Review Article

Using Participatory Crowdsourcing in South Africa to Create a Safer Living Environment

Bhaveer Bhana, Stephen Flowerday, and Aharon Satt

1 Department of Information Systems, University of Fort Hare, Private Bag X9083, East London 5200, South Africa
2 IBM R&D Labs in Israel, University of Haifa Campus, Mount Carmel, Haifa 31905, Israel

Correspondence should be addressed to Stephen Flowerday; sflowerday@ufh.ac.za

Received 5 July 2013; Revised 21 September 2013; Accepted 22 September 2013

Academic Editor: Yu Gu

Copyright © 2013 Bhaveer Bhana et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increase in urbanisation is making the management of city resources a difficult task. Data collected through observations (utilising humans as sensors) of the city surroundings can be used to improve decision making in terms of managing these resources. However, the data collected must be of a certain quality in order to ensure that effective and efficient decisions are made. This study is focused on the improvement of emergency and nonemergency services (city resources) through the use of participatory crowdsourcing (humans as sensors) as a data collection method (collect public safety data), utilising voice technology in the form of an interactive voice response (IVR) system. This study proposes public safety data quality criteria which were developed to assess and identify the problems affecting data quality. This study is guided by design science methodology and applies three driving theories: the data information knowledge action result (DIKAR) model, the characteristics of a smart city, and a credible data quality framework. Four critical success factors were developed to ensure that high quality public safety data is collected through participatory crowdsourcing utilising voice technologies.

1. Introduction

Local government accepts responsibility for maintaining the city's infrastructure, as well as for providing a safe living environment [1]. Currently, cities around the world are facing challenges in their attempts to keep up with the rate of urbanisation. Accordingly, city resources are unable to support all members of the public simply because the ratio of number of citizens to city resources (e.g., water and electricity distribution or public safety emergency and nonemergency units) is high. This is not because city resources are scarce, but because they are not managed effectively and efficiently [2]. Therefore, improving the management of city resources would assist cities in adapting to increased urbanisation.

In order to explain how one would improve the management of city resources, the data information knowledge action result (DIKAR) model can be used. This model, illustrated in Figure 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 if the management of city resources is to be improved. However, the data collected must be of sufficient quality to facilitate effective decision making regarding the management of city resources.

This paper is focused on the collection of high quality public safety data in order to facilitate decision making relating to the management of emergency and nonemergency units, ultimately leading to the creation of a safer living environment. Because of citizens’ knowledge of their surroundings, crowdsourcing has been found to be the most appropriate method for collecting public safety data. “Crowdsourcing” is a term that refers to the collection of large volumes of data or reports on certain events, by making use of the geographical dispersion of people [4]. Data can thus be collected and/or reported on using applications that make use of images collected from camera phones; by voice, through a phone
This paper will discuss how crowdsourcing (humans as sensors) can be used as a smart city initiative to reduce or mitigate the problems associated with urbanisation, specifically public safety. The paper will centre on efforts to ensure that high-quality public safety data is collected. Before discussing this, the relationship between the smart city and crowdsourcing was placed in context. As part of a smart city initiative, crowdsourcing is used to reduce or mitigate the challenges associated with urbanisation. The paper will discuss a specific public safety crowdsourcing project, which assisted in the identification of critical success factors for ensuring the collection of high-quality public safety data, using participatory crowdsourcing utilising voice technologies. In the following section, the research methodology used for this study will be discussed.

2. Materials and Methods

A mixed method approach was undertaken in this study guided by design science methodology. Design science guidelines were 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" [6, page 76]. Design science consists of seven guidelines, all of which must be considered [7]. Figure 2 graphically depicts the way in which the seven guidelines were followed throughout this study. Note that the design science guidelines presented by Hevner et al. [7] have been rearranged for the purpose of this study.

**Guideline 1: Problem Relevance.** The problem identified in this study refers to the presence of low-quality data, which would result in ineffective decisions being made on the management of city resources. This study will focus specifically on the data quality of public safety reports provided by citizens through participatory crowdsourcing.

**Guideline 2: Research Rigour.** Secondary data was used to create and support all logical conclusions during the development of the critical success factors. Theories were used to guide the research and validate any assumptions. Subsequently, two expert reviews were conducted for the purpose of assessing the critical success factors.

**Guideline 3: Design as an Artefact.** This study was intended 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 proposed critical success factors were evaluated by applying an expert review. Accordingly, the proposed artefact was developed after considering the content of the research, and the material was then presented to seven experts, with sufficient and relevant educational background and experience, for evaluation. After considering the feedback from the expert review process, the artefact was appropriately modified.

**Guideline 5: Design as a Search Process.** The proposed artefact was constructed from related primary and secondary data. The comments and recommendations from the expert groups allowed the proposed artefact to be refined by removing ambiguity and ensuring completeness.

**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 such 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, the way in
which high quality data can be achieved through the use of technology will be illustrated.

In the next section, the details and findings of this study will be discussed.

3. Discussion and Findings

The discussion and findings section will firstly define “smart city” and “crowdsourcing” before illustrating the relationship between them. The next section will discuss the crowdsourcing safety Initiative (CSI) participatory project and its data collection method before the public safety data quality criteria are presented. The presentation of the critical success factors will follow, and they will then be linked to the identified public safety data quality problems identified. The last section under the heading “Discussion and Findings” presents the relationship between the crowdsourcing areas and the critical success factors.

3.1. Conceptualisation of a “Smart City”. Nam and Pardo [8] explain that the most important consideration when defining a smart city is to view it as one “large organic system.” When conceptualising the smart city it is important to consider the stakeholders involved, smart technology, goals, and influential areas of the smart city. The “Conceptualisation of a smart city” section in Figure 3 incorporates all these components and can be used to explain the smart city concept. Note that the six boxes (smart economy, smart living, smart people, smart mobility, smart environment, and smart governance) refer to the six characteristics of a smart city, constructed by Gifinger et al. [9]. The characteristics were constructed on the basis of theories relating to traditional neoclassical 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 et al. [10], Lombardi et al. [11], and Desourdis [12]. The grey boxes in the “Conceptualisation of a smart city” section in Figure 3 indicate the areas relevant to public safety, which is the focus of this paper (public safety data quality).

Smart technologies are used to collect data on, or directly influence, certain areas of the city (economy, living, people, mobility, environment, and governance) [13]. 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, such technologies can act as sensors to report traffic accidents, which will assist the traffic department in creating safer roads and managing traffic flow. These initiatives contribute to making the city smarter and more knowledge driven. Participatory crowdsourcing is
Conceptualisation of a Smart City

Knowledge-intensive city

Smart City

Behavioural change/decision making

Citizens, businesses, and government

Data analysis

Smart technology

Data collection

Smart economy
Smart living
Smart people
Smart mobility
Smart governance
Smart environment

Components of crowdsourcing

Type of data

Data collection method (e.g., social networks, mobile phones)

Type of crowdsourcing

Participatory
Crowds as sensor operators

Opportunistic
Crowds as targets of sensing

Crowds as data sources

Crowdsourcing stakeholders

Evaluators (programmers)
Provide technical support
Set the campaign requirements and the system's required capabilities

Initiators (organisers)
Obtain the campaign requirements and useful information
Provide feedback and analyse

Gatherers (participants)
Submit the gathered data
Transfer the gathered data

Analysts (researchers)

Figure 3: Smart city and crowdsourcing relationship.
3.2. Crowdsourcing Defined. Various authors use the terms “crowdsourcing” and “crowdsensing” interchangeably. This has been accepted by many publishers, but to avoid any confusion, this paper will use the term “crowdsourcing.” Crowdsourcing refers to a group of individuals who collect data or report on certain events of a similar nature and pool all the data collected [4]. This is the underlying concept of crowdsourcing. Accordingly, tasks can be solved or data generated by taking advantage of the wide geographical dispersion of citizens [14]. The dynamic mobility of citizens and their observations and knowledge of their surroundings make them the perfect candidates for collecting data on activities occurring in their environment. In this manner, large volumes of data can be collected in a short period of time, at a low cost. This data can, in turn, be used to make decisions [3], identify patterns [15], solve problems [14], and even influence behavioural change [5].

Before a crowdsourcing initiative can be implemented, one must determine exactly what data will be collected from crowds. For example, data on weather conditions will entail temperature, wind speed, and a geographical location. The data that is expected to be collected will influence the method in which it is collected. Methods for collecting data can range from social networks to sophisticated sensor devices. In all such initiatives, a common decision that must be made, regardless of the method chosen, is the type of crowdsourcing initiative to be undertaken.

3.3. Types of Crowdsourcing. Crowdsourcing initiatives can be divided into three categories based on the role and extent of participation [16]. Srivastava et al. [16] strongly emphasize 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. The three types of crowdsourcing are tabulated in Table 1. When the crowdsourcing initiative has been clearly defined, the stakeholders involved can be selected.

3.4. Crowdsourcing Stakeholders. A typical crowdsourcing project has four stakeholder groups: (1) evaluators; (2) initiators; (3) gatherers; and (4) analysts [15]. The “Crowdsourcing Stakeholders” section in Figure 3 graphically illustrates the four stakeholder groups and their general roles and responsibilities in a crowdsourcing project. 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. The next section will discuss how all the considerations of crowdsourcing are combined to create an effective crowdsourcing initiative.

3.5. Components of Crowdsourcing. The discussion above shows that there are four components of a typical crowdsourcing initiative. These four components include (1) the type of data collected; (2) the method used to collect the data; (3) the type of crowdsourcing initiative; and (4) the stakeholder groups involved in the crowdsourcing initiative. The “Components of crowdsourcing” section in Figure 3 incorporates all four components to illustrate what needs to be considered for a typical crowdsourcing initiative to facilitate its data collection purpose successfully. Note that the bottom box (labelled “Crowdsourcing Stakeholders”) was constructed by Yang et al. [15].

Before one can decide on the most appropriate and feasible crowdsourcing option and the stakeholders required, one must first understand the type of data that needs to be collected. For example, traffic accidents will require a description of the incident, time, and geographical location. One must then determine the best data collection method which is feasible and will allow high quality data to be collected. The grey boxes in Figure 3 (in the “Components of crowdsourcing” section) indicate the decision adopted by the CSI participatory project. Before discussing the CSI participatory project in more detail, the link between smart city and crowdsourcing will be provided.
The next step was to choose the data collection method to be used, for example, social networks, mobile phones, and blogs. The CSI participatory project utilised mobile and landline phones for the collection of voice data. This was deemed the most appropriate data collection method based on the countries technological infrastructure (access to internet and network connection speeds), access to computers and smartphones, 11 official languages, low literacy rates, and low computer literacy rates. 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 were directed by message prompts (computer interaction).

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

The CSI participatory project made use of an interactive voice response (IVR) system for an audio user interface (in the form of message prompts). IVR systems comprise an interactive telephonic interface [20] in terms of which pre-coded messages are provided to the user, who in turn supplies audio input to the system [21]. The IVR system serves as a substitute for the World Wide Web by sending and receiving data to people through their telephones [22]. The CSI project took advantage of the IVR system's voice recognition functionality when developing the message prompts to direct the caller to respond effectively.

The message prompts are intended to instruct the caller to provide the correct information so that the public safety data collected can be used appropriately. After a number of discussions on and iterations of potential message prompts, a free-flowing data reporting method (see Figure 5) was agreed on. The audio interface is vital for data collection as this directed the user to provide the required information on public safety issues. Greeff et al. [20] found that at times users have difficulty understanding the message prompts. Accordingly, the presentation of message prompts to the user has to be clear, ensuring that no ambiguity is present.

3.6. Relationship between Smart City and Crowdsourcing. It is important to link smart city and crowdsourcing to emphasise the significance of this study and the contribution it makes to the body of knowledge. This will also add to an understanding of the way the critical success factors were decided upon and why they were deemed important (critical) to the research area of this study. Figure 3 illustrates the relationship between smart city and crowdsourcing, but more specifically how participatory crowdsourcing can contribute to the smart city.

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 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 details on the CSI participatory project.

3.7. CSI Participatory Project. The University of Fort Hare (located in East London, South Africa) and IBM have pooled their resources in order to run a public safety crowdsourcing pilot study in East London, South Africa. East London is a small district situated in Buffalo City, which is located in the developing country of South Africa. Figure 4 indicates the location of East London with the letter “A.”

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 project, public safety data will be gathered. Therefore, the initiative will contribute towards the smart living area of a smart city. Because of the vast array of public safety issues, general data common to all safety incidents should be collected so as to avoid data overload. At the same time, the data should be sufficient 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 the type of incident, the date and time of the incident, and the geographical location of the incident.
Wang and Strong’s [26] Data Quality Framework was used to develop criteria for public safety data quality for the CSI participatory project.

Figure 3 illustrates that there are four types of stakeholder groups in a typical crowdsourcing initiative. As mentioned above, an entity can perform the role of more than one stakeholder group, which is the case in many crowdsourcing initiatives. The CSI participatory project also follows this stakeholder map, with the University of Fort Hare acting as the initiator, while both the University of Fort Hare and IBM play the joint role of analysts and evaluators (strictly in terms of setting up the IVR system); the participants comprise the citizens of East London.

3.8. Data Quality Framework. Wang and Strong’s [26] 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 include only those deemed important to a data consumer [26]. A data consumer is described as a person or organisation that accesses or uses the data [26]; therefore, the data consumer within the CSI participatory project would be the entity responsible for acting on the data collected (in this case, emergency and nonemergency services). The Data Quality Framework is used to construct criteria for public safety data quality and to assess the presence of the data quality attributes in the data collected from the citizens through the CSI project.

The Data Quality Framework was used to construct “yes or no” data quality assessment questions. 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. 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. Note that no weights will be allocated to questions used in the public safety data quality criteria as all questions are considered equally important.

Table 2 illustrates the questions that were constructed and indicates the data quality attributes (and the data quality dimension to which it belongs based on Wang and Strong’s [26] Data Quality Framework) that were considered when constructing each individual question. This collection of questions is referred to as the public safety data quality criteria. These quality criteria were used to assess 100 public safety reports collected from the CSI project. A benchmark of 94 or fewer was found to be appropriate in assuming quality problems in the public safety reports. Therefore, any question that scored lower than 95 is seen as a data quality issue or a potential data quality issue. These questions that have scored below 95 are highlighted in bold in Table 2.

Following the assessment of the public safety reports, nine problems were found that affect the quality of the data. These problems were grouped and are presented in Table 5 together with the question number/s from Table 2 in brackets to indicate the question/s that identified the problems. These problems were used as a guide when compiling the critical success factors.
Table 2: Public safety data quality criteria.

| DQ dimension | DQ attribute | Question description | Total |
|--------------|--------------|----------------------|-------|
| Contextual   | Value-added  | Is it possible to determine the response team responsible for responding to the report? | 99    |
| Intrinsic    | Accuracy     |                       |       |
| Contextual   | Value-added  | Can one determine the number of response teams required to resolve the issue? | 99    |
| Intrinsic    | Accuracy     |                       |       |
| Contextual   | Value-added  | Can the type of incident be determined? | 100   |
| Intrinsic    | Believability|                       |       |
| Intrinsic    | Accuracy     |                       |       |
| Representational | Consistency |                       |       |
| Representational | Concise    |                       |       |
| Contextual   | Value-added  | Can the date the incident occurred/was observed be determined? | 76    |
| Intrinsic    | Completeness |                       |       |
| Intrinsic    | Believability|                       |       |
| Representational | Consistency |                       |       |
| Contextual   | Value-added  | Can the time the incident occurred/was observed be determined? | 37    |
| Intrinsic    | Believability|                       |       |
| Representational | Consistency |                       |       |
| Contextual   | Value-added  | Does the report mention the area (suburb or highway) in which the incident occurred/was observed? | 71    |
| Contextual   | Completeness |                       |       |
| Intrinsic    | Believability|                       |       |
| Intrinsic    | Accuracy     |                       |       |
| Representational | Consistency |                       |       |
| Representational | Concise    |                       |       |
| Contextual   | Value-added  | Does the report mention the street/landmark where the incident occurred/was observed? | 86    |
| Contextual   | Completeness |                       |       |
| Intrinsic    | Believability|                       |       |
| Intrinsic    | Accuracy     |                       |       |
| Representational | Understandability | |       |
| Representational | Consistency |                       |       |
| Representational | Concise    |                       |       |
| Contextual   | Relevancy    | Does the reported incident show any sign of endangering citizens or does it have the potential to endanger citizens? | 99    |
| Contextual   | Timeliness   | 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    |
| Contextual   | Timeliness   | 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? | 95    |
| Contextual   | Timeliness   | Are all the audio files available for transcription when required? | 96    |
### Table 2: Continued.

| DQ dimension | DQ attribute | Question description | Total |
|--------------|--------------|----------------------|-------|
| Contextual   | Completeness | Is there an audio version and textual version of the report? | 100   |
| Contextual   | Completeness | Is it possible to understand the extent of the public safety incident? | 100   |
| Contextual   | Accuracy     | Can the specific time period (a.m. or p.m.) be determined? | 28    |
| Intrinsic    | Believeability | 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 (e.g., heavy rain caused floods)? | 100   |
| Intrinsic    | Objectivity  | Is the caller’s voice consistently calm and soft throughout the report, with the absence of background noise? | 94    |
| Representational | Concise | Does all the additional data provided influence the dispatch decision? | 96    |
| Representational | Understandability | Are system glitches absent? (e.g., lagging) | 96    |
| Representational | Understandability | Is background noise absent? | 97    |
| Representational | Understandability | Is radio interference absent? | 99    |
| Representational | Understandability | Is wind created noise absent? (including caller breathing distortion) | 86    |
| Representational | Understandability | Does the caller have a clear accent? | 89    |
| Representational | Understandability | Is the use of slang absent? | 99    |
| Representational | Understandability | Can the type of public safety issue be interpreted in a single point of view? | 96    |
| Representational | Understandability | Is the caller’s use of acronyms appropriate to comprehend the report? | 93    |
| Representational | Understandability | Are speech impediments absent? | 94    |
| Representational | Understandability | Is the caller’s vocabulary and use of grammar appropriate to comprehend the report? | 100   |
| Accessibility | Accessibility | Are the naming conventions of files and folders logical? | 100   |
| Accessibility | Accessibility | Can all authorised files be accessed? | 100   |
| Accessibility | Accessibility | Can all authorised folders be accessed? | 100   |
| Accessibility | Security | Has all the personal data been removed from the report? | 84    |
| Accessibility | Security | Is the report still readable after the removal of all personal data? | 100   |

### 3.9. Critical Success Factors

Based on Wang and Strong’s [26] Data Quality Framework, four critical success factors were proposed. These were compiled from both the secondary and the primary data and are presented in Table 3. After the feedback from the first expert review, the names of the critical success factors were made more descriptive, and a fifth critical success factor was added. After the second expert review, the critical success factors were repositioned as design principles [6], and the five refined critical success factors were reduced back to four. The finalised critical success factors indicate what should be in place and considered to ensure quality data is collected when implementing a crowdsourcing initiative. The final critical success factors are presented in Table 3 (the finalised critical success factors) to illustrate changes between the proposed critical success factors and the final (refined) critical success factors.

A detailed explanation of each refined critical success factor is provided in Table 4.

### 3.10. Critical Success Factors and Problem Areas

The critical success factors discussed above are the most important (critical) areas that need to be considered to ensure success in their respective environments. 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 the critical success factors. Table 5 emphasises the importance of all the critical success factors by illustrating the problems that they address.
The data quality problems were identified by assessing 100 public safety reports. This is also helpful for understanding which critical success factors require more attention than others.

Table 5 illustrates the areas (critical success factors) that need to be addressed to ensure that the problems affecting public safety data quality can be solved. This shows that if these factors are considered, nine problems can be solved and, subsequently, the quality of public safety data in participatory crowdsourcing through audio data collection can be increased. This will make this method of participatory crowdsourcing suitable for use as a smart city initiative and for reducing the public safety problem. Although Table 5 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.

### 3.11. Critical Success Factors and the Theoretical Background.

Since the construction of the critical success factors was supported by the Wang and Strong's [26] Data Quality Framework, it is valuable to illustrate the relationship between a typical crowdsourcing project and the supporting factors of the Data Quality Framework. This is presented in Figure 6.

Through conversational analysis on the CSI participatory project, it was found that the specific data that had to be collected included the type of incident and the date, time, and location (street/landmark and suburb/highway) of the incident. Based on the city's (East London) limitations, such as technological accessibility and literacy rate, voice was chosen as the data collection method using 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 project is a joint effort by Fort Hare University and IBM, these two entities undertook the responsibility for multiple stakeholders (gatherers—East London citizens, initiators—Fort Hare University, analysts—Fort Hare University and IBM, and evaluators—Fort Hare University and IBM).

### 4. Conclusion

The CSI participatory project aims at providing an environment in which the public may 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 (participant successfully reporting quality public safety data), through participatory crowdsourcing (using humans as sensors), will result in an increased ability to resolve the issues raised by the public [27]. This will enable a significant contribution to be made to the smart city.

The study firstly established the components of a smart city and crowdsourcing and explained how crowdsourcing can be used as a smart city initiative (see Figure 3). This created context for the data quality problem experienced when using crowdsourcing. Subsequently, public safety data quality criteria were developed to assess the public safety reports collected from the CSI project. Following the data quality assessment of 100 public safety reports, it was found that there were nine common problems affecting the quality of data. These problems can be reduced or mitigated by considering the four critical success factors. These factors will also ensure that any other problems that could affect data quality are reduced or eliminated.

The critical success factors indicated what areas need to be considered to ensure quality data is collected. In terms of future research, it would be interesting to examine the implementation of participatory crowdsourcing in order to determine how these areas can be optimally addressed to ensure that high quality public safety data is collected. These solutions can then be applied in practice to test the effectiveness and efficiency of the data collected. In addition, all participatory crowdsourcing initiatives rely on participation to be successful. This study assumes (supported by the Rational Choice Theory [18]) that if the participants’ reporting is effortless (simplification of the crowdsourcing process), participants will be more likely to participate. In addition, if participants perceive a crowdsourcing project as an effective initiative and anonymity is ensured, they will be more likely to cooperate and even take the extra effort to provide high quality data. Measuring the extent of these conclusions will be interesting research for future studies.

### Acknowledgments

This work is based on the research supported in part by the National Research Foundation (NRF) of South Africa, the International Business Machine Corporation (IBM), the Govan Mbeki Research and Development Centre (GRMDGC), the University of Fort Hare, and the citizens of East London. The authors acknowledge that the opinions, findings, and
Table 4: Refined critical success factors.

| Code | Name                          | Description                                                                                                                                                                                                 |
|------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CSF1 | 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. |
| CSF2 | Data collection method        | There is no such thing as the "best" data collection method. Like most methods, each data collection method has its advantages and disadvantages. The most appropriate method will be influenced by numerous factors, since technology usually supports the collection of data with regards to the crowdsourcing environment, the technological infrastructure, internet speed, user experience, accessibility, and costs have the potential to limit the collection of certain data (e.g., an advanced smartphone is required to record noise or air pollution). |
| CSF3 | Type of crowdsourcing         | The type of crowdsourcing method chosen depends on numerous factors such as the type of data required and the level of detail. In addition to the data, the most appropriate crowdsourcing method chosen will be influenced by the level of involvement of the data source. The three types of crowdsourcing mentioned above have a clearly defined level of user involvement within the process of collecting data through crowds. The less user involvement will result in a higher likelihood of collecting consistent and accurate data. However, some data collection requires a high level of human involvement. In this case, one must ensure clear and unambiguous instruction be provided to ensure quality data is collected. |
| CSF4 | Crowdsourcing stakeholders    | Data source: 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 overemphasised based on the data source's perception. Clear and unambiguous instructions must be given 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 project, the instructions are given to the data sources on the CSI website (http://csi.ufh.ac.za/) and through message prompts.     |

**Analysis:** For the purpose of analysing the data collected, it 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 done by giving the data source language options to provide the data and stating that the use of acronyms and slang is forbidden. In terms of presentation, the data collected must be in the same format or be capable of being converted to the same format. This will allow one to compare and contract data and identify patterns and anomalies with little effort. |

**Evaluators:** As mentioned above, evaluators are responsible for developing and implementing the system and any other technology for the crowdsourcing initiative. 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 that the user who modified the data can be identified. In addition, ethical issues can be avoided if all personal data is removed before use. It is also suggested that periodic backups of audio and textual data be conducted. |

**Initiators:** Initiators are responsible for the planning, setup, and other managerial aspects of the crowdsourcing initiative. 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 result in callers skipping the instructions, or they will 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 good quality public safety data. System availability should be tested regularly 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 straightforward as possible, as long and complex processes increase the risk of the quality of the public safety data being negatively affected. This critical success factor can be supported by the Rational Choice Theory used by a number of authors ([17–19]) to explain that people weigh up 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 to use, the benefits of reporting will outweigh the costs. Although public safety has a larger scope than crime, the concept behind the theory is relevant to public safety reporting. Taylora [18] adds that the simplicity of the system will encourage citizens to provide a high quality report.
Table 5: Critical success factors and problem areas.

| Description                                                                 | Critical success factors |
|----------------------------------------------------------------------------|--------------------------|
| Inconsistent collection of public safety incident date and time data (4, 5, and 14). | X  X  X  X               |
| Inconsistent collection of public safety incident location data (6 and 7).   | X  X  X  X               |
| Reports which required an immediate response were reported over a month or few weeks later (9). | X  X  X  X               |
| Caller's frustration overstated the nature of the issue (17).               | X  X  X  X               |
| Heavy breathing into the mouthpiece of the phone (22).                      | X  X  X  X               |
| Accent made it difficult to comprehend the intended message in terms of the pronunciation of certain words (23). | X  X  X  X               |
| Caller's use of acronyms made it difficult to comprehend the public safety report (26). | X  X  X  X               |
| Speech impediments resulting from language barriers in describing the event (27). | X  X  X  X               |
| Data source found it difficult to follow data collection instructions, or the instructions were not clear (32). | X  X  X  X               |

Total problems: 3 3 7 9

Figure 6: Critical success factors and the Wang and Strong's [26] Data Quality Framework
conclusions expressed are those of the authors and that the NRF, IBM, GMRDC, and UFH accept no liability whatsoever in this regard.

References

[1] L. Fuzile, Local Government Budgets and Expenditure, National Treasury, Pretoria, South Africa, 2011.
[2] J. P. Exner, P. Zeile, and B. Streich, “Urban Monitoring Laboratory: new benefits and potential for urban planning through the use of urban sensing, Geo- and Mobile-Web,” in Real Corp, pp. 1087–1096, 2011.
[3] J. Ward and J. Peppard, Strategic Planning for Information Systems, John Wiley & Sons, England, UK, 2002.
[4] H. Väätäjä, T. Vainio, E. Sirkkunen, and K. Salo, “Crowdsourced news reporting: supporting news content creation with mobile phones,” in Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI ’11), pp. 435–444, ACM, New York, NY, USA, September 2011.
[5] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, “A survey on privacy in mobile participatory sensing applications,” Journal of Systems and Software, vol. 84, no. 11, pp. 1928–1946, 2011.
[6] S. Gregor and A. R. Hevner, “Positioning and presenting design science research for maximum impact,” MIS Quarterly, vol. 37, no. 2, pp. 337–355, 2013.
[7] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.
[8] T. Nam and T. A. Pardo, “Conceptualizing smart city with dimensions of technology, people, and institutions,” in Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp. 282–291, ACM, New York, NY, USA, June 2011.
[9] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. P. Milanović, and E. Meijers, Smart Cities: Ranking of European Medium-Sized Cities, Centre of Regional Science, Vienna, Austria, 2007.
[10] A. Caragliu, C. del Bo, and P. Nijkamp, “Smart cities in Europe,” Journal of Urban Technology, vol. 18, no. 2, pp. 65–82, 2011.
[11] P. Lombardi, S. Giordano, H. Farouh, and W. Yousef, “An analytic network model for smart cities,” in Proceedings of the International Symposium on Analytic Hierarchy Process, pp. 1–6, AHP Academy, Sorrento, Italy, 2011.
[12] R. I. Desourdis, Achieving Smart City Resilience: A Sense-and-Respond Architecture For Emergency Preparedness and Disaster, Pacific Telecommunications Council, Curran Associates, Honolulu, Hawaii, USA, 2010.
[13] A. H. Bartels, Smart Computing Drives the New Era of IT Growth, Forrester Research, Cambridge, UK, 2009.
[14] F. Ali, A. S. Shirazi, A. Schmidt, U. Kramer, and Z. Nawaz, “Location-based crowdsourcing: extending crowdsourcing to the real world,” in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 13–22, ACM, New York, NY, USA, October 2010.
[15] H. Yang, J. Zhang, and P. Roe, “Using reputation management in participatory sensing for data classification,” Procedia Computer Science, vol. 5, no. 1, pp. 190–197, 2011.
[16] M. Srivastava, T. Abdelzaher, and B. Szymanski, “Human-centric sensing,” Philosophical Transactions of the Royal Society A, vol. 370, no. 1958, pp. 176–197, 2012.
[17] J. Kääriäinen and R. Sirén, “Trust in the police, generalized trust and reporting crime,” European Journal of Criminology, vol. 8, no. 1, pp. 65–81, 2011.
[18] N. Taylor, “Under-reporting of crime against small businesses: attitudes toward police and reporting practices,” Policing and Society, vol. 13, no. 1, pp. 79–89, 2002.
[19] L. Zhang, S. F. Messner, and J. Liu, “An exploration of the determinants of reporting crime to the police in the city of Tianjin, China,” Criminology, vol. 45, no. 4, pp. 959–984, 2007.
[20] M. Greeff, L. Coetzee, and M. Pistorius, “Usability evaluation of the South African National Accessibility Portal interactive voice response system,” in Proceedings of the 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, ACM, New York, NY, USA, October 2008.
[21] J. Dawes and J. Rowley, “Enhancing the customer experience: contributions from information technology,” Management Decision, vol. 36, no. 5, pp. 350–357, 1998.
[22] S. K. Agarwal, A. Kumar, A. A. Nanavati, and N. Raijput, Spoken Web: A Parallel WWW in Developing Regions, Edited by: V. Kunj, IBM, 2011.
[23] S. Lobo, P. Doke, and S. Kimbahune, “GappaGoshiti—a social networking platform for information dissemination in the rural world,” in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI ’10), pp. 727–730, isl, October 2010.
[24] R. Farrell, N. Raijput, R. Das, C. Danis, and K. Dhanesha, “Social navigation for the spoken web,” in Proceedings of the 4th ACM Recommender Systems Conference (RecSys ’10), pp. 333–336, ACM, New York, NY, USA, September 2010.
[25] K. Abu-Hasaballah, A. James, and R. H. Aseltine Jr., “Lessons and pitfalls of interactive voice response in medical research,” Contemporary Clinical Trials, vol. 28, no. 5, pp. 593–602, 2007.
[26] R. Y. Wang and D. M. Strong, “Beyond accuracy: what data quality means to data consumers,” Journal of Management Information Systems, vol. 12, no. 4, pp. 5–33, 1996.
[27] S. E. Solomon and C. D. Uchida, 311: Building a 3-1-1 System for Police Non-Emergency Calls: A Case Study of the City of Austin Police Department, 21st Century Solutions, Austin, Tex, USA, 2003.
Submit your manuscripts at
http://www.hindawi.com