

SURVEY PAPER

Open Access



The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: a review and synthesis

Simon Elias Bibri^{1*} and John Krogstie²

*Correspondence:
simoe@ntnu.no

¹ Department of Computer and Information Science and Department of Urban Planning and Design, NTNU Norwegian University of Science and Technology, Sem Saelands veie 9, 7491 Trondheim, Norway
Full list of author information is available at the end of the article

Abstract

Data sensing, information processing, and networking technologies are being fast embedded into the very fabric of the contemporary city to enable the use of innovative solutions to overcome the challenges of sustainability and urbanization. This has been boosted by the new digital transition in ICT. Driving such transition predominantly are big data analytics and context-aware computing and their increasing amalgamation within a number of urban domains, especially as their functionality involve more or less the same core enabling technologies, namely sensing devices, cloud computing infrastructures, data processing platforms, middleware architectures, and wireless networks. Topical studies tend to only pass reference to such technologies or to largely focus on one particular technology as part of big data and context-aware ecosystems in the realm of smart cities. Moreover, empirical research on the topic, with some exceptions, is generally limited to case studies without the use of any common conceptual frameworks. In addition, relatively little attention has been given to the integration of big data analytics and context-aware computing as advanced forms of ICT in the context of smart sustainable cities. This endeavor is a first attempt to address these two major strands of ICT of the new wave of computing in relation to the informational landscape of smart sustainable cities. Therefore, the purpose of this study is to review and synthesize the relevant literature with the objective of identifying and distilling the core enabling technologies of big data analytics and context-aware computing as ecosystems in relevance to smart sustainable cities, as well as to illustrate the key computational and analytical techniques and processes associated with the functioning of such ecosystems. In doing so, we develop, elucidate, and evaluate the most relevant frameworks pertaining to big data analytics and context-aware computing in the context of smart sustainable cities, bringing together research directed at a more conceptual, analytical, and overarching level to stimulate new ways of investigating their role in advancing urban sustainability. In terms of originality, a review and synthesis of the technical literature has not been undertaken to date in the urban literature, and in doing so, we provide a basis for urban researchers to draw on a set of conceptual frameworks in future research. The proposed frameworks, which can be replicated and tested in empirical research, will add additional depth and rigor to studies in the field. In addition to reviewing the important works, we highlight important applications as well as challenges and open issues. We argue that big data analytics and context-aware

computing are prerequisite technologies for the functioning of smart sustainable cities of the future, as their effects reinforce one another as to their efforts for bringing a whole new dimension to the operating and organizing processes of urban life in terms of employing a wide variety of big data and context-aware applications for advancing sustainability.

Keywords: Smart sustainable cities, Urban sustainability, Big data analytics, Context-aware computing, Sensors, Models, Data processing, Cloud computing, Middleware, Big data and context-aware applications

Introduction

The contemporary city is evolving into becoming computerized on a hard-to-imagine scale due to the rapid development of ICT. This is increasingly fueled by new discoveries in computer science and data science, coupled with the quick-paced ubiquity and massive use of computational and data analytics within a variety of urban domains to address the complex challenges of sustainability and urbanization facing the city. This is manifested in the ongoing large-scale design, development, deployment, and implementation of sensor technologies, data processing platforms, cloud computing infrastructures, middleware architectures, and wireless communication networks across urban environments. In parallel, the increasing convergence, prevalence, and advance of urban ICT is giving rise to new faces of cities that are quite different from what has been experienced hitherto on many scales. This is increasingly boosted by data acquisition and storage, information processing, data networking, and intelligence decision support increasingly infiltrating urban systems as operating and organizing processes of urban life. Accordingly, it has been suggested that the potential of monitoring, understanding, and analyzing the city through advanced ICT can well be leveraged in advancing its contribution to the goals of sustainable development. Indeed, the cities that are engaging on the new transition in ICT are getting smarter in how to become more sustainable (e.g. [1–8]). Besides, cities as complex systems, with their domains becoming more interconnected and their processes being highly dynamic, rely more and more on sophisticated technologies to realize their potential for responding to the challenge of sustainability and urbanization. Among these technologies are big data analytics and context-aware computing, which are rapidly gaining momentum and generating worldwide attention in the realm of smart sustainable urban development (e.g. [1, 2, 5, 6, 9–11]). Big data and context information constitute the fundamental ingredients for the next wave of urban functioning and planning, especially in relation to sustainability. There indeed is a variety of potential uses of big data analytics and context-aware computing to address urban sustainability issues from the source thanks to the deep insights, intelligent decision-making processes, and efficient services delivery enabled by data mining, machine learning, and statistics and related modeling, simulation, and prediction methods. This points to new opportunities and alternative ways to develop, operate, and plan future cities.

The prospect of smart sustainable cities is becoming the new reality with the recent advances in and integration of ICT of various forms of pervasive computing and the underlying cutting-edge enabling technologies. Smart sustainable cities typically rely on the fulfillment of the prevalent ICT visions of the new wave of computing, where everyday objects communicate with each other and collaborate across heterogeneous and

distributed computing environments to provide information and services to urbanites and diverse urban entities. The most prevalent forms of pervasive computing in relation to the urban domain are UbiComp, AmI, the IoT, and SenComp [6]. Context-aware behavior and big data capability are considered as prerequisites for realizing the novel applications pertaining to such technologies (e.g. [2, 5, 6, 9–14]). In all, the expansion of these computing trends as to the underlying technologies and applications are increasingly stimulating smart sustainable city initiatives and projects in ecologically and technologically advanced nations [3].

The past 5 years have seen extensive investments in ICT infrastructure in cities, which have improved the ability to collect and process large amounts of data throughout urban systems. Virtually every urban aspect, process, activity, and domain is now open to data collection and processing and often even instrumented for data collection and processing: operations, functions, and services in terms of management, control, optimization, enhancement, planning, and so on. At the same time, information is now widely available on external states and events such as urban trends, environmental dynamics, socio-economic patterns, and so on. This broad availability of data has led to increasing interest in methods and techniques for inferring context knowledge as well as extracting useful knowledge from various forms and sources of data—the realm of context-aware computing and data science—for knowledgeable and strategic decision-making purposes. In all, data are being produced and warehoused, the computing power is available and affordable, the environmental pressures and socio-economic concerns are alarming, and urbanization challenges are enormous.

The need to understand what constitutes the informational landscape of smart sustainable cities in terms of big data analytics and context-aware computing technologies presents an important topic and new direction of research in the field of smart sustainable cities of the future. The prominence lies in identifying the core enabling technologies and related key techniques and processes required to design, develop, deploy, and implement big data and context-aware applications for advancing urban sustainability. Topical studies tend to only pass reference to such technologies or to largely focus on one particular technology as part of big data and context-aware ecosystems in the realm of smart cities. Moreover, empirical research on the topic, with some exceptions, is generally limited to case studies without the use of any common conceptual frameworks. In addition, relatively little attention has been given to the integration of big data analytics and context-aware computing as advanced forms of ICT in the context of smart sustainable cities. This topic is a significant research area that merits attention, and this endeavor is a first attempt to address these two major strands of ICT of the new wave of computing in relation to the informational landscape of smart sustainable cities. This is to highlight that computers have become far more powerful, networks have become ubiquitous, and techniques and algorithms have been developed that can combine a large number and variety of sensors and connect various datasets to enable broader and deeper computational and analytical solutions than previously possible. The convergence of these phenomena is increasingly enabling many applications of smart computing and data science principles and big data analytics techniques.

The original contribution we make with this paper is to review and synthesize the relevant literature with the objective of identifying and distilling the core enabling

technologies of big data analytics and context-aware computing as ecosystems in relevance to smart sustainable cities, as well as to illustrate the key computational and analytical techniques and processes associated with the functioning of such ecosystems. In doing so, we develop, elucidate, and evaluate the most relevant frameworks pertaining to big data analytics and context-aware computing in the context of smart sustainable cities, bringing together research directed at a more conceptual, analytical, and overarching level to stimulate new ways of investigating their role in advancing urban sustainability. In terms of originality, a review and synthesis of the technical literature has not been undertaken to date in the urban literature, and in doing so, we provide a basis for urban researchers to draw on a set of conceptual frameworks in future research. The proposed frameworks, which can be replicated and tested in empirical research, will add additional depth and rigor to studies in the field. In addition to reviewing the important works, we highlight important applications as well as challenges and open issues.

The main motivation for this endeavor is to provide the necessary material to inform relevant research communities of the state-of-the-art research and the latest development in the field of smart sustainable cities in terms of the major technological components of their informational landscape, as well as a valuable reference for researchers and practitioners who are seeking to contribute to, or working towards, the design, development, and implementation of smart sustainable city applications. Especially, with vast amounts of urban data being now available, diverse entities in connection with every urban domain are focused on exploiting data for sustainable advantage.

The remainder of this paper is structured as follows. In “[Conceptual and theoretical background](#)” section, we introduce and describe the main conceptual and theoretical constructs that make up the study. “[Related work](#)” section presents a survey of related work in terms of the state-of-the-art research, technological developments, issues, debates, gaps, and challenges. In “[The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities of the future](#)” section, we identify, distill, describe, and discuss the core enabling technologies of big data analytics and context-aware computing necessary for the functioning of smart sustainable cities of the future, as well as touch upon some key related issues. “[The state-of-the-art analytical and computational processes](#)” section describes the state-of-the-art analytical and computational processes: data mining and context recognition, and also points out basic issues of context-aware applications. In “[Context-aware computing and its computational, technical, and urban dimensions](#)” section, we delve into the urban, computational, technical, and conceptual dimensions of context-aware computing. “[The key applications of big data analytics and context-aware computing technologies for urban sustainability](#)” section presents, enumerates, and documents the key applications of big data analytics and context-aware computing technologies for urban sustainability. “[The main scientific challenges of big data analytics and context-aware computing](#)” section identifies and enumerates the key scientific and intellectual challenges of big data analytics and context-aware computing. Finally, we provide our conclusions together with some thoughts in “[Conclusions](#)” section.

Conceptual and theoretical background

Smart sustainable cities

The concept of smart sustainable cities has emerged as a result of three important global trends at play across the world, namely the diffusion of sustainability, the spread of urbanization, and the rise of ICT [5]. As echoed by Höjer and Wangel [17], the inter-linked development of sustainability, urbanization, and ICT has recently converged under what is labelled ‘smart sustainable cities.’ Accordingly, smart sustainable cities is a new techno–urban phenomenon that materialized around the mid–2010s (e.g. [5, 6, 15–17, 72]). The idea revolves around leveraging the advance and prevalence of ICT in the transition towards the needed sustainable development in an increasingly urbanized world. Therefore, the development of smart sustainable cities is gaining increasing attention worldwide from research institutes, universities, governments, policymakers, and ICT companies as a promising response to the imminent challenges of sustainability and urbanization.

The term ‘smart sustainable city,’ although not always explicitly discussed, is used to denote a city that is supported by a pervasive presence and massive use of advanced ICT, which, in connection with various urban systems and domains and how these intricately interrelate and are coordinated, enables the city to control available resources safely, sustainably, and efficiently to improve economic and societal outcomes. Here ICT can be directed towards, and effectively used for, collecting, processing, analyzing, and synthesizing data on every urban domain and system in terms of forms, structures, infrastructures, networks, facilities, services, and citizens. The resulting knowledge can then be employed to develop urban intelligence functions and build urban simulation models for strategic decision-making associated with sustainability. Further, the combination of smart cities and sustainable cities, of which many definitions are available, has been less explored as well as conceptually difficult to delineate due to the multiplicity and diversity of the existing definitions (see [5] for an overview). ITU [18] provides a comprehensive definition based on analyzing around 120 definitions, ‘a smart sustainable city is an innovative city that uses...ICTs and other means to improve quality of life, efficiency of urban operation and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social and environmental aspects.’ Another definition put forth by Höjer and Wangel ([17], p. 10), which is deductively crafted and based on the concept of sustainable development, states that ‘a smart sustainable city is a city that meets the needs of its present inhabitants without compromising the ability for other people or future generations to meet their needs, and thus, does not exceed local or planetary environmental limitations, and where this is supported by ICT.’ This entails unlocking and exploiting the potential of ICT of the new wave of computing as an enabling, integrative, and constitutive technology for achieving the environmental, social, and economic goals of sustainability due to the underlying transformational, substantive, and disruptive effects [5, 6].

Context-aware computing

Context awareness has been defined in multiple ways depending on the application domain in terms of the number and nature of the subsets of the context of a given entity

(e.g. traffic system, energy system, healthcare system, education system, information system, human user, etc.) that can be integrated in the design and development of a given computational artifact. Originated in pervasive computing the term ‘context awareness’ is used to describe technology that ‘is able to sense, recognize, and react to contextual variables, that is, to determine the actual context of its use and adapt its functionality (and behavior) accordingly or respond appropriately to features of that context.’ ([19], p. 76). Another definition of context proposed by Dey [20] states: ‘context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.’ Context-aware applications and systems in the urban domain entail the acquisition of contextual urban data using sensors of many types to perceive situations of urban life, the abstraction of contextual urban data by matching sensory readings to specific urban context concepts, and application behavior through firing actions based on the outcome of reasoning against contextual urban information, i.e. the inferred context, to draw on Schmidt [21].

In recent years, the concept of context awareness has been expanded beyond the ambit of HCI applications to include urban applications, such as energy systems, transport systems, communication systems, traffic systems, power grid systems, healthcare systems, education systems, security systems, and so on (e.g. [1, 6, 9–11, 21–24]). Here context denotes, drawing on Chen and Kotz [25], the environmental conditions within the urban landscape that either determine applications’ behavior or in which application events occur and are interesting to different classes of users, including citizens, urban administrators, urban operators, urban authorities, and urban departments.

Big data analytics: characteristics and techniques

The term ‘big data’ is used to describe the growth, proliferation, heterogeneity, complexity, availability, temporality, changeability, and utilization of data across many application domains, which renders the processing of these data exceed the computational and analytical capabilities of standard software applications and conventional database infrastructure. In short, the term essentially denotes datasets that are too large for traditional data processing systems. Traditional analytic systems are not suitable for handling big data (e.g. Katal et al. [26]; [27]). This implies that big data entails the use of tools (classification, clustering, regression algorithms, etc.), techniques (data mining, machine learning, statistical analysis, etc.), and technologies (Hadoop, HBase, MongoDB, etc.) that work beyond the limits of the data analytics approaches that are used to extract useful knowledge from large masses of data for timely and accurate decision-making and enhanced insights. As a common thread running through most of the definitions of big data, the related information assets are of high-volume, high-variety, and high-velocity and thus require cost-effective, innovative forms of data processing, analysis, and management for enhanced decision-making and insight. While there is no canonical or definitive definition of big data in the context of smart sustainable cities, the term can be used to describe a colossal amount of urban data, typically to the extent that their manipulation, analysis, management, and communication present significant computational, analytical, logistical, and coordinative challenges. It is near on impossible to humanly make sense of or decipher big urban data based on existing computing practices. Important

to note is that such data are invariably tagged with spatial and temporal labels, largely streamed from various forms of sensors, and mostly generated automatically and routinely. Regardless of the lack of agreement about the definition of big data, there seems to be consensus that big data will lead to, in light of the projected advancements and innovations, immense possibilities and fascinating opportunities in the coming years. Moreover, big data solutions requires novel technologies to proficiently process large volumes of data emanating from multiple sources, in unprecedented quantities, and in quick time.

Big data are often characterized by a number of Vs. The main of which—identified as the most agreed upon Vs—are volume, variety, and velocity (e.g. [28, 29]). Additional Vs include veracity, validity, value, and volatility (e.g. [27]). The emphasis here is on the main characteristics of big data, namely the huge amount of data, the velocity at which the data can be analyzed, and the wide variety of data types.

The term 'big data analytics' refers commonly to any vast amount of data that has the potential to be collected, stored, retrieved, integrated, selected, preprocessed, transformed, analyzed, and interpreted for discovering new or extracting useful knowledge, which can subsequently be evaluated and visualized in an understandable format prior to its deployment for decision-making purposes (e.g. a change to or enhancement of operations, strategies, practices, and services). Other computational mechanisms involved in big data analytics include search, sharing, transfer, querying, updating, modeling, and simulation. In the context of smart sustainable cities, big data analytics denotes a collection of sophisticated and dedicated software applications and database systems run by machines with very high processing power, which can turn a large amount of urban data into useful knowledge for well-informed decision-making and enhanced insights pertaining to various urban domains, such as transport, mobility, traffic, environment, energy, land use, planning, and design.

The common types of big data analytics include predictive, diagnostic, descriptive, and prescriptive analytics. These are applied to extract different types of knowledge or insights from large datasets, which can be used for different purposes depending on the application domain. Urban analytics involves the application of various techniques based on data science fundamental concepts—i.e. data-analytic thinking and the principles of extracting useful knowledge (hidden patterns and meaningful correlations) from data, including machine learning, data mining, statistical analysis, regression analysis (explanatory modeling versus predictive modeling), database querying, data warehousing, or a combination of these. The use of these techniques depends on the urban domain as well as the nature of the urban problem to be tackled or solved.

Related work

Research on big data analytics and context-aware computing has been active for more than 2 decades, resulting in the development of many concepts, approaches, and systems spanning a large number and variety of application domains. Context-aware computing has been researched extensively by the HCI community from various perspectives, including conceptual (e.g. [19, 20, 30, 31]), theoretical (e.g. [25, 32, 33]), critical (e.g. [12, 34, 35]), and philosophical (e.g. [36]). The notion of intelligence alluded to in pervasive computing, in which context awareness has been given a prominent role, has generated a

growing level of criticism over the past decade, questioning its feasibility in terms of the inherent complexity surrounding the modeling of all kinds of situations of life (based on the cognitive, affective, emotional, social, behavioral, conversational, and physical subsets of context), as well as challenging its added value as to transforming the way people live (e.g. [19, 35–39]). The whole premise is that it is too difficult to identify and model the specifics of context in real life given their extreme subtlety, subjectivity, and fluidity. In addition, the failure of the original promise of intelligence points to a two-sided problem: the persistent elusiveness of ordinary human reasoning and knowing what people really want and the permissiveness of the definitional looseness of intelligence in terms of what can be expected of the role and scope of artificial reasoning in context-aware interaction paradigms ([38], p. 12). Context awareness research and development continues to grapple with the problem of what the intelligence in context-aware computing can stand for. Nonetheless, the notion of intelligence as enabled by context awareness capabilities has inspired a whole generation of scholars and researchers into a quest for the immense, fascinating opportunities enabled by the incorporation of computer/machine intelligence into our everyday lives, as well as a large body of research into new techniques and methods for enhancing the sensing, analysis, reasoning, inference, and modeling processes. These have been of extreme value to several other applications (industrial, urban, and organizational) than those directed for human users, in which these processes are inapt to handle the complexity of the nature and scope of inferences (context knowledge) generated by computationally constrained reasoning mechanisms and oversimplified models and on the basis of limited, uncertain, incomplete, or imperfect data collected through sensors. However, the issues stemming from these challenges are under scrutiny and investigation by the research community towards alternative directions (e.g. [39]), most notably situated intelligence which entails that the cognitive processes and behavior of a situated system should be the outcome of a close coupling between the system (agent) and the environment (user) (see [40]). This form of intelligence entails ‘assisting people in better assessing their choices and decisions and thus enhancing their actions and activities’, and the ‘quest for situated forms of intelligence is seen by several eminent scholars as an invigorating alternative for artificial intelligence research within context-aware computing’ ([19], p. 9).

However, the emphasis in this paper is on the notion of intelligence as enabled by context awareness capabilities but in relation to urban applications rather than human-inspired HCI applications. Urban intelligence in this sense involves enhancing the efficiency of energy systems, communication systems, traffic systems, transportation systems, and so on, as well as the delivery of several classes of city services (utility, healthcare, safety, learning, etc.), based mainly on the physical, situational, spatiotemporal, and socio-economic subsets of context. Especially, building and maintaining complex models of smart sustainable cities functioning in real time from routinely sensed data has become a clear prospect [2, 5, 6].

In addition, there are several thorough surveys of context modeling and reasoning in pervasive computing (e.g. [19, 39–43]). While these surveys tend to differ as regards to both technical emphases (e.g. machine learning techniques, ontological methods, and logical approaches) as well as comparative views on research into modeling and reasoning techniques applied in context awareness, the focus of the analysis and evaluation

revolves around the most common approaches into context representation and reasoning and their integration. Integrated approaches have been mainly proposed to overcome the shortcomings (information incompleteness and uncertainty, lack of expressiveness, inflexibility, lack of scalability, etc.) associated with the application of a single approach. For instance, context recognition methods based on probabilistic reasoning inherently suffer from ad-hoc static models, scalability, and data scarcity, and the ontology approach allows easy incorporation of machine understandability and domain knowledge, which provide rich expressiveness and facilitates reusability and intelligent processing at a higher level of automation [42]. However, the ontology approach falls short in handling information uncertainty and vagueness (e.g. [41]). Important to underscore is that most reviews focus on context awareness in relation to the HCI domain, while the literature on context awareness in relation to the urban domain remains scant, in particular as to large-scale applications in the context of smart sustainable cities.

Furthermore, several studies (e.g. [42–48]) have addressed middleware technologies associated with pervasive computing environments and distributed applications. Middleware plays a key role in the functionality of distributed context-aware applications, as it represents the logic glue in a distributed computing system by connecting and coordinating many components constituting distributed applications. Among the key topics addressed in the literature include architectures for pervasive context-aware services in smart spaces in terms of middleware components and prototype applications, middleware for context representation and management in pervasive computing, middleware-based development of context-aware applications with reusable components, middleware for real-time systems, and so forth. There is a need for further research in the area of middleware with regard to the use of large-scale context-aware applications as part of the informational landscape of smart sustainable cities, as well as to the modeling and management of context information in distributed pervasive applications and in open and dynamic pervasive environments.

Research on big data analytics has been active since the mid-1990s (e.g. [29, 47–51]), and several books have been written on the topic from a business intelligence perspective. As a prerequisite for realizing the IoT as an ICT vision of pervasive computing, big data analytics entails extracting useful knowledge from large masses of data for enhanced decision-making and insights pertaining to a large number and variety of domains. In recent years, the concept and application of big data analytics has been expanded beyond the ambit of business intelligence (e.g. banking, customer relationship management, targeted marketing, fraud detection, and manufacturing) to include the area of urban development as to such domains as energy, environment, transport, mobility, traffic, power grid, buildings, planning and design, healthcare, education, safety, the quality of life, socio-economic forecasting, and so on in the context of sustainability (e.g. [1, 2, 5, 6, 50–54]). Moreover, big data analytics has become a key component of the ICT infrastructure of smart sustainable cities [5, 6]. In this context, big data analytics targets optimization and intelligent decision support pertaining to the control, optimization, automation, management, and planning of urban systems as operating and organizing processes of urban life, as well as to the enhancement of the associated ecosystem and human services related to utility, healthcare, education, safety, and so on. Additionally, it targets the improvement of practices, strategies, and policies by changing them based

on new trends and emerging shifts. In all, the analytical outcomes of data mining/knowledge discovery (see [55] for an overview with relevant use cases) serve to improve urban operational functioning, optimize resources utilization, reduce environmental risks, and enhance the quality of life and well-being of citizens.

Furthermore, many reviews or surveys have been conducted in recent years on big data analytics. While they offer different perspectives on, and highlight various dimensions of, the topic, they overlap in many computational, analytical, and technological aspects. Also, they are more often than not oriented towards business intelligence (e.g. [56, 57]), and tend to put emphasis on different components of big data analytics, such as techniques, algorithms, software tools, platforms, and applications. Chen et al. [58] provide a systematic review of data mining in technique view, knowledge view, and application view, supported with the latest application cases related mostly to business intelligence. In their survey, Zhang et al. [59] explore new research opportunities and provide insights into selecting suitable processing systems for specific applications, providing a high-level overview of the existing parallel data processing systems categorized by the data input as stream processing, machine learning processing, graph processing, and batch processing. Singh and Singla [60] provide an overview of the leading tools and technologies for big data storage and processing, throw some light on other big data emerging technologies, as well as cover the business areas from which big data are being generated. In their review, Tsai et al. [61] discuss big data analytics and related open issues, focusing on how to develop a high performance data processing platforms to efficiently analyze big data and to design an appropriate mining algorithm to extract useful knowledge from big data, in addition to presenting some research directions. One of the aspects emphasized in their work is the steps (selection, preprocessing, transformation, mining, and interpretation/evaluation) of the whole process of knowledge discovery in databases (KDD), as summarized by Fayyad et al. [49]. Most of research articles focus typically more on data mining than other steps of KDD process. Tsai et al. [61] simplify the whole process into three parts (input, data analytics, and output) and seven steps (collection, selection, preprocessing, transformation, mining, evaluation, and interpretation). Katal et al. [26] provide a varied discussion covering several big data issues, challenges, tools, characteristics, sources, and best practices in relation to such applications as social media, sensor data, log storage, and risk analysis. Karun and Chitharanjan [62] deliver a whole review on Hadoop in terms of HDFS infrastructure extensions, making a comparison of Hadoop Infrastructure Extensions (HadoopDB, Hadoop++, Co-Hadoop, Hail, Dare, Cheetah, etc.) on the basis of scalability, fault tolerance, load time, data locality, and data compression. Chen et al. [63] reviews the big data background and the associated technologies, including applications and challenges (in relation to data generation, acquisition, storage, and analysis). However, the literature on the core enabling technologies of big data analytics is scant in relation to smart sustainable cities and related sustainability applications.

In addition, a number of smart city infrastructures (e.g. [53, 62–70]) have been proposed and some of them have been applied in recent years as part of case studies. These infrastructures are based on cloud computing and tend to focus on technological aspects (especially big data analytics, context-aware computing, development and monitoring, etc.), urban management, privacy and security management, or citizen services in

terms of the quality of life. There have been no research endeavors undertaken thus far to develop comprehensive or integrated infrastructures for smart sustainable cities as a holistic urban development approach. But there have been some attempts to address some aspects of environmental sustainability in the context of smart cities. For example, Lu et al. [71] propose a framework for multi-scale climate data analytics based on cloud computing. Speaking of the climate in this context, there is still a risk of a mismatch between urban climate targets and the opportunities offered by ICT solutions (e.g. [7]).

In all, despite the recent increase of research on big data analytics and context-aware computing, the bulk of work tends to deal largely with the domain of business intelligence and the field of HCI respectively in terms of techniques, algorithms, processes, architectures, platforms, and services, thereby barely exploring their relevance and role in the urban domain in terms of advancing sustainability and integrating its dimensions. Especially, a new research wave has started to focus on how to enhance smart city approaches as well as sustainable city models by combining the two urban development strategies in an attempt to achieve the required level of urban operations, functions, designs, and services in line with the goals of sustainable development (e.g. [5, 6, 72]). In particular, this holistic urban development approach emphasizes the combination of big data analytics and context-aware computing as a set of advanced technologies, techniques, processes, and applications and related platforms, architectures, and infrastructures [5, 6]. In other words, these two advanced forms of ICT are being given a prominent role in smart sustainable cities, and the evolving data-centric and context-aware approach is seen to hold great potential to address the challenge of sustainability under what is labelled 'smart sustainable cities' of the future [5, 6]. The way forward for future cities to advance sustainability and provide the quality of life to their citizens is through advanced ICT that ensures the utilization of big data and the access to contextual information (see, e.g. [1, 2, 5, 6, 11]). Local city governments are investing in advanced ICT to provide technological infrastructures supporting AmI and UbiComp, as well as to foster respect for the environmental and social responsibility [11].

The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities of the future

Like other application areas to which big data analytics and context-aware computing as advanced strands of ICT of the new wave of computing are applied, smart sustainable cities require these two related digital ecosystems and their components to be put in place, spanning different spatial scales in the form of enabling technologies necessary for designing, developing, deploying, and implementing the diverse applications that support, and ideally integrate, the dimensions of urban sustainability. As scientific and technological areas, these two strands involve low-level data collection, intermediate-level information processing, and high-level application action and service delivery (e.g. [19]). Worth noting is that as a result of the ongoing effort to realize and deploy smart sustainable cities, which are evolving due to the advance and prevalence of the enabling technologies of ICT of the new wave of computing, all the three areas are under vigorous investigation in the creation of urban environments merging the informational and physical landscapes of such cities for advancing sustainability.

There are many permutations of the core enabling technologies underlying big data analytics and context-aware computing. However, they all pertain to ICT of the new wave of computing, an integration of UbiComp, AmI, the IoT, and SenComp, which will in the near future be the dominant mode of monitoring, understanding, analyzing, and planning smart sustainable cities to improve sustainability [5, 6]. It is worth iterating that both big data analytics and context-aware computing share the same core enabling technologies because they are an integral part of ICT of the new wave of computing, as we will elucidate below. As such, they involve unobtrusive and ubiquitous sensing technologies and networks, sophisticated data management and analysis approaches, data processing platforms, cloud computing and middleware infrastructures, and advanced wireless communication technologies. These are to provide solutions in the form of useful and context knowledge for the purpose of achieving the required level of sustainability in the context of smart sustainable cities. Moreover, to have effective and successful solutions on the basis of core enabling technologies, it is required to select a number of design and development priorities in a planned manner prior to any deployment and implementation. For example, it is essential to consider flexible design, quick deployment, extensible implementation, more comprehensive interconnections, and more intelligence (e.g. [73]). However, while most of the core enabling technologies are general and apply to many application domains, others remain specific to the urban application domain, specifically to the special requirements and objectives of smart sustainable cities.

Pervasive sensing for urban sustainability

Collecting and measuring urban big data

In the emerging field of smart sustainable urban planning (e.g. [5, 6]), many scholars in different disciplines and practitioners in different professional domains advocate particularly the inclusion of ubiquitous sensing. Sensor ubiquity is a core feature of smart sustainable cities of the future, which rely on the fulfillment of the prevalent ICT visions of pervasive computing. Within the next 15 years or so, most of the data that will be used to monitor, understand, analyze, and plan the systems of smart sustainable cities will come from digital sensing of observations, transactions, and movements associated with the operating and organizing processes of urban life, which can provide readings on many environmental, social, economic, and physical phenomena. These data will be available in various forms, with temporal tags and geotags, coupled with a variety of data mining methods and data visualization techniques for displaying and presenting patterns and correlations. A large number of methods for collecting and capturing urban big data from new varieties of digital access are being fashioned and deployed across urban environments. Examples of digital access include the satellite-enabled GPS in vehicles and on citizens, traces left from online transactions processing and related demand-supply situations, online interactions (e.g. social media sites), numerous kinds of web sites, and online interactive data systems pertaining to crowd-sourcing. Satellite remote-sensing data are also becoming widely deployed, in addition to a variety of scanning technologies associated with the IoT. The convergence of these phenomena are increasingly paving the way for big data analytics (and context-aware computing) to become the dominant mode of urban analytics in relation to urban operational functioning and planning,

as well as for exploiting and extending a variety of data mining and machine learning techniques through which the generation of models will be essential in a wide range of engineering solutions for advancing urban sustainability, i.e. improving the contribution of smart sustainable cities to the goals of sustainable development. Such cities are to be monitored, understood, analyzed, and planned across several spatial levels mostly on the basis of data routinely and automatically collected by sensors. With the flourishing smart sustainable urban planning approach (e.g. [2, 5]), pervasive sensing is gaining increased momentum and prevalence as to measuring and collecting data on urban functioning and change in a new way, from the ground up, by means of powerful sensing technologies (motion, behavior, orientation, location, etc.). At present, for instance, sensing urban change from the ground up occurs 'through new sensing technologies that depend on hand-held and remote devices through to assembling transactional data from online transactions processing which measure how individuals and groups expend energy, use information, and interact' ([2], p. 492) with respect to resources. Linking and meshing data from various types of sophisticated measuring devices (RFID, NFC, GPS, laser scanners, etc.) with the automation of standard secondary sources of data and unconventional data no doubt provides a rich nexus of possibilities as to providing new and open sources of data necessary for monitoring and understanding how smart sustainable cities will function in a more effective and efficient way.

At present, the urban environment is pervaded by huge quantities of active devices of diverse kinds and forms to particularly automate routine decisions. The fabric of smart sustainable cities is expected to be, arguably, enveloped with an electronic skin, which can be sewed together and entrenched with even more advanced embedded measuring devices, information processing systems, and communication networks. These include countless intelligent sensing and computing devices and related sophisticated and dedicated techniques and algorithms, as well as widespread diffusion of wirelessly ad-hoc, mobile network infrastructures and related protocols. The primary aim is to build an entirely new holistic system which supports the following:

- The acquisition and coordination of data from multiple distributed sources.
- The management and organization of data streams.
- The integration of heterogeneous data into coherent databases and their warehousing.
- The preprocessing and transformation of data.
- The management and seamless composition of extracted models and patterns respectively.
- The evaluation of the quality of the extracted models and patterns.
- The visualization and exploration of behavioral patterns and models.
- The simulation of the mined patterns and models.
- The deployment of the obtained results for decision support and efficient service provision.

Regardless of their scales, new sensing and computing devices are projected to be equipped with quantum-based processing capacity, unlimited memory size, and high performance communication capabilities, all linked by mammoth bandwidth and

wireless (internet) connectivity as well as middleware architectures connecting several kinds of distributed, heterogeneous hardware systems and software applications [19]. All of the above is to be directed for advancing the contribution of smart sustainable cities to the goals of sustainable development. Explicitly, future urban ICT driven by the new wave of computing will result in a blend of advanced applications, services, and computational (data) analytics enabled by constellations of instruments across several spatial scales linked via multiple networks, which can provide a fertile environment conducive to monitoring, understanding, analyzing, evaluating, and planning the sustainability of future cities.

Recent advances in sensor technology have given rise to a new class of miniaturized devices characterized by advanced signal processing methods, high performance, multi-fusion techniques, and high-speed electronic circuits. The trends toward ICT of the new wave of computing, coupled with the evolving concept of smart sustainable cities, are driving research into ever-smaller sizes of sensors capable of powerfully sensing complex and varied aspects of urban life and environment at very low cost. The production of sensing devices with a low cost-to-performance ratio is further driven by the rapid development of sensor manufacturing technologies (e.g. [19]). The increasing miniaturization of computer technology is making it possible to develop miniature on-body and remote sensors that allow registering various human and urban parameters without disturbing citizens or interfering with urban activities, thereby the commonsensical infiltration of sensors into daily urban life and environment. This is instrumental in enhancing the computational understanding and data processing of human mobility, urban dynamic processes, and urban operational functioning, a process that entails analysis, interpretation, modeling, and evaluation of big data for enhanced decision-making and deep insights. The new wave of urban computing is about the omnipresence of invisible technology in urban environments and thus citizens' everyday life. Countless tiny, distributed, networked sensor devices will be invisibly embedded in cities for data collection. The research in the area of micro- and nano-engineering [74] is expected to yield major shifts in ICT performance and the way mechatronic components and devices are manufactured, designed, modeled, and implemented, thereby radically changing the nature and structure of sensing devices and thus the way cities will be monitored, understood, analyzed, probed, and planned in the near future.

Sensor-based urban sustainability mining

As part of urban reality mining (e.g. [2, 75]), urban sustainability mining, which pertains to sensing complex environmental and socio-economic systems by means of ubiquitous sensors embedded throughout urban environments, is a key determinant of how cities developing and responding to the challenge of sustainability are becoming smarter. Mining of urban sustainability depends on dedicated, powerful software applications to log urban infrastructures, spatial organizations and interactions, and mobility and travel behavior as well as ecosystem and public services. The analysis of derived large datasets helps to extract computationally complex activity, behavior, process, and environment models to identify and gain predictive insights into new forms, structures, systems, and processes as to how smart sustainable cities can increase their contribution to sustainability through enhancing urban intelligence functions for decision-making in this

regard. Therefore, sensor-based big data have enormous potential to gain new insights into and drive decisions about how sustainability can be better translated into the built, infrastructural, operational, and functional forms of smart sustainable cities across several spatial scales. Further studies in this direction are most likely to enhance mobility, transport engineering, energy engineering, planning, spatial and physical structures, and data-driven characterization of urban functioning in the context of sustainability.

Sensor technologies in context-aware computing

Sensor types and sensing areas in context-aware applications As with big data analytics, context-aware computing involves a wide variety of sensors. A sensor can be described as a device that detects or measures a physical property or some type of input from the physical environment, and then indicates or reacts to it in a particular way (e.g. [19]). The output is a signal in the form of human-readable display at the sensor location or a recorded data that can be transmitted over a network for further processing. Commonly, sensors can be classified according to the type of energy they detect as signals, and include, but are not limited to, the following types:

- Location sensors (e.g. GPS, active badges).
- Optical/vision sensors (e.g. photo-diode, color sensor, IR and UV sensor).
- Light sensors (e.g. photocells, photodiodes).
- Image sensor (e.g. stereo-type camera, infrared).
- Sound sensors (e.g. microphones).
- Temperature sensors (e.g. thermometers).
- Heat sensors (e.g. bolometer).
- Electrical sensors (e.g. galvanometer).
- Pressure sensors (e.g. barometer, pressure gauges).
- Motion sensors (e.g. radar gun, speedometer, mercury switches, tachometer).
- Orientation sensors (e.g. gyroscope).
- Physical movement sensors (e.g. accelerometers).
- Biosensors (e.g. pulse, galvanic skin response measure).
- Vital sign processing devices (heart rate, temperature).
- Wearable sensors (e.g. accelerometers, gyroscopes, magnetometers).
- Identification and traceability sensors (e.g. RFID, NFC).

While there are different ways of sensing that could be utilized for detecting various features of context, in the realm of smart sustainable cities not all the above are of use in relation to context-aware applications in terms of optimization, control, management, operation, and service delivery associated with sustainability dimensions. How many and what types of sensors can be used in relation to a given context-aware application is determined by the way in which context is operationalized (defined so that it can be technically measured and thus conceptualized) in terms of the number of the entities of context that are to be incorporated in the system based on the application domain, and also whether and how these entities can be combined to generate a high-level abstraction of context (e.g. the physical, situational, behavioral, and social dimension of

context). Too often, in relation to both citizens and urban systems, various kinds of sensors are used to detect context.

Acquisition of sensor data about citizens and urban systems (energy, traffic, transport, mobility, etc.) and their behavior and functioning is an important factor in addition to the knowledge domain for analysis of such data by data processing units. In relation to context-aware applications pertaining to citizens, data can be generated from multiple sources, including software equivalents in relation to citizens' devices, such as smartphones, computers, laptops, and other everyday objects. In other words, data are collected and captured from a variety of digital sensors as well as online interactive applications. Observed information about the citizen and urban system' states or situations in conjunction with the dynamic models for the citizen and system' relevant processes serve as input for the process of computational understanding. This entails the analysis and estimation of what is going on in the surrounding environment in the context of smart sustainable cities. Accordingly, for a context-aware application or system to be able to infer high-level context abstraction based on the interpretation of and reasoning on context information, it is first necessary to acquire low-level data from physical sensors (and other sources). Researchers from different application domains within the field of context-aware computing have investigated context recognition for the past 2 decade or so by developing a diversity of sensing devices (in addition to methods and techniques for signal and data processing, pattern recognition, modeling, and reasoning tasks). Thus, numerous types of sensors are currently being used to detect various attributes of context.

Multi-sensor data fusion and its application in context-aware applications and systems In context-aware computing, underlying the multi-sensor fusion methodology is the idea that an abstraction of context as an amalgam of different, interrelated contextual elements can be generated or inferred on the basis of information detected from multiple, heterogeneous data sources, which provide different, yet related, sensor information. Thus, sensors should be integrated to yield optimal context recognition results, i.e. provide robust estimation of context. A given dimension of the context, a higher level of the context, can be deduced by using a number of external or internal contexts as an atomic level of the context. Figure 1 illustrates multisensor fusion for context awareness.

The use of multi-sensor fusion approach in context-aware applications and systems allows gaining access simultaneously to varied information necessary for accurate estimation or inference of context. Multi-sensor fusion systems have the potential to enhance the information gain while keeping the overall bandwidth low [19]. Figure 1 illustrates a multi-sensor fusion approach.

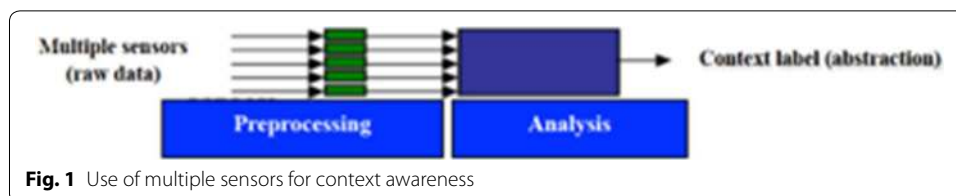


Fig. 1 Use of multiple sensors for context awareness

Wireless communication network technologies and smart network infrastructures In the context of smart sustainable cities, wireless solutions are set to proliferate in ways that are hard to imagine, as ICT continues to be fast embedded and interwoven into the very fabric of current smart and sustainable cities in terms of their systems and processes in an increasingly computerized urban society. This is a future world of pervasive computing infrastructures and communication networks. Countless sensors will use various wirelessly ad-hoc and mobile networks to provide cities with all kinds of data necessary for a wide variety of applications and services. In particular, the widespread diffusion of wireless network technologies will, as a by-product of their normal operations, enable to sense, collect, and coordinate massive repositories of spatiotemporal data pertaining to urban systems, which represent city-wide proxies for all kinds of activities and operating and organizing processes.

Also, smart networks are necessary for big data applications in terms of connecting the components and entities of smart sustainable cities, including diverse citizens' everyday objects (computers, smart phones, cars, house devices, etc.) and city infrastructures and facilities as well as urban departments, authorities, and enterprises. Such networks are intended to provide efficient means for transferring the collected data from heterogeneous and distributed sources to data warehouses where big data are to be stored, coalesced, organized, and integrated for processing and analysis in connection with intelligent decision support systems. This involves transferring responses back to the different citizens' devices and urban entities' systems for the purpose of improving different aspects of sustainability.

In relation to ICT of the new wave of computing, networking is a core enabling technology, in addition to cheap, low-power sensing and computing devices. In this context, the role of networking lies in tying hardware and software systems all together for the functioning of ubiquitous applications and services in urban areas, to draw on Bibri [19]. Accordingly, many heterogeneous components and devices across dispersed infrastructures and disparate networks need to interconnect as part of vast architectures enabling big data analytics, context-aware computing, intelligence functions, and service provisioning on a hard-to-imagine scale [19]. To put it differently, wireless network technologies are prerequisite for coordinating data as well as linking up many diverse distributed sensing devices and computing components and enabling them to interact in the midst of a variety of hardware and software systems necessary for realizing smart urban environments for advancing sustainability. Wireless technologies, especially satellite-enabled GPS, Wi-Fi, and mobile phone networks, enable to sense, collect, and coordinate massive environmental and socio-economic data representing enormous proxies for the operations, functions, and services of smart sustainable cities and thus powerful physical-environmental and socio-behavioral microscopes (e.g. [6]). This may facilitate, by means of big data analytics (data mining and database integration capabilities) which offer the prospect for adding value in terms of massive data analysis and integration, discovering the hidden patterns, correlations, and models that characterize, on the one hand, human mobility and movement as part of daily trajectories and activities of citizens and, on the other hand, physical structures and spatial organizations, which can be instrumental in strategic decision-making associated with urban sustainability planning (see [6]). In all, while pervasive sensing and computing infrastructures allow for

monitoring, understanding, and analyzing urban life in terms of infrastructure, built form, administration, and ecosystem and human services, pervasive networking infrastructures allow for collecting and coordinating extensive data in terms of how these data are stored, made accessible, and utilized.

In the context of smart sustainable cities, advanced digital networks are crucial to urban operational functioning and planning due to the interrelationships between urban components and domains that are too many to catalogue (transport, mobility, communication, building, energy, environment, water, waste, land use, healthcare, etc.). These are planned to be further heavily networked while the activities relating to these domains to be linked up. The key domains 'which currently are being heavily networked involve: transport systems of all modes in terms of operation, coordination, timetabling, utilities networks which are being enabled using smart metering, local weather, pollution levels and waste disposal, land and planning applications, building technologies in terms of energy and materials, health information systems in terms of access to facilities by patients the list is endless. The point is that we urgently need a map of this terrain so that we can connect up these diverse activities' ([2], p. 493). Especially, the evolving technourban contexts are opening spaces for smart sustainable initiatives in domain networking at current times of tension as alternative trajectories are actively being sought due to the challenge of sustainability, which entails creating innovative solutions that further facilitate collaboration among urban domains and hence integrate urban systems.

In parallel, the aim of emerging technological platforms such as UbiComp, AmI, the IoT, and SenComp is to orchestrate and coordinate the various computational entities in the informational landscape of smart sustainable cities and merging it with their physical landscape into an open system that helps diverse urban entities cope with and plan their activities in relation to improving sustainability. Besides, the growing depth, scale, and complexity of urban networks in terms of both domains and technological infrastructures call for developing and coordinating such networks and enhancing their digital capabilities in ways that increase and sustain the contribution of smart sustainable cities to the goals of sustainable development. Advanced wireless technologies are extremely placed to initiate this development and coordination. Moreover, with their ever-growing volume, variety, velocity, and timeliness, data on the state of urban networks as built artifacts as well as on that of their use as part of urban activities and processes provide enormous potential to improve urban operational functioning and planning (see, e.g. [1, 2]) in terms of sustainability, efficiency, and the quality of life by exploiting the analytical power of big data for deep insights and enhanced decision-making. To effectively use these data when implementing big data applications in smart sustainable cities requires fostering these data by advanced wireless technologies, especially in relation to real-time applications. The rationale is that such applications entail that the data from distributed sources should be aggregated and fused prior to being transferred in real-time to cloud computing infrastructures or data processing platforms for stream processing and decision-making. Important to note is that the aggregation and fusion should be carried out in ways that enable data to remain reliable, accurate, and correct for more effective results and thus beneficial knowledge in terms of decision-making processes. This is in turn of critical importance for maintaining the quality and performance of real-time big data applications in terms of decision-making processes [76].

Data processing platforms for big data analytics

There is a variety of available data processing platforms for big data analytics, which provide the stream processing required by real-time big data applications in relation to various urban domains. Therefore, data processing platforms are a key component of the ICT infrastructure of smart sustainable cities of the future with respect to big data applications. Among the leading platforms for big data storage, processing, and management include Hadoop MapReduce, IBM Infosphere Streams, Stratosphere, Spark, and NoSQL-database system management (e.g. [1, 28, 53, 60, 62, 63]). These platforms work well on cluster systems to meet the requirements of big data applications for smart sustainable cities; entail scalable, evolvable, optimizable, and reliable software and hardware components; and provide high performance computational and analytical capabilities (namely selection, preprocessing, transformation, mining, evaluation, interpretation, and visualization), in addition to storage, coordination, and management of large datasets across distributed environments. As ecosystems, they perform big data data analytics related to a wide variety of large-scale applications intended for different uses associated with the process of sustainable urban development, such as management, control, optimization, assessment, and improvement, thereby spanning a variety of urban domains and sub-domains. In all, they are prerequisite for data-centric applications for smart sustainable cities of the future. The focus on Hadoop MapReduce is justified by the suitability of its functionalities as to handling urban data as well as to its advantages associated with load balancing, cost effectiveness, flexibility, and processing power compared to other data processing platforms. Hadoop MapReduce has become the primary big data storage and processing system given its simplicity, scalability, and fine-grain fault tolerance [59]. For example, it is capable of handling all data types collected from multiple sources to derive actionable insights. However, it does pose issues regarding processing efficiency, rigid data flow, and low-level abstraction. NoSQL (e.g. Mongo DB and Cassandra) is also fast becoming a choice for storing and sorting structured and unstructured data and cluttering them with greater efficiency and scalability.

Cloud computing for big data analytics: characteristic features and benefits

Big data analytics can also be performed in the cloud. This involves both big data platform as a service (PaaS) and infrastructure as a service (IaaS) (e.g. [77]). Having attracted attention and gained popularity worldwide, cloud computing is becoming increasingly a key part of the ICT infrastructure of both smart cities and sustainable cities (e.g. [1, 5, 7, 53, 66, 67, 71]) as an extension of distributed and grid computing due to the prevalence of sensor technologies, storage facilities, pervasive computing infrastructures, and wireless communication networks. Especially, most of these technologies have become technically mature and financially affordable by cloud providers. By commoditizing services, low cost open source software, and geographic distribution, cloud computing is becoming increasingly an attractive option [78].

Big data analytics is associated with cloud computing (e.g. [1, 77]; [79]), an Internet-based computing model that is increasingly seen as the most suitable solution for highly resource intensive and collaborative applications as an on-demand network access to a shared pool of computing resources (memory capacity, energy, computational power, network bandwidth, interactivity, etc.) [1, 7, 80]. This entails that computer-processing

resources, which reside in the cloud, are virtualized and dynamic, which implies that only display devices for information and services need to be physically present in relation to urban domains where diverse stakeholders (administrators, planners, landscape architects, sustainability strategists, authorities, citizens, etc.) can make use of software applications and services to improve sustainability. Such stakeholders can access cloud-based software applications through a web browser and a lean client (a computer program that depends on its server to fulfill its computational roles) or mobile devices while software tools and urban data of all kinds are stored on servers at a remote location. Indeed, cloud computing model is based on hosted services in the sense of application service provisioning running client server software locally. In this respect, smart sustainable city applications pertaining to transport, traffic, mobility, energy, public health, civil security, education, and so on reside 'in the cloud' and can be accessible per demand. Moreover, the software development platform can be offered in a public, private, or hybrid network, where the cloud provider manages the platform that runs the applications and relieves the cloud clients from the burden of securing dedicated platforms, which would otherwise be very demanding and costly in terms of resources and time. The cloud clients can accordingly benefit from tested, scalable, reliable, and maintainable platforms offered by the cloud provider. Another advantage involves service process optimization through advanced functionalities of software development platforms, namely flexibility, interoperability, reusability, scalability, and cooperation. There is also a great opportunity to slash or minimize energy consumption associated with the operation of ICT infrastructure, especially when it comes to large-scale deployments like in the case of smart sustainable cities as to different departments and service agencies. Beloglazov et al. [81] develop policies and algorithms that aim at increasing energy efficiency in cloud computing. Energy consumption is way too lower than if all urban entities have their own software development platforms. These are indeed shared by multiple users as well as dynamically reallocated per demand. This approach maximizes the use of computational power and reduces energy usage and thus mitigate GHG emissions associated otherwise with powering a variety of functions as well as data centers dispersed throughout the departments and service agencies of smart sustainable cities. Whether public or private, the cloud provider includes the cloud environment's servers, storage, networking, and data center operations. This implies that the cloud provider has the actual energy-consuming computational resources; users or clients can simply log on to the network without installing anything, thereby curbing energy usage and making the best of the available computational power. Energy efficiency in cloud computing can result from energy-aware scheduling and server consolidation [82]. Mastelic et al. [83] provides a survey on energy efficiency in cloud computing. Also, cloud computing is seen as a form of green computing, especially if it is based on renewable energy like solar panels. It has other intuitive benefits because it relies on sharing of resources and maximizing the effectiveness of the shared resources, thereby reducing the costs otherwise incurred by ICT operations as to human, technical, and organizational resources. In cloud computing, supercomputers in large data centers as a distributed system of many servers are used to deliver services in a scalable manner as well as to enable the storage and processing of vast quantities of data. Cloud computing offers great opportunities for streamlining data processing [84]. In all, cloud computing constitutes an efficient

and elegant solution in terms of facilitating the huge demand for computing resources associated with big data analytics for decision-making processes in relation to the operational functioning and planning of cities in terms of sustainability. Through the use of cloud computing, smart sustainable cities can accordingly have higher possibilities to perform more effectively and efficiently thanks to the advanced technological features underlying the functioning of cloud computing model.

In addition, cloud computing performs service-oriented computing. In this regard, it can rapidly process large and complex data produced from urban activities and simultaneously serve citizens in relation to healthcare, education, housing, utility, and so on, providing a kind of integrated and specialized center for information services to both the general public and urban departments across various urban domains. In light of this, with reference to smart sustainable cities, cloud computing has the ability to run smart applications on many connected computers and smartphones at the same time for different purposes associated with increasing sustainability performance.

In sum, among the key advantages provided by cloud computing technology include cost reduction, location and device independence, virtualization (sharing of servers and storage devices), multi-tenancy (sharing of costs across a large pool of cloud provider's clients), scalability, performance, reliability, and maintenance. Therefore, opting for cloud computing to perform big data analytics in the realm of smart sustainable cities remains thus far the most suitable option for the operation of infrastructures, applications, and services whose functioning is contingent upon how urban domains interrelate and collaborate, how efficient they are, and to what extent they are scalable as to achieving and maintaining the required level of sustainability.

Middleware infrastructure for context-aware computing: characteristics and functions

Middleware infrastructure is associated with pervasive computing environments and distributed applications. These encompass UbiComp, AmI, and SenComp environments and applications. Middleware infrastructure (e.g. [44, 47, 48]) plays a key role in the functionalities of complex distributed applications, including context-aware applications. Thus, context-aware computing, which is associated with UbiComp, AmI, and SenComp, requires middleware infrastructure to operate. This infrastructure can also run on cloud computing [platform as a service (PaaS) and infrastructure as a service (IaaS)]—i.e. cloud middleware.

Middleware infrastructure represents the logic glue in a distributed computing system, as it connects and coordinates many components constituting distributed applications. This occurs, more specifically, 'in the midst of a variety of heterogeneous hardware systems and software applications needed for realizing smart environments and their proper functioning. To put it differently, in order for the massively embedded, distributed, networked devices and systems, which are invisibly integrated into the environment, to coordinate require middleware components, architectures, and services. Middleware allows multiple processes running on various sensors, devices, computers, and networks to link up and interact to support (and maintain the operation of context-aware applications needed by citizens and urban entities to cope with and perform their) activities wherever and whenever needed.' ([19], p. 50). Indeed, it is the ability of multiple, heterogenous hardware and software systems to cooperate, interconnect, and

communicate seamlessly across disparate networks that create smart environments rather than just their ubiquitous presence and massive use. In the context of smart sustainable cities, such systems in their various forms (e.g. sensors, smartphones, computers, databases, data warehouses, application integration methods, application servers, web servers, context management systems, and messaging systems) are highly distributed, interoperable, and dynamic, involving a myriad of embedded devices and information processing units 'whose numbers are set to increase by orders of magnitude and which are to be exploited in their full range to transparently provide services on a hard-to-imagine scale, regardless of time and place' [19]. This in turn allows for the functioning of context-aware applications across the diverse domains of smart sustainable cities.

There are different approaches to conceptualizing middleware. According to Schmidt [46], middleware consists of the following four distinct layers based on their intended functionality:

- (1) Host-infrastructure middleware
- (2) Distribution middleware
- (3) Common middleware services
- (4) Domain-specific middleware services.

Another conceptualization of middleware entails a common multilayer architecture that provides particular functionalities and constitute the basis for upper layers of more abstraction. It includes the following components:

- Infrastructure and communications (messaging services) pertaining to entities of the upper layer
- Services and agents related to semantic descriptions
- Middleware services concerned with the software environment
- Intelligence associated with the coordination of application actions and involving a number of devices in the environment.

As regards to some of its characteristic features compared to cloud computing, middleware-based architectures entail reusable software infrastructure that resides between the application programs (in this case context-aware applications) and the underlying hardware and operating systems. That is to say, middleware sits between the kernel and applications. Incidentally, the functionality of network protocol stacks (TCP/IP) was previously provided separately by middleware, but nowadays is integrated in every operating system. Moreover, middleware simplifies and supports the development of complex distributed applications, using such tools as web servers, application servers, messaging systems, and content management systems. These applications collaborate with, or leverage services from, other disparate applications that are systematically tied using methods of application integration. In addition to handling the distribution and heterogeneity of computing resources associated with the logic of context-aware applications in this context, middleware is intended to bridge the gap between the applications and the underlying lower-level hardware and software infrastructure to ensure and boost coordination, cooperation, interconnection, dynamicity (e.g. sensors join and leave Aml

infrastructure in a dynamic fashion), and interoperability of the different components of distributed applications (e.g. [19, 45, 85]). These functionalities are in fact necessary for supporting scalable systems as well as highly heterogeneous and distributed components, such as agents and services. In relation to this, middleware support and deploy data-centric distributed systems (e.g. network-monitoring systems, sensor networks, and dynamic web) whose ubiquity creates large application networks spreading over large geographical areas [19]. Especially, AmI and UbiComp infrastructures are highly dynamic and involve high degree of heterogeneity (e.g. [86]). As to interoperability, for instance, context-aware applications run on different operating systems, thereby the role of middleware in enabling interoperability between applications by supplying services for exchanging data in a standard way. Indeed, in the realm of context-aware computing, which entails distributed processing in the sense of multiple applications being connected to create larger applications over a network, middleware provides services beyond or more than those available from the operating system of these applications to enable the various elements of the underlying distributed system to communicate and manage data, thereby serving as a kind of a software glue. Therefore, distributed processing is empowered by middleware for transferring signals from various sources and for realizing information fusion from multiple perceptive components [44].

Middleware and cloud computing infrastructures differ in their technical details as to how they provide application services and which kind of services they are concerned with, as well as in the characteristic features of their operation and complexity. Yet, they denote computing models where machines in large data centers across distributed environments can be used to deliver a variety of services and meet the needs of different urban constituents in terms of the use of big data and context-aware applications for improving sustainability. Hence, both are prerequisite for in the operation of smart sustainable cities. This is anchored in the underlying assumption that big data and context-aware applications are an integral part of ICT of the new wave of computing, and smart sustainable cities typically rely on the fulfillment of its underlying visions.

Big data management

Given the volume, variety, and velocity characterizing big data, effective and suitable big data management tools are extremely important to ensure a useful utilization of big data in terms of analytics and the related results and inferences. Accordingly, as smart sustainable cities involve the generation of large, varied, and time-based data pertaining to such urban domains as transport, traffic, mobility, energy, environment, land use, healthcare, education, and so on, huge data management capabilities are necessary to allow to make sense of these data. Especially the field of urban sustainability necessitates these domains to be interrelated and coordinated to collaborate and inform one another. In this respect, the urban data are generated on a regular basis in the form of massive repositories, i.e. huge amounts of data on environmental and socio-economic aspects of urban areas, which provide a powerful microscope of, and a real-time view of what is happening in, the city as to sustainability performance across several spatial scales and over multiple temporal scales. A successful utilization of these valuable data in smart sustainable cities requires advanced big data management tools and methods. This entails the development and implementation of scalable and powerful architectures,

best practices, and dedicated computational processes for properly managing data life-cycle throughout various phases of data use, particularly in terms of addressing the issue of variety and velocity, i.e. recognizing their different formats and sources as well as organizing, cataloguing, classifying, and controlling all classes and structures of data. In addition, for smart sustainable city applications, big data management should provide tools for scalable handling of massive data to serve real-time applications and support offline applications (see [1]). For the interested reader, there are several studies that have addressed the topic of big data management (e.g. [84–89]) in terms of concepts, approaches, techniques, and challenges.

Advanced big data analytics techniques and algorithms

In smart sustainable cities, big data analytics should involve highly sophisticated and dedicated techniques and algorithms (data mining, machine learning, statistics, database query, etc.) that can perform complex computational processing of data for timely and accurate decision-making purposes. Traditional techniques and algorithms are inadequate for handling big data associated with smart sustainable city applications due to their high-volume, high-variety, and high-velocity. Urban big data necessitate high speed processing power and high performance to obtain useful results necessary to enhance decision-making pertaining the urban operational functioning and planning of smart sustainable cities. Therefore, existing techniques and algorithms need to be improved in ways that can handle the extreme volume of data, the wide variety of data types, and the time constraints on data processing. In particular, data mining algorithms and techniques are by far unfit for handling big data because they are designed to deal with limited and well-defined datasets (e.g. [90]). In the context of smart sustainable cities, such techniques and algorithms need to be exploited, enhanced, and extended in order to yield the desired outcomes in terms of extracting the useful knowledge (patterns and correlations) necessary for improving sustainability performance (see, e.g. [2, 5, 6]). Alternative or novel solutions in this regard are required to be designed with more scalability and flexibility to handle dynamic and real-time aspects of big data applications for smart sustainable cities, among other things. Moreover, they are to operate as an integral part of cloud computing (PaaS) and thus collaborate across diverse networks for aggregating, fusing, processing, analyzing, and visualizing data collected from countless sensing devices from multiple sources, stored in massive repositories, and coordinated through smart networks. In other words, they need to work effectively across disparate networks, dispersed infrastructures, distributed geographical locations, and heterogeneous computing environments, as well as to be capable of operating in highly scalable and dynamic settings, to reiterate. New approaches to storing, managing, coordinating, and analyzing big data, in particular in relation to smart sustainable city applications should rely on advanced artificial intelligence programs and machine learning techniques. This is in contrast to loading big data into traditional relational databases for analysis, a process that relies on data schema and is time consuming and computationally expensive. For a detailed account of big data analytics techniques and algorithms from a general perspective, the interested reader might want to read Provost and Fawcett [57]. For a relevant account of data mining techniques and algorithms, the interested reader can be

directed to Barbi [55]. Also, Chen et al. [58] provide a thorough survey on data mining techniques and algorithms.

Privacy mechanisms and security measures

It is highly important to ensure that all technological components associated with big data and context-aware applications for smart sustainable cities are supported by security measures and privacy mechanisms. It is essential to control big data [91] and context data [92]. Massive repositories of urban data are at stake, and failure to protect these data will pose risks and threats to the functioning of smart sustainable cities as well as to the safety and well-being of their citizens on several scales. Therefore, security measures and privacy mechanisms should be at the core of urban policy and governance practice associated with the design, development, deployment, and implementation of big data and context-aware applications within smart sustainable cities. Any attempt of an unauthorized access, malicious attack, or abuse of information on citizens, infrastructures, networks, and facilities can compromise the integrity of such applications and related services. Smart sustainable cities generate colossal amounts of data on virtually every urban process, which are to be stored, processed, and shared. Urban environments 'are now being continually forged and re-forged in (sensorial), informational, and communicative processes. It is a world where...cities think of us, where the environment reflexively monitors our behavior' ([93], p. 1), including whether and the extent to which we behave in a sustainable way through the activities we perform in cities.

However, it is commonly held views that the more cities think of and know about us and technologies monitor urban environments and collect information, 'the larger becomes the privacy threats, and the larger...the networks, the higher the security risks' ([92], p. 218). When sensing, computing, and networks become ubiquitous, 'when everything is embedded with intelligence and connected to everything else via the internet and other networks, the threats and vulnerabilities will become even greater than they are nowadays' ([92], p. 218). There is a need for technological safeguards as a response to the risks posed by the emerging urban trends of big data analytics and context-aware computing. Clear guidelines, recommendations, and requirements must be identified and put in place in relation to big data and context-aware applications for smart sustainable cities. Among the privacy mechanisms proposed thus far for addressing the issue of privacy include 'anonymity, pseudonymity, unlinkability, and unobservability'; yet they need to be 'fully developed, evaluated, and instantiated in their operating environment to test their performance—how well they work' [92]. Big data and context-aware applications for smart sustainable cities require the development of more robust, if not unconventional, privacy-protecting safeguards by considering the most likely ways through which the information from different urban domains can be leaked and breached. As regards to the security, the scientific challenges 'include methods supporting the evaluation of risk exposure..., security design principles to enable control of the risk exposure, methods for...security analysis, security of big (and context) data..., secure cloud of physical and smart things, cyber physical systems security, lightweight security solutions, authentication and access control..., identification and biometrics..., cyber-attacks detection and prevention, and so on' ([92], p. 223–4). While information security risks are of diverse nature, including 'modification, destruction, theft, or lack of availability of

computer assets such as hardware, software, data, and services' ([94], p. 442), integrity and confidentiality—i.e. protection of information from modification and unauthorized use—should be more of focus as categories of security threats in the ICT of the new wave of computing networks than in the traditional networks [92]. This is due to the fact that there are 'possible conflict of interests between communicating entities; network convergence; large number of ad-hoc communications; small size and autonomous mode of operation of devices; and resource constraints of mobile devices' ([95], p. 50). Of critical importance, nevertheless, is to develop a new security paradigm which supports advanced features of context-aware technology, as conventional password entry schemes using traditional input devices have proven to be vulnerable to attacks. To address these issues, new research endeavors are focusing on such new techniques as authenticating with minds; pointing and selection using gaze and keyboard; and gaze-based user authentication [92]. In relation to this, Wright et al. [95] suggest some research directions, including 'improving access control methods by multimodal fusion, context-aware authentication and unobtrusive biometric modalities', and 'increasing security by detection of unusual patterns.'

Standards and open standards

It is important to follow standards when it comes to data integration to make sense of data proliferation as well as to ensure data quality. Standard rules are also needed for evaluating the accuracy and correctness of data and for dealing with such issues as uncertainty and incompleteness of data, especially in relation to real-time big data (and context-aware) applications which require the data to be described using advanced models of the very urban systems that that data are associated with in case of missing and inconsistent data. It is of equal importance to set and comply with standard rules with respect to new applications for advancing urban sustainability to achieve seamless integration between the available urban systems (in terms of infrastructural, physical, operational, and functional forms) and the introduced big data (and context-aware) applications across diverse urban domains. In this regard, the way forward is to carry out a thorough investigation of the different urban entities and actors as well as the infrastructure, built form, administration, and ecosystem and human services as to their operation as urban systems to strategically assess the benefits of new solutions and the readiness of urban stakeholders to join any smart movement associated with improving urban sustainability. In light of such investigation, new practices, regulations, and standard models of design and rules can be developed for big data and context-aware applications for smart sustainable cities.

Concerning other areas related to big data and context-aware applications for smart sustainable cities, it can be advantageous to pursue open standards for designing and implementing solutions with respect to various urban domains, as such applications involve large-scale and heterogeneous data systems. The rationale behind open standards in this respect is to provide some flexibility for scaling up, upgrading, improving, and maintaining applications for smart sustainable cities, as new challenges are most likely to emerge and thus operative solutions may become inadequate to handle potential complexities and difficulties as to translating sustainability into the built, infrastructural, operational, and functional forms of such cities.

The state-of-the-art analytical and computational processes

The process of data mining

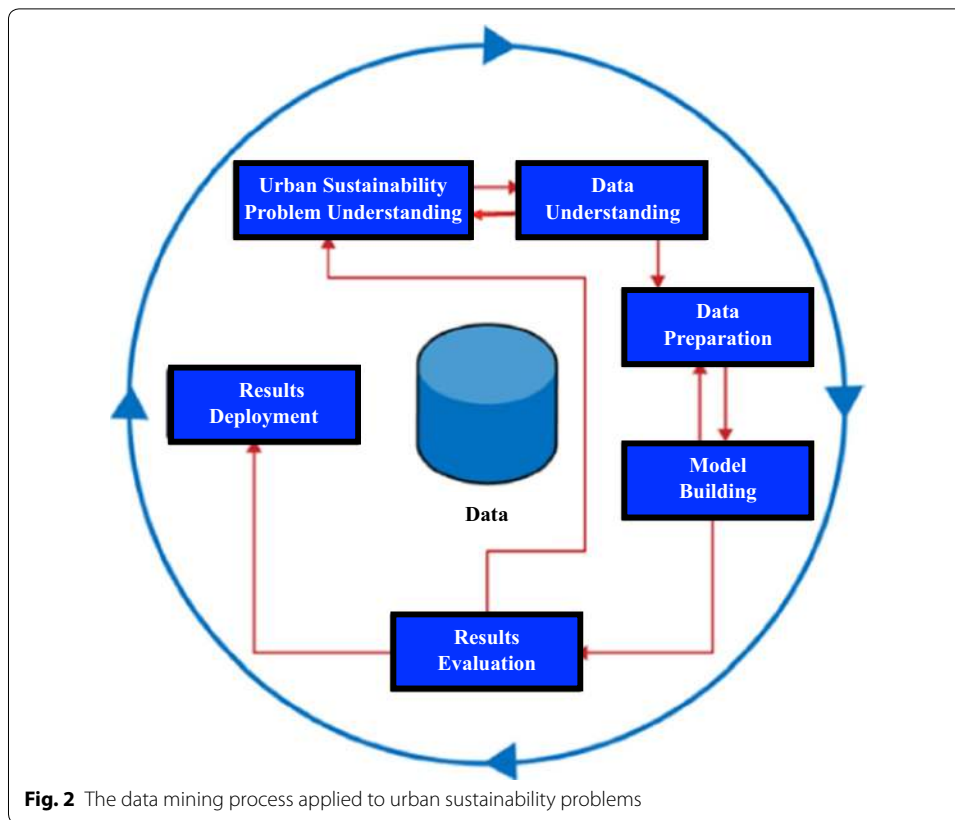
One of the fundamental concepts of data science is the automated extraction of useful knowledge from large masses of urban data to solve urban sustainability problems (physical, environmental, social, and economic) related to diverse urban domains, which can be treated systematically by following a set of reasonably well-defined stages, i.e. several codifications of the process of data mining, most notably the cross industry standard process for data mining (CRISP-DM). This process provides a framework to structure urban thinking about data analytics problems related to different dimensions of sustainability, as in smart sustainable urban planning practice, it is important to devise analytical solutions based on careful analysis and evaluation of the relevant problems using high-powered analytical and evaluative tools, as well as creativity, common sense, and specialized knowledge. Therefore, structured thinking about urban data analytics is of significance to supporting decision-making processes concerning different dimensions of sustainability within almost all urban domains. A codification of the data mining process, drawing on Shearer [50, 51], involves the following steps:

1. Urban sustainability problem understanding
2. Data understanding
3. Data preparation
4. Model building
5. Results evaluation
6. Results deployments.

For a detailed description and discussion of these steps, including an account of data mining techniques, algorithms, and tasks based on supervised and unsupervised methods, the interested reader can be directed to [55]. Also, Chen et al. [58] provide a systematic review of data mining in technique view, knowledge view, and application view, supported with the latest application cases related mostly to business intelligence. A well-understood process of urban analytics places a structure on the problems pertaining to urban sustainability, allowing reasonable consistency, repeatability, and objectiveness. The codification of the data mining process espoused here is adapted from CRISP-DM [50, 51], as illustrated in Fig. 2. The deep technical details of the sub-processes of the data mining process and how they relate to urban domains and subdomains in the context of sustainability dimensions is beyond the scope of this paper.

Applicable to the domains of smart sustainable cities, the process of data mining emphasizes the idea of iteration. This implies that solving a particular urban sustainability problem may require going through the process more than once.

The objective of the process of data mining is to discover new knowledge in large masses of urban data pertaining to different sustainability dimensions to improve the environmental and socio-economic performances of smart sustainable cities. Accordingly, such process is concerned with solving problems related to the spatial, physical, infrastructural, operational, and functional forms of such cities in the context of



sustainability. For an optimal outcome in the case of spatial data mining, for example, it is invaluable to integrate different methods (spatial analysis, spatial statistics, fuzzy logic, probability theory, cluster analysis, etc.), especially these methods are not mutually exclusive [65]. As to the spatial data, they include such features as spatial, massive temporal, massive multidimensional, and complex [96].

The process of context recognition

Smart sustainable cities as intelligent urban environments provide important contextual information that should be exploited in such way that the intelligent actions taken by diverse context-aware applications related to both citizens and systems within such environments must be relevant to the current context. Context recognition is the process whereby various contextual features of the urban environment (physical, environmental, spatiotemporal, socio-economic, situational, and behavioral) are detected, monitored, analyzed, and interpreted to generate relevant inferences through reasoning processes. It encompasses many different tasks, namely context modeling in terms of knowledge representation and reasoning, contextual features monitoring, data processing and pattern recognition, and intelligent decision-making and action-tacking. The whole process involves the following steps:

1. Create computational models of contexts pertaining to the citizen and urban system (energy, traffic, transport, building, mobility, healthcare, utility, etc.) in a way that allows software agents/systems to perform reasoning and manipulation
2. Monitor and capture relevant contextual features depending on the application domain
3. Process observed information (low-level contextual data) through aggregation and fusion to generate a high-level abstraction of context
4. Decide which algorithm or a set of algorithms to use. This is based on the way in which contexts as related to some aspects of sustainability are modeled, represented, and reasoned about (often based on a hybrid modeling approach)
5. Carry out pattern recognition and generate inferences
6. Make a timely decision and take the most appropriate action accordingly.

These steps can be applicable to the recognition of different dimensions of context, e.g. situational, spatiotemporal, behavioral, and environmental. Researchers from different application domains in the field of context-aware computing have investigated context recognition for the past 2 decades by developing and enhancing a variety of approaches, techniques, and algorithms in relation to a wide variety of context-aware applications [19].

Basic issues of context-aware applications

Placing reliance on context knowledge through recognizing, interpreting, and reasoning about contextual data from sensors to infer a high-level abstraction of a context and react to it by performing application actions to support citizens or operate urban systems is a process that is non-trivial and often extremely difficult to realize, regardless of the application domain in the context of smart sustainable cities. A central concern, in particular, is the issue of linking the perceived context (observed contextual data) to application behaviors—firing context-dependent actions. Drawing on Bibri [19], there are four basic issues related to generic contextual model that are central and necessary to be addressed to create context-aware applications pertaining to smart sustainable cities:

1. *Perception as precondition.* To create context-aware applications it is inevitable to provide them with perception capabilities as to various types of urban context, including the domains of sensing, abstraction, and modeling (conceptualization and representation)
2. *Finding and analyzing situations pertaining to citizens and urban systems that are of relevance to context-aware applications.* When such applications are based on some kind of implicit interaction, it becomes a central problem to find the situations that should have an effect on their behavior
3. *Abstracting from situations to context.* Describing a situation is already a high-level abstraction of context. To describe what should have an influence on application classes of situations have to be selected which will influence the behavior of context-aware applications

4. *Linking context to behavior.* To describe application classes of situations and in a more abstracted way, contexts must be linked to actions carried out by context-aware applications.

Context-aware computing and its computational, technical, and urban dimensions

Context awareness technology for urban sustainability

The focus in this paper is on the physical, situational, spatiotemporal, and socio-economic features of the citizen context as to the interactive applications pertaining to healthcare, education, learning, security, accessibility, utility, and so forth, as well as on the physical, operational, environmental, and spatiotemporal aspects of the urban system context as to the operating and managing processing relating to energy, transport, traffic, mobility, power grid, and so on. Thus, the capabilities of context awareness technology directed for enhanced decision-making and service delivery processes are of high relevance to smart sustainable cities in terms of urban analytics associated with sustainability. The adaptive and responsive features of context-aware applications and systems constitute forms of urban intelligence in its wider sense. One of the cornerstones of ICT of the new wave of computing (UbiComp, AmI, and SenComp) is the intelligent behavior of applications and systems in response to the different contexts of citizens and urban systems (situations, events, locations, settings, behaviors, activities, etc.).

Context-aware computing is becoming increasingly a key component of the infrastructure of smart sustainable cities (e.g. [5, 6, 11, 97] and future smart cities (e.g. [98]). Having access to context information in smart sustainable city applications plays a key role in supporting decision-making processes pertaining to sustainability (e.g. [1, 5, 6, 11]). As one of the prerequisites technologies for enabling ICT of the new wave of computing and for realizing the ICT visions of pervasive computing, context awareness aims to 'support human action, interaction, and communication in various ways wherever and whenever needed' ([19], p. 1) by enabling sensorily and computationally augmented urban environments to provide the most efficient services to citizens and intelligent support to urban actors within a variety of settings (e.g. [5, 6]). Entailing a set of advanced computational functionalities, context-aware applications are able to, by relying on context knowledge, control over processes and automate operative tools as well as provide services and support decision-making needs, thereby exhibiting intelligent behavior. In short, context awareness technology is associated with control, automation, optimization, and management, as well as with the adaptation of services. The system behavior's adaptation is based either on pre-programmed heuristics or real-time reasoning capabilities. The purpose of machine learning and reasoning is to monitor the behaviors of urban systems and citizens and the changes in their environment using sensors of many types to generate inferences (high-abstractions of contexts) based on reasoning mechanisms, and then use physical actors (actuators) or application actions to react and pre-act accordingly in ways that are more constructive in terms of enhancing urban operations, functions, and services in line with the goals of sustainable development. The widespread adoption of diverse sensors within cities provides interactions through opportunistic and people-centric sensing [99, 100]. In this regard, context-aware applications can monitor what is happening in urban environments, analyze, interpret, and

respond to them in a variety of ways—be it in relation to smart energy, smart street lights, smart traffic, smart transport, smart mobility, smart education, smart health-care, or smart safety—across several spatial and temporal scales (e.g. [4]). It is becoming increasingly evident that smart urban environments based on context-aware technologies—especially within smarter cities as future forms of smart cities, namely ambient cities, sentient cities, ubiquitous cities, Internet-of-everything, and real-time cities (e.g. [2, 5, 9, 10, 98–105]; Kyriazis et al. 106)—which can support sustainable urban living in various ways through intelligent service provision and decision support, will be commonplace in the near future.

Context awareness and its feasibility in urban intelligence

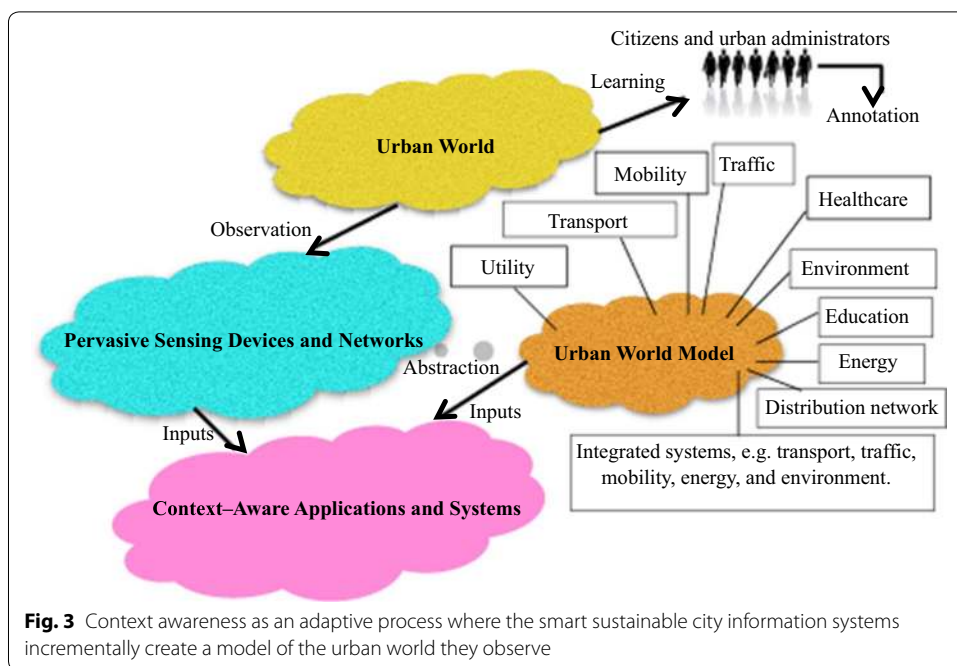
Being of prominence and significance to ICT of the new wave of computing in the context of smart sustainable cities [5, 6, 55], context awareness is grounded in the idea that it becomes possible to, through the use of artificial intelligence, detect, monitor, analyze, and model situations of urban life in ways that enable a wide variety of applications and systems to adaptively and proactively take more knowledgeable actions. This relates to the prevailing idea in AmI, UbiComp, and SenComp that the environment can sense and intelligently react to contextual features of urban systems and citizens in the realm of smart sustainable cities of the future [5, 6]. This constitutes also a core characteristic aspect of what is labelled ‘smarter cities’ [5, 6]. However, in contrast to the context-aware applications directed to humans (cognitive, emotional, social, and conversational features of context), where the feasibility issues are essentially linked with the inherent complexity surrounding the modeling of situations of human life [19], the context-aware applications directed for urban systems are not expected to involve such feasibility issues since it is possible to model situations of urban systems as intelligent operating and organizing entities. However, these models should be sophisticated enough and well suited to include the evolution of urban phenomena under major changes. In other words, new models should focus on addressing the problems of sustainability and urbanization. Investigating approaches to context modeling (representation and reasoning techniques) constitutes a large part of a growing body of research on the use of context awareness as a technique for developing context-aware applications and systems that can adapt to and act autonomously on behalf of citizens and urban systems. ICT of the new wave of computing will enhance the quality and speed of urban models in terms of both computational capabilities and structures, as well as the scope of the evolving issues that new modeling approaches can address [2]. Moreover, with the huge potential of artificial intelligence programs and machine learning and data mining techniques being under development for smart sustainable cities of the future, novel methods will emerge to address the issues related to most of the reasoning processes suggested for urban scenarios as to the complexity of generating inferences based on relatively limited and imperfect sensor data. Explicitly, new development are expected to provide ways of filling in missing data and addressing data uncertainty and vagueness using dynamic models of the very systems that these data pertains to.

Sensor observations and dynamic urban models

Context-aware applications and systems involve sensors to monitor or observe different contextual aspects of intelligent entities in urban environments, analyze and interpret the collected contextual data, generate inferences by reasoning against context models, and then react (and pre-act) accordingly. These processes pertain to both urban systems and information systems used by citizens. Regarding urban systems, multilevel integrated modeling is central to the endeavor of merging real-time data with traditional data across urban domains as sectional sources in a way that link real-time issue to long term strategic planning in terms of sustainability, i.e. energy systems, transport systems, and traffic systems. However, the idea of smart sustainable cities of the future is that, based on context-aware computing, related applications and systems can be made more perceptive and responsive by becoming aware of their surroundings (monitoring or observing the urban environment in its various forms, including physical, environmental, operational, behavioral, spatiotemporal, and socio-economic) and reacting to this awareness that can be attained by means of multiple, diverse sensors embedded throughout the urban environment that help build and maintain diverse models that represent the state of the dynamically changing and evolving urban world. Both the observed information about and the existing dynamic models for this world serve as input for the process of computational understanding, which involves the analysis and estimation of what is happening in the urban environment. The dynamic models can be viewed as interpretations of the urban environment, and thus are to be continuously improved to inform intelligent decision-making through machine reasoning about the meaning of all kinds of relevant situations taking place in that environment. Hence, they stand between the urban environment to be sensed and analyzed based on context data and the abstract notion of application actions. Dynamic models represent situations as problem and solution statements or sets of propositions expressing relationships among urban constructs which form the vocabulary of, and used to describe problems and to specify their solutions within urban sustainability domain, to draw on Bibri [19]. They therefore serve as an important input together with the collected sensor data from multiple sources to context-aware applications and systems. Important to note is that new dynamic models should be developed based on what constitutes smart sustainable cities in terms of their systems, processes, and forms. They are key in context information processing associated with real-time (and sometimes offline) applications in relation to such cities. Crucially, such models must be comprehensive, consistent, flexible, robust, and have a high degree of fidelity with real world urban phenomena.

A common use of the sensing and computing devices in smart sustainable cities of the future is to build and maintain an urban world model (see Fig. 3), which allows various context-aware applications and systems to be constructed and operate intelligently to catalyze and boost the process of sustainable development using such strategies as optimization, control, and management concerning urban systems as well as service efficiency and enhancement with regard to citizens.

Figure 3 illustrates an adaptive context awareness process where the smart sustainable city information and urban systems incrementally create the urban world they observe using sensors (and datasets). In this process, the users of such systems (citizens and urban operators/administrators) can be taken into the loop to train and detect contexts



on the spot; the systems learn new situations by example, with their users as the teachers. This flexible learning scheme, which is essential to the evolution of urban models, is known as incremental learning, that is, old classes can be retrained should they have changed, and new classes can be trained with no need to retrain the ones that were already trained (see [19]). In this way, urban contexts can be learned and recognized automatically, that is, sensor observations can be associated to human-defined context labels using machine learning and reasoning techniques.

Urban context recognition techniques and algorithms

Machine learning and hybrid modeling in context-aware computing

Context-aware computing as a prerequisite enabling technology for AmI, UbiComp, and SenComp is heralding new ways of interaction and applications in the context of smart sustainable cities. Pattern recognition algorithms in context-aware computing are under vigorous investigation in the development of smart urban environments. A multitude of such algorithms and their integration are being proposed and studied on the basis of the way in which urban contexts are operationalized, modeled, represented, and reasoned about. This can be done during a specification process whereby either concepts of context and their interrelationships are described based on urban knowledge (from city-directed disciplines) and represented in a computational format that can be used as part of reasoning processes to generate inferences, i.e. ontologies are used to represent and reason about context information, or contexts are learned and recognized automatically, i.e. machine learning techniques are used to build context models and perform further means of pattern recognition, i.e. probabilistic and statistical reasoning. While several context recognition algorithms have been applied in the area of context-aware computing, the most commonly used ones are those that are based on machine

learning techniques and ontological approaches. Such techniques and approaches have been integrated in various context-aware applications. This falls under what is referred to as hybrid context modeling and reasoning approaches, which involve both knowledge representation formalisms and reasoning mechanisms. (See [19] for examples of hybrid approaches developed in relation to various application domains). Hybrid approaches involve other methods, such as case-based methods, rule-based methods, logic programming, and database modeling techniques. The interested reader can be directed to Bibri [19] for a detailed overview of ontological approaches, to Bettini [41] for case-based methods, to Chen and Nugent [42] for a short account of logical modeling and reasoning and related algorithms in terms of logical theories and representation formalisms, and to Strimpakou et al. [48] for database modeling techniques.

Hierarchical hybrid models are assumed to bring clear advantages in terms of the set of the requirements defined for a generic context model used by context-aware applications. For example, they can provide solutions to overcome the weaknesses associated with the expressive representation and reasoning in description logic. Bettini et al. [41] contend that there is likelihood to satisfactorily address a larger number of the identified requirements by hierarchical hybrid context model if hybrid approaches can be further extended to design such model. The authors propose a model that is intended to provide a more comprehensive solution in terms of expressiveness and integration of different forms of reasoning. In this model, the representation formalism used to represent data retrieved from a module executing some sensor data fusion technique should, in order to support the scalability requirements of context-aware services, enable the execution of efficient reasoning techniques to infer high-level context data on the basis of raw ones by, for example, executing rule-based reasoning in a restricted logic programming language. As suggested by the authors, a more expressive, ontology-based context model is desirable on top of the respective representation formalism, as it inevitably does not support a formal definition of the semantics of context descriptions. See Bibri [19] for an illustration of the corresponding framework and a description of its multiple layers. This framework can be adapted so to fit in with context-aware applications pertaining to smart sustainable cities.

Supervised versus unsupervised methods

The concepts of supervised and unsupervised were inherited from the area of machine learning, a subfield of artificial intelligence that deals with artificial systems that are able to improve their performance overtime, in response to their experience in the world. Specifically, machine learning is the subfield of computer science that is concerned with the development of software programs that provide computer systems with the ability to learn from experiences without pursuing explicitly programmed instructions—that is, to teach themselves to grow and change when exposed to new data [19]. As a widely quoted, more formal definition provided by Mitchell ([107], p. 2), 'A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E '. Such improvement often involves analyzing data from the environment and making predictions about unknown variables.

Making the computer systems deployed throughout urban environments able to compute, communicate, and share data does not make them intelligent; rather, the key and challenge to really equipping these systems with intelligence and augment such environments with it lies in the way the systems learn and keep up to date with the needs and requirements of citizens and urban system respectively by themselves in the context of sustainability. In fundamentally operational terms, this description resonates with the idea that computer systems can think, a technological feature that underlies the prevalent ICT visions of pervasive computing, most notably UbiComp, AmI, and SenComp in terms of context-aware applications. However, it is computationally unfeasible to build models for all sorts of situations of urban life concerning context-aware applications. The underlying assumption is that training sets are finite, urban life situations are dynamic, and behavioral patterns of urban systems are uncertain, and so on, adding to the limited routinely generated sensor data in the context of real-time urban functioning. Besides, 'notwithstanding the huge potential of machine learning techniques, the underlying probability theory usually does not yield assurances of the performance of algorithms; rather, probabilistic (and statistical) reasoning limits to the performance are quite common. This relates to computational learning theory, a branch of theoretical computer science that is concerned with the computational analysis of machine learning algorithms and their performance in relation to different application domains' ([19], p. 172). In relation to context-aware applications, machine learning involves a wide variety of algorithms which can be classified into different categories based on the following methods [19]:

- Supervised and unsupervised learning (explained below).
- Semi-supervised learning (combines both labeled and unlabelled examples to generate an appropriate classifier).
- Transductive inference (attempts to predict new outputs on specific test cases from observed training cases).
- Learning to learn (learns its own inductive bias based on previous experience).
- Reinforcement learning (executes actions which trigger the observable state of a dynamic environment to change, and attempts to gather information about how the environment reacts to actions as well as to synthesize a sequence of actions that maximizes some notion of cumulative reward).

Indeed, issues of agency and cognition in terms of how an intelligent agent use learned knowledge to reason and act in its environment are characteristic to machine learning.

For a descriptive account of supervised and unsupervised methods for predictive and descriptive data mining in the context of smart sustainable cities, the reader can be directed to Bibri [55]. This article also covers a number of use cases for each method, in addition to new insights into the next wave of urban analytics in light of data science and the application of data mining and knowledge discovery to urban sustainability.

Context recognition techniques and algorithms

Context recognition algorithms based on supervised and unsupervised learning methods primarily use probabilistic and statistical reasoning. Supervised learning requires the

use of labelled data on which an algorithm is trained, and training sets are called labeled data because the value for the class label is known. Following training the algorithm can classify unknown and future data. Thus, the basic idea of supervised learning is to classify data in formal categories that an algorithm is trained to recognize. In this sense, the machine learning process examines a set of atomic contexts which have been pre-assigned to categories, and makes inductive abstractions based on these data that assist in the process of classifying unknown and future atomic contexts into high-level contexts. Supervised learning algorithms require an important training period during which several examples of each context and related concepts are collected and analyzed. The quality of training influences the outcome of the classification critically. And the granularity of the learned context concepts is influenced by the availability and nature of the low-level contextual data from sensors [41]. In all, supervised learning algorithms enable context-aware applications and systems to keep a trace of their previous observed experiences in the form of trained classes of context and employ them to dynamically learn the parameters of the stochastic context models (a pattern that may be analyzed statistically but not predicted precisely). This enables them to generate predictive models based on the observed agents' context profiles. The general process using a supervised learning algorithm for context recognition encompasses several steps, namely, borrowing Chen and Nugent's [42] terminology:

- (1) To acquire sensor data representative of relevant context features/attributes pertaining to citizens or urban systems, including related labelled annotations.
- (2) To determine the input data features and their representation.
- (3) To aggregate data from multiple data sources and transform them into the application-dependent features, e.g. through data fusion, noise elimination, dimension reduction and data normalization.
- (4) To divide the data into a training set and a test set.
- (5) To train the recognition algorithm on the training set.
- 6) To test the classification performance of the trained algorithm on the test set.
- (7) To apply the algorithm in the context of context recognition. It is common to repeat steps (4) to (7) with different partitioning of the training and test sets in order to achieve better generalization with the recognition models.

There are a wide range of algorithms and models for supervised learning and context recognition. In the area of context recognition, the basic idea of classification is to determine different context labels on the basis of a set of context categories (training examples) learned from the real world as models. The algorithm is presented with a set of inputs and their desired outputs, e.g. association of sensor data with real world contexts, and the goal is to learn a general rule that maps inputs (e.g. sensor data) to outputs (context labels), so that the algorithm can map new sensor data into one of these context labels. The quality of classification, how well a classifier performs, is inextricably linked to the richness of the learning experience of the algorithm, and also depends critically on the features of the contextual data to be classified. Building new models, training new classes, during the analysis of the collected sensor data is important for making future inductive abstractions in terms of classifying unknown and future contextual data

(i.e. performance) through gaining experience. Classification of contexts is done using a classifier that is learned from a comprehensive training set of annotated context examples. Classifiers represent tasks entailing the use of pattern matching to determine a best match between the features extracted from sensor data and a context description. This is about classifying sensor cues into a known category and storing general patterns of context. There are various supervised learning classifiers, and they vary in terms of performance, which depends on the application domain to which they can be applied. For example, binary decision tree uses a decision tree as a predictive model which maps sensor observations to inferences about the context's target value. Support vector machine (SVM) builds a model that predicts into which of two categories a new example falls, assuming that each training example is marked as belonging to one of these two categories. Other classifiers include neural network, k-nearest neighbor, dynamic and naive Bayes, and Hidden Markov Models (HMMs). They have been applied in a wide variety of context awareness domains within both laboratory-based as well as real-world environments (see [19] for a set of selected examples). Important to note is that no one classifier is superior to another, nor is there a single classifier that works best for all on all given problems. It follows that to determine a suitable classifier for a given problem domain is linked to the complexity and nature of that problem domain. Still, among the supervised learning algorithms, HMMs and Bayes networks are thus far the most commonly applied methods in the area of context recognition. While both of them have been shown to be successful in context-aware computing, they are both very complex and require lots of a large amount of labelled training and test data. This is in fact the main disadvantages of supervised learning algorithms in the case of probabilistic methods, adding to the fact that it could be computationally costly to learn each context in a probabilistic model for an infinite richness or large diversity of contexts in real world application scenarios (see [42]). Moreover, given that context-aware applications usually incorporate different contextual features that should be combined in the inference of a particular dimension of context, adding to the fact that one feature may involve different types of sensor data, the repetitive diversification of the partitioning of the training and test sets may not lead to the desired outcome with regard to the generalization with the context recognition models. This has implication for the accuracy of the estimation of context, that is, the classification of dynamic contextual data into relevant context labels. Machine learning methods in the case of probabilistic methods 'choose a trade-off between generalization and specification when acquiring concepts from sensor data recordings, which does not always meet the correct semantics, hence resulting in wrong detections of situations' ([41], p. 11). A core objective of a learning algorithm is to generalize from its experience whereby generalization denotes the ability of a learning mechanism to perform accurately on not previously seen context examples after having experienced a learning data set—the combination of context patterns and their class labels. While this is a decision that should be made, the resulting context models are often ad-hoc and not reusable. In fact, supervised learning algorithms inherently suffer from several limitations, namely scalability, data scarcity, inflexibility, ad-hoc static models; these methods 'should tackle technical challenges in terms of their robustness to real-world conditions and real-time performance' [42]. New research endeavors should focus on creating alternative theories based on new discoveries in human-directed sciences in terms of developing less

complicated, computationally elegant, and, more importantly, effective and robust algorithms with wider applicability, irrespective of the application domain.

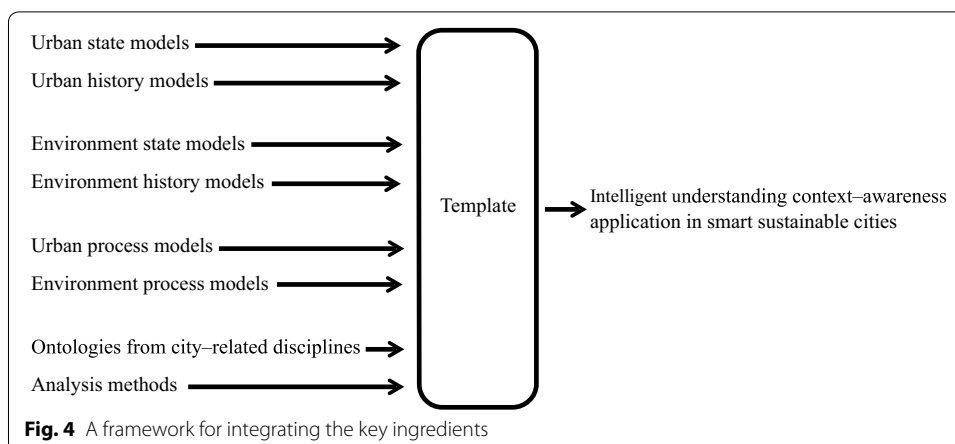
Distinct from supervised learning, unsupervised learning tries to directly build recognition models from unlabeled data. With having no labels, the learning algorithm is left on its own to group similar inputs or density estimates that can be visualized effectively [108]. Thus, unsupervised learning provides context-aware applications with the ability to find context patterns in cues as abstraction from raw sensor data—i.e. features extracted from the data stream of multiple, diverse sensors. Probabilistic algorithms can be used for finding explanations for streams of data, helping recognition systems to analyze processes that occur over time [126]. The basic idea of unsupervised learning algorithm is to manually assign a probability to each possible context and to use a pre-defined stochastic model to update these likelihoods on the basis of both new sensor readings and the known state of the system (see [42]). The general process of unsupervised learning algorithms for context recognition includes the following steps Chen and Nugent ([42], p. 414):

- (1) To acquire unlabeled sensor data.
- (2) To aggregate and transform the sensor data into features.
- (3) To model the data using either density estimation (to estimate the properties of the underlying probability density) or clustering methods (to discover groups of similar examples to create learning models).

There exist several algorithms for unsupervised learning that are based on probabilistic reasoning, such as Bayes networks, graphical models, multiple eigenspaces, and different variants of HMMs. Further, unsupervised learning probabilistic methods are capable of handling the uncertainty and incompleteness of sensor data. Probabilities can be used to serve various purposes in this regard, such as modeling uncertainty, reasoning on uncertainty, and capturing domain heuristics (see, e.g. [41, 42]). However, unsupervised learning probabilistic methods are usually static and highly context-dependent, adding to their limitation as to the assignment of the handcrafted probabilistic parameters (e.g. modeling uncertainty, capturing heuristics) for the computation of the context likelihood (see [42]). Indeed, they seem to be less applied than supervised learning in the domain of context recognition.

Conceptual context models and a framework for integrating the key ingredients

As high-level abstractions of contexts can be semantically abstracted from contextual cues extracted from low-level context data obtained from physical sensors, human knowledge and interpretation of the urban world must be formally conceptualized and modeled according to certain formalisms. In smart sustainable cities, conceptual context models are concerned with what constitutes various types of contexts and their conceptual structures depending on the application domain. While the semantics of what constitutes ‘context’ has been widely discussed in the literature and defining what constitutes context information has been studied extensively in relation to humans (e.g. [19]), little, if no, attention has been given to what constitutes context information in relation to urban systems (energy, traffic, transport, etc.). This is a fertile research area in the



realm of smart sustainable cities. Generally, context information in the urban domain refers to the representation of the situation of an entity (energy system in a building, a traffic system in a district, a transport system in an urban area characterized by mixed-land use and density, etc.) in some computer system run by, for example, an urban department, where a set of contextual features are of interest to a provider of operational functioning services for assessing the timeliness and context-dependent aspects of the system behavior. Works that can identify qualitative features of urban system context information remain scant. And those related to citizens are numerous and diverse (see [19] for a detailed survey). However, most of the latter class of works does not provide formal representations of the proposed models.

However, one of the challenges in smart sustainable cities is to provide frameworks that cover the class of context-aware applications that exhibit computational understanding and intelligent behavior in relation to sustainability performance. Here computational understanding entails performing the analysis and interpretation of context data, including reasoning about such data, and estimating what is happening in the urban environment (high-level abstractions of context pertaining to various urban systems), a process for which input are observed information about the situations of urban systems over time (i.e. urban monitoring) and dynamic models for the spatial, spatiotemporal, physical, infrastructural, environmental, and socio-economic processes of smart sustainable cities. Note that different types of models are needed. Intelligent behavior entails context-aware applications coming up with and firing situation-dependent actions that provide support for different aspects of sustainability. With the above in mind, a basic framework can be suggested, which combines different models and methods, as illustrated in Fig. 4.

Figure 4 shows a basic framework which combines urban state and history models, environment state and history models, dynamic process models (about urban functioning), dynamic environment process models, ontologies and knowledge from city-related disciplines (including urban sustainability), and analysis methods on the basis of such models, such as spatial analysis, spatiotemporal analysis, environmental analysis, behavioral analysis, and so on. As a template for the class of context-aware applications showing computational understanding and intelligent behavior, the framework can encompass slots where the content of applications specific to various urban systems

together with the generic operative methods can be filled in order to obtain an executable design for a working application. Accordingly, context-aware applications can show advanced computational understanding of the urban environment and react from this understanding in a knowledgeable manner through intelligent decision-making for improving different aspects of sustainability through optimization, control, management, and planning of urban systems.

A basic multilayered architecture of context information processing

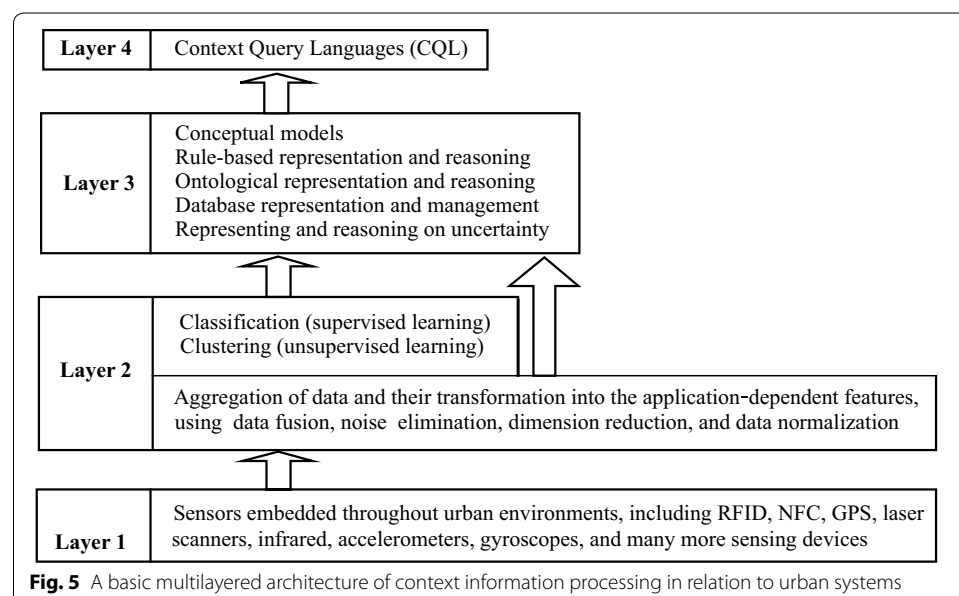
Researchers from different application domains have investigated context awareness for the past 2 decades or so by developing a diversity of architectures (see [19] for an overview). Context awareness involves a wide range of architectures that basically aim to provide the appropriate infrastructures for context-aware applications pertaining to diverse application domains. Context-aware applications are based on a multilayered architecture, as shown in Fig. 5. Here the focus is on urban systems.

Layer 1—physical sensors

Signals in the urban environment are detected from multiple sources using sensors of many types. This sensor layer is usually defined by open-ended (unrestricted) collection of sensors embedded in urban systems and spread in their surrounding environment. The data supplied by sensors, irrespective of the application domain (traffic, energy, transport, mobility, etc.), are usually very different, ranging from slow sensors to fast and complex sensors that provide larger volume and higher velocity of data.

Layer 2—context data processing

This layer is dedicated to aggregate, fuse, organize, and propagate contextual data. At this stage, signal processing and machine learning techniques (pre-processing, analysis, and pattern recognition) are used to recognize situations of urban systems from sensor



signals and labelled annotations in the case of classification, for example. The recognition can also occur through mapping sensor readings to matching properties described in ontologies of particular urban domains in the case of hybrid models, and using these ontologies to aggregate and fuse sensor observations to generate situations of urban systems. Which recognition approach to espouse is determined by how context is modeled and thus encoded and reasoned about. However, as contextual data need at this stage to be organized in the form of information, this layer introduces abstractions of real world situations.

Layer 3—context representation and reasoning

As the core of context modeling, this layer involves the application of representation and reasoning techniques, which usually entail an integration of different modeling approaches for an effective outcome as to the estimation of the situation of urban systems. Indeed, this hybrid method allows to address issues such as uncertainty, incompleteness, and vagueness associated with contextual information.

Layer 4—application action

At this layer actions are fired based on the generated inferences. Specifically, decisions are taken as to what actions are to be executed on the basis of high-level abstractions of situations produced through reasoning processes. The type of actions to be performed is typically dependent on the application domain. For example, the actions taken at the application level can be oriented towards supporting energy efficiency, traffic congestion reduction, transport management, and so on. Furthermore, there is variety of context query languages that can be used by context-aware applications, as there is no specific requirement in context awareness architectures with respect to which query languages to use. However, the selection of the query language is contingent upon the applied modeling approaches in terms of representation and reasoning techniques. ‘The meaning of the queries must be well-specified because in the implementation the queries are mapped to the representations used in layer 3. An important role of the middle layer and the query language is to eliminate direct linking of the context providing components to context consuming components... Thus, the query language should support querying a context value regardless of its source... It should be noted that since the CQL acts as a facade for the applications to the underlying context representation, the context information requirements of the applications are imposed as much on the query language as on the context representation and context sources’ ([43], p. 2).

Figure 5 illustrates four layers of context information processing. The architecture also includes examples of existing techniques and methods that can be used in context-aware applications, depending on the application domain and its complexity and scale.

However, while some advanced design alternatives for context awareness architectures are being deployed in smart sustainable cities, it will take some time before standard, interoperable context information modeling and management become a reality.

The key applications of big data analytics and context-aware computing technologies for urban sustainability

The prospect of developing and implementing smart sustainable cities based on big data analytics and context-aware computing as a set of technologies and applications is becoming increasingly a reality. This new techno-urban phenomenon is opening entirely new windows of opportunity for smart cities to explicitly incorporate sustainability and for sustainable cities to smarten up their contribution to sustainability. Smart sustainable cities as a techno-urban innovation represent transformative processes that have been fueled by the increasing infiltration of information intelligence into urban systems in terms of operations, functions, services, designs, practices, and policies. This information intelligence enabled and driven by big data analytics and context-aware computing could be leveraged in the advancement of urban sustainability by enhancing and integrating urban systems as well as by facilitating coordination and collaboration among diverse urban domains.

Accordingly, the two identified classes of applications of ICT of the new wave of computing pertain to big data analytics and context-aware computing in the context of smart sustainable cities. In other words, data-centric and context-aware applications constitute key components of the informational landscape of various models of smart sustainable city (see [6] for a detailed account of these models). As noted by Bibri and Krogstie [4], their effects reinforce one another as to their efforts for transforming urban life sustainably by employing and merging smart solutions to improve urban sustainability. The basic idea is that the opportunities for the deployment of the advanced solutions being offered by ICT of the new wave of computing are tremendous in the context of urban sustainability. Indeed, the applications associated with big data analytics and context-aware computing are compatible with the goals of sustainable development. They include, but are not limited to, the following:

- Data-centric and context-aware transport and mobility (e.g. [2, 4, 14, 22, 23, 98, 102, 105–112]).
- Data-centric and context-aware traffic lights and signals (e.g. [1, 22, 54]).
- Data-centric and context-aware energy systems (e.g. [1, 2, 22, 113, 114]).
- Data-centric and context-aware grid (e.g. [1, 23, 76, 109–115]).
- Data-centric and context-aware environment (e.g. [22, 116, 117]).
- Data-centric and context-aware buildings (e.g. [3, 4, 23]).
- Data-centric and context-aware public safety and civil security [10, 22].
- Data-centric and context-aware planning and design (e.g. [2, 6, 8, 118]).
- Data-centric and context-aware healthcare [11, 14, 92, 116, 119].
- Data-centric and context-aware education and learning (e.g. [1, 3, 22, 102]).
- Data-centric and context-aware citizen services (the quality of life) (e.g. [1, 11, 53, 67, 97]).
- Data-centric and context-aware urban infrastructures and facilities monitoring and management (e.g. [113, 120, 121]).

In all, the application of big data analytics and context-aware computing in smart sustainable cities offers the prospect of significantly improving different aspects of

sustainability. One of the core ideas underlying the use of these advanced technologies is to integrate and harness solutions and approaches through coordinating, coupling, and integrating urban domains. Hence, exposing big data and context information via a sustainable, socially synergistic, evolvable, dynamic, extensible, scalable, and reliable ecosystem offers a wide range of benefits and opportunities with respect to urban sustainability. As noted by Bibri and Krogstie [5] with reference to sustainable urban forms, ICT of the new wave of computing as a set of enabling and constitutive technologies can make substantial contributions—not only in terms of catalyzing and boosting the development processes of sustainable urban forms, but also in terms of planning such forms in ways that continuously evaluate, forecast, and thus strategically optimize their contribution to sustainability.

The main scientific challenges of big data analytics and context-aware computing

The rising demand for big data analytics and context-aware computing as disruptive technologies, coupled with their potential to serve many urban domains in terms of sustainability, comes with major scientific and intellectual challenges that need to be addressed and overcome with regard to the design, development, and deployment of data-centric and context-aware applications in the context of smart sustainable cities. In terms of context-aware computing, the challenges center around system engineering, design, and modeling. They include, but are not limited to, the following, to draw on Bibri [19]:

- Constraints of design science and engineering.
- Paradigms that govern the assembly of context-aware applications pertaining to diverse urban domains and their integration in connection with sustainability, as well as to dynamic models of their knowledge representation and run-time behavior.
- Tailored methodologies and tools for engineering urban context awareness.
- General methods for acquiring, storing, processing, analyzing, mining, modeling, querying, and making sense of context data for context-aware applications.
- The performance of real-time context-aware applications given that they need to be timely in taking actions.
- Provision of context data as a service to a wide range of applications within diverse urban domains based on integration with the service computing paradigm.
- Modeling and management of contextual information in large-scale distributed pervasive applications and in open and dynamic pervasive environments (e.g. [41, 48]).

Handling context-aware information in smart sustainable cities is a tremendous challenge (see [11]). Indeed, urban context awareness is a complex, multilevel problem as to its functioning and implementation, from low-level sensor data acquisition, through intermediate-level information processing and modeling, to high-level application action and service delivery.

In terms of big data analytics, the challenges are mostly scientific, computational, and analytical in nature. They include, but are not limited to, the following (e.g. [2, 5, 6, 101, 117–125]; Katal et al. [26]).

- Constraints of design science and engineering.
- Data analysis and evaluation.
- Management of IoT data produced in dynamic and volatile environments.
- Database integration across urban domains.
- Privacy and security.
- Establishing context (e.g. geolocation and time).
- Data growth and sharing.
- Data uncertainty and incompleteness.
- Data quality and veracity.
- Intelligence functions and simulation models.
- Fault tolerance and scalability.
- Storage and processing.

The main challenges of big data analytics arise from the nature of the data being generated in terms of their large, diverse, and time-evolving character. To put it differently, the scale, heterogeneity, and velocity of urban data makes it difficult to manage, integrate, process, analyze, evaluate, and deploy. Adding to these primarily technical challenges are the financial, organizational, institutional, regulatory, and ethical ones, which are associated with the implementation, retention, and dissemination of big data across the domains and entities of smart sustainable cities. In addition, controversies over the application and benefit of big data analytics relate to limited access and related divide and ethical concerns about accessibility [28]. Nevertheless, understanding, exploiting, and extending, or simply advancing knowledge of, the available computation, analysis, and management capabilities associated with big data analytics and context-aware computing in terms of conceptions, tools, principles, paradigms, methodologies, and risks, great opportunities could be realized in terms of improving, harnessing, and integrating urban systems and thus facilitating collaboration, coordination, and coupling among urban domains through data-centric and context-aware applications in the context of smart sustainable cities. It is safe to say that as long as big data and context data in urban analytics are driven by sustainable development agenda and thus utilized and implemented strategically for the purpose of monitoring, understanding, probing, and planning smart sustainable cities, ICT of the new wave of computing will drastically change the way such cities function as to increasing their contribution to the goals of sustainable development over the long run. This requires the current open issues stemming from the aforementioned challenges to be under rigorous investigation and scrutiny by the socio-technical systems involved in the underlying technological innovation system of big data analytics and context-aware computing, namely industry consortia, business communities, research institutes, universities, policy makers and networks, and governmental agencies.

Conclusions

Big data analytics and context-aware computing are rapidly growing areas of ICT that are becoming even more important to smart sustainable cities with respect to their operational functioning and planning to improve their contribution to the goals of sustainable development. These concepts were crystallized into realist notions in the

domain of sustainable urban planning not too long ago—until UbiComp, AmI, the IoT, and SenComp as the most prevalent ICT visions of pervasive computing have become achievable and deployable paradigms and also matured thanks to the recent advances in sensor technologies, data processing platforms, cloud computing and middleware infrastructures, and wireless communication networks. This major technological transition is drastically changing how smart sustainable cities can be monitored, understood, analyzed, and planned to advance sustainability. Accordingly, big data analytics and context-aware computing are opening unique windows of opportunity for enabling such cities to leverage their informational landscape by developing, deploying, and implementing a variety of advanced applications to enhance their operational functioning, planning, and design in line with the vision of sustainability.

The aim of this paper was to review and synthesize the relevant literature with the objective of identifying and distilling the core enabling technologies of big data analytics and context-aware computing as ecosystems in relevance to smart sustainable cities, as well as to illustrate the key computational and analytical techniques and processes associated with the functioning of such ecosystems. The main contribution of this paper lies in developing, elucidating, and evaluating a number of relevant frameworks pertaining to big data analytics and context-aware computing in the context of smart sustainable cities by bringing together research directed at a more conceptual, analytical, and overarching level to stimulate new ways of investigating their role in advancing urban sustainability. The proposed frameworks, which can be replicated and tested in empirical research, will add additional depth and rigor to studies in the field. Big data analytics and context-aware computing share basically the same core enabling technologies in the realm of smart sustainable cities, especially their effects overlap in many aspects with regard to advancing the process of sustainable development. The underlying enabling technologies and related key computational and analytical techniques and processes consist of the following:

- Data collection and preprocessing, e.g. data sensing methods and signal processing techniques
- Data repositories or storage facilities, e.g. database and data warehouse servers
- Data processing, e.g. data analytic systems, cloud computing models, middleware architectures, including software tools and database systems
- Analysis techniques and algorithms, e.g. data mining, machine learning, statistics, and database query and related computational mechanisms
- Wireless network technologies, e.g. the satellite-enabled GPS, mobile phone, LPWAN, and Wi-Fi networks, for collecting and coordinating data in terms of the data themselves and how that data are stored and made accessible
- Data visualization for representing and displaying useful and context knowledge in understandable formats for human interpretation.

Adding to the above components are privacy and security mechanisms, open standards and standard rules, as well as conceptual frameworks (data mining process, context recognition process, context information processing, etc.) and related methods (supervised and unsupervised learning).

Furthermore, the availability of the various permutations of the core enabling technologies underlying big data analytics and context-aware computing is justified by the varied technical details of the application domains pertaining to smart sustainable cities in terms of their complexity, scale, requirement, and objective. Regardless, to facilitate an effective functioning of big data and context-aware applications, it is important to ensure a seamless amalgamation of their core enabling technologies. This is of equal importance to better understand, monitor, analyze, and plan smart sustainable cities for the purpose of catalyzing and boosting the process of sustainable development towards achieving the long-term goals of sustainability.

The emerging ability to use big data and context-aware techniques and methods for advancing sustainability promises to revolutionize various urban domains. The key applications enabled by big data analytics and context-aware computing include transport, mobility, traffic lights and signals, energy systems, power grid, environment, buildings, public safety and civil security, planning and design, healthcare, education and learning, the quality of life, and urban infrastructures and facilities monitoring and management. Besides all the benefits, the large-scale deployment and amalgamation of big data analytics and context-aware computing as advanced forms of ICT is beset with several challenges due to the massive size, diverse nature, and fast-changing pace of big data and to the constraints of system engineering, design, and modeling of context awareness.

In all, the use of big data analytics and context-aware computing entails that smart sustainable cities take the form of constellations of architectures, platforms, applications, and computational and data analytics capabilities connected through wirelessly ad-hoc and mobile networks with a modicum of intelligence across several spatial scales, which provide and coordinate continuous data regarding the physical, infrastructural, spatial, spatiotemporal, operational, functional, and socio-economic forms of such cities. We argue that big data analytics and context-aware computing are prerequisite technologies for the functioning of smart sustainable cities of the future, as their effects reinforce one another as to their efforts for bringing a whole new dimension to the operating and organizing processes of urban life in terms of employing a wide variety of smart and data-driven applications for advancing sustainability.

Authors' contributions

SEB and JK have both made substantive intellectual contributions to the study. SEB has made substantial contributions to the collection, review, and synthesis of the literature. They have both been involved in drafting the manuscript and revising it critically for important intellectual content, as well as given the final approval of the current version to be published. Both authors read and approved the final manuscript.

Authors' information

Simon Elias Bibri is a Ph.D. scholar in the area of smart sustainable cities of the future and a professor assistant at the Norwegian University of Science and Technology (NTNU), Department of Computer and Information Science and Department of Urban Design and Planning, Trondheim, Norway. His intellectual pursuits and endeavors have hitherto resulted in an educational background encompassing knowledge from, and meta-knowledge about, different academic and scientific disciplines. He holds a Bachelor of science in computer engineering and 10 Masters of science in diverse areas, including computer science, system science, innovation, strategic sustainability, sustainable urban development, eco-technology, business administration, and communication and media. Bibri has earned all his Master's degrees from different universities in Sweden, namely Lund University, West University, Blekinge Institute of Technology, Malmö University, Stockholm University, and Mid Sweden University.

Bibri's current areas of research work include smart sustainable cities; Ambient Intelligence (AmI), Ubiquitous Computing (UbiComp), the Internet of Things (IoT), and Sentient Computing (SenComp), as well as how these computing paradigms relate to urban sustainability, sustainability science, urbanization, and urban design and planning; and big data analytics and context-aware computing and the associated core enabling technologies, namely sensor technologies, data processing platforms, cloud computing infrastructures, middleware architectures, and wireless communication networks.

Bibri is the author of two academic books in the field of pervasive computing and in the process of publishing his third book. Also, he has occasionally been working on his fourth book in parallel with his doctoral studies.

Author details

¹ Department of Computer and Information Science and Department of Urban Planning and Design, NTNU Norwegian University of Science and Technology, Sem Saelands veie 9, 7491 Trondheim, Norway. ² Department of Computer and Information Science, NTNU Norwegian University of Science and Technology, Sem Saelands veie 9, 7491 Trondheim, Norway.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable.

Funding

The study is an integral part of a Ph.D. research project being carried out at NTNU.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 October 2016 Accepted: 20 September 2017

Published online: 07 November 2017

References

1. Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J. Applications of big data to smart cities. *J Internet Serv Appl*. 2015;6(25):1–15.
2. Batty M, Axhausen KW, Giannotti F, Pozdnoukhov A, Bazzani A, Wachowicz M, Ouzounis G, Portugali Y. Smart cities of the future. *Eur Phys J*. 2012;214:481–518.
3. Bibri SE, Krogstie J. On the social shaping dimensions of smart sustainable cities: a study in science, technology, and society. *Sustain Cities Soc*. 2016;29:219–46.
4. Bibri SE, Krogstie J. Big data analytics and context-aware computing for smart sustainable cities of the future, NOBIDS conference (in press). vol.1818. 2016; pp.417.
5. Bibri SE, Krogstie J. Smart sustainable cities of the future: an extensive interdisciplinary literature review. *Sustain Cities Soc*. 2017;31:183–212.
6. Bibri SE, Krogstie J. ICT of the new wave of computing for sustainable urban forms: their big data and context-aware augmented typologies and design concepts. *Sustain Cities Soc*. 2017;32:449–74.
7. Kramers A, Höjer M, Lövehagen N, Wangel J. Smart sustainable cities: exploring ICT solutions for reduced energy use in cities. *Environ Model Softw*. 2014;56:52–62.
8. Shahrokni H, Årman L, Lazarevic D, Nilsson A, Brandt N. Implementing smart urban metabolism in the Stockholm Royal Seaport: smart city SRS. *J Ind Ecol*. 2015;19(5):917–29.
9. Böhlen M, Frei H. Ambient intelligence in the city: overview and new perspectives. In: Nakashima H, Aghajan H, Augusto JC, editors. *Handbook of ambient intelligence and smart environments*. New York: Springer; 2009. p. 911–38.
10. Shepard M, editor. *Sentient city: ubiquitous computing, architecture and the future of urban space*. Cambridge: MIT Press; 2011.
11. Solanas A, Patsakis C, Conti M, Vlachos IS, Ramos V, Falcone F, Postolache O, Pérez-Martínez PA, Di Pietro R, Perrea DN, Martínez-Balleste A. Smart health: a context-aware health paradigm within smart cities. *IEEE Commun Mag*. 2014;52(8):74–81.
12. Coutaz J, Crowley JL, Dobson S, Garlan D. Context is key. *Commun ACM*. 2005;48(3):49.
13. Schmidt C. Context-aware computing, Berlin Inst. Technology Tech, 1–9. 2011. http://diuf.unifr.ch/pai/education/2002_2003/seminar/winter/ubicomp/02_Pervasive.pdf. Accessed 16 Sept 2016.
14. Vongsingthong S, Smanchat S. Internet of things: a review of applications and technologies. *Suranaree J Sci Technol*. 2014;21(4).
15. Al-Nasrawi S, Adams C, El-Zaart A. A conceptual multidimensional model for assessing smart sustainable cities. *J Inf Syst Technol Manag*. 2015;12(3):541–58.
16. Kramers A, Wangel J, Höjer M. Governing the smart sustainable city: the case of the Stockholm royal seaport. In: *Proceedings of ICT for sustainability 2016*, Amsterdam: Atlantis Press, vol 46. 2016; p. 99–108.
17. Höjer M, Wangel S. Smart sustainable cities: definition and challenges. In: Hilty L, Aebischer B (eds.), *ICT innovations for sustainability*, Springer-verlag, Berlin, 2015. pp. 333–349.
18. International Telecommunications Union (ITU). Agreed definition of a smart sustainable city. Focus Group on Smart Sustainable Cities, SSC-0146 version Geneva, 5–6 March; 2014.
19. Bibri SE. *The human face of ambient intelligence, cognitive, emotional, affective, behavioral, and conversational aspects*. Berlin: Springer-Verlag; 2015.
20. Dey AK. *Providing architectural support for building context-aware applications*, PhD thesis, College of Computing, Georgia Institute of Technology; 2000.
21. Schmidt A. *Ubiquitous computing: computing in context*, PhD Dissertation, Lancaster University; 2003.

22. ISTAG 2003, Ambient Intelligence: from vision to reality (for participation—in society and business). http://www.ideo.co.uk/DTI/CatalIST/istag-ist2003_draft_consolidated_report.pdf. Accessed 23 Oct 2009.
23. ISTAG 2008, Working Group on ICT and sustainability (including energy and environment). <http://www.cordis.lu/ist/istag.htm>. Accessed 20 Mar 2012.
24. ISTAG 2012, Towards horizon 2020—recommendations of ISTAG on FP7 ICT work program 2013. http://cordis.europa.eu/fp7/ict/istag/reports_en.html. Accessed 15 Mar 2012.
25. Chen G, Kotz D. A survey of context-aware mobile computing research, Paper TR2000–381, Department of Computer Science, Dartmouth College; 2000.
26. Katal A, Wazid M, Goudar R Big Data: issues, challenges, tools and good practices. In: Proceedings of 6th International Conference on Contemporary Computing (IC3), Noida, August 8–10, IEEE, US, 2013. pp. 404–409.
27. Khan M, Uddin MF, Gupta N. Seven V's of big data understanding: big data to extract value. American society for engineering education (ASEE Zone 1), 2014 Zone 1 conference of the IEEE; 2014. p. 1–5.
28. Fan W, Bifet A. Mining big data: current status, and forecast to the future. *ACM SIGKDD Explor Newslett.* 2013;14(2):1–5.
29. Laney D. 3-D data management: controlling data volume, velocity and variety. *META Group Res Note.* 2001;6:70.
30. Dey AK. Understanding and using context. *Pers Ubiquit Comput.* 2001;5(1):4–7.
31. Dey AK, Abowd GD, Salber D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Hum Comput Interact.* 2001;16(2–4):97–166.
32. Schilit B, Adams N, Want R. Context-aware computing applications. In: Proceedings of IEEE workshop on mobile computing systems and applications, Santa Cruz, CA, USA; 1994. p. 85–90.
33. Schmidt A, Beigl M, Gellersen HW. There is more to context than location. *Comput Graph.* 1999;23(6):893–901.
34. Criel J, Claeyls L. A transdisciplinary study design on context-aware applications and environments, a critical view on user participation within calm computing. *Observatorio (OBS*).* 2008;5:057–77.
35. Lueg C. Operationalizing context in context-aware artifacts: benefits and pitfalls. *Hum Technol Interface;* 2002;5(2).
36. Ulrich W. Information, context, and critique: context awareness of the third kind, The 31st Information Systems Research Seminar in Scandinavia, Keynote talk presented to IRIS 31; 2008.
37. Crutzen CKM. Intelligent ambience between heaven and hell. *Inf Comm Eth Soc.* 2005;3(4):219–32.
38. Gunnarsdóttir K, Arribas-Ayllon M. Ambient intelligence: a narrative in search of users. Cesagen: Lancaster University and SOCSL, Cardiff University; 2012.
39. José R, Rodrigues H, Otero N. Ambient intelligence: beyond the inspiring vision. *J Univers Comput Sci.* 2010;16(12):1480–99.
40. Lindblom J, Ziemke T. Social situatedness: Vygotsky and beyond. In: 2nd international workshop on epigenetic robotics: modeling cognitive development in robotic systems, Edinburgh, Scotland; 2002. p. 71–8.
41. Bettini C, Brdiczka O, Henriksen K, Indulska J, Nicklas D, Ranganathan A, Riboni D. A survey of context modelling and reasoning techniques. *J Pervasive Mob Comput.* 2010;6(2):161–80.
42. Chen L, Nugent C. Ontology-based activity recognition in intelligent pervasive environments. *Int J Web Inf Syst.* 2009;5(4):410–30.
43. Perttunen M, Riekkilä J, Lassila O. Context representation and reasoning in pervasive computing: a review. *Int J Multimed Eng.* 2009;4(4).
44. Azodolmolky S, Dimakis N, Mylonakis V, Souretis G, Soldatos J, Pnevmatikakis A, Polymenakos L. Middleware for in-door ambient intelligence: the polyomaton system. In: Proceedings of the 2nd international conference on networking, next generation networking middleware (NGNM 2005), Waterloo; 2005.
45. Paspallis N. Middleware-based development of context-aware applications with reusable components. PhD Thesis, University of Cyprus; 2009.
46. Schmidt DC. Middleware for real-time and embedded systems. *Commun ACM.* 2002;45(6):43–8.
47. Soldatos J, Dimakis N, Stamatis K, Polymenakos L. A breadboard architecture for pervasive context-aware services in smart spaces: middleware components and prototype applications. *Pers Ubiquit Comput.* 2007;11(3):193–212.
48. Strimpakou M, Roussak I, Pils C, Anagnostou M. COMPACT: middleware for context representation and management in pervasive computing. *Pervasive Comput Commun.* 2006;2(3):229–45.
49. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *Artif Intell Mag.* 1996;17(3):37–54.
50. Shearer C. The CRISP-DM model: the new blueprint for data mining. *J Data Warehous.* 2000;5(4):13–22.
51. Shearer C. The CRISP-DM model: the new blueprint for data mining. *J Data Warehous.* 2000;5(4):13–22.
52. Bettencourt LMA. The uses of big data in cities Santa Fe Institute. New Mexico: Santa Fe; 2014.
53. Khan Z, Anjum A, Soomro K, Tahir MA. Towards cloud based big data analytics for smart future cities. *J Cloud Comput.* 2015;4(2).
54. Kumar A, Prakash A. The role of big data and analytics in smart cities. *Int J Sci Res (IJSR).* 2014;6(14):12–23.
55. Bibri SE Smart sustainable cities of the future: the untapped potential of big data analytics and context-aware computing for advancing sustainability. *Urban Book Series.* Springer; 2017 (in press).
56. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. *MIS Q.* 2012;36(4):1165–88.
57. Provost F, Fawcett T. Data science for business. Sebastopol: O'Reilly Media Inc; 2013.
58. Chen F, Deng P, Wan J, Zhang D, Vasilakos AV, Rong X. Data mining for the internet of things: literature review and challenges. *Int J Distrib Sens Netw.* 2015:1–14.
59. Zhang Y, Cao T, Tian X, Li S, Yuan L, Jia H, Vasilakos AV. Parallel processing systems for big data: a survey. In: Proceedings of the IEEE, special issue on big data, Nov 2016; 2016.
60. Singh J, Singla V. Big data: tools and technologies in big data (0975–8887). *Int J Comput Appl.* 2015;112(15).
61. Tsai CW, Lai CF, Chao HC, Vasilakos AV. Big data analytics: a survey. 2015;2(21).
62. Karun KA, Chitharanjan K. A review on hadoop—HDFS infrastructure extensions. In: IEEE, information & communication technologies (ICT); 2013. p. 132–137.
63. Chen M, Mao S, Liu Y. Big data: a survey. *Mob Netw Appl.* 2014;19(2):171–209.

64. Al-Hader M, Rodzi A. The smart city infrastructure development and monitoring. *Theor Empir Res Urban Manag*. 2009;4(2):87–94.
65. DeRen L, JianJun C, Yuan Y. Big data in smart cities. *Sci China Inf Sci*. 2015;58:1–12.
66. Khan Z, Kiani SL. A cloud-based architecture for citizen services in smart cities. In: ITAAC workshop 2012, 315–320. IEEE fifth international conference on utility and cloud computing (UCC), Chicago, IL, USA. IEEE; 2012.
67. Khan Z, Kiani SL, Soomro K. A framework for cloud-based context-aware information services for citizens in smart cities. *J Cloud Comput Appl*. 2014;3(14):1–17.
68. Khan Z, Ludlow D, McClatchey R, Anjum A. An architecture for integrated intelligence in urban management using cloud computing. *J Cloud Comput Appl Adv Syst Appl*. 2012;1(1):1–14.
69. Khan Z, Pervez Z, Ghafoor A. Towards cloud based smart cities data security and privacy management. In: 2014 7th IEEE/ACM international conference on utility and cloud computing—SCCTSA workshop, 8th–11th December, London, UK; 2014. p. 806–11.
70. Nathalie M, Symeon P, Antonio P, Kishor T. Combining cloud and sensors in a smart city environment. *EURASIP J Wirel Commun Netw*. 2012;247:1–10.
71. Lu S, Li MR, Tjhi CW, Leen KK, Wang L, Li X, Ma D. A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data. In: Proceedings of the 3rd IEEE international conference on cloud computing technology and science, Nov 29–Dec 1 2011, 618–622, Divani Caravel, Athens, Greece; 2011.
72. Ahvenniemi H, Huovila A, Pinto-Seppä I, Airaksinen M. What are the differences between sustainable and smart cities? *Cities*. 2017;60:234–45.
73. Chourabi H, Nam T, Walker S, Gil-Garcia JR, Mellouli S, Nahon K, Pardo TA, Scholl HJ. Understanding smart cities: an integrative framework. In: The 245th Hawaii international conference on system science (HICSS), HI, Maui; 2012. p. 2289–97.
74. Lyshevski SE. Nano- and microelectromechanical systems: fundamentals of nano- and microengineering. Boca Raton: CRC Press; 2001/2005.
75. Eagle N, Pentland AS. Reality mining: sensing complex social systems. *Pers Ubiquit Comput*. 2006;10:255.
76. Mohamed N, Al-Jaroodi J. Real-time big data analytics: applications and challenges. In: High performance computing & simulation (HPCS), 2014 international conference; 2014. p. 305–10.
77. Ji C, Li Y, Qiu W, Awada U Li K. Big data processing in cloud computing environments. In: Pervasive systems, algorithms and networks (ISPAN), 2012 12th international symposium on IEEE; 2012. p. 17–23.
78. Kalyvas JR, Overly MR, Karlyn MA. Cloud computing: a practical framework for managing cloud computing risk-part I. *Intell Prop Technol Law J*. 2013;25(3):7.
79. Khan Z, Anjum A, Kiani SL. Cloud based big data analytics for smart future cities. In: Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, IEEE Computer Society, 2013. pp. 381–386.
80. Voorsluys W, Broberg J, Buyya R. Introduction to cloud computing. In: Buyya R, Broberg J, Goscinski A, editors. *Cloud computing: principles and paradigms*. New York: Wiley Press; 2011. p. 1–44.
81. Beloglazov A, Abawajj J, Buyya R. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Gener Comp Syst*. 2012;28(5):755–68.
82. Berl A, Gelenbe E, Di Girolamo M, Giuliani G, De Meer H, Dang MQ, Pentikousis K. Energy-efficient cloud computing. *Comput J*. 2010;53(7):1045–51.
83. Mastelic T, Oleksiak A, Claussen H, Brandic I, Pierson JM, Vasilakos AV. Cloud computing: survey on energy efficiency. *ACM Comput Surv*. 2014;47(2):1–36.
84. Buyya R, Beloglazov A, Abawajj J. Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. In: Paper presented at the 2010 international conference on parallel and distributed processing techniques and applications, PDPTA 2010, Las Vegas, USA; 2010.
85. Gokhale A, Schmidt DC, Nataralan B, Wang N. Applying model-integrated computing to component middleware and enterprise applications. *Commun ACM*. 2002;45(10):65–70.
86. Johanson B, Fox A, Winograd T. The interactive workspaces project: experiences with ubiquitous computing rooms. *IEEE Pervasive Comput Mag*. 2002;1(2):67–75.
87. Borkar V, Carey MJ, Li C. Inside big data management: ogres, onions, or parfais? In: Proceedings of the 15th international conference on extending database technology. ACM; 2012. p. 3–14.
88. Chaudhuri S. What next?: a half-dozen data management research goals for big data and the cloud. In: Proceedings of the 31st symposium on principles of database systems. ACM; 2012. p. 1–4.
89. Xiaofeng M, Xiang C. Big data management: concepts, techniques and challenges. *J Comput Res Dev*. 2013;1(98):146–69.
90. Wu X, Zhu X, Wu GQ, Ding W. Data mining with big data. *IEEE Trans Knowl Data Eng*. 2014;26(1):97–107.
91. Forrester Research, Inc. The future of data security and privacy: controlling big data, Forrester US; 2012.
92. Bibri SE. The shaping of ambient intelligence and the internet of things: historico-epistemic, socio-cultural, politico-institutional and eco-environmental dimensions. Berlin: Springer-Verlag; 2015.
93. Crang M, Graham S. Sentient cities: ambient intelligence and the politics of urban space. *Inf Commun Soc*. 2007;10(6):789–817.
94. Straub D, Welke R. Coping with systems risk: security planning models for management decision making. *MIS Q*. 1998;22(4):441–69.
95. Wright D, Gutwirth S, Friedewald M. Shining light on the dark side of ambient intelligence. *Foresight*. 2007;9(2):46–59.
96. Li DR, Wang S, Li DY. Spatial data mining theories and applications. Beijing: Science Press; 2006.
97. Kamberov R. Using social paradigms in smart cities mobile context-aware computing. NOVA IMS: New University of Lisbon; 2015.
98. Riva G, Vatalaro F, Davide F, Alcañiz M. Ambient intelligence: the evolution of technology, communication and cognition towards the future of human-computer interaction. Amsterdam: IOS Press; 2005.

99. Lane ND, Eisenman SB, Musolesi M, Miluzzo E, Campbell AT. Urban sensing systems: opportunistic or participatory? In: *HotMobile'08: proceedings of the 9th workshop on mobile computing systems and applications*, ACM, NY, USA; 2008. p. 11–6.
100. Manzoor A, Patsakis C, Morris A, McCarthy J, Mullarkey G, Pham H, Clarke S, Cahill V, Bourouche M. Citywatch: exploiting sensor data to manage cities better. *Trans Emerg Telecommun Technol.* 2014;25:64–80.
101. Kitchin R. The real-time city? Big data and smart urbanism. *Geo J.* 2014;79:1–14.
102. Lee SH, Han JH, Leem YT, Yigitcanlar T. Towards ubiquitous city: concept, planning, and experiences in the Republic of Korea. In: Yigitcanlar T, Velibeyoglu K, Baum S, editors. *Knowledge—based urban development: planning and applications in the information era*. Hershey: IGI Global, Information Science Reference; 2008. p. 148–69.
103. Perera C, Zaslavsky A, Christen P, Georgakopoulos D. Sensing as a service model for smart cities supported by internet of things. *Trans Emerg Telecommun Technol.* 2014;25(1):1–12.
104. Shin D. Ubiquitous city: urban technologies, urban infrastructure and urban informatics. *J Inf Sci.* 2009;35(5):515–26.
105. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of things for smart cities. *IEEE Internet Things J.* 2014;1(1):22–32.
106. Kyriazis D, Varvarigou T, Rossi A, White D, Cooper J. Sustainable smart city IoT applications: heat and electricity management and eco-conscious cruise control for public transportation. In: *Proceedings of the 2013 IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Madrid, Spain, 2014. pp. 1–5.
107. Mitchell T. *Machine Learning*. New York: McGraw Hill; 1997.
108. Bishop CM. *Pattern recognition and machine learning*. Heidelberg: Springer; 2006.
109. Dlodlo N, Foko T, Mvelase P, Mathaba S. The state of affairs in internet of things research. *Electron J Inf Syst Eval.* 2012;15(3):244–58.
110. Ghose A, Biswas P, Bhaumik C, Sharma M, Pal A, Jha A. Road condition monitoring and alert application: using in-vehicle smartphone as internet-connected sensor. In: *Proceedings of the 2012 IEEE 10th international conference on pervasive computing and communications (PERCOM workshops)*, 19–23 March, Lugano, Switzerland; 2012. p. 489–91.
111. Ren X, Jiang H, Wu Y, Yang X, Liu K. The internet of things in the license plate recognition technology application and design. In: *Proceedings of the 2012 2nd international conference on business computing and global informatization (BCGIN)*, 12–14 October, Shanghai, China; 2012. p. 969–72.
112. Shang J, Zheng Y, Tong W, Chang E. Inferring gas consumption and pollution emission of vehicles throughout a city. In: *Proceedings of the 20th SIGKDD conference on knowledge discovery and data mining (KDD 2014)*; 2014.
113. Ersue M, Romascanu D, Schoenwaelder J, Sehgal A. Management of networks with constrained devices: use cases, IETF internet; 2014.
114. Parello J, Claise B, Schoening B, Qutteck J. Energy management framework. IETF Internet; 2014.
115. Yin J, Sharma P, Gorton I, Akyoli B. Large-scale data challenges in future power grids. In: *Service oriented system engineering (SOSE)*, 2013 IEEE 7th international symposium on IEEE; 2013. p. 324–8.
116. Zheng Y, Liu F, Hsieh H. U-Air: when urban air quality inference meets big data. In: *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM; 2013.
117. Li S, Wang H, Xu T, Zhou G. Application study on Internet of Things in environment protection field. *Lecture Notes in Electrical Engineering.* 2011;133:99–106.
118. Nielsen M. *Reinventing discovery: the new era of networked science*. Princeton: Princeton University Press; 2011.
119. Zheng Y, Chen X, Jin O, Chen Y, Qu X, Liu X, Chang E, Ma W, Rui Y, Sun W. A cloud-based knowledge discovery system for monitoring fine-grained air quality, MSR-TR-2014-40; 2014.
120. Chui M, Löffler M, Roberts R. The internet of things. McKinsey Q. <https://www.mckinsey.com/industries/high-tech/our-insights/the-internet-of-things>. Accessed 10 Jul 2014.
121. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of things (IoT): a vision, architectural elements, and future directions. *Future Gener Comput Syst.* 2013;29(7):1645–60.
122. Kaisler S, Armour F, Espinosa JA, Money W. Big data: issues and challenges moving forward. In: *Proceedings of 46th Hawaii international conference on systems sciences (HICSS)*, IEEE, Wailea, Maui; 2013. p. 995–1004.
123. Qin Y, Sheng QZ, Falkner NJ, Dustdar S, Wang H, Vasilakos AV. When things matter: a survey on data-centric internet of things. *J Netw Comput Appl.* 2016;64:137–53.
124. Porambage P, Ylianttila M, Schmitt C, Kumar P, Gurtov A, Vasilakos AV. The quest for privacy in the internet of things. *IEEE Cloud Comput.* 2016;3(2):36–45.
125. Townsend A. *Smart cities—big data, civic hackers and the quest for a new utopia*. New York: Norton & Company; 2013.
126. Russell S, Norvig P. *Artificial intelligence: a modern approach*. New Jersey: Prentice-Hall, Inc; 1995.