size of the expert group usually ranges from 3 to 8 experts in information systems research. The experts chosen should be professionals in the field of research undertaken to ensure that valid feedback is obtained (Molich & Jeffries, 2003). The integration of the expert reviews conducted through the process of finalising the Critical Success Factors is illustrated in Figure 6-2.

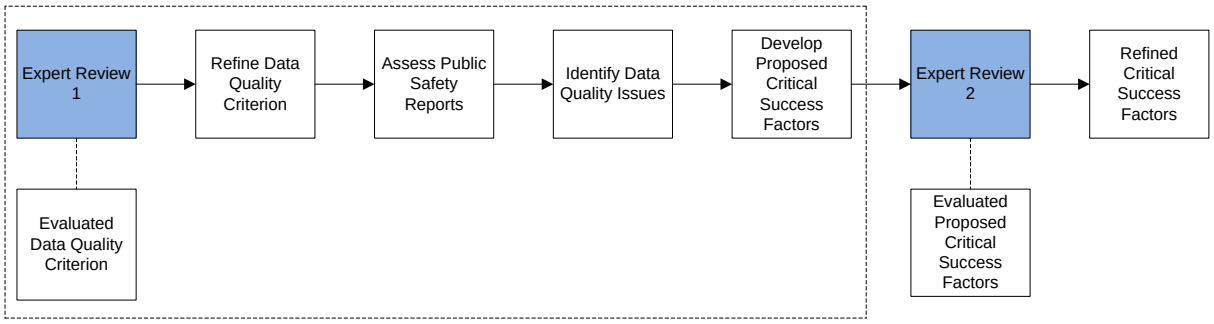


Figure 6-2: Expert Review Process

The first expert review assessed the Public Safety Data Quality criteria that was developed in Chapter 4. The second expert review assessed the proposed Critical Success Factors. To ensure a logical flow throughout the chapters in this study, the results for the second expert review will be discussed in section 6.8. The area included within the dotted rectangle in Figure 6-2 will be discussed in the remainder of this chapter. Five open-ended questions were attached to the Public Safety Data Quality criteria and the proposed Critical Success Factors to ensure that the two segments were thoroughly assessed by the expert group.

It was found that open-ended questions were best suited for the expert groups in terms of receiving quality feedback. Open-ended questions allow respondents to support their answers through further explanation (Bentley & Whitten, 2007). In addition, respondents are encouraged to voice their own opinions, ideas and concerns more openly and honestly (Bentley & Whitten, 2007). Open-ended questions usually begin with phrases such as; "Why", "How", "I would like to know", "Tell me about" and "I am interested in hearing about". Structuring a question in this manner restricts the use of one or two word answers such as yes or no as participants are forced to explain their view in addition to indicating agreement or disagreement. This was taken into consideration when constructing the questions addressed to the expert group. Six experts were used to assess both the Public Safety Data Quality criteria and the proposed Critical Success Factors. The following lists the relevancy of each chosen expert group member.

Table 6-1: Expert Review Group

Number of Experts	Expert Relevancy
2	- Educational background on crowdsourcing. - CSI participatory crowdsourcing project team member.
1	- Educational background on IVR and the Spoken Web.
1	- CSI participatory crowdsourcing data transcriber.
1	- Educational background and experience in data integrity.
1	- Educational background on crowdsourcing, data security and integrity, IVR and Spoken Web.

Table 6-1 indicates that all members in the expert group are capable of assessing both the Public Safety Data Quality criteria and the proposed Critical Success Factors, and of providing constructive feedback. The responses for the first expert review will be discussed in the next section.

6.4.1 Expert Review 1 – Public Safety Data Quality Criteria

The knowledge and experience of six carefully selected experts were used to assess the Public Safety Data Quality criteria (see Table 4-4). The criteria will be used to assess public safety reports collected from the CSI participatory crowdsourcing project. The main comments and recommendations collected from the expert group are listed below.

- All experts confidently explained that the Public Safety Data Quality criteria is a highly effective tool for assessing the quality of public safety data as it is based on a sound theoretical background and considers all areas which could affect the quality of data. In addition, the questions in the Public Safety Data Quality criteria relate specifically to public safety data, making them capable of identifying all data quality problems.
- Five experts stated that the length of the Public Safety Data Quality criteria (number of questions) was not acceptable in terms of the amount of time taken to conduct quality assessments. The reason for the lengthy list of questions was because certain questions related to the assessment of a number of data quality attributes. To indicate this, certain questions were asked a number of times in different ways. Therefore, to reduce the length (number of questions) of the Public Safety Data Quality criteria, all repeat questions were removed.
- One expert suggested that questions relating to data quality attributes and the data quality dimension should not be presented to the person assessing the public safety data. This would ensure a more objective assessment is conducted as the data quality assessor would be unable to determine the question relationship to specific

areas. For example, questions 6.1 to 8.3 relate to the data source (intrinsic data quality dimension); the data assessor can easily assign blame to the data source by answering questions 6.1 to 8.3 in a subjective manner. Taking this suggestion into consideration, both the data attributes and dimensions were removed.

- One expert explained that the "week" and "72 hours" in question 3.3 should be removed as it limits the flexibility of the Public Safety Data Quality criteria. This suggestion was considered valid, which changed the question to read "Are all audio files available for transcription when required?" This allows one to change the required time without having to modify the Public Safety Data Quality criteria.
- Two experts stated that questions 2.1 and 2.2 be merged together as well as questions 4.1 and 4.2, questions 6.6 and 6.7 and questions 9.1 and 9.2. One expert added that questions 8.1, 8.2 and 8.3 seemed to assess the same thing.
    - Questions 2.1 and 2.2 were asked separately to indicate whether the reported public safety issue was a current problem or potential problem. It was explained that this was unnecessary as the Public Safety Data Quality criteria assessed quality and not the type of public safety issues being reported. Therefore, questions 2.1 and 2.2 were merged together.
    - Questions 4.1 and 4.2 were similar to questions 2.1 and 2.2 both indicating whether an audio version and a text version were available. The questions were merged together for this reason and to improve the practicality of the Public Safety Data Quality criteria (reduces the length of questions).
    - Questions 6.6 and 6.7 can be merged together as both questions determined if data was realistic based on time and location (question 6.6) or the occurrence of another event (question 6.7). If a "yes" was provided in question 6.6 or 6.7, then the data provided was considered to be realistic. Therefore, these two questions should be merged together.
    - Questions 9.1 and 9.2 were used to indicate whether the whole public safety report was provided in another language besides English or if partial English was used. It was explained by an expert that regardless of full non-English or partial English, both affect the quality of data as one problem. If different problems were associated with full non-English and partial English, then it would be acceptable to have separate questions. For this reason, questions 9.1 and 9.2 were merged together.

- Questions 8.1, 8.2 and 8.3 measure the same thing. It was found that these questions could be merged as questions 8.1 and 8.2 test the caller's objectivity based on their voice tone. Questions 8.2 and 8.3 were used to indicate if a caller's voice was aggressive (question 8.2) or calm (question 8.3). Determining whether the caller's voice tone was aggressive or calm would be interesting to note, but not relevant to assess the quality of data. For this reason questions 8.1, 8.2 and 8.3 were merged together to identify aggressive tone because as suggested in Chapter 4, this would influence a caller's objectivity when reporting a public safety issue.

Following the considerations of the experts' comments and recommendations, the Public Safety Data Quality criteria was modified appropriately. Table 6-2 illustrates the refined Public Safety Data Quality criteria.

Table 6-2: Refined Public Safety Data Quality Criteria

	Question Description	**Ref #**	**DQ Attribute**	**DQ Dimension**
1	Is it possible to determine the response team responsible for responding to the report?	1.1	Value-added	Contextual
		7.1	Accuracy	Intrinsic
2	Cannot determine the number of response teams required to resolve the issue.	1.2	Value-added	Contextual
		7.2	Accuracy	Intrinsic
3	Can the type of incident be determined?	1.3	Value-added	Contextual
		6.1	Believability	Intrinsic
		7.5	Accuracy	Intrinsic
		11.1	Consistency	Representational
		12.1	Concise	Representational
4	Can the date the incident occurred/observed be determined?	1.4	Value-added	Contextual
		4.4	Completeness	Contextual
		6.2	Believability	Intrinsic
		11.2	Consistency	Representational
5	Can the time the incident occurred/observed be determined?	1.5	Value-added	Contextual
		6.3	Believability	Intrinsic
		11.3	Consistency	Representational
6	Does the report mention the area (suburb or highway) at which the incident occurred/observed?	1.6	Value-added	Contextual
		4.6	Completeness	Contextual
		6.4	Believability	Intrinsic
		7.4	Accuracy	Intrinsic
		11.4	Consistency	Representational
		12.3	Concise	Representational
7	Does the report mention the street/landmark at which the incident occurred/observed?	1.7	Value-added	Contextual
		4.7	Completeness	Contextual
		6.5	Believability	Intrinsic
		7.4	Accuracy	Intrinsic
		10.9	Understandability	Representational
		11.5	Consistency	Representational
		12.4	Concise	Representational
8	Does the reported incident show any sign of endangering citizens or has the potential to endanger citizens?	2.1	Relevancy	Contextual
		2.2	Relevancy	Contextual

9	Is the time mentioned in the report less the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	3.1	Timeliness	Contextual
10	Is the time the data was captured less the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	3.2	Timeliness	Contextual
11	Are all audio files available for transcription when required?	3.3	Timeliness	Contextual
12	Is there an audio version and textual version of the report?	4.1	Completeness	Contextual
		4.2	Completeness	Contextual
13	Is it possible to understand the extent of the public safety incident?	4.3	Completeness	Contextual
		7.3	Accuracy	Intrinsic
14	Can the specific time period (am or pm) be determined?	4.5	Completeness	Contextual
		10.8	Understandability	Representational
		12.2	Concise	Representational
15	Does all the additional data provided influence the dispatch decision?	5.1	Amount of Data	Contextual
16	Is the reported incident realistic and likely to have occurred given its expressed location and time or considering the occurrence of a certain unrelated incident (example: heavy rain caused floods)?	6.6	Believability	Intrinsic
		6.7	Believability	Intrinsic
18	Is the caller's voice consistently calm and soft throughout the report, with the absence of background noise?	8.1	Objectivity	Intrinsic
		8.2	Objectivity	Intrinsic
		8.3	Objectivity	Intrinsic
19	Is the report absent of any words or phrases from a language other than English?	9.1	Interpretability	Representational
		9.2	Interpretability	Representational
20	Are system glitches absent? (example: lagging)	10.6	Understandability	Representational
21	Is background noise absent?	10.1	Understandability	Representational
22	Is radio interference absent?	10.2	Understandability	Representational
23	Is wind created noise absent? (including caller breathing distortion)	10.3	Understandability	Representational
24	Does the caller have a clear accent?	10.4	Understandability	Representational
25	Is the use of slang absent?	10.5	Understandability	Representational
26	Can the type of public safety issue be interpreted in a single point of view?	10.7	Understandability	Representational
27	Is the caller's use of acronyms appropriate to comprehend the report?	10.10	Understandability	Representational
28	Are speech impediments absent?	10.11	Understandability	Representational
29	Is the caller's vocabulary and use of grammar appropriate to comprehend the report?	10.12	Understandability	Representational
30	Are the naming conventions of files and folders logical?	13.1	Accessibility	Accessibility
31	Can all authorised files be accessed?	13.2	Accessibility	Accessibility
32	Can all authorised folders be accessed?	13.3	Accessibility	Accessibility
33	Has all the personal data been removed from the report?	14.1	Security	Accessibility
34	Is the report still readable after the removal of all personal data?	14.2	Security	Accessibility

Based on the feedback from the experts, the most common problem was the length of the Public Safety Data Quality criteria. As indicated in Table 6-2; a number of questions have been grouped together to reduce the total number of questions from 64 to 34, after refinement. Note that a person using the Public Safety Data Quality criteria should only be presented with the first two columns (question number and question description) in Table

6-2. This will reduce the risk of any subjectivity during the public safety data quality assessment process. The column labelled "Ref #" illustrates which repetitive questions were grouped together. The last two columns indicate the question's influence on the respective data quality attributes and data quality dimensions. This concludes the refinement of the Public Safety Data Quality criteria based on feedback from the expert review. The Public Safety Data Quality criteria is ready to be used in practice to assess the quality of public safety data collected from the CSI participatory crowdsourcing project.

6.5 Public Safety Data Quality Issues

The refined Public Safety Data Quality criteria presented was used to assess the data quality in the public safety reports. The criteria consists of straightforward, yes or no questions that were used to assess 14 data quality attributes deemed important to a data consumer. All questions that were answered with a "yes" indicated that the specific assessment areas (data quality attributes) on which the question was based had quality. As previously mentioned, the audio reports were transcribed into a text version and can be found in Appendix D. The Public Safety Data Quality criteria was used to assess 100 public safety reports. Note that the nature of the questions requires data quality to be assessed by using the transcribed version and audio version of reports, as well as conducting system tests. The results were totalled and presented in Table 6-3 below.

Table 6-3: Public Safety Data Quantity Assessment Results

	Question Description	**Total**
1	Is it possible to determine the emergency or non-emergency unit responsible for responding to the report?	99
2	Cannot determine the number of response teams required to resolve the issue.	99
3	Can the type of incident be determined?	100
4	Can the date the incident occurred/observed be determined?	**76**
5	Can the time the incident occurred/observed be determined?	**37**
6	Does the report mention the area (suburb or highway) at which the incident occurred/observed?	**71**
7	Does the report mention the street/landmark at which the incident occurred/observed?	**86**
8	Does the reported incident show any sign of endangering citizens or has the potential to endanger citizens?	99
9	Is the time mentioned in the report less the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	**69**
10	Is the time the data was captured less than the report time stamp, an appropriate time to assume that the nature of the public safety incident remained unchanged?	95
11	Are all audio files available for transcription when required?	96
12	Is there an audio version and textual version of the report?	100
13	Is it possible to understand the extent of the public safety incident?	100
14	Can the specific time period (am or pm) be determined?	**28**
15	Does all the additional data provided influence the dispatch decision?	96
16	Is the reported incident realistic and likely to have occurred given its expressed location and time or considering the occurrence of a certain unrelated incident (example: heavy rain caused floods)?	100

18	Is the caller's voice consistently calm and soft throughout the report, with the absence of background noise?	**94**
19	Is the report absent of any words or phrases from a language other than English?	100
20	Are system glitches absent? (example: lagging)	96
21	Is background noise absent?	97
22	Is radio interference absent?	99
23	Is wind created noise absent? (including caller breathing distortion)	**86**
24	Does the caller have a clear accent?	**89**
25	Is the use of slang absent?	99
26	Can the type of public safety issue be interpreted in a single point of view?	96
27	Is the caller's use of acronyms appropriate to comprehend the report?	**93**
28	Are speech impediments absent?	**94**
29	Is the caller's vocabulary and use of grammar appropriate to comprehend the report?	100
30	Are the naming conventions of files and folders logical?	100
31	Can all authorised files be accessed?	100
32	Can all authorised folders be accessed?	100
33	Has all the personal data been removed from the report?	**84**
34	Is the report still readable after the removal of all personal data?	100

A benchmark of 95 or less was found to be appropriate in assuming quality problems in the public safety reports. Therefore, any question that scored 95 or lower is seen as a data quality issue or potential data quality issue. These are highlighted in Table 6-3. Following the explanation of how the results were presented, the next section provides the problem areas affecting the quality of data in the CSI participatory crowdsourcing project.

6.6 Public Safety Quality Problems

Following the assessment of the public safety reports, 9 problems were found that affect the quality of the data. These problems are presented below with the question number/s in brackets to indicate the question/s that identified the problems. The finalised (grouped) list of problems affecting the quality of public safety data is listed below.

1. Inconsistent collection of public safety incident date and time data (4 , 5 and 14).
2. Inconsistent collection of public safety incident location data (6 and 7).
3. Reports which required an immediate response were reported over a month or few weeks later (9).
4. Callers frustration overstates the nature of the issue (18).
5. Heavy breathing into the talking piece of the phone (23).
6. Accent makes it difficult to comprehend the intended message in terms of the pronunciation of certain words (24).
7. Caller's use of acronyms makes it difficult to comprehend the public safety report (27).
8. Speech impediments are due to language barriers and describing the event (28).

9. Data source is finding it difficult to follow data collection instructions or the instructions are not provided clearly (33).

The 9 problems which were grouped above, will be used as a guide when constructing the Critical Success Factors. Table 6-5 indicates which Critical Success Factors address these problems. The next section will explain how the Critical Success Factors were developed.

6.7 Construction of Critical Success Factors

Critical Success Factors are the limited number of areas which must be considered in order to ensure success in an organisation or project (Lin & Fu, 2012). Critical Success Factors should be measurable, controllable and limited in number (Puri, 2012). The Critical Success Factors should also be clear and unambiguous to ensure consistent consideration. The Critical Success Factors identified for this study will assist the participatory crowdsourcing public safety projects in developing cities such as East London, to collect large amounts of high quality public safety data. It was found that the most effective approach to identify the Critical Success Factors for this study was through critical thinking. This section will explain how the Critical Success Factors were constructed through the use of critical thinking throughout this study.

6.7.1 Critical Thinking

"Critical thinking is the ability to identify a problem, select and evaluate pertinent information, recognize assumptions, formulate appropriate hypotheses, and draw valid conclusions and critical inferences" (Ross, Loeffler, Schipper, Vandermeer, & Allan, 2013, p. 724). In other words, any statement provided is assessed before accepting its validity. This approach was conducted throughout this study to ensure that valid conclusions and sub-conclusions were realised. These conclusions and sub-conclusions led to the development of the Critical Success Factors. The approach of critical thinking will now be discussed.

Rainbolt and Dwyer (2012) explain graphically the concept of critical thinking. This can be found in Figure 6-3 below.

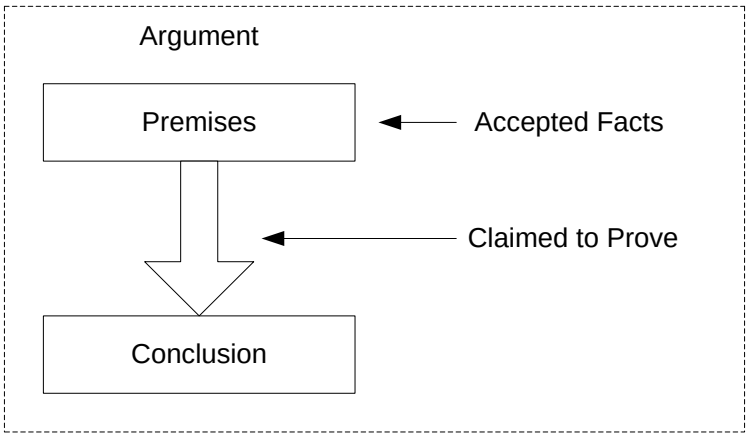


Figure 6-3: Critical Thinking Process (Rainbolt & Dwyer, 2012)

The argument illustrates the process of critical thinking. Premises are statements that are used to logically support conclusions (Rainbolt & Dwyer, 2012). The arrow between the premises and the conclusion represents logical arguments which either agree, disagree or partially agree with the premises (Rainbolt & Dwyer, 2012). The conclusion represents the results of the critique (Rainbolt & Dwyer, 2012). When conducting this research, the aim was to create a set of logical arguments which would direct the research in creating the Critical Success Factors. Figure 6-4 illustrates the template that will be used to show how the Critical Success Factors for this study were constructed.

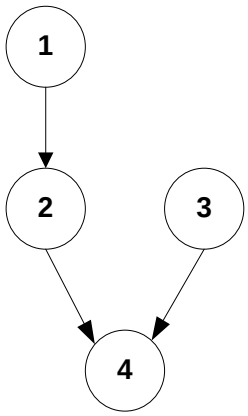


Figure 6-4: Critical Thinking Template (Rainbolt & Dwyer, 2012)

In Figure 6-4, 2 is the conclusion made from assessing statement 1, and 4 is the conclusion made after considering statements 2 and 3. This format was used in Figure 6-5 to illustrate the argument line which lead to the development of the Critical Success Factors. Note that certain premises are linked (illustrated by a dotted arrow) to comply with the thought process which led to the construction of the Critical Success Factors. The numbers 5, 12, 27 and 33 in Figure 6-5 are shaded to indicate that these are the proposed Critical Success

Factors. It should be noted that the proposed Critical Success Factors were constructed from secondary data (relevant literature on crowdsourcing, public safety and data quality) and primary data (conversational analysis and first expert review feedback) that aligned with Wang and Strong's (1996) Data Quality Framework (see Chapter 4).

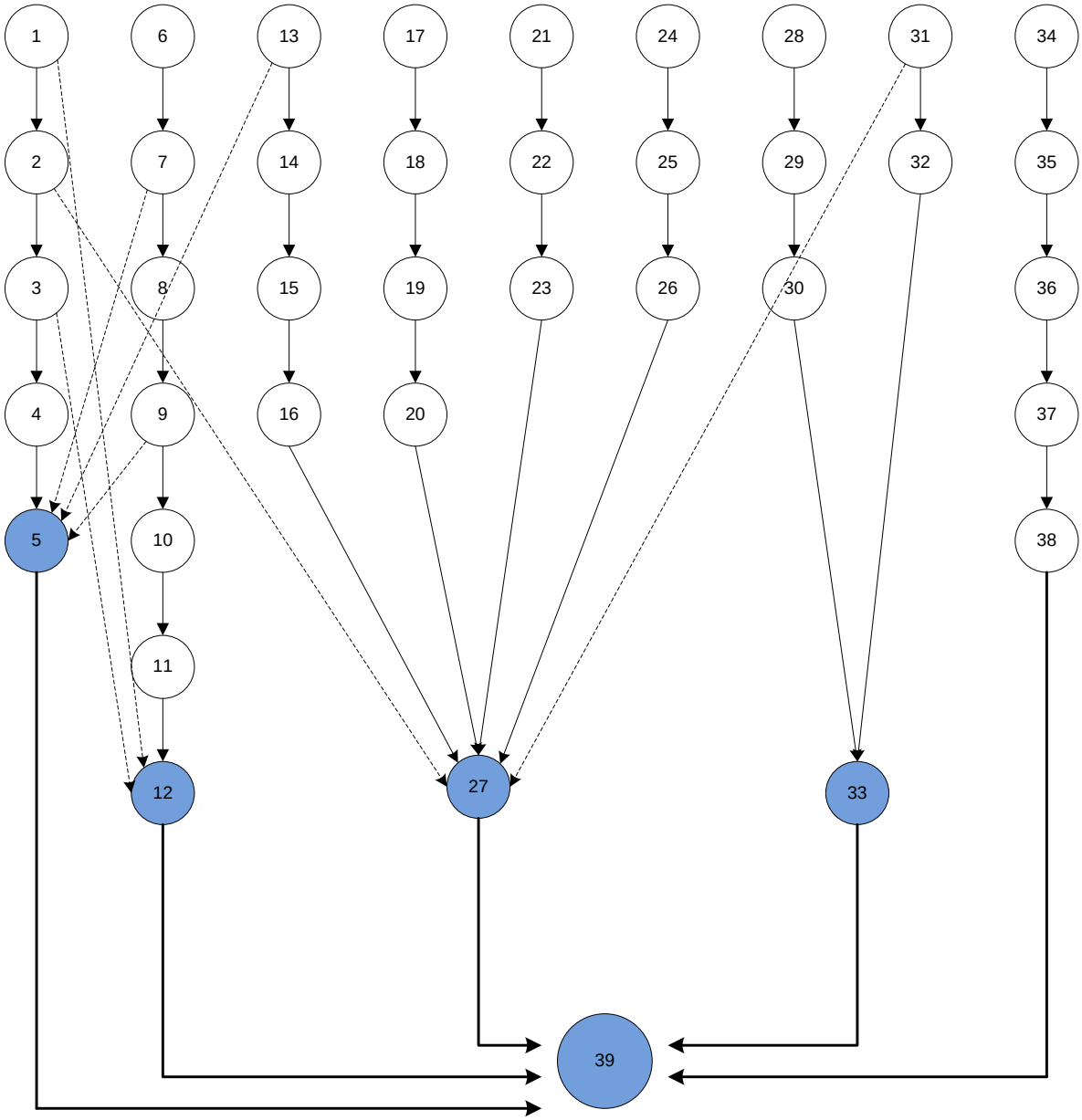


Figure 6-5: Argument Line Leading to Critical Success Factors

The numbers in Figure 6-5 above are allocated to statements listed below. The numbers 5, 12, 27 and 33 below are in bold to indicate that these are the proposed Critical Success Factors.

1. Public safety issues which are not reported as soon as they occur/spotted make it difficult for emergency units or non-emergency units to respond to the issue, as the nature of the incident may have changed by the time the emergency units or non-emergency units receive the report.
2. Public safety reports are used to respond to the issue and the collection of reports are used to identify patterns.
3. The data in public safety reports are usually deficient in some important data or alternatively, too much data are provided which obscures the important data.
4. Some issues reported, fall outside the scope of responsibility of emergency units or non-emergency units.
5. **Contextual Relevance.**
6. Public's inexperience in responding to public safety issues affects their capability to provide the necessary data about the issue.
7. One should be able to determine the emergency unit (example: policemen or firemen) or non-emergency units (for example: road maintenance team) responsible for responding to the incident, as well as the number of response teams.
8. The public fails to indicate the location of the public safety issue.
9. The nature of the public safety issue may change if delays occur between the time that the incident occurred/spotted and the time that the report was provided.
10. Anger and frustration created from frequent public safety issues re-occurring or that issues are not addressed in a timely manner causes the public safety issues to be over stated when described in a report.
11. Occasional prank (fake) public safety issues are reported.
12. **Data Source.**
13. Emergency and non-emergency units must analyse the public safety reports before responding or use the reports to identify patterns.
14. Language barriers affect the ability to interpret the public safety data.
15. Speech impediments (for example, stuttering) make it difficult to comprehend the public safety reports.
16. Vocabulary and grammar affects the ability to understand the report.
17. Ambiguity in public safety reports makes it difficult to respond to and clearly understand the extent of the issue.
18. Unclear description of the public safety incident affects understandability.

19. The exclusion of AM or PM when mentioning the time that the public safety issue occurred/spotted makes it difficult to determine the precise time period.
20. When landmarks are used to describe the location of the public safety issue, the precise location is difficult to find when the landmark covers a large geographical area.
21. Informal communication in a public safety report makes it difficult to understand.
22. Uncommon slang (usually local slang) affects the understandability of public safety reports.
23. The use of acronyms makes it difficult to understand the public safety report.
24. Noise affects the clarity of a public safety report.
25. Noise is created by humans (caller and third party disturbance).
26. Noise is created from technological interferences.
27. **Data Interpretation and Presentation.**
28. The integrity of public safety data is at risk if unauthorised people have access to the data.
29. Occasionally, authorised people do not have access to the data.
30. Certain content in the audio reports (for example, stuttering and long pauses) are not represented in the transcribed version.
31. All public safety data used to identify patterns must be absent of personal information which can identify the data source, so that data can be used by various sources without creating any ethical problems.
32. Removal of personal data may affect the integrity of the data.
33. **Data Integrity and Security.**
34. The Design Science research methodology was used in this study.
35. Design Science guideline 5 (design as a search process) was achieved by refining the Public Safety Data Quality criteria through an expert review.
36. Design Science guideline 5 (design as a search process) was also used to identify data quality problems within participatory crowdsourcing.
37. Design Science guideline 2 (research rigor) was achieved by assesing the Public Safety Data Quality criteria and the Critical Success Factors with an expert group.
38. Recommendations and comments from the expert review group assisted in refining the research artefact.
39. Proposed Critical Success Factors.

This section illustated how the proposed Critical Success Factors for this study were developed. Four proposed Critical Success Factors were found from the secondary data and primary data; these are provided in Table 6-4. After the feedback from the second expert review, the naming conventions of the Critical Success Factors were made more descriptive and a fifth Critical Success Factor was added. The final Critical Success Factors are presented in Table 6-4 (refined Critical Success Factors) to illustrate changes between the proposed Critical Success Factors and the final (refined) Critical Success Factors.

Table 6-4: Critical Success Factors for this Study

Proposed Critical Success Factors	Refined Critical Success Factors
Contextual Relevance	Relevant Public Safety Data
Data Source	Public Safety Reporting Instructions
Data Interpretation and Presentation	Public Safety Data Interpretation and Presentation Format
Data Integrity and Security	Public Safety Data Integrity and Security
	Simple Participatory Crowdsourcing System Setup

The proposed Critical Success Factors listed in Table 6-4 indicate the areas which need to be considered to ensure quality data is collected in participatory crowdsourcing projects focusing on the collection of public safety data. The next section will discuss the feedback from the second expert review to explain the refined list of Critical Success Factors.

6.8 Expert Review 2

It was mentioned above that the first expert review focused on the assessment of the Public Safety Data Quality criteria which was then used to assess the public safety reports for data quality problems. The problems identified assisted in constructing the proposed Critical Success Factors to ensure Quality Public Safety Data were collected in a participatory crowdsourcing project. The second expert review was conducted to evaluate the proposed Critical Success Factors. This is illustrated within the dotted rectangle in Figure 6-6.

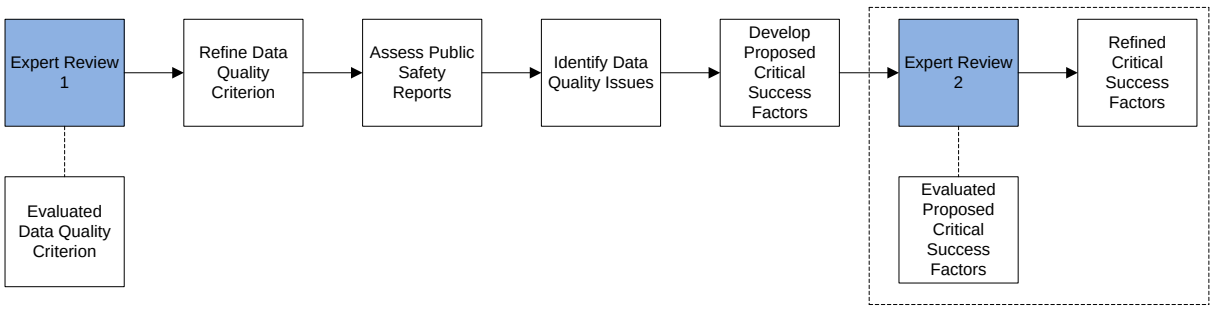


Figure 6-6: Second Expert Review Process

The second expert review was conducted after the proposed Critical Success Factors were developed (see section 6.7). As mentioned above, the same expert group was used to assess the proposed Critical Success Factors. Description of the expert group members are illustrated in Table 6-1. From their educational background and experience, these experts are capable of assessing the proposed Critical Success Factors and providing quality feedback. In addition, it should be noted that all experts have obtained one or more technologically related postgraduate degree. Therefore, they do have sufficient research knowledge to provide adequate feedback on the proposed Critical Success Factors. The feedback from the second expert review concerns the proposed Critical Success Factors.

6.8.1 Expert Review 2 – Proposed Critical Success Factors

The second expert review provided suggestions relating to the proposed Critical Success Factors for this study which focused on ensuring quality public safety data collected through participatory crowdsourcing. The main comments and recommendations collected from the second expert review are provided below.

- All experts agreed that the study was original as it provided focus areas to ensure quality data was collected from an automated (human-to-computer interaction replaced human-to-human interaction) public safety data collection method, using voice (audio data collection). They added that the study related to the implementation of participatory crowdsourcing in South Africa (developing country), adding to the originality of this study.
- Two experts explained that this study was significant as it contributed to the crowdsourcing body of knowledge. Three experts stated that this study would assist Governments in improving public safety services. One expert explained that the study contributed to the public safety data quality research by utilising an innovative data collection method and considering various developing country characteristics such as dominant languages, literacy rates and economically active citizens.
- The majority of the experts stated that the naming conventions of the proposed Critical Success Factors were too vague, explaining that it would be acceptable if they related to crowdsourcing data collection in general. However, the proposed Critical Success Factors were intended particularly to ensure quality public safety

data was collected. The naming conventions of the proposed Critical Success Factors will be made more specific so one can clearly understand their importance.

- Each proposed Critical Success Factor subtly touches on what processes in the participatory crowdsourcing system affect the quality of public safety data in a negative manner. One expert suggested that a new Critical Success Factor be added to acknowledge that the system setup has a risk of affecting data quality. The importance of a system setup in terms of its influence on the quality of data is significant enough to be added as a Critical Success Factor. The problems emphasising the significance of the system setup can be linked to the data collected from conversational analysis between IBM and the University of Fort Hare CSI participatory crowdsourcing team (see section 6.3).

The comments and recommendations collected from the second expert review which assessed the proposed Critical Success Factors, were taken into consideration when refining the proposed Critical Success Factors. The naming conventions of the proposed Critical Success Factors were made more specific and a fifth Critical Success Factor was added. The refined Critical Success Factors are provided in Table 6-5.

Table 6-5: Refined Critical Success Factor

Code	Name	Description
CSF01	Relevant Public Safety Data	The data collected from a participatory crowdsourcing project must be relevant to the task at hand. When attempting to determine what data is relevant to the task/s, one must first identify how the data will be used and the intended decisions produced from reviewing the data. When considering what data is relevant, it is important to find the correct level of detail. Little detail will limit its use while too much detail will increase the risk of collecting incomplete data. One must also consider the length of time that the data will remain relevant from its collection to its use. This is a common problem if data requires modification, analysis, interpretation or presentation before it can be used.
CSF02	Public Safety Reporting Instructions	The data source in all participatory crowdsourcing projects is the public who participate in the initiative. Since certain participatory crowdsourcing (excluding the use of devices as sensors) initiatives require human observation, the data provided may be over emphasised based on the data source's perception. Clear and unambiguous instructions must be provided to all data sources to ensure that they are capable of providing the correct data required by the data consumer. In terms of the CSI participatory crowdsourcing project, the instructions are provided to the data sources on the CSI website (http://csi.ufh.ac.za/) and through message prompts.
CSF03	Public Safety Data Interpretation and Presentation Format	The data collected must be capable of being interpreted and presented with little effort; this should be considered before data collection. Increasing the probability that the data will be interpretable can be achieved by giving the data source language options to provide the data, and expressing that the use of acronyms and slang is forbidden. In terms of presentation, the data collected must be in the same format or capable of being converted to the same format.
CSF04	Public Safety Data Integrity and Security	The integrity and security of data can be addressed by ensuring that only authorised users have access to the data. Two groups of authorised users should be created; one set of users should only be allowed to view the data and the other

		set of users should be allowed to view and modify the data. It is suggested that a tracker be applied to users who can modify data so one can determine the user who modified the data. In addition, ethical issues can be avoided if all personal data is removed before use. It is also suggested that periodical backups of audio and textual data be conducted.
CSF05	Simple Participatory Crowdsourcing System Setup	The participatory crowdsourcing system should be set up in a simple manner. The public safety instructions must be clear and concise as lengthy instructions will either cause callers to skip the instructions or perceive reporting as too much of an effort, leading to reduced participation and increases in dropped calls. Additionally, this will improve the data source's user experience which will encourage them to take the added time and effort to provide a quality public safety data. System availability should be constantly tested to ensure that it is accessible when required. It is suggested that data sources are also provided with a repository to communicate system problems. The processes from data collection to data use should be as short and straight forward as possible, as long and complex processes increase the risk of negatively affecting the quality of the public safety data. This Critical Success Factor can be supported by the Rational Choice Theory used by a number of authors (Kääriäinen and Sirén, 2011; Taylora, 2002; Zhang, Messner and Liu, 2007) to explain that people weigh the costs and benefits of reporting crime. An example of the costs would be the inconvenience of reporting the crime while the benefits may be creating a safer environment or making an insurance claim. If the participatory crowdsourcing System is simple and effortless for the data source, the benefits of reporting will outweigh the costs. Although public safety is a larger scope than crime, the concept behind the theory is relevant to public safety reporting. Taylora (2002) adds that the simplicity of the system will encourage citizens to provide a higher quality report.

Table 6-5 provides the finalised version of the Critical Success Factors after refinement. The next section will discuss the relationship between the Critical Success Factors and the problem areas.

6.9 Critical Success Factors and Problem Areas

It was emphasised above that Critical Success Factors are the most important (critical) areas that need to be considered to ensure success in its respective environment. In other words, all major problems jeopardising the success of collecting quality public safety data in a participatory crowdsourcing project will be mitigated or reduced if one considers its Critical Success Factors. Table 6-6 emphasises the importance of all the Critical Success Factors by illustrating the problems that they address. Recall that the data quality problems (indicated by the code P01 to P09 in Table 6-6) were identified through assessing the 100 public safety reports (see sections 6.5 and 6.6). This also allows one to understand which Critical Success Factors require more attention than others.

Table 6-6: Critical Success Factors and Problem Areas

Problem Areas		Critical Success Factors				
Code	**Description**	**CSF01**	**CSF02**	**CSF03**	**CSF04**	**CSF05**
P01	Inconsistent collection of public safety incident date and time data.	X	X	X	X	X
P02	Inconsistent collection of public safety incident location data.	X	X	X	X	X
P03	Reports which required an immediate response were reported over a month or few weeks later.	X	X			
P04	Callers frustration overstates the nature of the issue.		X			
P05	Heavy breathing into the talking piece of the phone.		X	X		
P06	Accent makes it difficult to comprehend the intended message in terms of the pronunciation of certain words.		X	X		
P07	Caller's use of acronyms makes it difficult to comprehend the public safety report.		X	X		
P08	Speech impediments are due to language barriers and describing the event.		X	X		
P09	Data source is finding it difficult to follow data collection instructions or the instructions is not being provided clearly.				X	X
Total Problems		3	8	6	3	3

Recall that these 9 problems were considered significant enough to affect the quality of public safety data as these problems affected 95 or more out of 100 public safety reports which were assessed with the Public Safety Data Quality Criteria. All other potential problems (also assessed with Public Safety Data Quality Criteria) did not affect the quality of public safety data as these problems did not exist in 95 or more out of 100 public safety reports assessed. Therefore, public safety data quality (in the CSI participatory crowdsourcing project's context) is compromised due to these 9 problems.

Table 6-6 illustrates the areas (Critical Success Factors) which need to be addressed to ensure that the problems affecting public safety data quality can be solved by considering the 5 Critical Success Factors. This emphasises that the consideration of the Critical Success Factors are capable of solving these 9 problems and therefore, increasing the Quality of Public Safety Data in participatory crowdsourcing through audio data collection. This will make this method of participatory crowdsourcing capable of being used as a Smart City initiative and contribute to reducing the public safety problem. Although Table 6-6 shows which Critical Success Factors address the specific problems affecting Data Quality, it does not illustrate which stakeholders are responsible for the specific Critical Success Factors. This will be discussed in the next section.

6.10 Critical Success Factors and the CSI Project Areas

Since the Critical Success Factors relate to public safety data quality within participatory crowdsourcing, it is valuable to illustrate the relationship between a typical crowdsourcing project and Critical Success Factors. This is presented in Figure 6-7 below. For consistency, the CSI participatory crowdsourcing project was used as an example in Figure 6-7. An explanation of the CSI participatory crowdsourcing project illustrated in Figure 6-7 is provided to clarify one's understanding of the Critical Success Factors focus areas within the CSI participatory crowdsourcing project's context.

Through conversational analysis it was found that specific data that needed to be collected included the type of incident, date, time and location (street/landmark and suburb/highway). Based on the city's (East London) limitations such as technological accessibility and literacy rate, voice was chosen as a data collection method through mobile and landline phones. The same constraints were considered when participatory crowdsourcing (crowds as data sources) was selected. A typical crowdsourcing project, regardless of the type of crowdsourcing option selected (participatory or opportunistic), involves four stakeholder groups. Since the CSI participatory crowdsourcing project is a joint effort by Fort Hare University and IBM, these two entities undertook the responsibility of multiple stakeholders, as indicated in Figure 6-7. Note that the third Critical Success Factor (Simple Participatory Crowdsourcing System Setup) relates to the entire participatory crowdsourcing process as this Critical Success Factor refers to the simplification of processes from data collection to use.

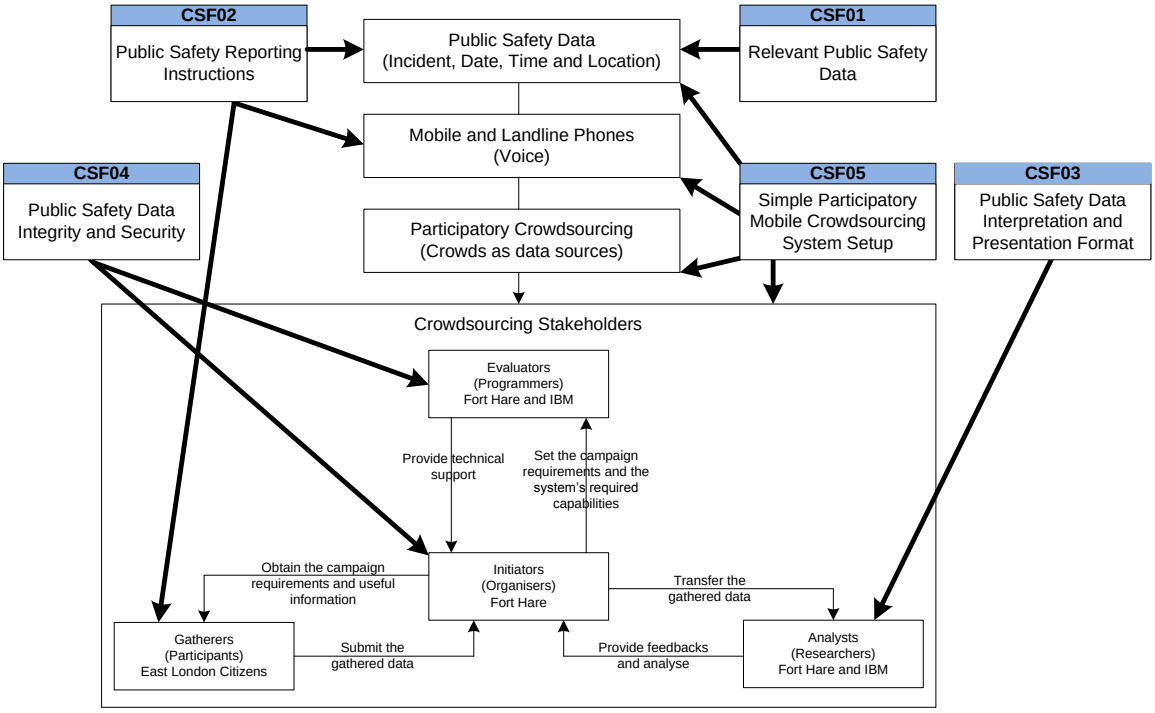


Figure 6-7: Critical Success Factors and the CSI Participatory Crowdsourcing Project

It is emphasised in Figure 6-7 that a collaborative effort is required from all crowdsourcing stakeholders to ensure that quality data is collected. The next section discusses the relationship between the Critical Success Factors and this study's research questions.

6.11 Critical Success Factors and Research Questions

This section will discuss the areas of this study which contribute towards the construction of the Critical Success Factors. It was found that three research questions needed to be answered in order to answer the main research question for this study. As mentioned before, the main research question asked: What factors are required to ensure high Quality Public Safety Data in a participatory crowdsourcing initiative, using an IVR system? Table 6-7 illustrates how the Critical Success Factors address the three research sub-questions of this study.

Table 6-7: Critical Success Factors and Research Questions

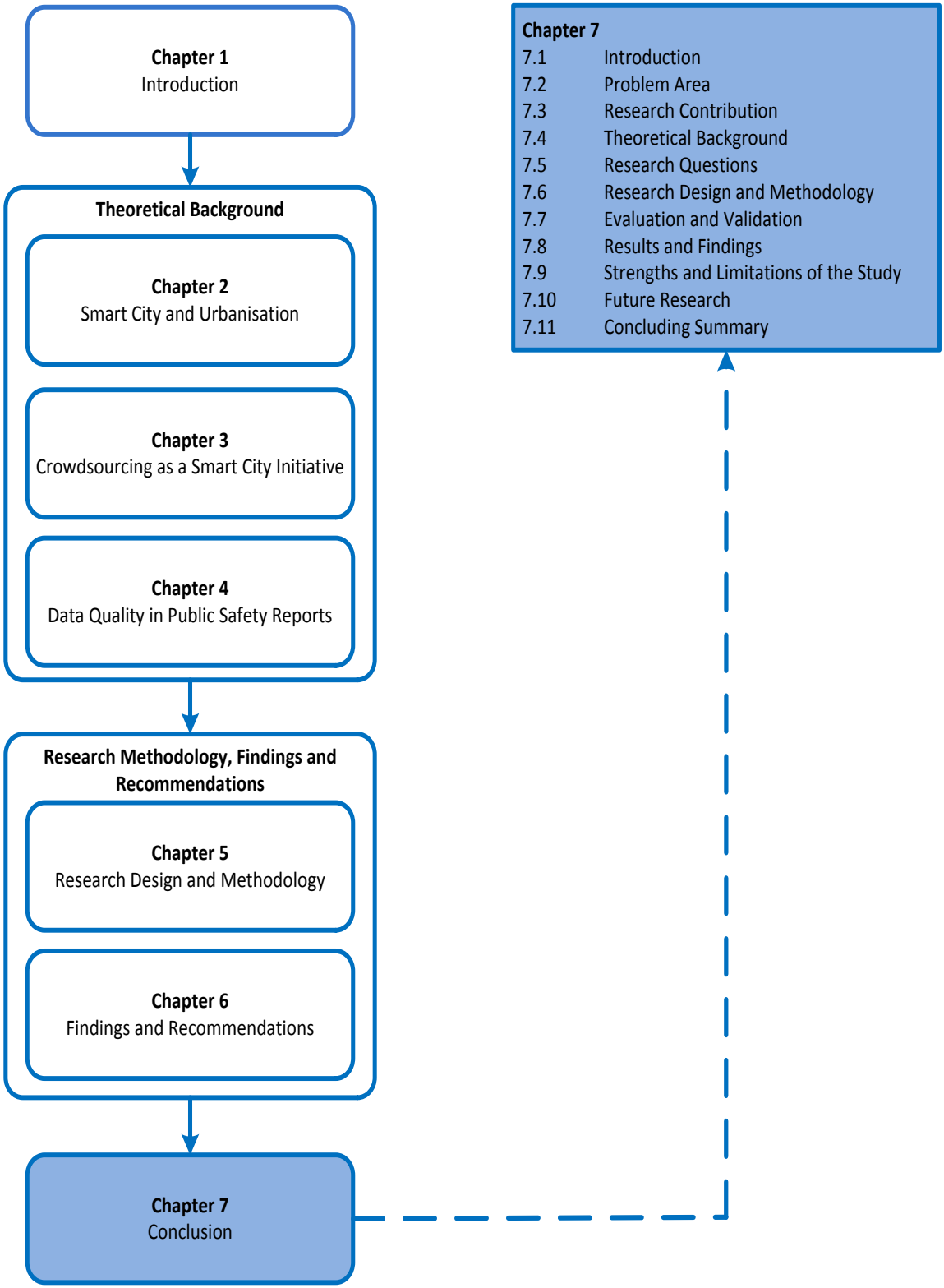
Sub-question	Research Questions	Critical Success Factors				
		CSF01	CSF02	CSF03	CSF04	CSF05
1	How can the Smart City concept be used to reduce problems associated with urbanisation?	X		X		
2	How can crowdsourcing be used as a Smart City initiative to reduce public safety issues?		X		X	X
3	What factors constitute data quality within a public safety participatory crowdsourcing context?	X	X	X	X	

Table 6-7 illustrates which Critical Success Factors address the three research sub-questions which assisted in answering the main research question. A conclusion for Chapter 6 will follow.

6.12 Conclusion

The chapter firstly provided the relationship between Smart City, crowdsourcing and data quality to indicate how secondary and primary data are tied together in developing the findings for this study. The next section explained the data collected from conversational analysis before the first expert review was discussed. The first expert review was used to assess the Public Safety Data Quality criteria developed in Chapter 4. Once the expert group's responses were taken into consideration, the Public Safety Data Quality criteria was appropriately modified before assessing the public safety reports for data quality problems. The data quality problems identified through the Public Safety Data Quality criteria were then presented. A graphical illustration of how the proposed Critical Success Factors were constructed followed. Thereafter, the second expert review was presented, and used to assess the proposed Critical Success Factors. The proposed Critical Success Factors were then refined after considering the comments and recommendations from the expert group. An explanation of each refined Critical Success Factor was provided to express what areas must be considered when addressing Critical Success Factors. The Critical Success Factors were then presented in line with the data quality problems identified in the public safety reports to illustrate their relationship. The areas within a typical crowdsourcing project were presented to indicate the areas that influence the Critical Success Factors. The last section presented the relationship between the Critical Success Factors and the research questions to illustrate the questions that contributed to the construction of each Critical Success Factor. The last chapter will provide a conclusion summarising all chapter discussion areas as well as future research suggestions.

Chapter 7 – Conclusion



7.1 Introduction

This concluding chapter will summarise and finalise what was introduced in Chapter 1 (Hofstee, 2006). The research problem area was introduced, followed by the research contribution. The next section stated the theoretical background on which this study was based. The research questions were then provided along with an explanation of how the questions were answered as well as an indication of the sections in this study which contributed to answering the research questions. The research design and methodology adopted for this study were then presented, followed by this study's evaluation and validation methods. The results and findings of the study were then presented. The next section discussed this study's strengths and limitations, followed by future research directions before a concluding summary was provided. A discussion of the problem will follow.

7.2 Problem Area

The high quality of life created in cities has encouraged people to seek refuge in urban areas. This has caused a drastic increase in urbanisation, which is expected to continue in the future (Washburn & Sindhu, 2010). Although this is in fact a positive movement, it is coupled with various problems. Most of these problems involve the scarcity of resources. The most common problems facing urbanisation include: resource depletion, infrastructure deterioration, energy shortage, environmental issues, health concerns, unemployment and public safety. Through further investigation it was found that the reason why city resources were seen as scarce was because they were not managed effectively and efficiently (Exner, Zeile, & Streich, 2011), due to the lack of information regarding the city environment. Therefore, collecting data on various occurrences around the city will improve decision-making on the distribution of city resources. This is a difficult task as it requires constantly monitoring a large geographical area over long periods of time. The use of crowdsourcing as a Smart City initiative was found to be an effective method of collecting large volumes of data in a short space of time.

Crowdsourcing is the act of utilising the public to collect data on certain events, occurrences or anomalies within their surroundings (Väätäj, Vainio, Sirkkunen, & Salo, 2011). Due to the specific barriers faced within the developing country of South Africa (technology accessibility, computer literacy and language barriers), participatory crowdsourcing, utilising voice technologies was found to be the most appropriate and effective crowdsourcing solution. Utilising participatory crowdsourcing as a Smart City

initiative allows citizens to assist their local Government and business, and contribute to the growth of their city. This study has focused on a specific participatory crowdsourcing project (CSI participatory crowdsourcing project) relating to the collection of public safety data from the citizens of East London, South Africa. The data collected will be utilised by the city's emergency and non-emergency units to improve their service delivery. However, the success of any crowdsourcing project (participatory or opportunistic) relies on the ability to collect high quality data (Burke, *et al*., 2006).

Low quality data provided by participants is a common problem in most crowdsourcing projects (Yang, Zhang, & Roe, 2011). The review of low quality data will result in ineffective decisions. More specifically, it could endanger the public and the respective response unit, leading to respondents appearing in the incorrect location or wasting time searching for the correct location, applying a diminutive work force to a problem requiring a large workforce, or appearing once the problem has been resolved or grown to an extent that the current solution becomes ineffective. In addition to response problems, inaccurate patterns could lead to ineffective distribution of emergency and non-emergency units. The research contribution section explains how this problem was addressed through this study.

7.3 Research Contribution

This research contributes to the body of information systems knowledge regarding the use of technology to overcome problems associated with urbanisation. This study explained how participatory crowdsourcing can be used as a Smart City initiative to reduce urbanisation problems by firstly determining the requirements of a Smart City initiative. It was found that a collaborative effort from citizens, businesses and local government is required (see section 2.3.1) in addition to the use of technology in a "smart" manner (see section 2.3.2). These activities can be focused within six contribution areas: Smart Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment and Smart Living (see section 2.4).

Secondly, a crowdsourcing road map (see Figure 3-2) was developed to assist the decision-making process for determining the most optimal crowdsourcing initiative based on a number of factors. The crowdsourcing road map was developed by firstly discussing the use of mobile phones as they are the most common method of data collection (see section 3.1.1). Secondly, the types of crowdsourcing options were discussed (see section 3.1.2), followed by the stakeholders involved in a typical crowdsourcing project (see section

3.1.3). Smart City and crowdsourcing was aligned by linking a number of crowdsourcing initiatives currently in progress in six contribution areas of the Smart City (see section 3.2).

Thirdly, a Public Safety Data Quality criteria was constructed to assess the quality of public safety reports from participatory crowdsourcing. The Public Safety Data Quality criteria was developed in a manner which allowed the data consumer to identify the problem areas jeopardising quality, in line with the four dimensions of Wang and Strong's (1996) Data Qaulity Framework. The four dimensions include: Contextual Data Quality (see section 4.5), Intrinsic Data Quality (see section 4.6), Representational Data Quality (see section 4.7) and Accessibility Data Quality (see section 4.8).

Fourthly, the Public Safety Data Quality criteria was tested by assessing public safety data from a specific participatory crowdsourcing project (CSI participatory crowdsourcing project) focused on the collection of public safety data (see section 6.5). The quality problems identified after the assessment assisted in determining the Critical Success Factors (see section 6.7). The Critical Success Factors provided insight into 5 important (critical) areas which needed to be considered within a public safety participatory crowdsourcing project (see Table 6-5).

In summary, the overall contribution of this study provides Critical Success Factors to ensure that high quality public safety data is collected from a public safety participatory crowdsourcing project. This assists in providing an additional technique for collecting large volumes of public safety data in an effective and efficient manner.

7.4 Theoretical Background

The secondary data relating to Smart City, crowdsourcing and data quality were directed by two theories. The theories were used to support assumptions and ensure discussions were aligned to the research problem addressed in this study. The two theories include, the characteristics of a Smart City (see sections 2.4 and 3.2), and Wang and Strong's (1996) Data Quality Framework (see section 4.3). A discussion explaining these theories and how they were used in this study will follow.

The characteristics of a Smart City were developed by Giffinger, Fertner, Kramar, Kalasek, Milanović and Meijers (2007) to illustrate the contribution areas for the Smart City concept (see section 2.4). The characteristics of a Smart City include: Smart Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment and Smart Living

(Giffinger *et al*., 2007). These six characteristics were developed from theories on urban development and growth. The six characteristics assisted the understanding of how problems associated with urbanisation can be addressed. Smart Living was the focus area of this study as public safety formed part of Smart Living (see section 2.4.6.3).

The Data Quality Framework developed by Wang and Strong (1996) illustrates 15 data quality attributes deemed most important to the data consumer (see section 4.3). These 15 data quality attributes were grouped into four distinctive categories: Contextual Data Quality, Intrinsic Data Quality, Representational Data Quality and Accessibility Data Quality. The 15 data quality attributes organised within the four categories were used as a guide when determining constructing a Public Safety Data Quality criteria to assess the public safety reports. In addition, four of the five Critical Success Factors (CSF01 TO CSF04) were developed based on this Data Quality Framework (see sections 4.5, 4.6, 4.7, 4.8 and 6.7.1). A fifth Critical Success Factor was suggested by experts in the field of study and supported by the Rational Choice Theory.

The two theories discussed above assisted in answering the research sub-question. The next section explains how the main research question and research sub-questions were answered.

7.5 Research Questions

The main research question asks: **What factors are required to ensure high quality public safety data in participatory crowdsourcing used as a Smart City initiative?** In order to answer this question, three sub-questions needed to be addressed first. These sub-questions are discussed below.

1. **How can the Smart City concept be used to reduce problems associated with urbanisation?**

   The literature relating to the Smart City assisted in finding how the Smart City concept can be used to address various problems created from urbanisation. It was found that a collaborative effort between a city's local Government, business and citizens is vital for a city to become "smarter". In addition, technology must be utilised in a "smart" manner by aligning it with the technological infrastructure of the city. It was found that the Smart City concept can address urbanisation problems in six broad areas. These areas were labelled: Smart Economy, Smart

People, Smart Governance, Smart Mobility, Smart Environment and Smart Living. A number of Smart City initiatives currently in progress were discussed to understand how these initiatives contribute to problems associated with urbanisation.

2. **How can participatory crowdsourcing be used as a Smart City initiative to reduce public safety issues?**

   Publications on crowdsourcing and various ongoing crowdsourcing projects assisted in determining how crowdsourcing could be used as a Smart City initiative. This allowed one to understand how crowdsourcing could be used as an initiative to reduce public safety issues. It was found that any crowdsourcing initiative requires the utilisation of technology which is focused on the collection of data. Based on reviewing various examples of crowdsourcing, it was found that the majority of crowdsourcing projects can be organised into one of three categories: opportunistic sourcing (1) and two types of participatory sourcing; crowds as data sources (2) and crowds as sensor operators (3). The technology chosen should be aligned with the type of crowdsourcing. This decision is commonly driven by the availability of technologies, the users' ability to use the technology (computer literacy level) and the type of data that will be collected (for example; audio, text, images and/or videos). The stakeholders involved in a typical crowdsourcing project were then discussed. The crowdsourcing stakeholders include, the data sources, the initiator, the evaluator and the analysts. These four stakeholder groups need to work together to ensure that the crowdsourcing project runs smoothly.

   It was found that the most practical public safety crowdsourcing project to be implemented in South Africa would be participatory crowdsourcing, using mobile or landline phones as a data collection device. Audio data collection was chosen over text-based due to the low literacy rate in South Africa. Audio data collection is driven by pre-recorded message prompts which are used to assist the caller to provide the correct data. Mobile and landline phones were deemed most appropriate as citizens would have the option of reporting a public safety incident as soon as it had occurred or seen.

3. **What factors constitute data quality within a public safety participatory crowdsourcing context?**

Secondary data on various attributes of data quality assisted in answering this research sub-question. The CSI participatory crowdsourcing project was used as a case study to develop a Public Safety Data Quality criteria. The factors that constitute data quality attributes were determined in the process of developing the Public Safety Data Quality criteria to assess the quality of public safety reports collected from the CSI participatory crowdsourcing project. The criteria was constructed by using 15 data quality attributes found by Wang and Strong (1996) to be important to a data consumer. A discussion on each data quality attribute was provided explaining what factors should be in place to ensure its presence. These factors assisted in the construction of questions to assess the public safety reports for data quality.

The above discussion briefly summarised how the research sub-questions were answered. Table 7-1 presents a more precise presentation of the discussion sections which assisted in answering the main question and the sub-questions.

Table 7-1: Sections That Addressed the Respective Research Questions

Research Number	Research Question	Sections	
Main Research Question	**What factors are required to ensure high quality public safety data in participatory crowdsourcing used as a Smart City initiative?**	**2.2, 2.3, 2.4, 2.5, 2.6, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8**	
Note that three research questions needed to be addressed to answer the main research question. These 3 questions were considered as research sub-questions and are indicated below, along with their contribution areas.			
Research Sub-question 1	How can the Smart City concept be used to reduce problems associated with urbanisation?	2.2 2.3 2.4 2.5 2.6 6.3 6.7 6.8	Problem with Urbanisation Smart City Defined Citizens, Businesses and Government Participation Technology Smart City Characteristics Conversational analysis Construction of Critical Success Factors Expert Review 2 - Proposed Critical Success Factors
Research Sub-question 2	How can crowdsourcing be used as a Smart City initiative to reduce public safety issues?	3.2 3.3 3.4 3.5 3.6 3.7 3.8 6.3 6.4 6.5 6.6 6.7	Crowdsourcing Defined Mobile Phones as Sensors Types of Crowdsourcing Crowdsourcing Stakeholders Components of Crowdsourcing Crowdsourcing Initiatives CSI Participatory Crowdsourcing Project Conversational Analysis Expert Review 1 - Public Safety Data Quality Criteria Public Safety Data Quality Issues Public Safety Quality Problems Construction of Critical Success Factors

		6.8	Expert Review 2 - Proposed Critical Success Factors
Research Sub-question 3	What factors constitute data quality within a public safety participatory crowdsourcing context?	4.2 4.3 4.4 4.5 4.6 4.7 4.8 6.3 6.4 6.5 6.6 6.7 6.8	Data Collection Data Quality Framework Public Safety Data Quality Criteria Contextual Data Quality Intrinsic Data Quality Representational Data Quality Accessibility Data Quality Conversational Analysis Expert Review 1 - Public Safety Data Quality Criteria Public Safety Data Quality Issues Public Safety Quality Problems Construction of Critical Success Factors Expert Review 2 - Proposed Critical Success Factors

The above discussion explained how the research questions were answered and in which sections the discussion areas can be found. The next section summarises the research design and methodology.

7.6 Research Design and Methodology

Details of the research design and methodology can be found in Chapter 5. After assessing various research paradigms *Reality as a Contextual Field of Information* was found to be most appropriate for this study. This research paradigm combines characteristics of both the positivism and the interpretivism philosophy (Collis & Hussey, 2009) so that one can have a more extensive view of the research context.

This study generally adopted the mix method approach, involving a combination of qualitative and quantitative research. The qualitative research approaches used in this study included case studies on the Smart City and participatory crowdsourcing initiatives, theories on data quality, and two sets of expert reviews to assess the Public Safety Data Quality criteria and the proposed Critical Success Factors. The quantitative research approach involved the assessment of 100 public safety reports using the Public Safety Data Quality criteria developed in Chapter 4 (and presented in Table 5-4).

The Design Science methodology was followed throughout this study. This methodology was found to be the most appropriate method to conduct this research and ensure quality results were achieved. Although the 7 guidelines of the Design Science methodology were rearranged to accommodate this study, ensuring all 7 guidelines was considered when conducting this research ensuring that the Critical Success Factors optimally addressed the problem area. The next section will discuss the research evaluation and validation.

7.7 Evaluation and Validation

Evaluation and validation ensure that research is credible and trustworthy. The evaluation and validation of this study complies with the Design Science methodology. Primary data and secondary data were used during the evaluation and validation processes. Primary data were collected from two expert reviews. An expert review is an ad hoc method used to evaluate research findings by using a group of relevant professionals (Molich & Jeffries, 2003). Secondary data were collected from journals, conference proceedings, books, websites, theories, methodologies, frameworks, models, case studies and reports. This will be discussed in further detail below.

A Public Safety Data Quality criteria was developed to assess the Quality of Public Safety Data collected through participatory crowdsourcing utilising voice technologies. The development of the Public Safety Data Quality criteria was guided by Wang and Strong's (1996) Data Quality Framework (see section 4.3). In addition, relevant literature (journals, conference proceedings, books, websites, theories, frameworks, models and reports) on Data Quality and Public Safety assisted the development of the Public Safety Data Quality criteria. The Public Safety Data Quality criteria was developed in Chapter 4 with the consolidated view provided in Chapter 5. Once the Public Safety Data Quality criteria was developed, it was presented to 6 experts for assessment. The experts were carefully selected based on their educational background and experience in order to ensure quality feedback was collected (see Table 6-1). The general feedback from the experts referred to the lengthiness of the Public Safety Data Quality criteria, making it inappropriate to assess Quality of Public Safety Data (see section 6.4.1). For this reason, the Public Safety Data Quality criteria was appropriately refined before it was used to assess 100 public safety reports. The second expert review referred to the assessment of the proposed Critical Success Factors.

The proposed Critical Success Factors were supported by Wang and Strong's (1996) Data Quality Framework (see section 4.3). These proposed Critical Success Factors were presented to 6 experts for assessment. The same expert group was used as they were capable of assessing the proposed Critical Success Factors and providing useful feedback. The experts generally explained that the proposed Critical Success Factors needed to be more specific while one expert suggested that another Critical Success Factor be added to the existing 4 proposed Critical Success Factors. The four proposed Critical Success

Factors were made more descriptive to improve one's understanding of their focus area. The suggestion of adding a fifth Critical Success Factors was deemed appropriate due to its significant influence on Data Quality. In addition, the Rational Choice Theory supported that the fifth Critical Success Factor was an appropriate addition. In addition to the use of primary data to evaluate and validate research findings, secondary data were also used.

The majority of the secondary data used in this research was utilised in Chapters 2, 3 and 4. Secondary data included a collection of journals, conference proceedings, books, websites, theories, methodologies, frameworks, models, case studies and reports.

7.8 Results and Findings

The results and findings of this study assisted the development of five Critical Success Factors which need to be considered to ensure quality public safety data is collected. The data quality problems identified after assessing the 100 public safety reports assisted when indicating the areas of focus required to ensure these problems were addressed (see Table 6-6). An explanation on each Critical Success Factor is provided below. A more detailed explanation of each Critical Success Factor can be found in Table 6-5.

1. Relevant Public Safety Data
    - The data collected from a participatory crowdsourcing project must be relevant to the task at hand. One must determine how the data collected will be used and the level of detail required.
2. Public Safety Reporting Instructions
    - Instructions provided to the data source (citizens) must be clear and unambiguous to ensure that the data is not influenced by the data source's subjective view of the incident.
3. Public Safety Data Interpretation and Presentation Format
    - To address interpretation, the data source should be provided with language options and express that the use of acronyms and slang is forbidden. In terms of presentation, the data collected must be in the same format or capable of being converted to the same format.
4. Public Safety Data Integrity and Security
    - The integrity and security of data can be addressed by ensuring that only authorised users have access. In addition, ethical issues can be avoided if

all personal data is removed before data are used. Periodic backups should also be conducted on both the audio and textual data.

5. Simple Participatory Crowdsourcing System Setup
    - The processes from data collection to data use should be as short and straightforward as possible, as long and complex processes increase the risk of negatively affecting the quality of the public safety data.

These Critical Success Factors are supported by two theories, Wang and Strong's (1996) Data Quality Framework and the Rational Choice Theory. The Data Quality Framework presents four Data Quality dimensions, deemed to influence the Quality of Data. These dimensions include: Contextual Data Quality, Intrinsic Data Quality, Representational Data Quality and Accessiblity Data Quality. In terms of the Rational Choice Theory, a number of authors (Kääriäinen and Sirén, 2011; Taylora, 2002; Zhang, Messner and Liu, 2007) explain that people weigh the costs and benefits of reporting crime. An example of the costs would be the inconvenience of reporting the crime and the benefits may be creating a safer environment or making an insurance claim. If the participatory crowdsourcing System is simple and effortless for the data source, the benefits of reporting will outweigh the costs and callers will be more willing to provide a higher quality public safety report.

7.9 Strengths and Limitations of the Study

This study has contributed to the body of knowledge by developing Critical Success Factors to ensure that quality public safety data is collected through participatory crowdsourcing. This was achieved by developing three sub-artefacts, which assisted the development of the Critical Success Factors. Firstly, a model was developed to illustrate what was required to contribute to the Smart City. Secondly, a roadmap in the form of a model was developed to assist decision-making when selecting the optimal crowdsourcing initiative. Thirdly, a public safety data quality diagram was developed to assess the data quality of public safety data collected through participatory crowdsourcing utilising voice technologies. These three sub-artefacts assisted the development of the Critical Success Factors for this study.

These findings will be published in relevant journals and conference proceedings. In addition, this dissertation will be accessible in the library of the University of Fort Hare. The findings will also be presented to IBM for future research purposes.

In terms of the limitations of this study, instructions on reporting a public safety issue are provided to the caller in the message prompts and on the website (http://csi.ufh.ac.za/). Reporting instructions are provided on the website to reduce the length of the message prompts in order to increase user experience. The instructions on the website basically explain that the data collected will only be used for research purposes at this stage, therefore, one should not expect the public safety issue to be resolved. The instructions also state that the caller should not provide any personal details when making a public safety report. The message prompts explain that a caller must read these instructions before providing a report. This was necessary to comply with the ethical considerations of this study and the CSI participatory crowdsourcing project. Therefore, internet access and a basic level of computer literacy were required to read these instruction before a public safety report could be provided. This restricted callers who did not have internet access and a basic level of computer literacy from providing a public safety report. Therefore, the 100 public safety reports assessed for quality excluded callers who had no internet access and a basic level of computer literacy. If not for these ethical restrictions, additional public safety data quality problems may have been identified. In addition, due to the high financial costs of obtaining a toll-free line for reporting on the CSI participatory crowdsourcing project, call costs were charged for calls. Although these costs were lower than the standard calls costs and conveyed to the caller before reporting, this may have restricted citizens who did not wish to incur call costs, from reporting.

7.10 Future Research

The purpose of this section is to encourage further research within this research area. The future research suggestions are based on the problem statement and research gaps identified throughout this study.

- This study has developed Critical Success Factors which ensure that high quality public safety data is collected through participatory crowdsourcing, used as a Smart City initiative. The Critical Success Factors indicated **what** areas need to be considered to ensure quality data is collected. It would be interesting to examine the implementation of participatory crowdsourcing by determining **how** these areas can be optimally addressed to ensure quality public safety data is collected through participatory crowdsourcing. These solutions can then be applied in practice to tests its effectiveness and efficiency.

- The majority of research on participatory crowdsourcing and the Smart City are applied in the context of developed countries with sophisticated and accessible technology infrastructures, high literacy rates, and high computer literacy rates. The existing findings of such research need to be adapted to suit the context of developing countries with low technological impact, low literacy rates and low computer literacy rates, as this will have an impact on Data Quality.

These suggestions will provide a useful contribution to the existing body of knowledge. The last section will provide a summary of all chapters in this study.

7.11 Concluding Summary

Chapter 7 provided a summary of the research problem area followed by the research contribution. This was intended to highlight the significance and originality of this study. The next section mentioned the theoretical background which directed this study. The research questions indicated how the theoretical background was used and explained, how the questions were answered, and also indicated which sections in this study assisted in answering the research questions. The research design and methodology summarised how this research was conducted. The evaluation and validation methods used for this study were then summarised, followed by the results and findings of the study. The next section provided the strengths and limitations of the study before future research directions were provided.

References

Abu-Hasaballah, K., James, A., & Aseltine Jr., R. H. (2007). Lessons and pitfalls of interactive voice response in medical research. *Contemporary Clinical Trials, 28*(5), 593–602.

Agarwal, S. K., Jain, A., Kumar, A., & Rajput, N. (2010). The World Wide Telecom Web browser. *Proceedings of the First ACM Symposium on Computing for Development (ACM DEV '10)* (p. article no. 4). New York: ACM.

Agarwal, S. K., Kumar, A., Nanavati, A. A., & Rajput, N. (2011). *Spoken Web: A Parallel WWW in Developing Regions.* Vasant Kunj: IBM.

Aldama-Nalda, A., Gil-Garcia, J. R., Alawadhi, S., Chourabi, H., Mellouli, S., Nam, T., *et al*. (2012). Smart Cities and Service Integration Initiatives in North American Cities: A Status Report. *The Proceedings of the 13th Annual International Conference on Digital Government Research* (pp. 289-290). College Park: ACM.

Alt, F., Shirazi, A. S., Schmidt, A., Kramer, U., & Nawaz, Z. (2010). Location-based Crowdsourcing:Extending Crowdsourcing to the Real World. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (pp. 13-22). New York: ACM.

Angeles, M. d., & García-Ugalde, F. J. (2012). Relevance of Quality Criteria According to the Type of Information Systems. *The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA)* (pp. 175-181). Saint Gilles: IARIA.

Bakıcı, T., Almirall, E., & Wareham, J. (2012). A Smart City Initiative: the Case of Barcelona. *Journal of the Knowledge Economy*, 1-14.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modelling Information Manufacturing Systems to Determine Information Product Quality. *Management Science, 44*(4), 462-484.

Barkhuus, L., & Jørgensen, T. (2008). Engaging the crowd: studies of audience-performer interaction. *Extended Abstracts on Human Factors in Computing Systems (CHI EA)* (pp. 2925-2930). New York: ACM.

Bartels, A. H. (2009). *Smart Computing Drives The New Era Of IT Growth.* Cambridge: Forrester Research.

Bell, S. (2002). *Economic Governance & Institutional Dynamics.* South Melbourne: Oxford University Press.

Benkert, R., & Peters, R. M. (2005). African American Women's Coping with Health Care Prejudice. *Western Journal of Nursing Research, 27*(7), 863-889.

Bentley, L. D., & Whitten, J. L. (2007). *Systems Analysis and Design for the Global Enterprise 7th Edition.* New York: McGraw-Hill.

Blanche, M. T., Durrheim, K., & Painter, D. (1999). *Research in Practice: Applied Methods for the Social Sciences.* Cape Town: University of Cape Town Press.

Borek, A., Woodall, P., Oberhofer, M., & Parlikad, A. K. (2011). A classification of data quality assessment methods. *Proceedings of the 16th International Conference on Information Quality (ICIQ 2011)* (pp. 189-203). Adelaide: Cambridge University Engineering Department.

Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory Sensing. *Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications* (pp. 117–134). Colorado: ACM.

Cai, K., & Wang, J. (2008). Urban design based on public safety - Discussion on safety-based urban design. *Architecture Journal, 3*(2), 219–227.

Cappiello, C., Francalanci, C., & Pernici, B. (2002). A Model of Data Currency in Multi-Channel Financial Architectures. *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)* (pp. 106-118). Boston: MIT.

Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart Cities in Europe. *Journal of Urban Technology, 18*(2), 65-82.

Carley, M., Jenkins, P., & Small, H. (2001). *Urban Development and Civil Society: The Role of Communities in Sustainable Cities.* London: Earthscan.

Chapman, A. D. (2005). *Report for the Global Biodiversity Information Facility.* Copenhagen: GBIF Secretariat.

Christin, D., Reinhardt, A., Kanhere, S. S., & Hollick, M. (2011). A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software, 84*(11), 1928-1946.

Coe, A., Paquet, G., & Roy, J. (2000). E-goverance and smart communities: a social learning challenge. Faculty of Administration, University of Ottawa.

Collis, J., & Hussey, R. (2009). *Business Research: a practical guide for undergraduate and postgraduate students.* Basingstoke: Palgrave Macmillan.

Dalvi, N., & Suciu, D. (2007). Management of probabilistic data: Foundations and challenges. *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '07)* (pp. 1-12). New York: ACM.

Dawes, J., & Rowley, J. (1998). Enhancing the customer experience: contributions from information technology. *Management Decision, 36*(5), 350-357.

Debba, P., Koen, R., Holloway, J. P., Magadla, T., Rasuba, M., Khuluse, S., & Elphinstone, C. D. (2010). *Forecasts for electricity demand in South Africa (2010 – 2035) using the CSIR sectoral regression model.* Johannesburg: Council for Scientific and Industrial Research.

Deng, L., & Cox, L. P. (2009). LiveCompare: grocery bargain hunting through participatory sensing. *Proceedings of the 10th workshop on Mobile Computing Systems and Applications (HotMobile)* (p. Article 4). New York: ACM.

Department of Economic and Social Affairs. (2007). *World Urbanization Prospects.* New York: Department of Economic and Social Affairs.

Desourdis, R. I. (2010). Achieving Smart City Resilience: A Sense-and-respond Architecture for Emergency Preparedness and Disaster. *Pacific Telecommunications Council* (pp. 1-20). Honolulu: Curran Associates.

Dodgson, M., & Gann, D. (2011). Technological Innovation and Complex Systems in Cities. *Journal of Urban Technology, 18*(3), 101-113.

Dong, Y. F., Kanhere, S., Chou, C., & Bulusu, N. (2008). Automatic Collection of Fuel Prices from a Network of Mobile Cameras. *Proceedings of the 4th IEEE international conference on Distributed Computing in Sensor Systems (DCOSS)* (pp. 140-156). Heidelberg: Springer-Verlag.

Eastern Cape Socio Economic Consultative Council. (2012). *Eastern Cape Development Indicators.* Eastern Cape: Quantec.

Eger, J. (2003). Smart communities: becoming smart is not so much about developing technology as about engaging the body politic to reinvent governance in the digital age. *Urban Land, 60*(1), 50–55.

Exner, J.-P., Zeile, P., & Streich, B. (2011). Urban Monitoring Laboratory: New Benefits and Potential for Urban Planning through the Use of Urban Sensing, Geo- and Mobile-Web. *Real Corp*, (pp. 1087-1096). Wien.

Farrell, R. G., Rajput, N., Das, R., Danis, C., & Dhanesha, K. (2010). Social navigation for the spoken web. *Proceedings of the fourth ACM conference on Recommender systems (RecSys '10)* (pp. 333-336). New York: ACM.

Fiske, S. T. (2002). What We Know Now About Bias and Intergroup Conflict, the Problem of the Century. *Current Directions in Psychological Science, 11*(4), 123-128.

Fuzile, L. (2011). *Local Government Budgets and Expenditure.* Pretoria: National Treasury.

Gackowski, Z. J. (2006). Quality of Informing: Credibility Provisional Model of Functional Dependencies. *The International Journal of an Emerging Transdiscipline, 9*(1), 99-114.

Ganti, R. K., Ye, F., & Lei, H. (2011). Mobile Crowdsensing: Current State and Future Challenges. *Communications Magazine, 49*(11), 32-39.

Gendron, M., & Shanks, G. (2003). The Categorical Information Quality Framework (CIQF): A Critical Assessment and Replication Study. *Pacific Asia Conference on Information Systems (PACIS)* (pp. 626-638). Adelaide: Association for Information Systems AIS Electronic Library (AISeL).

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Milanović, N. P., & Meijers, E. (2007). *Smart cities: Ranking of European medium-sized cities.* Vienna: Centre of Regional Science.

Gloeckler, G. (2009, June 22). *Jack Welch Launches Online MBA*. Retrieved February 9, 2013, from Bloomberg Businessweek: http://www.businessweek.com/bschools/content/jun2009/bs20090622_962094.htm

Goldkuhl, G. (2003). Conversational analysis as a theoretical foundation for language action approaches? *8th International Working Conference on the Language Action Perspective (LAP2003)*, (pp. 1-14). Tilburg.

Gottdiener, M. (2001). *The Theming of America.* Boulder: Westview Press.

Grabe, M. E., Samson, L., Yegiyan, N. S., & Zelenkauskaite, A. (2009). Gender in Political News Reportage: Messenger and Message. *International Communication Association*, (pp. 1-36). Chicago.

Graham, S. (2002). Bridging UrbaDigital Divides? Urban Polarisation and Information and Communications Technologies (ICTs). *Urban Studies, 39*(1), 33-56.

Graves, A. (2010). Integrating and publishing public safety data using semantic technologies. *Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities (dg.o '10)* (pp. 193-199). Puebla: Digital Government Society of North America.

Greeff, M., Coetzee, L., & Pistorius, M. (2008). Usability evaluation of the South African National Accessibility Portal interactive voice response system. *Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology (SAICSIT '08)* (pp. 76-85). New York: ACM.

Griffith, J. (2000). Smart Governance for Smart Growth: The Need for Regional Governments. *Georgia State University Law Review, 17*(4), 1019-1062.

Grönlund, Å., Heacock, R., Sasaki, D., Hellström, J., & Al-Saqaf, W. (2010). *Increasing transparency and fighting corruption through ICT: empowering people and communities.* Stockholm: SPIDER, The Swedish Program for ICT in Developing Regions.

Haubensak, O. (2011). Smart Cities and Internet of Things. *Business Aspects of the Internet of Things*, 33-39.

Hester, V., Shaw, A., & Biewald, L. (2010). Scalable crisis relief: Crowdsourced SMS translation and categorization with Mission 4636. *Proceedings of the First ACM Symposium on Computing for Development* (p. 15). New York: ACM.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems research. *MIS Quarterly, 28*(1), 75-105.

Hofstee, E. (2006). *Constructing a Good Dissertation: A Practical Guide to Finishing a Master's, MBA or PhD on Schedule.* Johannesburg: EPE.

Hollands, R. G. (2008). Will the real smart city please stand up? City: analysis of urban trends, culture, theory, policy, action. *City: Analysis of Urban Trends, Culture, Theory, Policy, Action, 12*(3), 303-320.

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi Technique: Making Sense Of Consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8.

Hufty, M. (2011). Investigating Policy Processes: The Governance Analytical Framework (GAF). In U. Wiesmann, & H. Hurni, *Research for Sustainable Development: Foundations, Experiences, and Perspectives* (pp. 403-424). Bern: University of Bern.

International Bank for Reconstruction and Development. (2006). *A Decade of Measuring the Quality of Governance - Governance Matters 2006: World Governance Indicators.* Washington DC: International Bank for Reconstruction and Development.

Ishida, M. (2011). *Data Management in the United States and Canada: Academic Libraries' Contribution.* Vancouver: University of British Columbia.

Kääriäinen, J., & Sirén, R. (2011). Trust in the police, generalized trust and reporting crime. *European Journal of Criminology, 8*(1), 65-81.

Kahn, B. K., & Strong, D. M. (1998). Product and Service Performance Model for Information. *Proceedings of the 1998 Conference on Information Quality* (pp. 102-115). Cambridge: Massachusetts Institute of Technology.

Kanjo, E., Bacon, J., Roberts, D., & Landshoff, P. (2009). MobSens: Making Smart Phones Smarter. *IEEE Pervasive Computing, 8*(4), 50-57.

Kanter, R. M., & Litow, S. S. (2009, June). Informed and interconnected: a manifesto for smarter cities. Harvard Business School Working Paper.

Katerattanakul, P., & Siau, K. (1999). Measuring information quality of web sites: development of an instrument. *Proceedings of the 20th international conference on*

Information Systems (ICIS '99) (pp. 279-285). Atlanta: Association for Information Systems.

Klein, A., & Lehner, W. (2009). Representing Data Quality in Sensor Data Streaming Environments. *Journal of Data and Information Quality, 1*(2), no. 10.

Komninos, N., Pallot, M., & Schaffers, H. (2012). Smart Cities and the Future Internet in Europe. *Journal of Knowledge Economy, 3*(2), 1-16.

Lin, S.-W., & Fu, H.-P. (2012). Uncovering Critical Success Factors for Business-to-Customer Electronic Commerce in Travel Agencies. *Journal of Travel & Tourism Marketing, 29*(6), 566-584.

Lobo, S., Doke, P., & Kimbahune, S. (2010). Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NORDICHI). *GappaGoshti™ – A Social Networking Platform for Information Dissemination in the Rural World* (pp. 727-230). Reykjavik: ACM.

Lombardi, P., Giordano, S., Farouh, H., & Yousef, W. (2011). An Analytic Network Model for Smart Cities. *Proceedings of the International Symposium on Analytic Hierarchy Process* (pp. 1-6). Sorrento: AHP Academy.

Maisonneuve, N., Stevens, M., Niessen, M. E., & Steels, L. (2009). NoiseTube: Measuring and mapping noise pollution with mobile phones. *Information Technologies in Environmental Engineering, 4*(1), 215-228.

Massoud, A. S. (2005). Toward a smart grid: power delivery for the 21st century. *Power and Energy Magazine, 3*(5), 34-41.

McGilvray, D. (2008). *Executing Data Quality Projects:*. Burlington: Morgan Kaufmann Publishers.

Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing.* Gaithersburg: NIST Special Publication 800-145.

Molich, R., & Jeffries, R. (2003). Comparative expert reviews. *Extended Abstracts on Human Factors in Computing Systems (CHI EA 03)* (pp. 1060-1061). New York: ACM.

Monbiot, G. (2000). *The Captive State.* London: Macmillan.

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., *et al*. (2009). PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. *Proceedings of the 7th international conference on Mobile systems, applications, and services (MobiSys)* (pp. 55-68). New York: ACM.

Myers, M. D. (2009). *Qualitative Research in Business & Management.* London: Sage Publications.

Najjar, L., & Schniederjans, M. J. (2006). A Quality Information Benchmarking Methodology: A US Banking Industry Empirical Study. *Quality and Quantity, 40*(3), 419–433.

Nam, T., & Pardo, T. A. (2011). Conceptualizing smart city with dimensions of technology, people, and institutions. *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 282-291). New York: ACM.

Nam, T., & Sayogo, D. S. (2011). Government 2.0 collects the wisdom of crowds. *Proceedings of the Third international conference on Social informatics (SocInfo)* (pp. 51-58). Springer-Verlag: Berlin.

Niermann, S. (2005). Optimizing the Ordering of Tables With Evolutionary Computation. *The American Statistician, 59*(1), 41-46.

Oates, B. J. (2006). *Researching Information Systems and Computing.* London: Sage Publications.

Paddock, J. R., Terranova, S., & Giles, L. (2001). SASB Goes Hollywood: Teaching Personality Theories Through Movies. *Teaching of Psychology, 28*(2), 117-121.

Peng, Y. (2002). *Information quality of the Jordan Institute for Families web site.* Chapel Hill: University of North Carolina.

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

Prat, N., & Madnick, S. (2008). Measuring Data Believability: A Provenance Approach. *Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS '08)* (pp. 393-402). Washington: IEEE Computer Society.

Punch, K. F. (2006). *Developing Effective Research Proposals.* London: Sage Publications.

Puri, G. (2012). Critical Success Factors in E-learning: An Empirical Study. *International Journal of Multidisciplinary Research, 2*(1), 149-161.

Rainbolt, G. W., & Dwyer, S. L. (2012). *Critical Thinking: The Art of Argument.* Boston: Wadsworth Cengage Learning.

Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., & Hansen, M. (2007). Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype. *Proceedings of the 4th workshop on Embedded networked sensors* (pp. 13-17). New York: ACM.

Reddy, S., Shilton, K., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2008). Evaluating Participation and Performance in Participatory Sensing. *Proceedings of*

International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems (UrbanSense08) (pp. 1-5). Raleigh: Sensys.

Rosenstock, I. M. (2005). Why People Use Health Services. *The Milbank Quarterly, 83*(4), 1–32.

Ross, D., Loeffler, K., Schipper, S., Vandermeer, B., & Allan, G. M. (2013). Do Scores on Three Commonly Used Measures of Critical Thinking Correlate With Academic Success of Health Professions Trainees? A Systematic Review and Meta-analysis. *Academic Medicine, 88*(5), 724–734.

Russell, B. D. (2010). Intelligent Systems for Improved Reliability and Failure Diagnosis in Distribution Systems. *IEEE Trans. Smart Grid, 1*(1), 48-56.

Sackmann, S. A. (1991). Uncovering Culture in Organizations. *Journal of Applied Behavioral Science, 27*(3), 295-317.

Sample, I. (2004, March 8). *PCs: the latest waste mountain*. Retrieved from The Guardian: http://www.guardian.co.uk/science/2004/mar/08/sciencenews.theguardianlifesupplement

Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M., & Oliveira, A. (2011). Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation. *Future Internet Assembly, 6656*(1), 431-446.

Schneider, D., Moraes, K., de Souza, J. M., & Esteves, M. G. (2012). CSCWD: Five characters in search of crowds. *Computer Supported Cooperative Work in Design* (pp. 634 - 641). Wuhan: IEEE.

Seaton, A., MacNee, W., Donaldson, K., & Godden, D. (1995). Particulate air pollution and acute health effects. *The Lacent, 345*(8943), 176–178.

Shilton, K., Ramanathan, N., Reddy, S., Samanta, V., Burke, J., Estrin, D., *et al*. (2008). Participatory design of sensing networks: strengths and challenges. *Proceedings of the Tenth Anniversary Conference on Participatory Design (PDC '08)* (pp. 282-285). Indianapolis: Indiana University.

Solomon, S. E., & Uchida, C. D. (2003). *311: Building a 3-1-1 System for Police Non-Emergency Calls: A Case Study of the City of Austin Police Department.* Austin: 21st Century Solutions.

Srivastava, M., Abdelzaher, T., & Szymanski, B. (2012). Human-centric sensing. *Philosophical Transactions of the Royal Society, 370*(1), 176–197.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

Taylora, N. (2002). Under-Reporting Of Crime Against Small Businesses: Attitudes Toward Police And Reporting Practices. *Policing and Society: An International Journal of Research and Policy, 13*(1), 79-89.

The World Bank. (1991). *Managing Development - The Governance Dimension: A Discussion Paper.* Washington DC: The World Bank.

Väätäj, H., Vainio, T., Sirkkunen, E., & Salo, K. (2011). Crowdsourced news reporting: supporting news content creation with mobile phones. *Proceedings of the 13th Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2011)* (pp. 435-444). Stockholm: ACM.

Väätäjä, H., Sirkkunen, E., & Salo, K. (2011). Crowdsourced news reporting: supporting news content creation with mobile phones. *International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 435-444). New York: ACM.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Ward, J., & Peppard, J. (2002). *Strategic planning for information systems.* England: John Wiley & Sons.

Wasan, S. K., Bhatnagar, V., & Kaur, H. (2006). The impact of data mining techniques on medical diagnostics. *Data Science Journal, 5*(1), 119-126.

Washburn, D., & Sindhu, U. (2010). *Helping CIOs Understand "Smart City" Initiatives.* Cambridge: Forrester Research.

Wildeman, R. C., Gilpin, M., & Knoll, A. (2009). *Smart Grid Technologies: Coming To A Utility Near You.* Forrester Research.

Yang, H., Zhang, J., & Roe, P. (2011). Using Reputation Management in Participatory Sensing for Data Classification. *Procedia Computer Science, 5*(1), 190–197.

Ząbkowski, T. (2010). *Data Completeness Estimation in Management information Systems.* Warsaw: WULS Press Warsaw.

Zhang, L., Messner, S. F., & Liu, J. (2007). An Exploration of the Determinants of Reporting Crime to the Police in the City of Tianjin, China. *Criminology, 45*(4), 959–984.

Acronyms

Abbreviation	Description
IBM	International Business Machines
IVR	Interactive Voice Response
CSI	Crowdsourcing Safety Initiative

Appendix A – Participant Agreement

PARTICIPANT AGREEMENT

Thank you ("You") for agreeing to participate in a research activity undertaken jointly by the University of Fort Hare ("University") and International Business Machines Corp. and its subsidiaries ("IBM").

By signing below, You understand and agree as follows:

1. You will provide information, at Your discretion, about public safety related incidents observed by You in the [IDENTIFY APPROPRIATE GEOGRAPHIC REGION] of South Africa. You will provide information in English, over the telephone, using the phone number provided by the University to reach a recording system operated by IBM. You will not provide the phone number to any third party; it is only for Your use.

2. Before making a report, You will make sure that the incident has already been reported to the appropriate authorities.

3. You are at least 18 years of age.

4. You will not provide confidential or proprietary information belonging to You or any third party. You should not provide any personal identifying information.

5. Information You provide to the recording system will be reviewed by the University and/or IBM. You understand and agree that the University or IBM may, in their sole discretion, but are not obligated to, delete or remove any personal identifying information or any information that in the sole judgment of the University or IBM violates this Participant Agreement, or which might be offensive, illegal, or that might violate the rights, harm or threaten the safety of others.. The remaining information ("Usable Content") may be used by the University and IBM for any purpose, including research into the use of people as a source of information for cities concerning city affairs. When You make a report using the recording system, You automatically grant to the University and IBM an irrevocable, perpetual, non-exclusive, transferable, fully paid, worldwide license (with the right to sublicense) to use, copy, publicly perform, publicly display, reformat, translate, excerpt (in whole or in part) and distribute Usable Content for any purpose, to prepare derivative works of Usable Content or incorporate Usable Content into other works, and to grant and authorize sublicenses of the foregoing.

6. Under no circumstances, including, but not limited to, negligence, will the University, IBM, or their directors, employees, agents or subcontractors be liable to You or any third person for any indirect, incidental, punitive or consequential damages, including lost profits, arising from Your participation or the use of, or inability to use, the recording system, even if advised of the possibility of such damages. In no event will the University or IBM's total liability to You for all damages, losses, and causes of action (whether in contract, tort (including, but not limited to, negligence), or otherwise) exceed Five hundred US Dollars (500 US $).

7. The University or IBM can terminate this Participant Agreement, and Your authorized access to the recording system, immediately without notice if, in the University or IBM's discretion, You fail to comply with any term or provision of this Participant Agreement.

8. This Participant Agreement will be governed by the substantive laws of South Africa. Any proceedings to resolve disputes relating to this Participant Agreement will be subject to the exclusive jurisdiction of the courts of [IBM WILL INSERT JURISDICTION SHORTLY].

Accepted and Agreed to:

_____	_____	_____
Printed Name	Signature	Date

Appendix B – Characteristics of a Smart City (Giffinger *et al.*, 2007)

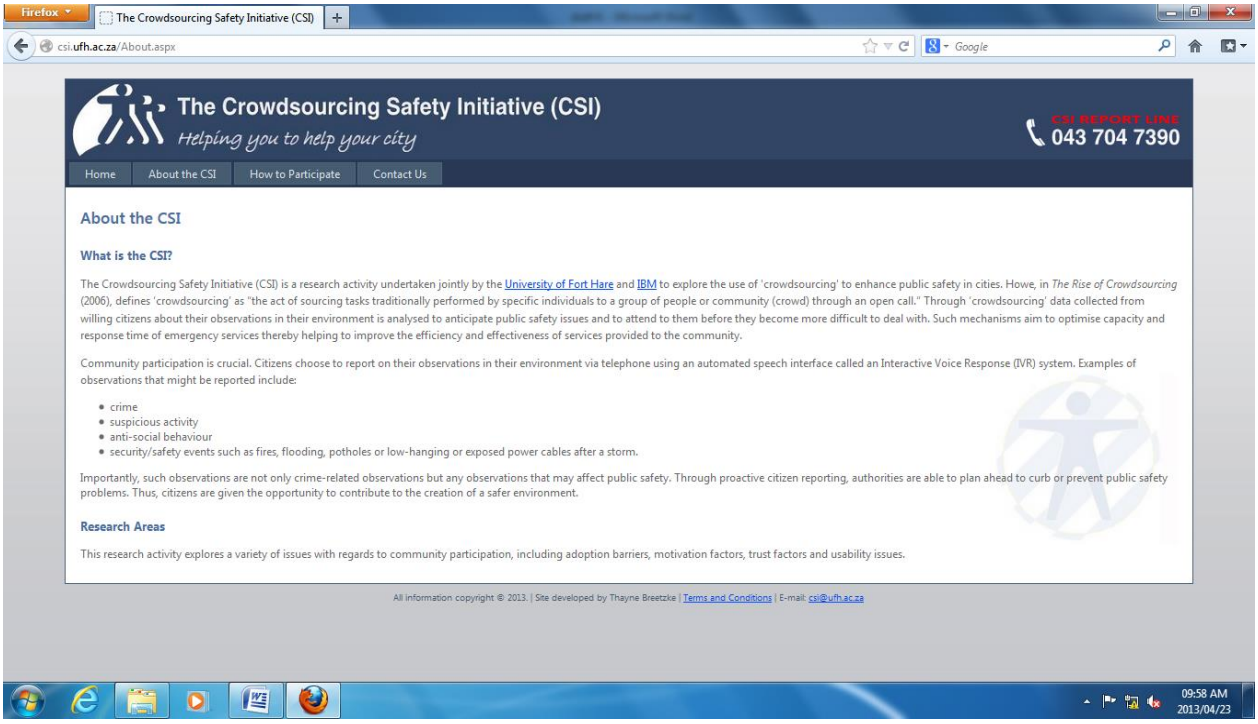
Smart Economy	
Factor	**Indicator**
Innovative spirit	R&D expenditure in % of GDP
	Employment rate in knowledge-intensive sectors
	Patent applications per inhabitant
Entrepreneurship	Self-employment rate
	New businesses registered
Economic image & trademarks	Importance as decision-making centre (HQ etc.)
Productivity	GDP per employed person
Flexibility of labour market	Unemployment rate
	Proportion in part-time employment
International embeddedness	Companies with HQ in the city quoted on national stock market
	Air transport of passengers
	Air transport of freight
Smart People	
Factor	**Indicator**
Level of qualification	Importance as knowledge centre (top research centres, top universities etc.)
	Population qualified at levels 5-6 ISCED
	Foreign language skills
Affinity to life long learning	Book loans per resident
	Participation in life-long-learning in %
	Participation in language courses
Social and ethnic plurality	Share of foreigners
	Share of nationals born abroad
Flexibility	Perception of getting a new job
Creativity	Share of people working in creative industries
Cosmopolitanism/ Open-mindedness	Voters turnout at European elections
	Immigration-friendly environment (attitude towards immigration)
	Knowledge about the EU
Participation in public life	Voters turnout at city elections
	Participation in voluntary work
Smart Governance	
Factor	**Indicator**
Participation in decision-making	City representatives per resident
	Political activity of inhabitants
	Importance of politics for inhabitants
	Share of female city representatives
Public and social services	Expenditure of the municipal per resident in PPS
	Share of children in day care
	Satisfaction with quality of schools
Transparent governance	Satisfaction with transparency of bureaucracy
	Satisfaction with fight against corruption
Smart Mobility	
Factor	**Indicator**
Local accessibility	Public transport network per inhabitant
	Satisfaction with access to public transport
	Satisfaction with quality of public transport
(Inter-)national accessibility	International accessibility
Availability of ICT infrastructure	Computers in households
	Broadband internet access in households
Sustainable, innovative and safe transport systems	Green mobility share (non-motorized individual traffic)
	Traffic safety

	Use of economical cars
Smart Environment	
Factor	**Indicator**
Attractivity natural conditions	Sunshine hours
	Green space share
Pollution	Summer smog (Ozone)
	Particulate matter
	Fatal chronic lower respiratory diseases per inhabitant
Environmental protection	Individual efforts on protecting nature
	Opinion on nature protection
Sustainable resource management	Efficient use of water (use per GDP)
	Efficient use of electricity (use per GDP)
Smart Living	
Factor	**Indicator**
Cultural facilities	Cinema attendance per inhabitant
	Museums visits per inhabitant
	Theatre attendance per inhabitant
Health conditions	Life expectancy
	Hospital beds per inhabitant
	Doctors per inhabitant
	Satisfaction with quality of health system
Individual safety	Crime rate
	Death rate by assault
	Satisfaction with personal safety
Housing quality	Share of housing fulfilling minimal standards
	Average living area per inhabitant
	Satisfaction with personal housing situation
Education facilities	Students per inhabitant
	Satisfaction with access to educational system
	Satisfaction with quality of educational system
Touristic attractivity	Importance as tourist location (overnights, sights)
	Overnights per year per resident
Social cohesion	Perception on personal risk of poverty
	Poverty rate

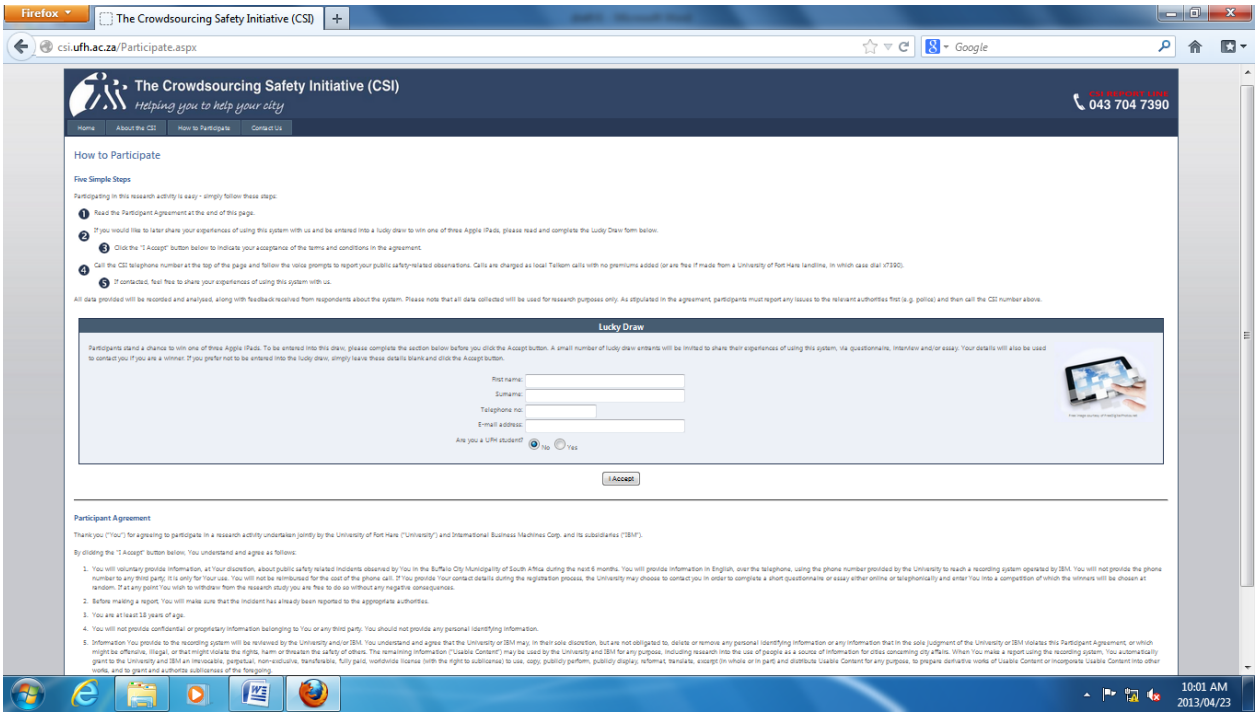
Appendix C – CSI Website (http://csi.ufh.ac.za/)

Home Page

About Us Page



How to Participate Page



Contact Us Page

Appendix D – Transcribed Public Safety Reports

	Public Safety Report
1	The incident is the state of the road. It is full of potholes. The road is Dick King road in Haven Hills or Amalinda. It's an ongoing problem and we have reported it repeatedly to the Municipal Road Division and to Mr. _____ who is in charge. The problem is so bad that to drive on the road is incredibly difficult especially as your car rocks from side to side and it can hurt your back. This has been ongoing for the past 3 years. Thank you
2	While travelling down Drake Road in Stirling on the third of January at about half past nine we noticed that manhole covers were off and that could be a danger to traffic
3	Hi, on the tenth of January I used Nina's and the Chinese shop at... in Epsom Road and there are about 3 car guards there. Two of which are probably drunk. They all 3 harass you as you step out of your vehicle vying for money from you. All 3 of them argue with each other, even if you say you are going into the shop for 5 minutes they are determined that they want to be paid. This... Their presence as you step out and into your vehicle is not very safe and you feel intimidated by their presence
4	It was on the 24 of December in Oxford Street in East London, I was in the taxi then the taxi driver was going… was stopping for someone but he didn't look at the cars that was behind him. He stopped, then… when this person get in the car he just went on without looking in the mirror to see if the car was near or what. Then that nearly caused an accident.
5	There are potholes at both sides of third Avenue in Gonubie making it impossible to access that road
6	Hi there. I would like to report a… many potholes in Gonubie, 17th Avenue. That was today at, 11 January 2013, Friday at about half past seven this morning
7	There are loiterers sitting on the corner of Third and eight of Gonubie. They are waiting for someone to come give them work but when you drive… pull into your driveway they come and harass you for money
8	I noticed potholes in Gonubie, 9th Avenue… on the corner of 9th avenue and 3rd street on… today's date is the 18th of January 2013. Time now is 12 minutes past 12
9	I am just reporting that garbage hasn't been collected in over 6 weeks in Muller street, Southwernwood
10	On Woolwash road in Amalinda we have frequent accidents on… at the intersection between Woolwash and Sunnyside Road. Many cars don't stop at those intersection and we've got frequent accidents. I would like to report that
11	Potholes in Gullsway, Gonubie close to Goubie High school. 7:20 big pothole has to drive through it and BCM never picks… pick up their phone. Never Ever
12	It's the 25th of January 2013. there is a manhole without a cover on the pavement on the corner of Cambridge and Commissioner Street which could pose a danger to any pedestrian
13	25th of January. Near fatal accident corner of Lukin and Gately Road. A Mercedes Benz slammed on brakes straight after going through a robot and a number of cars almost piled up into it.
14	The date today is the 29th of January 2013. The time is 2:03. I'm complaining about an empty house by Quinera Drive, Beacon Bay. Vagrants are living there. It is between number 31 and number 29 so it doesn't have a number on the gate. It belongs to the municipality or the Roads Department. There's a vacant house.
15	On Saturday the 19th of January in Nahoon there were people shooting flares off which is quite dangerous. Making a tremendous noise and I think causing a bit of a fire hazard. This was at the end of Beach Road
16	There is a broken...water drain cover in third avenue in Gonubie
17	Hi there I'm...referring to your template that I received today... In connection with the CSI report line...Am phoning here from Collandale, East London and I just wish to report that I have actually reported this numerous times to the Municipality, the BCM that there are potholes in our vicinity in Willow Park as well as in Collandale and issues such as the grass that never gets cut...the grass was last cut two years back it growing more to two meters already and its causing a crime issue and its causing flies and rats...and criminals and things to to to come and bother us here... and with that already crime...and...damage to our vehicles... in this area. I wish that you would be able to do something about it because as I say it is hopeless to talk to the BCM Municipality they just make promises which they never never never adhere to and never do anything about it. Thank you very much for this...
18	...Hosoff Hansen Avenue, 13 Greenfeild East London...the tarred road in front of our house is an absolute disgrace it's just potholes and they fill it up with...dirt ground and after the rains it's even worse. I would like somebody to come and seriously investigate...this situation... and solve the problem...this is still going to cause a major accident

19	There is a pothole where I live and exposed wire, where I live thank you
20	Its 13 February, the time is 11:50; I would like to report an open electrical box opposite Cash Crusaders in town.
21	There is a street light that's fallen down on Gardenia Streets in Braelyn Heights....I saw that event a week ago and it hasn't been addressed yet.
22	There is a broken concrete slab outside the University of Fort Hare... in town... The incident was spotted yesterday afternoon.
23	Mags have been stolen off my car in the Amalinda area last night. Thank you
24	I would like to report a housebreaking that occurred in Kennington road on the 2nd of February at around 1630 in the afternoon, in Nahoon. A flat screen TV, cellphones and other valuable items were stolen. We have reported the matter to the police.
25	On the 19th of February at around 2pm an ATM close to Abbotsford Spar was vandalised, in broad daylight, right next to the shops, nobody saw anything but we went on to notify the bank.
26	I would like to report a housebreaking that occurred on Monday the 18th at Hatfield Avenue in Greenfields, there two male suspects one was wearing a Blue tracksuit top they got away in a white Mercedes Benz. Thank you
27	Someone has just broken into my car outside Fort Hare University about 3 hours ago today, thank you.
28	Someone broke the burglar guards on my house in Quigney on Moore street this happened on Monday at 9.30. Thank you
29	My name ____ I was driving near Exel in Beacon Bay coming from Retail Park and there was a man who was having ... Supposedly a fit and apparently and ambulances have been trying to get to him over the past week. And it is a scam and he actually just cons people into giving him the luxuries of money to obviously survive and get along. Thank you
30	Hi my name is ____ and this happened around the 2nd of February there are ... Incidents of remote jamming if you could call it remote jamming were people are ...programming remotes to open... cars safety features...and they actually programming the remote such that when the person locks the car they can just press their button and get into the car and start the car. And the few incidents happened at Nahoon Spar where they had a stolen Mercedes and an Opel Corsa or something like that. Thank you
31	Hi... This is...I have phoned to tell you that one of my mother's friends from work, her work college received a call from an unknown male stating that he knew where she stayed and stuff like that. I think most people don't realise that personal information on social networking site are very....broadly distributed and people need... to be more careful of strangers...and weird and spiteful people out there. And people should be aware who they come into contact with when they are in the public and should always keep an eye out, although it doesn't mean people should live in fear. Just that they should take precautions....thank you
32	On January the 25th there was a house breaking in....Vincent where an old white lady was tied up by 3 African male occupants, and they tied her up at about 10am in the morning and they stole her laptop and jewellery and I think a play station. And I don't think many people are aware that leaving your.... elderly or young ones at home is a very inviting thing to criminals because they know that they are defenseless in their own way and people should always take precautions in ensuring that youngsters and elderly are safe at home. Whether they are workers, garden service or neighbors around. People need to be aware thank you
33	There is a nasty pothole at Retail Park intersection turning right from Eskom towards the N2 turn off
34	I would like to report an incident of an uneven road surface, on Fleet Street on the far left lane right opposite the fire station
35	I saw a big pothole on the 7th of February on Batting road on the traffic circle towards Retail Park
36	There is a bad pothole on the intersection of Meadow road and Western Avenue
37	Hi my name is ____ ____ my car was stolen on the 6th of February, at Beacon Bay retail Park, a white Nissan Sentra registration number _____. The car has a red fog light at the back and a scratch on the left side back door and fender as well as dashboard broken behind steering wheel. If anyone sees this car please contact me on this number _____. Thank you
38	Hi, I would like to report a deep pothole at Hemmingways on the second circle when u coming from the casino side. It's a very big pothole.
39	On the 7th of February I saw a huge a pothole on the edge...on Edge Rd, nearing Bird street intersection towards the circle. It's a big bad pothole.
40	This is a report of a house breaking that occurred on the 7th of February, during the day at Rowland Rd in Morningside. Windows and the back door were damaged
41	On the 14th of February I saw a big deep pothole at the edge of the road between Burmeisters and Pennypinchers, coming into Beacon Bay

42	....there is grass and things like that that need to be cut and... In Greenfield... Today's date is the 29th....we would like these things to be sorted out and things like that...things such as grass and things like that are high and... Even so potholes.... are causing problems for everyone here thank you bye bye.... I am ____ ____ and I would like to actually join...
43	Hi there...my public...safety issue is that... In the University of Fort Hare between the two buildings, which is Nkuhlu building. No between the Elco building and the main building there is a road in which cars pass by quickly and drive fast. Thus,....I think we should create, I think it will be a great precautionary measurement that we create... a zebra crossing where students will cross. They have a particular place to cross rather than them just crossing anywhere, sometimes when you cross there you don't see any car coming it can hit you....This incident happened, I almost got hit by a car....the date was...actually today...which is... let me check the date... 1st of March 2013.Thank you
44	...My safety initiative is that... On the 6th...at 7:30am....this morning... I was on my way to school and I saw...in Duncan Village I saw a couple of guys stealing copper from the danger. And this usually occurs weekends... And then Municipality replaces them and the next weekend it happens again. Thank you
45	Okay, it's the 4th of March early morning. I want to give a warning to motorist, MVA on NEX inbound, traffic is packed up all the way to Sasol. Take alternative route if possible.
46	Hi please look out for a Ford Laser registration number ____. It was stolen right outside Cambridge Spar on the 8th of February. Please contact ____ or ____ on ____
47	...In Scenary park... It's a.... in there is a location there of informal settlements and....I saw a couple of illegal connections on the pole....I didn't pay attention to the name of the street...yes...I hope this will help
48	It's the 8th of February I would like to report another pothole on NEX town direction just after the Stirling off ramp on the left hand lane
49	A vehicle was broken into on the e 27th of February on Marrison road on Cambridge West. The back window was smashed and the XTC car radio was stolen. If anyone has offered this car for sale, please contact ___ on ___
50	A house was broken into on the 7th of February in Fairfield Creasent. A laptop, play station and jewellery was stolen
51	Huge pothole in Bonza Bay road when you turn left from Baiting Road towards Retail Park. In is in the drive lane of most cars beware it will cause serious damage
52	On the 6th of February there was a reported stabbing incident at Checkers in Nahoon. Apparently a car guard stabbed another person. Civilians tried to assist the best way they could.
53	It's the 1st of March and a flat screen TV was stolen from a business in Kennington road, Nahoon last night. Security gate and front door kicked open.
54	There is a recently fixed pothole that opened up on Vincent road intersection with Western Avenue
55	It's the 27th of February and there is a nasty accident on N2 between Abbotsford and Cambridge, the vehicle rolled the emergency services on the scene, all the roads are wet please take care out there.
56	There seems to be a pothole across from Exclusive books...I mean Exclusive Foods on Beach Road I have seen some cars....
57	27 February house breaking in Beamish Crescent, Amalinda yesterday afternoon. It's the second house within two consecutive days. Stay alert and take note of suspicious persons and vehicles
58	No the 23rd of February there was an accident on Parkside bridge...This caused for the road to be closed off and no access from Black road to Parkside to town and from Parkside to town.
59	....My...Crowdsourcing...I mean my report...is that in Duncan Village...people steal...copper off the poles and electricity and then they burn them and then the sell them to a shop that sells copper....yes that is my thing...it happens randomly there is no specific date. Last week I saw some stealing and then ...yes thank you.
60	There is a bad pothole on the intersection of Meadow Rd at the police sports ground and Western Ave on the 13th of March.
61	This is a warning to all Greenfield residents; Hartsfield Ave there is a black male, with a stripped shirt caught in a yard. He is now jumping from yard to yard, the police and our guys are here now.
62	There is a big pothole on the left lane on Bird Street towards the industrial area...I saw this pothole on the 26th of February.
63	We received a report of a house breaking that occurred on the 4th of February in Kempton Rd in Nahoon a flat screen TV was stolen.
64	There is a bad pothole just before John's on Frere Rd, on the 21st of February.
65	Please be on the lookout for a red Citi Golf registration number ____. It was stolen last night between 6 and 8pm on the 10th of February in Ascot Rd just off Beach Rd Nahoon. If found please contact me

	on my BBM pin ____ or alternatively call _____. Thank you.
66	On the 6th of March, Monday evening, 5:30 in the afternoon, I was driving...to the beachfront the corner...the corner of Fleet and Currie Street. There is a huge pot hole there that might cause an accident. Thank you that's my public safety...yes
67	11th of March 2013, approximately 2 ARE in the morning. An attempted break in at... a house in Essex Rd in Vincent
68	Please look out for a Maroon BMW 330i ,registration number ____ ,its just high jacked on the 4th of March, from Greenfields
69	On the 10th of March, I driver through a pothole coming to work. I drive from Stirling to town, the University of Fort Hare and there are probably 4 or 5 potholes... To such an extent that I had to take my car for wheel alignment. So potholes and road safety is of big importance, and it's a public issue that I feel has to be addressed thank you
70	Yesterday morning coming home from work, fetching my daughter from school. On the corner of Fleet and Cambridge Street. There is a severe bad case of a hump in the road, which can cause a severe accident and damages to pedestrians in the future. The tar in the road is not even my husband had to severe, could have possible hit another car next to him and cause a server accident
71	Good day, there is a big pothole in the Southside of the University of Fort Hare, in Cambridge street .It has been there for quite a few weeks. In fact even months and every day I drive through it, thank you
72	The incident happened 3 weeks ago here at the University of Fort Hare. When one of the lecturers was going to teach the students. He had the data projector and the laptop when one of the tsotsies came straight to him wanting to take the laptop and the data projector; I guess the university has to fasten their security...
73	The incident happened here at Fort Hare parking when one of the student's cars was stolen. It really made us to be really always be in fright of what is going to happen. And if those criminals were able to steal a car like that and get away with it. Maybe the university can do something and...
74	I would like to report a Housebreaking that occurred on the 22nd of February at Moore Street, Quigney. Its same house they tried on Sunday evening. The suspect is wearing a white shirt and bleeding after cutting himself on the broken window, we are looking for him.
75	On the 13th of February at 6 o'clock a car rolled near the Hemmingway's N2.
76	Hi its ____ ____, this morning at 7:30am the day is the 14th of March. I was traveling from my home in Cambridge west,...and we have a traffic settle it's quite a complicated traffic circle along... Queens street into Kings way. And I was waiting for my turn... well I thought that the driver was going straight and as I was going into the circle when I was already in the circle. He decided to turn in my way, and I mean he didn't even indicate. I just want people to know how to properly use a traffic circle and they must indicate in advance because he could have knocked me...and this could...have caused a huge accident there. Thank you
77	It's the 5th of February round about 3 PM. This is a warning to Berea Residents look out for a white hatchback Ford Laser. It just broken into a house in Greenan Street and when confronted by owner they sped away. The owner was unable to get the registration Number .Its 2 black male suspects, one had earrings in both ears.
78	It's the 4th of February, right now as I am calling ,there are illegal electricity connections are being made at Orange Grove, and that's right next door to a creche .Just on Saturday the Daily Dispatch reported on a 4 year old boy that was electrocuted because of this .
79	Hi my name is ____ ____, and a pedestrian has been knocked down on the N2 below Hemmingways, on the 13th of February at …9:40
80	On the 3rd of February, a Silver/Grey Toyota Avanza ____, broke into a house in Iddleway Crescent, Dorchester Heights while owners home, front door kicked open and property stolen, 3 black male occupants.
81	On the 5th on March police said on Sunday a woman got arrested for making a bomb threat at the East London airport. I came across this report when reading the times live website.
82	We received a report a vehicle, a Renault Megane; ____ was abandoned in bushes on corners of Kenny Road and Grogan Road, Amalinda. If it is yours or you know the owner , now you know where it is.
83	Please be alert around the 20th of February in WILLOW PARK, COLLONDALE AREA. We got a message that there is a black male, wearing black clothes going into yards where people are not home.
84	On the 28th of February a White Bantam bakkie with canopy was stolen outside the business the registration number is ____. The name of the business is tech Engineering on Penkop road, Splash Park, Woodbrook
85	On the 1st of March, at the Riverside complex, which is at the bottom of Batting road. The Electric gate taken off runner and broke into they stole the car radio and battery my bakkie. This happened

	before 3 this morning. It's the 2nd time in 2 weeks that we come home and the gate is off the runner.
86	Hi my name is ____ ____ on the 4th of March, at 10:30 my car boot was broken into in Elizabeth road in Cambridge and they stole my lap top bag containing my laptop and camera .If you know of anyone who is trying to sell off a white Samsung laptop or a black canon camera, Please contact me.
87	We had a housebreak in in February 2013…clothes, property and wallets were stolen and the police took 14 hours to respond.
88	On the 9th of February Amalinda, Stoneydrift. There is a young boy, approximately 10 to 12 years old, riding a quad bike at high speed, with no helmet. This is a daily occurrence. At least 3 times he has pulled out of a side street in front of me. I tried to follow him earlier, but he rode away at an incredible speed and disappeared into Beamish Crescent, with no trace... I Hope there is a traffic officer who will see this.
89	On the 26th of February, I noticed new robots being put up close to Christian Centre , Abbotsford are almost complete. This raises a question.... The robots are there but not operational yet, we noticed some people treat it as a 4 way stop while others ignore it. What is the correct procedure to follow here? It looks like a potential accident hotspot at the moment. Should the authorities... not at least put up temporary 4 way stop signs until the robots are operational?
90	Hi CSI...yesterday at about 12:30 a grey 4x4 with four males stopped outside our home in Charles Avenue Greenfields.They were looking into the yard for a panel beater and pointing into the yard. When the maid asked what they wanted. They said they looking for a panel beater she pointed out that there was one down the road but they drove straight past it.
91	This issue occurred on the second of March please be AWARE... there is a scam currently targeting pharmacies' involving airtime... the con artist impersonates a doctor local to the pharmacy, he thoroughly researches and so answers any queries you may have. The impersonator takes thousands of Rands of airtime on his account and promises to call in and pay later, he never does and when you contact your local doctor he will think you have lost your mind. We have just been caught out as suckers by a guy who sounded like an Indian or Coloured impersonating a Dr Anderson from the number ____. Please be aware all retail companies, he could change his strategy.
92	22 February this is a warning to all men. If you are stopping anywhere in the Quigney area, specifically at night, please be aware of a new trend that has come to light. It seems that the ladies of the night have resorted to another way to supplement their income!!! At shops , Atm's , traffic lights or wherever you pull over to take a phone call or whatever reason , the ladies would just jump into your vehicle and demand payment and refuse to get out until they have been paid , if they don't get what they want they steal cellphones , wallets or other valuables from the vehicle. If this had to happen to you we advise that you proceed to your nearest police station, in this case Fleet Street, or flag down any passing police vehicle. We have had similar reports in the past, so this is not just another way for a guilty spouse to explain himself, if you were a victim of a similar incident in this area, please inbox us.
93	As if the service isn't poor enough. I have been watching my network status on blackberry for, hours now there has been no internet connection or limited connection. We are offline on the blackberry internet service most of the time and the times we are connected it is very slow. We are paying for a service that is not provided. Maybe someone from Vodacom can explain and rectify the poor service.
94	Today the 17th of March at 3pm white Isuzu just got stolen from Nahoon beach, registration number ____. If seen please contact the police or the owner on ____
95	I would like to report that On Friday the 15 of March at approximately 6 O'clock in the evening, I was harassed by. ....some teenage boys posing as car guards outside the Fleet street spar and I would also like to report that this is a re-occurring incident. And it is...and it is...
96	My sister's car was broken into again, in the Stirling Primary school grounds. Stole a black Dee M model agency bag. With all her belongings inside. I have given up hope.....
97	Red Toyota Corolla, ____, stolen 5 minutes ago from Spargs, Beacon Bay, any info please contact ____ on ____
98	Elderly gentleman was mugged in Oxford Street earlier. If anybody finds a wallet containing drivers license , I.D.Book or cards in the name of ____ ____, please contact ____
99	Warning: Darlington road Berea residents has been hit by thieves breaking in between 13 & 14h30, front door bashed down & flatscreen TV's stolen. Look out for either white or red Toyota Tazz/Mazda 323.
100	A business in Argyle Street in the CBD Area was broken into at approximately 04:00 this morning.