Addressing Big Data Challenges in Smart Cities: A Systematic Literature Review

Abstract

Purpose – Big data has become crucial for fulfilling the vision of smart cities, however, there are several challenges associated with its use. This study aims to provide a holistic view of challenges of using big data in smart cities and the solutions to deal with them by systematically selecting, reviewing, and synthesizing the literature in this area.

Design/methodology/approach – A systematic literature review was conducted by following a rigorous search protocol that identified a total of 310 peer-reviewed academic research papers. These papers were subsequently filtered and finally 38 relevant papers were selected.

Findings – A number of major challenges (Human Dynamics, Privacy, Security, Volume, Velocity, Variety, Veracity, and Value) associated with big data in smart city and the respective solutions to address them were identified. This study primarily contributes in providing the process of effectively managing big data in smart cities.

Practical implications – The study provides valuable information to the big data practitioners by illustrating the process of effective management of big data in smart cities. This study further provides the directions to the future researchers.

Originality/value – The study is original and is based on the existing literature and its interpretation.

Keywords: Smart Cities; Big Data; Challenges; Systematic Literature Review.

1. Introduction

The proportion of global population in cities is increasing with the current trend of movement of people from pre-urban to urban areas (Bohli *et al.*, 2015). By 2050, 70% of the world population is expected to reside in the cities (Lierow, 2014). The motivation for this mobility is attributed to the availability of better quality of life and opportunities in the cities (Bohli *et al.*, 2015). However, the increasing density of dwelling, utilities, and infrastructure in modern cities is leading to a number of challenges such as congested transport infrastructure, water and waste management, energy efficiency, climate change, and air pollution (Xiong *et al.*, 2014). Today, the Information and Communications Technologies (ICTs) have become increasingly pervasive in urban areas and are an essential basis for resilience and sustainability of smart cities. ICTs empower citizens in shaping their urban environment and to handle urban issues and co-create solutions (van der Graaf and Veeckman, 2014). Nevertheless, smart cities can be described as the cities utilizing ICTs to enhance the quality of life of its citizens and contributing to the sustainable development (Chourabi *et al.*, 2012).

A number of current and upcoming technologies such as Radio-frequency identification (RFID), Internet of things, and future Internet technologies contribute extensively in making the cities smarter (Su *et al.*, 2011; Roscia *et al.*, 2013; Dutton, 2014; Ballon and Schuurman, 2015). Increasing presence of these technologies are causing a large volume of data to be generated (Xiong *et al.*, 2014). Koh *et al.* (2015) articulated that approximately 2.5 quintillion bytes of data is generated every day and 90% of the data in the world has been created in last two years alone. If this large volume of data, also called big data, is managed and analyzed properly, it can create a major real impact on the functioning of smart cities (Aryal and Dutta, 2015). Therefore, the valuable information drawn from big data analytics is quickly transforming cities into the artificial ecosystem of interdependent, interconnected, and intelligent digital organisms (Xiong *et al.*, 2014; Dobre and Xhafa, 2014).

Big data has some important characteristics namely volume, velocity, variety, veracity, and value (Villanueva *et al.*, 2014). Volume indicates that the data is ever growing and expanding beyond terabytes. However, massive parallelism and data reduction techniques are still two of the main challenges with big data (Koh *et al.*, 2015). Velocity refers to the speed with which data must be produced and processed to meet the demand. Nevertheless, the need for quick and constant processing of big data streams poses a challenge to the smart city (Namiot and Sneps-Sneppe, 2012). Variety pertains to the fact that big data is often generated from a number of heterogeneous sources such as simulations, databases, sensors, and social media. However, discovering hidden patterns and creating knowledge base from city data poses a big challenge which is partially addressed by the big data paradigm (Koh *et al.*, 2015). Veracity emphasizes the significance of addressing and managing the data uncertainty, however big data often fails in unusual circumstances (Koh *et al.*, 2015). As high data quality is a crucial requirement for big data (Villanueva *et al.*, 2014), there should be a way to achieve unified data and avoid dirty data (Corradi *et al.*, 2015). Lastly, value refers to the necessity of technological innovation that enables people and organizations to integrate, analyze, and visualize different kinds of data at different temporal and spatial scales. However, sometimes the complexity of big data restricts it to be processed and analyzed using traditional ways (Villanueva *et al.*, 2014). As a result, though the big data has immense potential to improve the quality of urban life, its use in smart cities accompanies several challenges. Thus, an urban infrastructure capable of providing economical, unified, and simple access to a number of public services is still lacking (Zanella *et al.*, 2014).

Given that big data provides the opportunities to resolve the issues arising in smart cities, it is important to address the challenges related to it. Based on the aforementioned view, following research question is identified: What are the ways to address the big data challenges in smart cities? While there are quite a few studies that have provided the approaches to deal with such challenges, there is a scarcity of study which consolidates the existing studies and ground them into a framework that suggests the effective way to manage the big data in smart cities. Thus, this study is an attempt to provide the recommendations for dealing with big data challenges in smart cities as suggested by various researchers by conducting an in-depth systematic literature review. Most importantly, given the contemporary significance of the subject, this study also provides key insights to the practitioners and future researchers.

The rest of this research paper has been structured as follows: Section 2 provides the overview of the big data and smart cities. Section 3 discusses the systematic literature review process. Section 4 reports the results while section 5 discusses the findings of this study in detail. Section 6 presents the concluding remarks and potential areas for further research on this subject.

2. Smart City and Big Data

The term smart city has gained significant traction in academia and research. The smart city concept is related to many other concepts of digital city, knowledge city, information city, ubiquitous city, and intelligent city. All of these concepts are mainly focused on the application of ICTs in urban management (Lee and Lee, 2014). A number of smart city initiatives aim to utilize ICTs for enhancing the effectiveness, efficiency, transparency, and accountability of communications and transactions between government and citizens. As ICTs play a crucial role in designing smart cities, a substantial amount of city investment is used in the development of new technologies (Perboli *et al.*, 2014). As a result, smart city develops and manages a number of innovative services that help in delivering information related to all aspects of citizens' life via internet-based and interactive applications (Kuk and Jansson, 2011). There are six important characteristics of smart cities: smart economy, smart people, smart governance, smart mobility, smart environment, and smart living (Kumar, 2015; Chatterjee and Kar, 2015). Smart economy includes factors such as innovation, entrepreneurship, labor market flexibility, trademarks, productivity, and global integration. Smart people are described by their educational level, social and ethnic plurality, flexibility, creativity, and open-mindedness. Smart governance encompasses aspects of participation in decision-making, public and social services, political strategies, and transparent governance. Smart mobility includes local and global accessibility, ICT infrastructure, and innovative and sustainable transport systems. Attractive natural conditions, resource management, pollution, and environmental protection are the aspects of smart environment. Lastly, smart living includes the aspects of quality of life such as health, culture, housing, safety, tourism etc. (Giffinger *et al.*, 2007).

Smart city initiatives are mainly dependent on the collection and management of right kind of data, analysis of pattern, and optimization of system functioning. Two additional fundamental elements here are the large volume of data (i.e., big data) and the process of examining it (i.e., big data analytics). Big data is the large pool of unstructured data that can be captured, stored, analyzed, and managed. This large set of data is not useful until it is examined and assessed by the big data analytics which uncovers the unknown correlations, hidden patterns, and other valuable information (Tachizawa *et al.*, 2015).

New technologies provide the opportunities to gather and effectively use big data to enhance information awareness, facilitate prompt decision-making, and offer several opportunities for social interaction. Big data collected by sensors allows the automation of a number of real-time services which help in improving the urban management by using intelligent traffic-light patterns during peak hour, efficiently routing garbage collection trucks, reducing water consumption in parks, monitoring the use and condition of public infrastructure etc. (Domingo *et al.*, 2013; Kitchin, 2014). As depicted in table 1, several researchers have described big data projects that are helping in fulfilling the vision of smart cities.

Table 1: Big Data Projects for Smart Cities

Smart City Characteristic	Big Data Projects	Study
Smart Economy	Big Data techniques can be employed in manufacturing sector to enhance the intelligence and efficiency of design, production, and service process. This would help in reducing the development cycle, optimizing the assembly process, increasing the yields, and meeting the customer needs on time.	Li *et al.* (2015)
Smart people	The Japanese government has introduced a number of programs to utilize accumulated large scale data. The Ministry of Education, Culture, Sports, Science, and Technology in collaboration with research institutes and universities, operated the new Information Technology (IT) infrastructure for the Information-explosion Era project from 2005-2011.	Kim *et al.* (2014)
Smart Governance	Big data projects have been initiated by the local government in Syracuse, New York in collaboration with IBM in 2011. A smarter city project to use big data was launched for predicting and preventing vacant residential properties. Similarly, big data warehouse was constructed by seven Michigan's Department of Information technology for providing a single information source about the citizens to organizations and government agencies to enable them to provide better services.	Kim *et al.* (2014)
Smart Mobility	Countries such as UK, USA, and France are using big data applications on health care transport which help in analysing transport and distribution system data created by Global Positioning System (GPS), cell phone and other transactional data. Big data analytics facilitate to find the fastest route using various mechanisms such as O-Sense, and smart card data. It analyses the information gathered from GPS traces on traffic patterns and congestion.	Mehmood and Graham (2015), Niu *et al.* (2015)

Smart Environment	The smart grid project implemented in Los Angeles has augmented the physical infrastructure by using computational cyber infrastructure to ensure an efficient and reliable energy source. Similarly, the Peer Energy Cloud (PEC) project running in Germany is taking help of cloud based infrastructure that runs analytics on power consumption data gathered in every two seconds from the house-holds. This data analytics helps in measuring power consumption and forecasting the energy demand.	Zhou *et al.* (2016), Strohbach *et al.* (2015)
Smart Living	Big data analytics facilitate urban planning by using various types of sensors deployment (e.g., smart home sensors, vehicular networking, weather and water sensors, smart parking sensors, and surveillance objects). Eco-tourism services, virtual museums, digital art, new media co-creation, custom guides, live assisted translation and cultural mediation, exploration games are some of the ways to utilize big data for re-creation and ethnic diversity. Further, big data applications also help in attracting tourists, for example, Spain has increased its tourism rate by using geo-tagged photographs and tweets from visitors.	Rathore *et al.* (2016), Toppeta (2010), Sobolevsky *et al.* (2015)

3. Review Process

Based on the guidelines suggested by Kitchenham and Charters (2007), systematic literature review was carried out in this study. These guidelines include several activities such as development of review protocol, selection of review studies, extraction of data, synthesis of results from reviewed studies, and reporting the findings. Next, this section provides detailed elaboration of these activities and method used for conducting them.

3.1 Protocol development

This step involved the development of protocol for determining the search criteria for research papers. Research papers for review were obtained using advanced search option from the following prominent electronic databases: IEEEXplore, ProQuest, Sciencedirect, SpringerLink, JSTOR, ACM Digital Library, AIS eLibrary, EBSCO, Informs, Inderscience, Emerald, Sage, Wiley Interscience, and Google Scholar.

For extracting the research papers, the key search terms were decided. Table 2 provides the list of all the key categories and corresponding search term for that category. The Boolean "AND" operator was used for combining the search terms in each category together. Thus, all the possible combinations of search terms consisting of Category Type 1 AND Category Type 2 were used. This was done so that the text of extracted research papers contain the phrases related to "big data" and "smart city" together. This combination was used as the research papers required for this study should have addressed the big data challenges and their solutions in smart cities. The research papers extracted for this study were published on or before February 2016. The entire database search process in this step resulted into a total of 310 hits.

Table 2: Search Terms used in this Review

Category Type	**Search Terms**
Smart City	Smart City
	Smart Cities
	Digital City
	Digital Cities
	Knowledge City
	Knowledge Cities
	Information City
	Information Cities
	Ubiquitous City
	Ubiquitous Cities
	Intelligent City
	Intelligent Cities
	Urban
Big Data	Big Data
	Huge Data
	Large Scale Data
	Urban Data Management
	Large Volume of Data

3.2 Inclusion Decision on the basis of title and keywords

In this step, the research papers extracted via the search protocol were screened as per their title and keywords so that the inappropriate ones could be removed. Thus, this step also resulted into the exclusion of review articles, commentary, news, editorials, prefaces, non-English, and duplicate articles. Two researchers who were experts in the field of big data and smart city jointly worked on excluding the research papers that seemed irrelevant for fulfilling the objective of this study. This step resulted into a total of 190 research papers for further filtration.

3.3 Inclusion decision on the basis of abstract

This step involved the independent reading of abstract and conclusion of the papers by both the researchers. These researchers excluded the papers that they found irrelevant for this study. Though these research papers had the required search terms in their text, but they didn't focus on the big data challenges and their solutions in smart cities. Subsequently, third researcher read the abstract and conclusion of the excluded research papers so as to validate the reason for exclusion. Later, all three researchers had a discussion on the reasons for exclusion of particular research papers and jointly reached the agreement (Bennett, 1955). After this step, a total of 75 research papers were left.

3.4 Final Selection

Two researchers read the full text of the papers independently and screened them as per the following criteria:

a) Does the paper address the phenomenon of big data in context of smart cities?
b) Does the paper discuss the challenges of big data and the ways to deal with them?
c) Does the paper clearly state its research objective?

Later, third researcher reviewed and verified the excluded papers so as to achieve an agreement on reasons of exclusion for particular papers. Subsequently, the above criteria led to the identification of 38 potential research papers that could contribute in fulfilling the objective of this study. Later, all three researchers went through the entire reference list of the selected papers and could not discovered any other suitable paper. Thus, a total of 38 peer-reviewed research papers were shortlisted for the study being undertaken (see Appendix B). The selection process and the total number of research papers identified at each step are shown in figure 1.

Figure 2 depicts the year wise and publication channel wise distribution of research papers shortlisted for this study. It shows that the number of papers on the topic related to big data challenges and their solutions in smart cities have increased in the last few years. This publication trend can be a pointer of researchers' and practitioners' growing interest in this field. Moreover, it can be seen that the conference papers have 50% of share among all the publications.

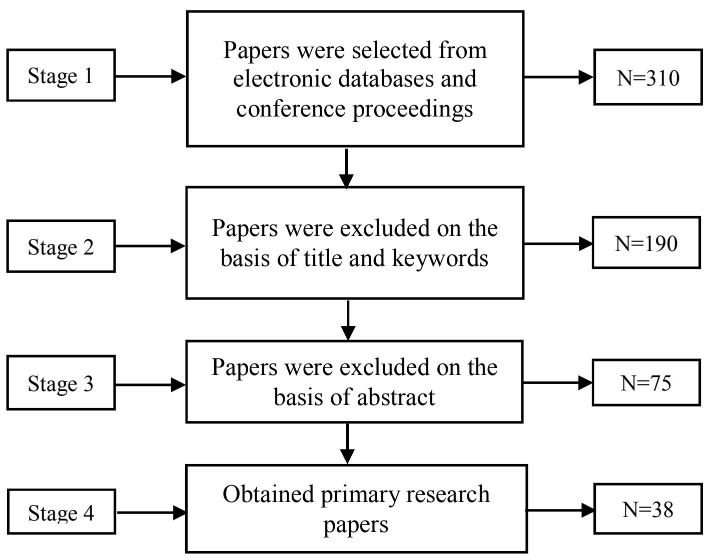

Figure 1: Stages of the study selection process

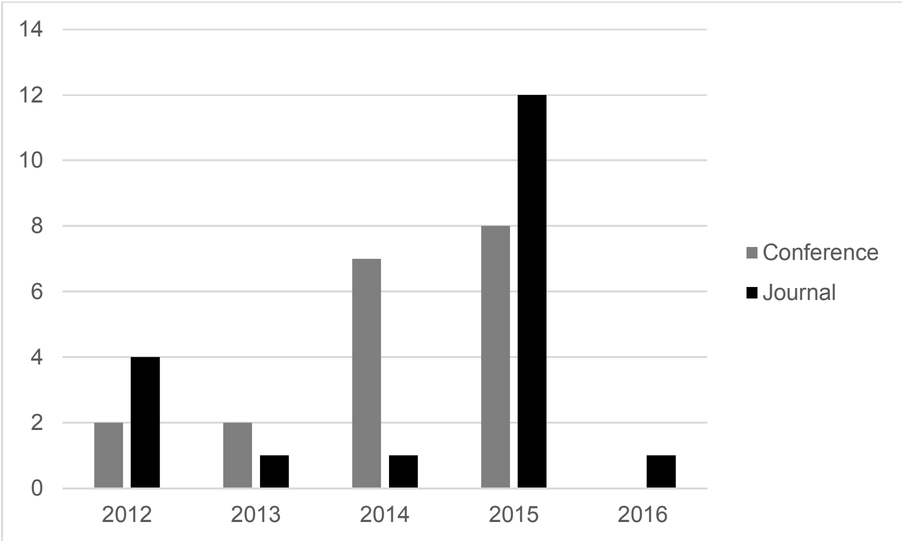

Figure 2: Year wise and Publication type wise distribution of research papers

3.5 Data Extraction and Synthesis

A data extraction form (see Appendix A) was created in order to extract the data from selected research papers. This was done to ensure that each research paper was analyzed consistently. All three researchers independently extracted the data from selected research papers for review. Later, they compared the extracted data and mutually resolved the mismatches. Lastly, the data was synthesized after identification of challenges associated with big data in smart cities and the respective solutions to deal with them.

4. Results

This section reviews the selected research papers and classifies the literature into eight broad categories of challenges faced by big data in smart cities (see Table 3). Further, it provides the possible solutions to deal with these challenges as suggested by various researchers. The rest of this section discusses each challenge and respective solutions one by one.

Table 3: Classification of Reviewed Articles

Study	Privacy	Security	Volume	Variety	Velocity	Veracity	Human Dynamics	Value
Ajiboye *et al.* (2015)		✓	✓		✓	✓		
Amato *et al.* (2014)			✓					✓
Artikis *et al.* (2013)			✓	✓		✓		
Aryal and Dutta (2015)	✓			✓		✓		✓
Baralis *et al.* (2014)								✓
Batty *et al.* (2012)	✓		✓	✓		✓	✓	✓
Bohli *et al.* (2015)	✓	✓						
Chen *et al.* (2012)			✓					

Cheng *et al.* (2015)				✓	✓			✓
Corradi *et al.* (2015)			✓					✓
d'Aquin *et al.* (2015)								✓
Dey *et al.* (2012)			✓					
Dobre and Xhafa (2014)	✓					✓		✓
Gupta *et al.* (2012)	✓		✓	✓		✓		
Jalali *et al.* (2015)	✓		✓					
Jara *et al.* (2013)							✓	
Jara *et al.* (2014)				✓				✓
Khan *et al.* (2013)					✓			✓
Khan *et al.* (2014)	✓	✓				✓		
Khan *et al.* (2015)								✓
Koh *et al.* (2015)					✓			✓
Lepri *et al.* (2015)			✓	✓	✓	✓		
Li *et al.* (2015a)								✓
Li *et al.* (2015b)	✓	✓	✓					
Li *et al.* (2015c)			✓	✓				✓
Lombardi *et al.* (2012)			✓	✓	✓			✓
Meijer and Bolívar (2015)							✓	
Namiot and Sneps-Sneppe (2012)						✓		
Parra *et al.* (2015)			✓	✓				
Sobolevsky *et al.* (2015)			✓	✓		✓	✓	✓
Strohbach *et al.* (2015)	✓	✓	✓	✓		✓		✓
Tachizawa *et al.* (2015)			✓			✓		✓
Truong and Dustdar (2014)								✓
Villanueva *et al.* (2014)					✓			
Wang and Wu (2016)							✓	
Wolff *et al.* (2015)			✓	✓		✓		
Xiong *et al.* (2014)						✓		
Zhu and Zuo (2015)		✓						

Privacy: Data privacy and protection issues are increasing as more data is being generated, collected, harvested, and processed (Batty, 2012). Analysis of improperly masked data can reveal the information of its users (Tien, 2013). Thus, big data needs to be ethically owned and used by the producers (David Chandler). However, if individual access to the data is made impracticable, data is likely to be de-identified to an extent sufficient to diminish the privacy concerns (Tene and Polonetsky, 2012). Online communication takes place via several channels and wider citizenry (Batty *et al.*, 2012). Therefore, sensitive information such as location needs to be implicitly altered in order to protect the communicator's privacy. This should be done at hardware as well as software levels (Tene and Polonetsky, 2012).

Privacy issues such as lack of viable anonymization technique, lack of data accessibility (Bohli *et al.*, 2015; Wehn and Evers, 2015), non-trustable data (Dobre and Xhafa, 2014), and data transfer over the unpredictable networks (Wolff *et al.*, 2015) have been mentioned in the literature. In order to deal with such privacy issues, researchers have suggested the solutions such as legal provisioning on using the data (Wolff *et al.*, 2015), use of data codes for ensuring safe travel of data over the networks (Wolff *et al.*, 2015), and development of context aware framework for managing the lack of predictability of wireless connectivity and data privacy (Dobre and Xhafa, 2014).

Security: Provision of diversified data by sensor networks, telescopes, scientific experiments, and high throughput instruments leads to exponential increase in the datasets (Batty *et al.*, 2012). Policies need to be formulated to keep the data accessibility fairly open to the data consumers while validating and monitoring all the parties involved in smart cities (Cheng *et al.*, 2015). Security of data processing and sharing, intellectual property, users, personal privacy, commercial secrets, and financial information are some of the concerns in smart cities. Further, fraudulent users and financial losses due to malicious data and legitimate services are other potential issues in the smart cities (Wolff *et al.*, 2015; Chen and Zhang, 2014). According to (Kaisler *et al.*, 2013), the unregulated accumulation of data by numerous social media organizations is the biggest threat to security as large set of data tempts the cyber attackers. Therefore, the security of data needs to be implemented at technological, government and business policy, and public levels through strict legal terms and conditions (Kaisler *et al.*, 2013).

For addressing the security issue, researchers have suggested the solutions such as forming functional groups ensuring security of communication and services (Wolff *et al.*, 2015), developing a mobile-cloud framework to solve the data over-collection problem, setting a benchmark to score various applications which could check data over-collection, and enforcing legitimate service provisioning for every user (d'Aquin *et al.*, 2015). A number of data pre-processing techniques such as data cleaning, data integration, data transformation, and date reduction can be applied to remove noise, correct inconsistencies, and find the relevant data (Chen and Zhang, 2014). Lastly, data should be modified only by an authorized party/owner to prevent its misuse and therefore, maintain its integrity (Hashem *et al.*, 2015).

Volume: The huge amount of data has emerged with the advent of social web and interconnected devices. Smartphones, Voice over Internet Protocol (VoIP), video conferencing, and Internet connected devices (sensors and computers) have allowed monitoring and sharing of every aspect of data in nanoseconds (Batty, 2012). Humans' daily activities are recorded and utilized for supporting day to day transactions without their knowledge. Thus, it raises the data storage issues (Kemp *et al.*, 2015).

Data overcrowding increases the data redundancy and makes the big data management as one of the serious challenges in smart era (d'Aquin *et al.*, 2015). In order to deal with this issue, researchers have suggested the solutions such as creating cloud-based data stores that provide reliable, scalable, and cost-effective way to store heavy chunks of data (Khan *et al.*, 2014) and adopting data literacy techniques for timely removal of undesirable data (Koh *et al.*, 2015). Further, a constant interaction between cloud servers is necessary to continuously process the data so that undesirable data can be discarded (Namiot and Sneps-Sneppe, 2012).

On recognizing the challenges arising due to high volume and complexity of defense-related big data, the U.S. Defense Advanced Research Project Agency (DARPA) initiated the XDATA program in 2012. The program engages applied mathematics, computer science, and data visualization communities to develop big data analytics and usability solutions for war fighters (Chen *et al.*, 2012).

Variety: Variety is the result of the growth of virtually unlimited heterogeneous data (Hashem *et al.*, 2015). This data is in various forms such as structured, semi-structured, unstructured, missing, text, video, image, and multimedia. The data is collected from various sources such as documents, emails, social media, text messages, videos, still images, audios, graphs, outputs from different types of machine-generated data from sensor devises, radio frequency identification (RFID) tags, machine logs, cell phone global positioning system (GPS) signals, and DNA analysis devices (Bertino *et al.*, 2011).

The incompatibility in data creates the heterogeneity problem (Li *et al.*, 2015b), recognition problem arising by merging of data from various sources (Xiong *et al.*, 2014), and interaction problem between big data and smart city (Khan *et al.*, 2013). For solving these issues, researchers have proposed a number of possible techniques such as automatic recognition of data to form patterns and remove outliers (Cheng *et al.*, 2015), use of data analytics to convert heterogeneous multi-level geographical data into useful information to derive the knowledge (Li *et al.*, 2015b), and construction of data models for deeply looking into the data for forming meaningful patterns (Khan *et al.*, 2013; Jara *et al.*, 2013).

Noordhuis applied PageRank Algorithm on Twitter user base to obtain user rankings. Computations in large amount of data were conducted in a two-phase process. In the first phase, i.e., crawling phase, the data was retrieved from Twitter and in the second phase, i.e., processing phase, the PageRank algorithm was applied to compute the data. Using this algorithm, Noordhuis reduced the variety of data arriving from 1.8 billion nodes and analyzed the user rankings from heterogeneous data within few seconds (Hashem *et al.*, 2015).

Velocity: This attribute is associated with the rate at which data is generated and the speed at which it has to be analyzed so that instantaneous decisions can be taken. The increasing velocity of data generation is attributed to the proliferation of sensing devices connected to the Internet (Hashem *et* *al.*, 2015). Further, this data should be converted into useful information on real time basis (Waterman and Bruening, 2014) as success of smart cities is based on the decisions taken on the relevant information gathered from different geographic locations (Baralis *et al.*, 2014).

Issues such as data visualization, i.e., failure of big data when unusual circumstances such as Tsunami takes place (Villanueva *et al.*, 2014), delay in data fetching from remote storage devices or due to geographical constraints (Li *et al.*, 2015b), lack of big data analytics platform between applications and services to provide data intelligence (Xiong *et al.*, 2014), and lack of appropriate tools and techniques for decision making (Truong and Dustdar, 2014) have been highlighted in literature. In order to deal with such issues, researchers have suggested techniques such as data pre-fetching using Bayesian algorithm which can timely predict the data required by the user and transfer it from remote location to the local cache (Li *et al.*, 2015b), data streaming in real time using glyphs to ensure scalability and modularity of data to overcome the visualization issues (Li *et al.*, 2015b), involvement of humans to visualize the patterns as they can successfully collect the data during unusual circumstances (Li *et al.*, 2015b), and the development of context aware platforms between data sources and services for effective decision making process (Xiong *et al.*, 2014).

Alacer Corporation, an online retailer in California suffered heavy revenue losses because of unreliable and untimely notifications of service problems within its e-commerce platform system. By using a monitoring platform, Alacer reduced the response time from hour to seconds. Alacer used big data algorithms for providing pro-active and reactive notifications on time, thus dramatically improving the satisfaction of customers and eliminated the service level shortfalls (Hashem *et al.*, 2015).

Veracity: Veracity deals with managing the quality, uncertainty, and trustworthiness of data (Buhl *et al.*, 2013). Data is affected by the lack of reliability and predictability (Ward and Barker, 2013). To maintain data veracity, it is important to discard noisy event sources and include reliable data sources (Artikis *et al.*, 2013). There are instances that put veracity in question, e.g., satellite signals are lost as they bounce off tall buildings and structures or GPS loses accuracy indoors (Li *et al.*, 2015a). Further, if data is objectively false, the analytical results are meaningless and unreliable (Dey *et al.*, 2012).

Veracity related issues arise because of a number of reasons such as data uncertainty leading to dirty data, non-unified data, and non-traceable data (Bohli *et al.*, 2015; Dobre and Xhafa, 2014; Cheng *et al.*, 2015; Dey *et al.*, 2012), process challenges in finding the right model for analysis and the ability to iterate quickly (Li *et al.*, 2015a), and management challenges such as governance and ethics ensuring the right use of data (Bohli *et al.*, 2015). In order to deal with these issues, researchers have suggested a number of techniques such as data vitalization for large-scale and heterogeneous dataset analysis (Corradi *et al.*, 2015), development of context-aware platform for ensuring the reliability of data (Bohli *et al.*, 2015; Dobre and Xhafa, 2014), cooperative location sensing system which is an adaptive location-sensing system enabling the devices to estimate their position in a self-organizing manner (Dey *et al.*, 2012).

In Sydney, Run-Time Event Calculus (RTEC) recognition engine uses city transport and traffic data. Data is collected from various sources such as sensors mounted on intersections, bus probe data stating their location, line, and delay, and data on the traffic arriving from various cities. Data streams from these multiple sources generate common composite events. These events are matched against each other to identify the mismatches. Based on these mismatches, temporal regions of uncertainty are identified and RTEC system allows to automatically estimate events to deal with uncertainty. Adoption of RTEC has increased the efficiency of the transport management system in Sydney and has emerged as a promising approach to deal with veracity in big data applications (Artikis *et al.*, 2013).

Human Dynamics: This attribute deals with the humans' role in data collection, data analysis, and decision-making process (Stankovic, 2014). Humans interrogate the data, formulate and test hypotheses, draw conclusions, and determine how big data can be successful or failure. As humans are inseparable entities, they act like prosumers, i.e., producers as well as consumers of the data. Therefore, they should abide by the policies of data collection and retrieval (Villanueva *et al.*, 2014). Interaction between humans and smart objects should be kept high as they are the ones who visualize data patterns (Shapiro, 2006). The continuous feedback by humans can help in improving the entire big data system (Amato *et al.*, 2014). During the technological development, considerable social and institutional transformations take place that require the human interface to support the structural changes. For example, when Singapore decided to become a "wise city", citizens started undertaking IT education and subscribing broadband at home (Gupta *et al.*, 2012).

Value: Large amount of data is worthless until it is converted into value. Value indicates extraction of intelligence from the data in order to gain competitive edge for effective performance (Jara *et al.*, 2014). There are certain value related issues such as challenges in mining the patterns from city data (Bohli *et al.*, 2015), inefficient handling of large amount of data (Dobre and Xhafa, 2014; Amato *et al.*, 2014) , inability to provide quality service to users by using fast processing engines to analyze large data sets on timely basis (Dobre and Xhafa, 2014; Koh *et al.*, 2015) , bottleneck in sharing of processed data across various applications (Corradi *et al.*, 2015; Jara *et al.*, 2013) , lack of legal provisions to enforce data policies in city (Dey *et al.*, 2012), and the high computational cost of data processing (Aryal and Dutta, 2015; Khan *et al.*, 2015).

Some of the suggested methods to handle these issues as suggested by researchers are data characterization using statistical tools (Aryal and Dutta, 2015) and models to determine the relationship between data and smart city (Amato *et al.*, 2014; Khan *et al.*, 2015) , continuous optimization of data to make it cost-effective for storage and processing (Aryal and Dutta, 2015;

Corradi *et al.*, 2015), data predictability to provide quality services to users (Dobre and Xhafa, 2014; Koh *et al.*, 2015), use of elasticity principle in big data analytics techniques to deal with diversity and distribution of data (Truong and Dustdar, 2014), and development of sophisticated operation center to integrate all type of real data which could provide customized services to users (Li *et al.*, 2015a).

5. Discussion

Figure 3 depicts the process of big data management in smart cities. It explains the challenges associated with the use of big data and the possible solutions to overcome those challenges so that smart decisions can be taken on time. These smart decisions favorably influence the functioning of smart cities. Next, this section explains the entire process in detail.

In smart cities, data from multiple sources such as antenna, RFID, sensors, and social media is collected, retrieved, and stored in cloud (Khan *et al.*, 2014; Giffinger *et al.*, 2007) which helps in reducing the cost of storage and operations (Li *et al.*, 2015b). Further, it is important to ensure the privacy and security of big data, however, the tools and techniques used for collecting the massive data sets are often deprived of adequate security and privacy measures (Perboli *et al.*, 2014). Lack of adequate compliance policies, existing technologies, and applications often breach security and privacy of data knowingly or unknowingly (Bohli *et al.*, 2015; Li *et al.*, 2015b; Buhl *et al.*, 2013). Therefore, it is essential to protect the data for reducing the risk of data theft and identity fraud which can lead to financial or structural damages (Bohli *et al.*, 2015; Aryal and Dutta, 2015).

The collected data sometimes lacks the basic structure and has missing values. However, this data becomes useful after efficient and timely processing (Bohli *et al.*, 2015). The four V's of big data: Volume, Variety, Velocity, and Veracity include a number of issues (Khan *et al.*, 2014; Jara *et al.*, 2014) such as data reduction, data structuring, data visualization, and dirty data. In order to combat these issues, it is essential to diagnose the data before-hand and make it context viable so that effective decisions can be taken (Koh *et al.*, 2015).

Data in high volume is created from many transactions and is stored as events in various files and databases (Demchenko *et al.*, 2013). It is important to draw unique data structures from these events based on data's proximity and users' needs (Namiot and Sneps-Sneppe, 2012). This would reduce big data into smaller, accessible, and manageable data records which are scalable and coherent in nature (Demchenko *et al.*, 2013). Commonly used data reduction strategies are filtering, sampling, binned aggregation, and model fitting (Liu *et al.*, 2013).

Variety in big data leads to data inconsistency and non-aligned data structures which result into analytic sprawl. Heterogeneous data gets generated when humans and their societies are monitored and studied comprehensively (Demchenko *et al.*, 2013). To address this issue, it is important to draw structures from big data using tagcloud visualization techniques (Chen *et al.*, 2012). Thus, it is crucial to streamline the data and create unique structures from it so as to make sense of the data and to take real time actions (Gupta *et al.*, 2012; Demchenko *et al.*, 2013). High velocity of big data generation creates a challenge of handling the speed with which new and/or updated data is constantly added to the data stores (Chen *et al.*, 2012). Lack of effective data processing techniques affects the real-time visualization of big data. In order to deal with this issue, human analysts should be kept in loop as their expertise can help in extracting relationships between the data (Assunção *et al.*, 2015)

Big data veracity has two aspects namely data consistency and data trustworthiness (Demchenko *et al.*, 2013). The data quality issue arises because of its ambiguity, incompleteness and inconsistency (Lukoianova and Rubin, 2014). Uncertainty management can be done by utilizing multiple reliable sources, checking data credibility by using various quantitative tools and identification frameworks (Lukoianova and Rubin, 2014; Kaisler *et al.*, 2013). The high quality feature adds value to the big data by allowing its effective utilization with improved accuracy and insights (Lukoianova and Rubin, 2014; Kaisler *et al.*, 2013).

In order to achieve the maximum benefits without compromising data authenticity, it is important to identify the registered users who can filter the data in order to draw valuable information out of it. The filtered data needs to be structured as per the geographic location and homogeneity of patterns (Li *et al.*, 2015b). Data services such as visualization techniques help in streamlining the data and drawing structures to extract meaningful information for further analysis (Chen *et al.*, 2012). Various computational devices mounted across the city provide timely information related to traffic and weather. Such information helps in taking timely decisions and saves money and time (Bohli *et al.*, 2015). Some of the functional devices such as energy harvesting wireless technological devices (Rathore et al., 2016), safety gadgets, and IoT sensing devices (Bohli et al., 2015) can further improve the quality of life of people in smart cities.

Various components of smart city such as smart living, smart governance, smart people, smart economy, smart mobility, and smart environment (Li *et al.*, 2015b; Corradi *et al.*, 2015), continuously interact with each other via feedforward intelligent information and feedbacks making the entire city competitive and effective (Bohli *et al.*, 2015). The continuous interactions between humans and machines provide opportunities to overcome the challenges by introducing new technologies and methodologies (Jara *et al.*, 2013).

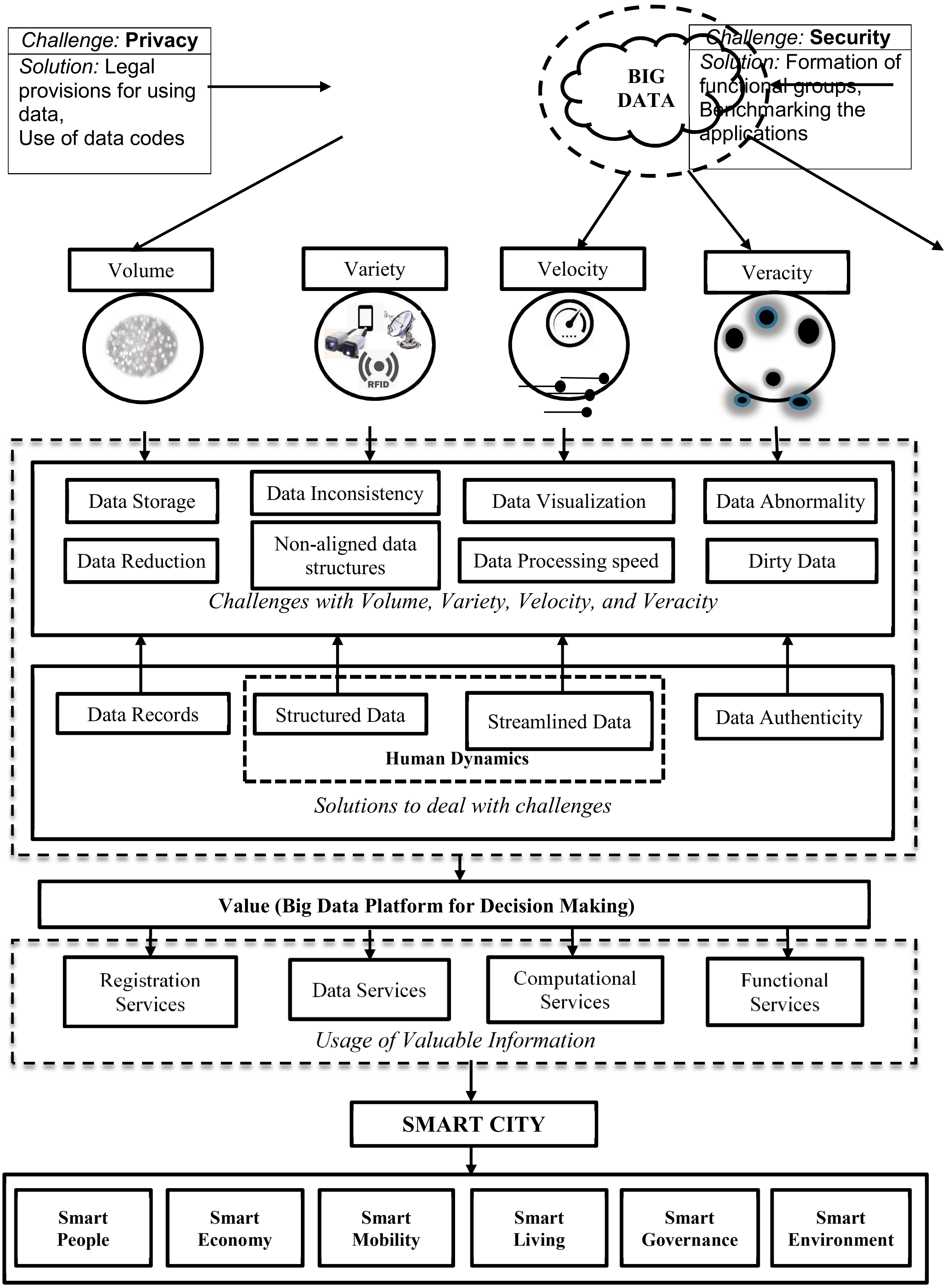

Figure 3: Management of Big Data in Smart City

6. Conclusion and Future Research

The key objective of this research was to discover the approaches to deal with challenges associated with big data in smart cities. In order to accomplish this objective, a systematic literature review was conducted to identify such challenges and solutions to deal with them as proposed by various researchers. This research papers follows a rigorous methodology for selecting papers for systematic literature review. The selected papers were reviewed, analyzed, and interpreted to identify the potential strategies to deal with big data challenges. The results of this paper contribute valuable information to the big data practitioners by illustrating the process of effective management of big data in smart cities. The process of managing big data in smart cities can facilitate in utilizing the benefits that can be drawn by analyzing the big data in real time.

This review clearly found that a majority of research papers in this area are conceptual in nature lacking empirical methodology. Therefore, empirical research aligned with big data findings may provide interesting and novel insights for academicians as well as practitioners. Secondly, existing research papers advocated the use of context aware platforms for fast data retrieval, processing, and decision making, however these platforms do not provide any solution for managing the large volume of data. Thus, data governance and policy research in the area of management of the information assets would be highly insightful. Thirdly, as more and more cities move towards urbanization, a huge volume of generated data can provide inputs for policy formation and process improvement. Thus, a lot of potential work is possible in these directions which may greatly augment the literature on smart cities. Fourthly, as discussed by various researchers, role of human beings is critically important in big data analytics. Future researchers need to identify the exact areas in which human cognitive system has strong connection with the various characteristics of smart cities. Lastly, many research papers focused on using big data in political, ecological, and economic contexts in smart cities, however, the areas such as social subsystems also needs attention.

Appendix A - Data Extraction Form

1. Date of extraction of the research paper
2. Bibliographic reference (title, author, year, and source) of the research paper
3. Selection of research paper type (journal article, conference paper, working paper, workshop paper or unclear)
4. Research methodology (empirical, experiential, experimental or review)
5. Findings of the research paper

Appendix B – Papers included in the Review

Ajiboye, S. O., Birch, P., Chatwin, C. and Young, R. (2015), "Hierarchical video surveillance architecture: a chassis for video big data analytics and exploration", In *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, pp. 94070K-94070K.

Amato, A., Di Martino, B. and Venticinque, S. (2014). Big Data Processing for Pervasive Environment in Cloud Computing, *Proceedings of the International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, pp. 598-603, IEEE, Salerno.

Artikis, A., Weidlich, M., Gal, A., Kalogeraki, V. and Gunopulos, D. (2013), "Self-adaptive event recognition for intelligent transport management", *Proceedings of the International Conference on Big Data*, pp. 319-325, IEEE, Silicon Valley, CA.

Aryal, J. and Dutta, R. (2015), "Smart City and Geospatabilty: Hobart Deeply Learned", *Proceedings of the 31st IEEE International Conference on Data Engineering Workshops (ICDEW)*, pp. 108 - 109, IEEE, Seoul.

Baralis, E., Cagliero, L., Cerquitelli, T., Chiusano, S., Garza, P., Grimaudo, L. and Pulvirenti, F. (2014), "Misleading generalized itemset mining in the cloud", *Proceedings of the International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, pp. 211-216, IEEE, Milan.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G. and Portugali, Y. (2012), "Smart cities of the future", *The European Physical Journal Special Topics*, Vol. 214 No. 1, pp. 481-518.

Bohli, J. M., Skarmeta, A., Victoria Moreno, M., Garcia, D. and Langendorfer, P. (2015), "SMARTIE project: Secure IoT data management for smart cities", *Proceedings of the International Conference on Recent Advances in Internet of Things (RIoT)*, pp. 1-6, IEEE, Singapore.

Chen, H., Chiang, R. H. and Storey, V. C. (2012), "Business Intelligence and Analytics: From Big Data to Big Impact", *MIS quarterly*, Vol. 36 No. 4, pp. 1165-1188.

Cheng, B., Longo, S., Cirillo, F., Bauer, M. and Kovacs, E. (2015), "Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander", *Proceedings of the International Congress on Big Data (BigData Congress)*, pp. 592-599, IEEE, New York, NY.

Corradi, A., Curatola, G., Foschini, L., Ianniello, R. and De Rolt, C. R. (2015), "Automatic extraction of POIs in smart cities: Big data processing in ParticipAct", *Proceedings of the International Symposium on Integrated Network Management*, pp. 1059-1064, IEEE, Ottawa, ON.

d'Aquin, M., Davies, J. and Motta, E. (2015). Smart Cities' Data: Challenges and Opportunities for Semantic Technologies, *IEEE Internet Computing*, (6), 66-70.

Dey, S., Chakraborty, A., Naskar, S. and Misra, P. (2012), "Smart city surveillance: Leveraging benefits of cloud data stores", *Proceedings of the 37th Conference on Local Computer Networks Workshops (LCN Workshops)*, IEEE, Clearwater, FL.

Dobre, C. and Xhafa, F. (2014), "Intelligent services for big data science", *Future Generation Computer Systems*, Vol. 37, pp. 267-281.

Gupta, R., Gupta, H. and Mohania, M. (2012), "Cloud computing and big data analytics: what is new from databases perspective?", In *Big Data Analytics*, Springer Berlin Heidelberg, pp. 42-61.

Jalali, R., El-Khatib, K. and McGregor, C. (2015), "Smart city architecture for community level services through the internet of things", *Proceedings of the 18th International Conference on Intelligence in Next Generation Networks (ICIN)*, pp. 108-113, IEEE, Paris.

Jara, A. J., Bocchi, Y. and Genoud, D. (2013), "Determining human dynamics through the Internet of Things", *Proceedings of the International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 3, pp. 109-113, IEEE, Atlanta, GA.

Jara, A. J., Genoud, D. and Bocchi, Y. (2014), "Big data in smart cities: from poisson to human dynamics", *Proceedings of the 28th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 785-790, IEEE, Victoria, BC.

Khan, Z., Anjum, A. and Kiani, S. L. (2013). Cloud based big data analytics for smart future cities. *Proceedings of the 6th international conference on utility and cloud computing*, pp. 381-386, IEEE Computer Society, Washington, DC, USA.

Khan, Z., Pervez, Z. and Ghafoor, A. (2014), "Towards cloud based smart cities data security and privacy management", *Proceedings of the 7th International Conference on Utility and Cloud Computing (UCC)*, pp. 806-811, IEEE, London.

Khan, Z., Anjum, A., Soomro, K. and Tahir, M. A. (2015), "Towards cloud based big data analytics for smart future cities", *Journal of Cloud Computing*, Vol. 4 No. 1, pp. 1-11.

Koh, J. M., Sak, M., Tan, H. X., Liang, H., Folianto, F. and Quek, T. (2015), "Efficient data retrieval for large-scale smart city applications through applied Bayesian inference", *Proceedings of the 10th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pp. 1-6, IEEE, Singapore.

Lepri, B., Antonelli, F., Pianesi, F. and Pentland, A. (2015), "Making big data work: smart, sustainable, and safe cities", *EPJ Data Science*, Vol. 4 No. 1, pp. 1-4.

Li, D., Cao, J., and Yao, Y. (2015a), "Big data in smart cities", *Science China Information Sciences*, Vol. 58, pp. 1-12.

Li, Y., Dai, W., Ming, Z. and Qiu, M. (2015b), "Privacy protection for preventing data over-collection in smart city", *IEEE Transactions on Computers*, Vol. PP No. 99, pp. 1.

Li, J., Tao, F., Cheng, Y. and Zhao, L. (2015c), "Big data in product lifecycle management", *The International Journal of Advanced Manufacturing Technology*, Vol. 81 No. 1-4, pp. 667-684.

Lombardi, P., Giordano, S., Farouh, H. and Yousef, W. (2012), "Modelling the smart city performance", *Innovation: The European Journal of Social Science Research*, Vol. 25 No. 2, pp. 137-149.

Meijer, A. and Bolívar, M. P. R. (2015), "Governing the smart city: a review of the literature on smart urban governance", *International Review of Administrative Sciences*, 0020852314564308.

Namiot, D. and Sneps-Sneppe, M. (2012), "Context-aware data discovery", In *Proceedings of the 16th International Conference on Intelligence in Next Generation Networks (ICIN)*, pp. 134-141, IEEE, Berlin.

Parra, L., Sendra, S., Lloret, J. and Bosch, I. (2015), "Development of a Conductivity Sensor for Monitoring Groundwater Resources to Optimize Water Management in Smart City Environments", *Sensors*, Vol. 15 No. 9, pp. 20990-21015.

Sobolevsky, S., Bojic, I., Belyi, A., Sitko, I., Hawelka, B., Murillo Arias, J. and Ratti, C. (2015), "Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity", *Proceedings of the International Congress on Big Data (BigData Congress)*, IEEE, New York, NY.

Strohbach, M., Ziekow, H., Gazis, V. and Akiva, N. (2015), "Towards a big data analytics framework for IoT and smart city applications", In *Modeling and Processing for Next-Generation Big-Data Technologies*, Springer International Publishing, pp. 257-282.

Tachizawa, E. M., Alvarez-Gil, M. J. and Montes-Sancho, M. J. (2015), "How “smart cities” will change supply chain management", *Supply Chain Management: An International Journal*, Vol. 20 No. 3, pp. 237-248.

Truong, H. L. and Dustdar, S. (2014), "Principles of software-defined elastic systems for big data analytics", *Proceedings of the International Conference on Cloud Engineering (IC2E)*, pp. 562-567, IEEE, Boston, MA.

Villanueva, F. J., Aguirre, C., Villa, D., Santofimia, M. J. and López, J. C. (2014), "Smart City data stream visualization using Glyphs", *Proceedings of the 8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pp. 399-403, IEEE, Birmingham, United Kingdom.

Wang, K. Y. and Wu, C. M. (2016), "Building an Intelligent Government, Intelligent City, and Intelligent Citizenry Through ICTs: Smart City Innovations in New Taipei City, Taiwan", In *Smarter as the New Urban Agenda*, Springer International Publishing, pp. 259-278.

Wolff, A., Kortuem, G. and Cavero, J. (2015), "Urban Data Games: creating smart citizens for smart cities", *Proceedings of the 15th International Conference on Advanced Learning Technologies (ICALT)*, IEEE, pp. 164-165, Hualien.

Xiong, Z., Zheng, Y. and Li, C. (2014), "Data Vitalization's Perspective Towards Smart City: A Reference Model for Data Service Oriented Architecture", *Proceedings of the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, pp. 865-874, IEEE, Chicago, IL.

Zhu, Y. and Zuo, J. (2015), "Research on Security Construction of Smart City", *System*, Vol. 9 No. 8, pp. 197-204.

References

Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A. and Buyya, R. (2015), “Big Data computing and clouds: Trends and future directions”, *Journal of Parallel and Distributed Computing*, Vol. 79-80, pp.3-15.

Ballon, P. and Schuurman, D. (2015), "Living labs: concepts, tools and cases", *info*, Vol. 17 No. 4.

Batty, M. (2012), "Smart cities, big data", *Environment and Planning B: Planning and Design*, Vol. 39 No. 2, pp. 191-193.

Bennett, E. B. (1955), “Discussion, decision, commitment and consensus in group decision”, *Human relations*, Vol. 8, pp. 251-273.

Bertino, E., Bernstein, P., Agrawal, D., Davidson, S., Dayal, U., Franklin, M., Gehrke, J., Haas, L., Halevy, A., and Han, J. (2011), "Challenges and Opportunities with Big Data", available at: A community white paper developed by leading researchers across the United States, http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf (accessed 10 January, 2016).

Buhl, H. U., Röglinger, M., Moser, F. and Heidemann, J. (2013), "Big data", *Business & Information Systems Engineering*, Vol. 5 No. 2, pp. 65-69.

Chatterjee, S. and Kar, A. K. (2015), "Smart Cities in developing economies: A literature review and policy insights", *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, pp. 2335-2340, Kochi.

Chen, C. P. and Zhang, C. Y. (2014), "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", *Information Sciences*, Vol. 275, pp. 314-347.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A. and Scholl, H. J. (2012), "Understanding smart cities: An integrative framework*", Proceedings of the 45th Hawaii International Conference on System Science (HICSS)*, IEEE, pp. 2289-2297, Maui, HI.

Demchenko, Y., Grosso, P., De Laat, C. and Membrey, P. (2013), "Addressing big data issues in scientific data infrastructure", *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, IEEE , pp. 48-55

Domingo, A., Bellalta, B., Palacin, M., Oliver, M. and Almirall, E. (2013), "Public open sensor data: Revolutionizing smart cities", *Technology and Society Magazine*, IEEE, Vol. 32 No. 4, pp. 50-56.

Dutton, W. H. (2014), "Putting things to work: social and policy challenges for the Internet of things", *info*, Vol. 16 No. 3, pp. 1 – 21.

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanović, N. and Meijers, E. (2007), "Smart Cities: Ranking of European Medium-Sized Cities. Vienna, Austria: Centre of Regional Science (SRF)", available at: www.smart-cities.eu/download/smart_cities_final_report.pdf (accessed 1 November 2015).

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A. and Khan, S. U. (2015), "The rise of "big data" on cloud computing: Review and open research issues", *Information Systems*, Vol. 47, pp. 98-115.

Kaisler, S., Armour, F., Espinosa, J. A. and Money, W. (2013), "Big data: Issues and challenges moving forward", *Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS)*, IEEE, pp. 995-1004, Wailea, Maui, HI.

Kemp, G., Vargas-Solar, G., Da Silva, C. F., Ghodous, P., Collet, C. and Lopezamaya, P. (2015), "Towards Cloud big data services for intelligent transport systems", In *Transdisciplinary Lifecycle Analysis of Systems: Proceedings of the 22nd ISPE Inc. International Conference on Concurrent Engineering*, Vol. 2, p. 377.

Kim, G. H., Trimi, S. and Chung, J. H. (2014), "Big-data applications in the government sector", *Communications of the ACM*, Vol. 57 No. 3, pp. 78-85.

Kitchenham, B.A. and Charters, S.M. (2007), "Guidelines for Performing Systematic Literature Reviews in Software Engineering", In *Version 2.3. EBSE Technical Report EBSE-2007-01*, Keele University and Durham University.

Kitchin, R. (2014), "The real-time city? Big data and smart urbanism", *GeoJournal*, Vol. 79 No. 1, pp. 1-14.

Kuk, G. and Janssen, M. (2011), "The business models and information architectures of smart cities", *Journal of Urban Technology*, Vol. 18 No. 2, pp. 39-52.

Kumar, T. V. (2015), *E-governance for Smart Cities* (pp. 1-43). Springer Singapore.

Lee, J. and Lee, H. (2014), “Developing and validating a citizen-centric typology for smart city services”, *Government Information Quarterly*, 31, S93-S105.

Lierow, M. (2014), "B2City: the next wave of urban logistics", available at: www.supplychain247.com/paper/b2city_the_next_wave_of_urban_logistics (accessed 1 November, 2015).

Liu, Z., Jiang, B. and Heer, J. (2013), “imMens: Real-time Visual Querying of Big Data”, *Computer Graphics Forum*, Vol. 32, No. 3-4, pp. 421-430.

Lukoianova, T. and Rubin, V.L (2014), “Veracity roadmap: Is big data objective, truthful and credible?”, *Advances in Classification Research Online*, Vol. 24 No. 1, pp. 4-15.

Mehmood, R. and Graham, G. (2015), “Big Data Logistics: A health-care Transport Capacity Sharing Model”, *Procedia Computer Science*, Vol. 64, pp. 1107-1114.

Niu, X., Zhu, Y., Cao, Q., Zhang, X., Xie, W. and Zheng, K. (2015), “An online-traffic-prediction based route finding mechanism for smart city”, *International Journal of Distributed Sensor Networks*, Vol. 2015, Article ID 970256, 16 page.

Perboli, G., De Marco, A., Perfetti, F. and Marone, M. (2014), “A new taxonomy of smart city projects”, *Transportation Research Procedia*, Vol. 3, pp. 470-478.

Rathore, M. M., Paul, A., Ahmad, A. and Rho, S. (2016), “Urban planning and building smart cities based on the internet of things using big data analytics”, *Computer Networks*.

Roscia, M., Longo, M. and Lazaroiu, G. C. (2013), “Smart city by multi-agent systems”, *Proceedings of the International Conference on Renewable Energy Research and Applications (ICRERA)*, IEEE, pp. 371-376, Madrid.

Shapiro, J. M. (2006), "Smart cities: quality of life, productivity, and the growth effects of human capital", *The review of economics and statistics*, Vol. 88 No. 2, pp. 324-335.

Stankovic, J. A. (2014), "Research directions for the internet of things", *Internet of Things Journal, IEEE*, Vol. 1 No. 1, pp. 3-9.

Su, K., Li, J. and Fu, H. (2011), “Smart city and the applications”, *Proceedings of the International Conference on Electronics, Communications and Control (ICECC),* IEEE, pp. 1028-1031, Zhejiang.

Tene, O. and Polonetsky, J. (2012), "Big data for all: Privacy and user control in the age of analytics", *Northwestern Journal of Technology and Intellectual Property*, Vol. 11 No. 5, pp. 239-273.

Tien, J. M. (2013), "Big data: Unleashing information", *Journal of Systems Science and Systems Engineering*, Vol. 22 No. 2, pp. 127-151.

Toppeta, D. (2010), “The smart city vision: How innovation and ICT can build smart, “liveable”, sustainable cities”, Innovation Knowledge Foundation (Think! Report 005/2010) available at: http://www.inta-aivn.org/images/cc/Urbanism/background%20documents/Toppeta_Report_005_2010.pdf (accessed 1 February, 2016).

van der Graaf, S. and Veeckman, C. (2014), "Designing for participatory governance: assessing capabilities and toolkits in public service delivery", *info*, Vol. 16 No. 6 pp. 74 – 88.

Ward, J. S. and Barker, A. (2013), "Undefined by data: a survey of big data definitions", *arXiv preprint arXiv:1309.5821*.

Waterman, K. K. and Bruening, P. J. (2014), "Big Data analytics: risks and responsibilities", *International Data Privacy Law*, Vol. 4 No. 2, pp. 89-95.

Wehn, U. and Evers, J. (2015), “The social innovation potential of ICT-enabled citizen observatories to increase eParticipation in local flood risk management”, *Technology in Society*, Vol. 42, pp. 187-198.

Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M. (2014), "Internet of things for smart cities", *Internet of Things Journal, IEEE*, Vol. 1 No. 1, pp. 22-32.

Zhou, K., Fu, C. and Yang, S. (2016), “Big data driven smart energy management: From big data to big insights”, *Renewable and Sustainable Energy Reviews*, Vol. 56, pp. 215-225.