

Article

The Sciences Underlying Smart Sustainable Urbanism: Unprecedented Paradigmatic and Scholarly Shifts in Light of Big Data Science and Analytics

Simon Elias Bibri ^{1,2}

¹ Department of Computer Science, The Norwegian University of Science and Technology, Sem Saelands veie 9, NO-7491 Trondheim, Norway; simoe@ntnu.no

² Department of Architecture and Planning, The Norwegian University of Science and Technology, Alfred Getz vei 3, Sentralbygg 1, 5th floor, NO-7491 Trondheim, Norway

Received: 2 April 2019; Accepted: 14 May 2019; Published: 23 May 2019



Abstract: As a new area of science and technology (S&T), big data science and analytics embodies an unprecedentedly transformative power—which is manifested not only in the form of revolutionizing science and transforming knowledge, but also in advancing social practices, catalyzing major shifts, and fostering societal transitions. Of particular relevance, it is instigating a massive change in the way both smart cities and sustainable cities are understood, studied, planned, operated, and managed to improve and maintain sustainability in the face of expanding urbanization. This relates to what has been dubbed data-driven smart sustainable urbanism, an emerging approach that is based on a computational understanding of city systems that reduces urban life to logical and algorithmic rules and procedures, as well as employs a new scientific method based on data-intensive science, while also harnessing urban big data to provide a more holistic and integrated view and synoptic intelligence of the city. This paper examines the unprecedented paradigmatic and scholarly shifts that the sciences underlying smart sustainable urbanism are undergoing in light of big data science and analytics and the underlying enabling technologies, as well as discusses how these shifts intertwine with and affect one another in the context of sustainability. I argue that data-intensive science, as a new epistemological shift, is fundamentally changing the scientific and practical foundations of urban sustainability. In specific terms, the new urban science—as underpinned by sustainability science and urban sustainability—is increasingly making cities more sustainable, resilient, efficient, and livable by rendering them more measurable, knowable, and tractable in terms of their operational functioning, management, planning, design, and development.

Keywords: smart sustainable urbanism; data science; urban science; sustainability science; urban sustainability; urban sustainability science; data-intensive science; paradigm shift; scholarly shift; big data science and analytics

1. Introduction

We are living at the dawn of what has been termed ‘the fourth paradigm of science’, a scientific revolution that is marked by the emergence of big data and analytics, and by the increasing adoption of the underlying technologies (large-scale compute, data-intensive techniques and algorithms, and advanced mathematical models) in scientific and scholarly research practices. This fourth paradigm, after the three interrelated paradigms of empirical, theoretical, and computational science, is known as data-intensive science, a data-driven, exploration-centered form of science, where big data computing and the underlying enabling technologies are heavily used to help scientists and scholars to manage, analyze, and share data for multiple purposes. Everything regarding science development or knowledge

production is fundamentally changing thanks to the ever-increasing deluge of data. This is the primary fuel of the new age, which powerful computational processes or analytical algorithms burn to generate valuable knowledge for enhanced decision-making and deep insights that are related to a wide variety of practical uses and applications (e.g., decisions to create more sustainable, efficient, resilient, liveable, and equitable cities). Big data science and analytics is a new area of science and technology (S&T). Data science is concerned with the collection, storage, management, processing, and analysis of massive-scale data. Big data denote collections of datasets, whose volume, velocity, variety, exhaustivity, relationality, and flexibility make it so difficult to store, manage, process, and analyze the data using the traditional database systems and software techniques. The scope and impact of big data science and analytics will continue to enormously expand in the upcoming decades as scientific data and data about all branches of science become overwhelmingly abundant and ubiquitously available. Especially, significant progress has been made within data science, information science, computer science, and complexity science, with respect to handling and extracting knowledge and insights from large masses of data, and these have been utilized within urban science.

Big data computing is an emerging paradigm of data science, a typical model that is of multidimensional data mining for scientific discovery over large scale infrastructure. It employs sophisticated computational methods to automatically extract useful knowledge and valuable insights from large masses of data—huge in volume, high in velocity, created in near or real-time, diverse in variety, exhaustive in scope, fine-grained in resolution, relational in structure, and extensible and scaleable in nature—while using data science methods, processes, and systems. Data mining as one of big data analytics techniques provides some of the clearest illustrations of the fundamental concepts and principles of data science. It denotes the computational process of probing colossal datasets in order to find frequent, hidden, and previously unsuspected and unknown patterns and subtle relationships; to make useful, meaningful, and valid correlations from these discoveries; and, to summarize the results in novel ways and then visualize them in understandable formats prior to their deployment for decision-making purposes [1,2], among other things. In recent years, the emphasis has been on the development of new data analytics that utilizes advanced techniques that are designed to manage, process, and analyse enormous datasets, such as data mining and pattern recognition, data visualization and visual analytics, statistical analysis, and prediction and simulation modeling [1,3–5]. These techniques are different from traditional statistical methods that were designed to perform data-scarce science; that is, to identify significant relationships from small, clean sample sizes with known properties. However, many big data analytics projects are still struggling to deliver useful results, often as a result of poor management and the utilization of resources and of inadequately trained data analysts and scientists. Indeed, data science is a heavily applied field where academic programs currently do not sufficiently prepare data scientists for the task [6,7].

There is a strong organizational, institutional, and governmental support for and commitment to big data technology-industry associations and consortia, business communities, scholarly and scientific research communities, policy bodies, and governmental agencies due to its tremendous (yet untapped) potentials and rapidly expanding success in relation to academic research and social practice. The increasing adoption of big data technology and its novel applications (e.g., urban domains), coupled with the ongoing, intensive research, development, and innovation within academic circles and industries, reflects its morphing power. In this respect, its productive and constitutive network operates on all scales of the city, not only in terms of revolutionizing science, transforming knowledge, advancing academic studies, and producing new discourses, but also in terms of creating new technological artifacts and environments, orientating technological innovations, captivating technological investments, shaping organizational and institutional developments, constituting policy bodies, catalyzing important shifts, and fostering major transitions.

Of particular relevance, a new era is presently unfolding, wherein smart sustainable/sustainable smart urbanism is increasingly becoming data-driven. A computational understanding of city systems that brings urban life to a set of logic, calculative, and algorithmic procedures, and that employs

new scientific methods and principles based on data-intensive science, as well as an endeavor of drawing together and interlinking urban big data to provide synoptic city intelligence are at the heart of such urbanism. This is increasingly directed for improving, advancing, and maintaining the contribution of both sustainable cities and smart cities to the goals of sustainable development in an increasingly urbanized world. This is underpinned by epistemological realism and instrumental rationality. Epistemological realism is a subcategory of objectivism that holds that what is known about an object exists independently of human mind. In this context, realist epistemology posits that there exists an external reality that operates independently of an observer and that can be objectively and accurately measured, tracked, analysed, modeled, simulated, and visualized to reveal the world as it actually is [8]. Instrumental rationality is a pursuit of any suitable means necessary to achieve a specific end. Specifically, it is practical reasoning serving for making decisions on how to efficiently perform technical tasks, solve problems, overcome challenges, and resolve conflicts by regarding the factors that are involved in a situation as variables to be controllable and knowable. As such, it underpins the conception that cities can be operated, managed, planned, and developed through a set of data levers and analytics and that urban issues can be solved through a range of technical solutions thanks to the ability of probing the deluge of urban data in neutral, value-free, and objective ways to reveal the truth about cities. Further, epistemological realism and instrumental reality sustain and are informed by urban science (a field in which data science and analytics is practiced), which seeks to make cities more measurable, knowable, and controllable in terms of their operational functioning, management, planning, and development, and thereby more sustainable, resilient, efficient, and equitable. These practices are indeed becoming highly responsive to a form of data-driven urbanism that is the key mode of production for what have widely been termed smart sustainable/sustainable smart cities, whose monitoring, understanding, and analysis are, accordingly, increasingly relying on the core enabling technologies of big data analytics.

However, cities are complex *par excellence*, and thus multifaceted, dynamically changing, contextual, contingent systems, and of wicked problems that are not easily steered and of contestation that is not easily captured, and that urban issues are often best solved through solutions that involve social and political interventions. From an analytically different angle, based on the premises of the social shaping of science-based technology or the social construction of scientific knowledge and its practical application, which relate to the philosophical framework of Science, Technology, and Society (STS) (e.g., [9]), the realist assumptions, which posit that urban science can reveal fundamental truths about the city, remain flawed [8].

Against the preceding background, this paper examines the unprecedented paradigmatic and scholarly shifts that the sciences underlying smart sustainable urbanism are undergoing in light of big data science and analytics and the underlying enabling technologies, as well as discusses how these shifts intertwine with and affect one another in the context of sustainability. I argue that the new urban science—as underpinned by sustainability science and urban sustainability—is increasingly making cities more sustainable, resilient, efficient, and livable by rendering them more measurable, knowable, and tractable in terms of their operational functioning, management, planning, design, and development.

This paper consists of six sections. Section 2 introduces, describes, and discusses the conceptual and theoretical constructs that make up this study. In Section 3, a survey of related work is provided. Section 4 examines the key scientific and paradigmatic shifts that are being instigated by big data science and analytics in relevance to smart sustainable urbanism. Section 5 addresses the scholarly shifts with respect to building the new urban science and establishing the related research domain, as well as urban knowledge discovery/data mining and big data studies. This paper ends, in Section 6, with discussion and conclusion

2. Conceptual and Theoretical Background

2.1. The Scientific Method

The scientific method—hypothesize, model, test—is an empirical method of knowledge acquisition and production, which has characterized the development of natural science for at least four centuries. It is constructed around testable hypotheses. In more detail, it involves formulating hypotheses, by means of induction, based on careful observations. These entail the rigorous skepticism (questioning doubt towards items of putative knowledge) about what is observed, given that cognitive assumptions regarding how the world works influence how scientists interpret an object of perception. The percept can bind sensations from multiple senses into a whole, or it describes a perception that is independent from perceivers. In addition, the scientific method involves experimental and measurement-based testing of deductions (a process of reasoning from premises to reach a logically certain conclusion) that are drawn from the hypotheses, and refinement or rejection of the hypotheses based on the experimental findings. In regard to making observations and asking questions about the natural world as part of the scientific method, scientists and researchers are naturally inquisitive and curiosity-driven, so they often develop hypotheses about why things are the way they are. The chosen hypothesis (based on certain selection criteria) leads to predictions that can be tested in various ways. Its most conclusive testing comes from reasoning based on carefully controlled experimental data. Depending on how well additional tests match the predictions, the chosen hypothesis may require alteration, expansion, or even elimination. If this hypothesis becomes very well supported, and thus survived testing, a general theory may be developed. In the coherentist approach to science development, a theory is validated if it makes sense of observations and experiments as part of a coherent whole.

The scientific method consists of nine steps, namely:

1. Make an observation
2. Ask a question
3. Conduct background research
4. Form a hypothesis
5. Test hypothesis: design and perform a study
6. Analyze the data: observe and record
7. Draw conclusion
8. Replicate the study in its entirety
9. Report results/share findings.

There are some difficulties in a formulaic statement of the scientific method. Indeed, the above steps are not always done in the same order, and not all of them take place in every scientific inquiry, nor to the same degree. Although they vary from one field of inquiry to another, they are frequently the same from one to another. Moreover, despite the scientific method that is being often presented as a fixed sequence of steps, the steps are better considered as general principles [10]. Nevertheless, the overall process of scientific inquiry involves making hypotheses, deriving predictions from them as logical consequences, and then carrying out experiments that are based on those predictions to determine whether the original hypothesis was correct [11]. To the discredit of the scientific method, some philosophers of science have argued that there is no scientific method, and the whole idea of theories of the scientific method is yester-year's debate. Next, I describe and discuss further the key elements of the scientific method

2.2. Hypothesis and Hypothesis Testing

The scientific method is based on the assumption that everything in the universe is linked by cause and effect—causality. There is a logical explanation for all the observed behavior of natural phenomena. Accordingly, a scientist or researcher using the scientific method of inquiry starts by

making assumptions about what to expect to find after conducting research. This initial assumption is called a hypothesis.

A hypothesis means an idea or explanation made from limited evidence, which thereby constitutes a starting point for further investigation through observation and experimentation. In other words, it is a postulation that is based on knowledge obtained while seeking answers to the question formed after making an observation. In science—in contrast to a theory which is a tested, well-substantiated, unifying explanation for a set of verified, proven factors—a hypothesis, whether be it very specific or broad is either a suggested explanation for an observable phenomenon, or a reasoned prediction of a possible causal correlation among multiple phenomena. With respect to the latter, the hypothesis statement is a prediction about what one think will happen or is happening in one's experiment that can be tested to answer one's question. A hypothesis is a novel suggestion that no one wants to believe is guilty, until found to be effective through testing by conducting an experiment. The purpose of an experiment is to determine whether observations agree with or conflict with the predictions derived from a hypothesis. The application of hypothesis testing is predominant in data science. It is imperative to simplify and deconstruct it, and hypothesis testing based on data leads us from a novel suggestion to an effective proposition. This pertains to what is termed a statistical hypothesis, one that is testable on the basis of observing a process that is modeled via a set of random variables [12]. The most common application of hypothesis testing is in the scientific interpretation of experimental data, which the philosophy of science naturally studies. This is equally true of hypothesis testing, which can justify conclusions even when no scientific theory exists. Commonly, two statistical data sets are compared, or a data set that was obtained by sampling is compared against a synthetic data set from an idealized model. A hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis that proposes no relationship between two data sets. A null hypothesis generally asserts that there is no meaningful relationship between two observed phenomena.

Performing a hypothesis testing consists of a seven-step process:

1. Make assumptions
2. Take an initial position
3. Determine the alternate position
4. Set acceptance criteria
5. Conduct fact-based tests
6. Evaluate the results and confirm whether the evaluation supports the initial position, and whether there is confidence in that the results are not due to chance
7. Reach one of the following conclusion: Reject the original position in favor of alternate position or fail to reject the initial position.

A thought experiment [13] considers some hypotheses or principles for the purpose of thinking through its consequences. The common goal of a thought experiment is to explore the potential consequences of the principle in question: 'A thought experiment is a device with which one performs an intentional, structured process of intellectual deliberation in order to speculate, within a specifiable problem domain, about the potential consequents (or antecedents) for a designated antecedent (or consequent)' [14] (p. 150). An explanatory thought experiment is put forward as explanation using relevant principles, (e.g., parsimony, the idea that, all other things being equal, we should prefer a simpler explanation over a more complex one), and is generally expected to seek consilience: fitting well with other accepted facts related to the phenomena [15]. This new explanation is used to make falsifiable predictions that are testable in nature. Disproof of a prediction is evidence of progress, which is done through the observation of natural phenomena, but also through experimentation that tries to simulate the natural events under controlled conditions as appropriate to the discipline. Experimentation is especially important in science to help establish causal relationships or to avoid the correlation fallacy. In addition, scientists may generate a model, an attempt to describe the phenomenon

in terms of a logical, physical, or mathematical representation, and to generate new hypotheses that can be tested on the basis of observable phenomena. However, a hypothesis is either modified or discarded based on the outcome of the observation and experimentation.

2.3. Hypotheses Versus Scientific Theories

If the hypothesis survives testing, then it may become adopted into the framework of a scientific theory. Commonly, a large number of hypotheses, surviving testing, can be logically bound together by a single theory. As often integrating and generalizing many hypotheses, a scientific theory is concise, coherent, systematic, predictive, and broadly applicable, as well as strongly supported by many different lines of evidence. Moreover, while a theory is formulated according to most of the same scientific principles as a hypothesis, it is a logically reasoned, self-consistent model for describing the behavior of much broader sets of phenomena. In general, theories are created to explain and predict phenomena and, in many cases, to challenge, enhance, and extend existing knowledge within the limits of critical bounding assumptions. A scientific theory is a well-substantiated explanation of some aspects of the natural world—based on a body of facts that have been repeatedly confirmed through observation and experiment. However, it may be modified, overturned, or disregarded if warranted by new evidence and perspectives (e.g., using inductive empiricism or data-driven science approaches based on big data analytics for knowledge discovery).

2.4. Paradigm and Paradigm Shift

The word ‘paradigm’ can be used to describe or indicate a typical pattern, archetype, or model of something (e.g., smart sustainable cities are a leading paradigm of urbanism). In this regard, a paradigm, arguably, does not impose a rigid approach, but can be taken more or less flexibly. Endeavoring to give this concept its contemporary meaning, the historian of science Kuhn [16] uses the word to refer to the set of concepts and practices that define a scientific discipline at any particular period of time. In his book, *The Structure of Scientific Revolutions* (1962, 1996), Kuhn defines a scientific paradigm as: universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of practitioners. In short, the existence of a single (historically contingent or conditioned) reigning paradigm, a world view that dominates science for a period of time during which that world view is (determined to be) extended is characteristic of the sciences. This implies that scientific facts are never really more than opinions, or theories are just beginnings, whose dominance is transitory and far from conclusive. However, in more detail, a paradigm denotes, in science and philosophy, a distinct set of thought patterns, which includes theories, research methodologies, assumptions, and standards for what constitutes legitimate contributions to a scientific domain. According to Kuhn (1962) [16], it entails the explanatory power—and hence, the universality of a theoretical model—and its broader institutional implications for the structure, organization, and practice of science. This power means the ability of a theory (or a group of related theories) to effectively explain the subject matter it pertains to, and a theoretical model is a theory that is designed to provide explanations within a scientific domain for a community of practitioners—in other words, a scientific discipline-shared intellectual framework encompassing the basic assumptions, ways of reasoning, and methodologies that are universally acknowledged by a scientific community [17].

Characterizing a paradigm shift is a fundamental change in the basic concepts and experimental practices of a scientific discipline. Following the Kuhnian paradigm shift, a scientific revolution represents an epistemological shift, or it denotes a significant change within belief systems. This occurs when the dominant paradigm (world view) becomes eroded, when scientists encounter inconsistencies and anomalies (that are typically brushed away as acceptable levels of error, or simply ignored and not dealt with, yet entailing various levels of significance to the practitioners of science at the time) that mount up to such an extent that researchers can no longer work within the philosophical and theoretical framework of a scientific discipline. This implies that these inconsistencies and anomalies cannot be explained by the universally accepted paradigm within which scientific progress has thereto

been made. The history of science has shown that the turbulence that sets in can lead to a paradigm shift that takes place over years or decades, rather than centuries. According to Kuhn ([18], p. 12), 'successive transition from one paradigm to another via revolution is the usual developmental pattern of mature science'. Accordingly, the sciences go through alternating periods of normal science when an existing model of reality dominates a protracted period of puzzle-solving, and revolution, when the model of reality itself undergoes sudden drastic change. This involves the science to be thrown into a state of crisis, according to Kuhn, during which new ideas are tried. Eventually, a new paradigm is formed, which gains its own new followers, and an intellectual battle takes place between them and the hold-outs of the old paradigm. Sometimes, the convincing force is just time itself and the human toll it takes, Kuhn stated, using a quote from Max Planck: 'a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it' ([18], p. 150). By and large, a paradigm shift in a scientific discipline should meet three conditions or encompass three criteria: it must be grounded in a meta-theory: theory about theory, be accepted by practitioners of a scientific community, and have a body of successful practices [16].

3. A State-of-the-Art Review: 'Small Data' and 'Big Data' Studies and City Analytics

Big data are referred to with respect to their humongous size and wide variety, with a particular focus on the deluge of urban data (i.e., datasets that were collected and coalesced through data warehousing for wide-city uses) that are directed towards transforming the knowledge of, and the sciences underlying, smart sustainable/sustainable smart urbanism, in particular, in relation to sustainability and the integration of its dimensions using urban science. This epitomizes a sea change in the kind of data that we generate about city systems across multiple urban domains as to what happened, and where, when, why, and how, as well as what will happen and what should be done, in the context of smart sustainable/sustainable smart cities. In this respect, Bibri and Krogstie (2018) [2] argue that data mining (or knowledge discovery) has innovative potential to revolutionize city analytics in the form of 'big data' studies by providing a novel way of thinking data-analytically about sustainable urbanism/urban sustainability problems. This entails enabling well-informed, knowledge-driven decision-making and enhanced insights pertaining to city operations, functions, strategies, designs, strategics, and policies across multiple urban domains for improving the contribution of smart sustainable/sustainable smart cities to the goals of sustainable development, thereby advancing smart sustainable urbanism. Especially, 'small data' studies using data collection and analysis methods, such as questionnaire surveys, focus groups, case studies, participatory or non-participatory observations, interviews, content analyzes, and ethnographies—'capture a relatively limited sample of data that is tightly focused, time and space specific, restricted in scope and scale, and relatively expensive to generate and analyze' [4] (p. 3) For example, as pointed out by Batty et al. (2012) [19], the mainstream analytical tools of transportation engineering, such as origin/destination matrices, are based on semantically rich data collected by means of field surveys and interviews. In a nutshell, much of what we know regarding cities to date has been gleaned from studies that are characterized by data scarcity [5] and based on traditional data and analysis methods. This form of academic and scientific research in the domain of sustainable urbanism has prevailed for three decades or so. This has consequently influenced the way that sustainability as underpinned by empirical investigations predominately based on such methods has been adopted as a set of practices in urban planning and development [20].

Big data can be used to overcome the constraints and limits of traditional data collection and analysis methods, namely their high cost, infrequent periodicity, quick obsolescence, incompleteness, inaccuracy, inflexibility, and inherent subjectivity and biases in the domain of smart sustainable/sustainable smart urbanism [1,4,19,21]. These issues have long affected the robustness and reliability of the research results (theories, generalizations, models, and other valid forms of knowledge) within the field of sustainable urbanism. This, in turn, has influenced urban practices in terms of the application of the principles of sustainability in urban planning and development. In the context of sustainable urbanism,

many studies investigating or referring to other research work been carried out on, the correlation between travel behavior, mobility, social equity, life quality, energy consumption, and other indicators of environmental and social sustainability performance, on the one hand, and density, compactness, diversity, mixed-land use, and other typologies through which sustainable urban forms can be achieved, on the other hand, point implicitly or explicitly to the disadvantages of the traditional data collection and analysis methods and how these compromise the value of the obtained research results [20,22,23]. These studies usually generate non-conclusive, weak, limited, unreliable, conflicting, or uncertain results [1]. This also relates the issue of sustainable urban forms being problematic and of a challenging nature to deal with. The interested reader can gain further insights by turning to Bibri and Krogstie (2017) [20], where a detailed discussion is provided on several topics that are related to sustainable urbanism, more specifically sustainable urban forms, including, in addition to big data analytics as an alternative to traditional data collection and analysis methods for investigating sustainable urban forms, the role of big mobility data in evaluating the environmental and socio-economic performance of sustainable urban forms, as well as urban simulation models as an innovative approach into strategically assessing and optimizing the contribution of sustainable urban forms to sustainability that is based on big data analytics. Overall, big data are seen as the most scalable and synergic asset or resource for smart sustainable/sustainable smart cities of the future, and they constitute the fundamental ingredient for the next wave of city analytics. The potential and hope of big data lies in transforming the knowledge of smart sustainable/sustainable smart cities through the creation of a data deluge that can, through analytics, provide much more sophisticated, finer-grained, wider-scale, real-time understanding and control of the kind of complex aspects and intractable issues of urbanity [2], just as smart cities [4].

There are many topical studies that provide or suggest new ways of mobility enhancement, transport management, energy management, environmental management, water management, waste management, healthcare, education, citizen services, and so forth across multiple urban domains, both in relation to sustainability and efficiency (e.g., [24]). In more detail, some of these studies offer a set of applications of data mining that are related to the different aspects of the systematic study of smart sustainable/sustainable smart cities and related problems that pertain to sustainability. In this regard, a recent study (see Bibri 2018 [24]) compares traditional methods and data mining process for energy consumption analysis, as well as discusses the energy consumption data with these parameters: public building, structure, construction, and behavior pattern [24]. This is an attempt to fill the existing gap by utilizing data mining in the energy efficiency evaluation of buildings. Here, data-driven smart energy management involves turning big data to valuable insights through advanced analytics. Khan et al. (2012) [25] compare the effectiveness and performances of several data mining techniques as for predicting irrigation water demand. Addressing various issues that are related to healthcare, Milovic and Milovic (2012) [26] state that data mining in health care aids in organizing large amount of data, using advanced technologies for automation, conducting early diagnosis, maintaining the security, predicting new trend, and so on. The authors further suggest the classification and regression, association rule, cluster analysis, and text mining techniques for an analysis of healthcare data. Benevolo et al. (2016) [27] discuss six objectives of mobility in the context of smart cities, namely pollution, noise pollution, traffic blocking, transfer speed, transfer cost, and people safety. The authors suggest four key factors: public companies, private companies, local government agencies, and the integration of these three to form Integrated Transport System (ITS), with a focus on ICT based ITS, which includes video surveillance for security, traffic control, and traffic data collection system, as well as other innovative solutions such as the use of ICT for smartphone-based integrated ticketing system, car sharing, and car reservation. Sin and Muthu (2015) [28] identify the challenges and issues facing education, including performance prediction in learning; designing courseware, assessment, and research; and, predicting future failures and finding solutions. The authors suggest the use of different open source data processing platforms, such as MongoDB, Hadoop, and Orange, in relation to data mining. Overall, the coupling of big data analysis

and computational modeling and simulation can open new horizons for city analytics and planning in the context of smart sustainable/sustainable smart urbanism.

However, the role of data mining technique in transforming the knowledge of smart sustainable/sustainable smart cities as a holistic approach to urbanism is largely ignored or barely explored to date [1,24], notwithstanding the relevance and the usefulness of this advanced data analytics technique in relation to the future form of smart sustainable/sustainable smart urbanism. Indeed, in the near future, the core enabling technologies of big data analytics, namely digital sensing networks, data processing platforms, cloud and fog/edge computing models, and wireless communication networks, will be the dominant mode of monitoring, understanding, and analyzing smart sustainable/sustainable smart cities, so as to improve, advance, and maintain their contribution to the goal of sustainable development through enhancing and optimizing their operations, functions, services, strategies, and policies, in line with the vision of sustainability [1].

4. Paradigmatic Shifts in Scientific Development and Discovery

Big data science and analytics is instigating a radical change in the basic concepts, assumptions, and experimental practices of science—thought patterns or ways of reasoning within the ruling theory of science, marking a paradigm shift from the dominant scientific way of looking at the world. The use of big data analytics as a framework has tremendous potential to advance or replace the prevailing scientific method. It has been argued that the unfolding and soaring deluge of data renders the scientific method obsolete, and it heralds the end (or decline) of theory resulting from the generalizations obtained from the experiments conducted by scientists as part of the scientific method. On the whole, the current model of reality, which has dominated a protracted period of puzzle-solving, is undergoing sudden drastic change, with wide ranging societal implications.

4.1. On the Old and New Way of Doing Science

One of the key questions the philosophers of science are concerned with and actively study is: Why do scientists continue to rely on models and theories which they know are partially inaccurate, among others? Indeed, how many of the theoretical models have been modified, overturned, or disregarded in light of new evidence and novel perspectives? How many of the theoretical models that have constantly been corrected or refined are still useful? And how many of the yet useful theoretical models are able to consistently, if imperfectly, explain the world around us? The attempt to answer these questions prompts us to question the accuracy of such models in the first place. The term 'model' enjoys a broad range of uses in science. Here, I shall examine one important use of this term: theoretical models that are quite distinct from other conceptions, sometimes called models. The philosophers of science have highlighted the importance of models, and they have claimed that their consideration will illuminate the structure, interpretation, and development of scientific thinking. A theoretical model is a group of related theories that are designed to provide explanations within a scientific domain for a community of practitioners. As a coherent whole, it is characterized by [17]:

1. involving a conceptual foundation for a scientific domain;
2. understanding and describing problems within that domain and specifies solutions;
3. being grounded in prior empirical findings and scientific literature;
4. being able to predict outcomes in situations where these outcomes can occur far in the future;
5. guiding the specification of a priori postulations and hypotheses;
6. using rigorous methodologies to investigate them; and,
7. providing a framework for the interpretation and understanding of the unexpected results of scientific investigations.

The history of science is replete with epistemological breaks, what Bachelard (1986) [29] refers to as unthought/unconscious structures that are immanent within the realm of the sciences. It, as asserted by Bachelard (1986) [29], consists in the formation and establishment of these epistemological

breaks, and then the subsequent tearing down of the obstacles. The latter stage is an epistemological rupture—where an unconscious obstacle to scientific thought is thoroughly ruptured or broken away from. Indeed, as asserted by Foucault (1970 [30], pp. xxi–xxii), knowledge is a matter of episteme: a pre-cognitive space that determines ‘on what historical *a priori*, and in the element of what positivity, ideas could appear, sciences be established, experiences be reflected in philosophies, rationalities be formed, only, perhaps, to dissolve and vanish soon afterwards’. In a nutshell, the history of science has shown that the turbulence that sets in can lead to an epistemological break or paradigm shift that takes place over varied periods of time. This implies that, among others, all of the theoretical models are flawed, if not wrong, and increasingly, we can do better or succeed without them. For example, while quantum mechanics is yet another theoretical model that is flawed, no doubt a caricature of a more complex underlying reality, quantum mechanics based on statistical analysis offers a way better picture of reality. The basic argument is that, the more we learn about natural phenomena, the further we find ourselves from a theoretical model that can explain them. Nevertheless, we do not have to settle for theoretical models, as we grow up in an era of massively abundant data, a deluge or corpus that is indeed being treated as a laboratory of the human condition, thereby providing the raw material for sifting through the most measured and tracked age in history. In the upcoming Exabyte/Zettabyte Age, the analysis of the deluge of data will generate valuable knowledge and deep insights, which will be good enough to enhance human decisions and thus practices, as well as advance and accelerate progress on science. In the data-intensive approach to scientific discovery, no causal analysis and assumptions about any kind of relationships are required; moreover, such an approach applies sophisticated methods (i.e., advanced simulation models informed by the science of complexity in terms of the common dynamical properties, processes, and behaviors that characterize complex systems) for predicting outcomes far in the future with unprecedented accuracy, as well as uses rigorous frameworks (e.g., data mining, statistical analysis, etc.) to answer challenging analytical questions. (See Bibri 2018 [24] for a detailed account and discussion of complexity science and complex systems in relation to smart sustainable cities of the future). With the deluge of the data that are available thanks to its new and extensive sources, the numbers (overwhelming data quantities) speak for themselves, and many complex phenomena and theories of human behavior can be tracked and measured with unprecedented fidelity in a world where large-scale computation, novel data-intensive techniques and algorithms, and advanced mathematical models—technologies underpinning big data computing/analytics—replace every other tool that might be brought to bear.

The big target is science where the scientific approach is built around testable hypotheses, a way of doing science that has prevailed for hundreds of years. Hypothesized models, as systems visualized in the minds of scientists, are tested, and subsequently experiments confirm or falsify the models of how the world works. In science, the term ‘model’ can have different meanings, depending on the context of its use (e.g., a physical model of a system that can be used for demonstrative purposes, an idea about how something works, an object or process that is used to describe and explain phenomena that cannot be experienced directly, etc.). Models are central to what scientists do in their research and when communicating their explanations related to new scientific theories. As a research method, scientific modeling means creating a mathematical or logical model—a set of equations that indirectly represents a real-world system or process—as a basis for simulation: an imitation or emulation of the operation of a real-world system. These equations (often characterizing the nature of the reciprocal relationships pertaining to the system) are based on relevant information regarding the system and on sets of hypotheses about how the system works. The act of simulating the system requires that a model be developed, which represents the key characteristics, behaviors, and functions of that system. The model represents the system itself, whereas the simulation represents the operation of the system over time. In this regard, given a set of parameters, a model can generate expectations about how the system will behave in a particular situation. A model and the hypotheses it is based upon are supported when the model generates expectations that match the behavior of its real-world counterpart. For what it entails in terms of oversimplification, modeling often involves idealizing

the system in some way—leaving some aspects of the real system out of the model in order to make the model computationally easier to work with in terms of simulation. Regardless, for a scientific hypothesis to be meaningfully tested, it must be falsifiable, which implies that it is possible to identify a possible outcome of an experiment or observation that conflicts with the predictions deduced from the hypothesis. In relation to hypothesis testing, it is important for scientists to understand the underlying mechanisms that connect two variables, since correlation does not imply causation, and hence, no conclusions should be drawn simply on the basis of correlation between two variables. A statistical hypothesis is testable on the basis of observing a process that is modeled via a set of random variables, and statistical hypothesis test is a method of statistical inference. In the statistics literature, statistical hypothesis testing plays a fundamental role [31] in statistical inference, as well as in the whole of statistics. As concluded by Lehmann (1992) [32] in a comprehensive review, despite its shortcomings, the new paradigm of statistical hypothesis, and the many developments carried out within its framework continue to play a central role in both the theory and practice of statistics. However, significance testing has particularly been the favored statistical tool in some experimental social sciences [33], while other fields have favored the estimation of parameters (e.g., effect size). It is used as a substitute for the traditional comparison of the predicted value and experimental result at the core of the scientific method. While hypothesis testing is of continuing interest to philosophers [34,35], statisticians, in other contexts, discuss much of their criticism of it; especially correlation does not imply causation. Covering a wide variety of issues, the criticism of statistical hypothesis testing fills volumes (e.g., [36–41]). As a bias related to science when performing experiments to test hypotheses, a scientist may have a preference for one outcome over another [42,43]. Nonetheless, eliminating this bias can be achieved by careful experimental design and transparency, as well as a thorough peer review [44,45]. A normal practice for independent researchers after the publication of the results of an experiment is to double-check how the research was carried out, and to follow up by performing similar experiments to determine how dependable the results might be [46]. Taken in its entirety, the scientific method allows for highly creative problem solving while minimizing any effects of the confirmation bias and other subjective biases [47].

The prevailing scientific approach is increasingly becoming obsolete as faced with the unfolding and soaring deluge of data related to a large number and variety of phenomena. Such deluge is available for scientific exploration within many different disciplines and fields. The data that were collected from various sensors, e.g., remote sensing technologies, are analyzed to extract useful knowledge and valuable insights for societal benefits, where a large number of scientists can collaborate in terms of designing, operating, and analyzing the products of sensor devices and networks for scientific studies. The data that were produced from several scientific explorations require advanced tools to facilitate the efficiency of their management, processing, analysis, validation, visualization, and dissemination, while preserving the intrinsic value of the data. Further, data-intensive approach to science is a promising epistemological shift, where colossal amounts of data allow for us to say that correlation is enough using data science systems, processes, and methods, specifically big data computing and the underlying enabling technologies. We can analyze the data regarding hypotheses about what they might show, and accordingly, we can stop looking for models. We can throw the numbers into huge computing clusters (dedicated data processing platforms) and let statistical or data mining algorithms discover patterns and make correlations from these discoveries where science cannot. It comes at no surprise that the application and use of big data analytics is increasingly gaining traction and foothold in many scientific research fields, taking over the prevailing scientific method.

4.2. Data-Intensive Science as an Epistemological Shift and Its Underpinnings

Data-intensive science/scientific development as a new paradigm has emerged as a result of the recent advances in data science systems, processes, and methods, and thus big data computing and the underlying enabling technologies. Turing award winner Jim Gray envisions data science as the new paradigm of science, and asserts that everything about science is changing, because of the impact of

advanced ICT and the evolving data deluge [48]. The Exabyte Age is upon us. Data-intensive scientific discovery is the fourth paradigm of science where science involves the exploration and mining of scientific data and using advanced big data analytics techniques to unify theory, simulation, and experimental verification [48]. The first paradigm is where science used empirical methods thousands of years ago; the second paradigm is where science became a theoretical field a few hundred years ago, involving the process of generating and testing hypotheses; and, the third paradigm is where science used calculation, conducting simulation and verification by computation in recent decades [48].

Data-intensive science as an epistemological shift mainly involves two positions. The first position is a form of inductive empiricism in which the data deluge, through analytics, as manifested in the data being wrangled through an array of multitudinous algorithms to discover the most salient factors regarding complex phenomena, can speak for itself free of human framing and subjectivism, and without being guided by theory (as based on conceptual foundations, prior empirical findings, and scientific literature). As argued by Anderson (2008) [49], 'the data deluge makes the scientific method obsolete' and that within big data studies 'correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all'. This relates to exploratory data analysis, which may not have pre-specified hypotheses; unlike confirmatory data analysis that was used in the traditional way of doing science that does have such hypotheses. The second position is data-driven science, which seeks to generate hypotheses out of the data rather than out of the theory, thereby seeking to hold to the tenets of the scientific method and knowledge-driven science ([50], p. 613). Here, the conventional deductive approach can still be employed to test the validity of potential hypotheses, but on the basis of guided knowledge discovery techniques that can be used to mine the data to identify such hypotheses. It is argued that data-driven science will become the new dominant mode of scientific method in the upcoming Exabyte/Zettabyte Age, because its epistemology is suited to exploring and extracting useful knowledge and valuable insights from enormous, relational datasets of high potential to generate more holistic and extensive models and theories of entire complex systems rather than parts of them, an aspect that traditional knowledge-driven science has failed to achieve [5,50].

The best practical example of inductive empiricism that is associated with the recent epistemological shift in science is the shotgun gene sequencing by John Craig Venter, while using statistical analysis as a big data analytics technique. Venter went from sequencing individual organisms to sequencing entire ecosystems, enabled by high-speed sequencers and supercomputers that statistically analyze the data they produce. As an alternative to the costly option of using supercomputers, advanced data processing platforms are designed for handling the storage, analysis, and management of large datasets directed for scientific and academic explorations. As an example of such platforms, Hadoop MapReduce is widely used in this regard due to the suitability of its functionalities with respect to dealing with colossal amounts of data, as well as to its advantages that are associated with load balancing, cost effectiveness, flexibility, and processing power. Hadoop allows for distributing the processing load among the cluster nodes, which enhances the processing power; to add or remove nodes in the cluster according to the requirements; to make the homogenous cluster with various group of machines; and, to handle unstructured data [24]. Further to the point, in his endeavor of sequencing the air in 2005, Venter discovered thousands of previously unknown species of bacteria and other life-forms. He can tell you almost nothing about the species he found; does not know what they look like, how they live, or much of anything else about their morphology; and, does not even have their entire genome. A statistical blip—a unique sequence that, being unlike any other sequence in the database, must represent a new species—is all he has, which would be impossible to achieve with the old way of conducting scientific research or doing science. The point is that, Venter has advanced biology more than anyone else of his generation by analyzing data with high-performance computing resources, thanks to big data computing and the underlying enabling technologies. The future potential of data-intensive science is so enormous that this kind of thinking is poised to go mainstream, pervading many different scientific and academic fields. While learning to use and mastering data processing

platforms or supercomputers may be challenging, the opportunity is tremendous in the sense that the new availability of huge amounts of data, along with the statistical analysis and data mining tools to crunch these numbers, offers a whole new way of explaining and understanding the world around us.

Following the Kuhnian paradigm shift, data-intensive science is a paradigmatic break with the current paradigm of science. As such, it represents universally recognized scientific achievements that, for the current period of time, provide model problems and solutions for a community of practitioners as associated with both inductive empiricism and data-driven science, that is:

- what is to be observed and scrutinized;
- the kind of questions that are supposed to be asked and probed for answers;
- how these questions are to be structured;
- what predictions made by the primary theory within the discipline;
- how the results of scientific investigations should be interpreted; and,
- how an experiment is to be conducted, and what equipment is available to conduct the investigation.

As to inductive empiricism, the scientific method becomes obsolete due to the massive data available for scientific exploration, correlation supersedes causation, and coherent models and unified are not required for scientific advancement. Concerning data-driven science, hypotheses are generated out of the data rather than out of the theory; the deductive approach is used to test the validity of potential hypotheses on the basis of guided knowledge discovery techniques; useful knowledge is explored and extracted from massive, interconnected datasets; and, holistic and extensive models and theories of entire complex systems can be generated.

In light of the above, data-intensive science represents a fundamental change in thought patterns, including theories, assumptions, and experimental practices. Kuhn (1962) [16] suggests that the history of science can be divided up into times of normal science and briefer periods of revolutionary science. He characterizes normal science (when scientists add to, elaborate on, and work with a central, accepted scientific theory) as the process of observation and puzzle solving that takes place within a paradigm, whereas revolutionary science occurs when one paradigm overtakes another in a paradigm shift (e.g., [51]). He asserts that, during times of revolutionary science, anomalies refuting the accepted theory have built up to such a point that the old theory is broken down and a new one is built to take its place in a paradigm shift. Each paradigm has its own distinct questions, aims, procedures, and interpretations. The choice between paradigms involves setting two or more depictions against the world and deciding which likeness is most promising. With the above in mind, it has been argued that scientists are currently encountering inconsistencies and anomalies, which have partly brushed away as acceptable levels of error for quite sometime, and they have partly been ignored and not dealt with, with different levels of significance to the practitioners of science. These inconsistencies and anomalies have been mounting up towards a point when researchers in many scientific fields are increasingly favoring the data-intensive approach to science, thereby no longer working within the existing framework of science. In short, a significant number of inconsistencies and anomalies are arising, and data-intensive science is making sense of them. As repeatedly shown by the history of science, there again is a turbulence setting in that is triggering a paradigm shift in science as being driven and shaped by data science, and hence the emerging advancements and innovations in big data computing and the underlying enabling technologies. Indeed, data-intensive science is taking an identifiable form and increasingly gaining its own new followers, which are currently in the phase of intellectual conflict with the hold-outs of the old paradigm of science. In this regard, this new scientific truth is not only making its opponents see the light, but eventually die, manifested in the new generation of data science advocates growing up that is familiar with this truth.

In addition, the rationale for the choice of the data-intensive approach to science as an exemplar is a specific way of viewing the current reality, where this view and the status of this exemplar are mutually reinforcing. This paradigm shift in science is so convincing that normally renders the possibility of new epistemological alternatives intuitive, thereby not obscuring the possibility of the existence of

other imageries that are hidden behind the current paradigm of science. Arguably, the conviction that the current paradigm of science is reality tends to disqualify evidence that might undermine the paradigm of science itself, which leads to a build-up of reconciled inconsistencies and anomalies that are determined to accumulate, and thus cause a paradigm shift in science. This is responsible for the eventual revolutionary overthrow of the incumbent paradigm of science, and its replacement by a new one [16]. Yet, the acceptance or rejection of a paradigm is a social process as much as a logical process, an argument that relates to relativism: the idea that knowledge and truth exist in relation to culture, society, or historical context, and are not absolute, or that views are relative to differences in perception and consideration. There is no universal, objective truth according to relativism; rather, each point of view has its own truth. Kuhn (1962 [16], p. 170) denies the accusation of being a relativist later in his postscript: 'scientific development is . . . a unidirectional and irreversible process. Later scientific theories are better than earlier ones for solving puzzles . . . That is not a relativist's position, and it displays the sense in which I am a convinced believer in scientific progress'.

The popularity of the term 'data science' has exploded in the academia where many critical academics see no distinction between data science and statistics. However, many statisticians envision data science as an increasingly inclusive applied field that grows out of traditional statistics and goes beyond traditional analytics. This implies that data science differs from statistics. One key difference is that statisticians are able, by means of data science methods, systems, and processes to develop models for highly complex systems that were unfathomable: incapable of being fully explored or understood, before. In addition, emerging in the wake of big data, data science, as argued by Donoho (2015) [7], does not equate to big data in that the size of the data set is not a criterion to distinguish data science and statistics. Additionally, data science is a heavily applied field where academic programs currently do not sufficiently prepare data scientists for the jobs in that many graduate programs misleadingly advertise their analytics and statistics training as the essence of a data science program [6,7]. From a technical perspective, while statistics emphasizes models grounded in probability theory to deal with data arising from real-world phenomena, and it provides principles and tools for the construction of statistical hypotheses as models that involve, such modeling processes as data generation, evaluation and assessment, prediction, and uncertainty quantification, data science brings to statistics large-scale compute (modern computational infrastructures), data-intensive techniques, algorithmic design and analysis, large datasets, and advanced mathematical models.

Data science, most often linked to the big data explosion, is the amalgamation of numerous parental disciplines, as mentioned above. As an example of capturing this, Blei and Smyth (2017) [52] describe data science as 'the child of statistics and computer science', where the 'child' metaphor appropriately depicts that data science inherits from both its parents, but eventually evolves into its own entity. They further elaborate: 'data science focuses on exploiting the modern deluge of data for prediction, exploration, understanding, and intervention. It emphasizes the value and necessity of approximation and simplification; it values effective communication of the results of a data analysis and of the understanding about the world and data that we glean from it; it prioritizes an understanding of the optimization algorithms and transparently managing the inevitable tradeoff between accuracy and speed; it promotes domain-specific analyses, where data scientists and domain experts work together to balance appropriate assumptions with computationally efficient methods' [52].

Data science is largely seen as the umbrella discipline that incorporates a number of other disciplines. As an interdisciplinary field, data science employs methodologies and practices from across several academic disciplines while morphing them into a new discipline. Data science is often said to include particularly the allure of big data, the fascination of unstructured data, the advancement of data-intensive techniques and algorithms, and the precision of mathematics and statistics. One implication of this is that data science is different from the existing practice of data analysis across all disciplines, which only focuses on explaining data sets. The practical engineering goal of data science: actionable knowledge and consistent patterns for generating predictive models takes it beyond traditional approaches to analytics. Currently, the data in those disciplines and applied

fields that lacked solid theories, like the social sciences and related disciplines, could be utilized to generate powerful predictive models [53]. Cleveland (2001) [54] urges prioritizing extracting applicable predictive tools over explanatory theories from colossal amounts of data. For the future of data science, Donoho (2015) [7] projects an ever-growing environment for open science where data sets used for academic publications are accessible to all researchers. Open science also involves making scientific research available to all levels of an inquiring society, as well as disseminating, sharing, and developing knowledge through collaborative networks. Several research institutes have already announced plans to enhance the reproducibility and transparency of research data [55]. Likewise, other big journals are following suit [56,57]. The future of data science not only exceeds the boundary of statistical theories in scale and methodology, but data science will revolutionize current academia and research paradigms [7]. The scope and impact of data science will, as concluded by Donoho (2015) [7], continue to enormously expand in the upcoming decades as scientific data and data about science itself become overwhelmingly abundant and ubiquitously available. Already, significant progress has been made within data science, information science, computer science, and complexity science with respect to handling and extracting knowledge and insights from big data and these have been utilized within urban science (e.g., [8,58,59]).

Data science, which is the new paradigm of science, employs scientific methods, systems, processes, and algorithms to extract useful knowledge and valuable insights from large masses of data in various forms, both structured and unstructured. It uses theories and techniques that are drawn from many fields within the context of statistics, mathematics, computer science, and information science. Data science (and thus big data analytics) techniques, such as data mining and pattern recognition, statistical analysis, machine learning, data visualization, and visual analytics, and optimization and simulation modeling are largely in the early stages of their development, given that the statistical methods that have prevailed over several decades were originally designed to perform data-scarce science, i.e., to identify significant correlations and relationships from small, clean sample data sizes with known attributes or properties. Nonetheless, recent years have witnessed a remarkable progress within computer science, information science, and data science with regard to handling and extracting knowledge from large masses of data, and these have been utilized in urban science. The evolving big data computing model represents the challenging task of data organization, processing, and analysis associated with the process of knowledge discovery from voluminous, varied, real-time, exhaustive, fine-grained, indexical, dynamic, flexible, and relational data. This approach is at the core of epistemology in terms of the nature of knowledge and how it can be generated, as well as involves the questionability of the existing knowledge claims, the criteria for knowledge and justification, and the sources and scope of knowledge as issues that are at the center of debate in epistemology.

Using the process of data mining/knowledge discovery as a systematic framework with well-defined stages for structuring data-analytic thinking and practice is increasingly pervading scientific disciplines in terms of research and innovation (e.g., [2]). Worth pointing out in this regard is that the best opportunity for using the data deluge is to harness and analyze data not as an end in itself, but rather to develop big theories, e.g., about how smart sustainable/sustainable smart cities can be operated, managed, planned, designed, developed, and governed in ways that overcome the challenges of sustainability and urbanization. In this context, big data analytics can be exploited to reveal hidden and previously unknown patterns and discover meaningful correlations in large datasets pertaining to natural and social sciences so to develop more effective ways of responding to population growth, environmental pressures, changes in socio-economic needs, global shifts/trends, discontinuities, and societal transitions in the form of new processes, systems, designs, strategies, and policies, as well as products and services. In the meantime, to really get a grip on the use of big data analytics to address the challenges of sustainability in an increasingly urbanized world, new theory about big data analytics theory—meta-theory—is necessary. From a general perspective, West (2013) [60] vividly argues that big data require big theories. Specific to smart sustainable/sustainable smart urbanism, discovering patterns and making correlations from the deluge of urban data can only

ever occur through the lens of a new kind of theory [3,24]. Especially, data-intensive science needs to meet three criteria in order to match the Kuhn's (1962, 1996) [16] notion of paradigm shift: it must be based on and provide a meta-theory, be acknowledged by a scientific community of practitioners, and possess a number of successful practices. The extant literature shows that data-intensive science as a paradigm shift is still evolving in terms of meta-theory—but has a large number of successful practices and it is acknowledged by the scientific community.

There are varied arguments about whether big data computing will herald the end of theory, and hence the extent to which it has the answers, as manifested in the number of the emerging epistemological positions pertaining to data-intensive science. This particularly pertains to the ways supercomputers: large-scale computation, data-intensive techniques and algorithms, and advanced mathematical models used in building and performing big data analytics, can potentially generate more useful, insightful, or accurate results than domain experts, scientists, specialists, and researchers who traditionally craft targeted hypotheses and devise research strategies. This revolutionary notion is increasingly entering the research practices of institutions, organizations, and governments. The idea being that the data deluge can reveal secrets to us that we now have the power and prowess to uncover. In other words, we no longer need to postulate and hypothesize; we simply need to let machines lead us to the patterns, correlations, trends, and shifts in social, economic, political, and environmental relationships. There is no denial of the significance of the analytical power of big data. The huge resources being invested in both the public and private sectors to further investigate and advance big data computing is a testament to this. Having recently, as a research wave and direction, permeated and dominated academic circles industries, coupled with its research status being consolidated as one of the most fertile areas of investigation, big data analytics has attracted researchers, scholars, scientists, experts, practitioners, policymakers, and decision makers from diverse disciplines and professional fields—given its importance and relevance for generating well-informed decisions and deep insights of highly useful value. Therefore, big data analytics is a rapidly expanding research area and is becoming a ubiquitous term in understanding and solving complex challenges and problems in many different domains. The big data movement has been propelled by the intensive R&D activities taking place in academic and research institutions, as well as in industries and businesses—with huge expectations being placed on the upcoming innovations and advancements in the field. In particular, a large part of ICT investment is being directed by giant technology companies, such as Google, IBM, Oracle, Microsoft, SAP, and CISCO, towards creating novel computing models and enhancing existing practices pertaining to the storage, management, processing, analysis, modelling, simulation, and evaluation of big data, as well as to the visualisation and deployment of the analytical outcome for different purposes [58]. Big data computing is undoubtedly useful for addressing and overcoming many important issues that are faced by society, including sustainability and urbanization, but it is important to ensure that we are far from being seduced by the promises and claims of big data computing to render theory unnecessary. Especially, there is a risk that big data leads to a shift in focus towards short-term, predictive, non-explanatory models, abandoning theory. Nevertheless, there are some projections, though, that most, if not all, of the social questions that are of most concern to us will be answered based on sifting through and harvesting sufficient quantities of big data. However, several skeptical views challenge the achievability of this vision or the realization of this prospect, being predicated on the assumption that there will always be uneven data shadows and inherent biases in how information is used and technology is produced. Therefore, it is equally important not to overlook the important role of expert domains or specialists to offer insights into what the data deluge can do, but perhaps they do not reveal it.

4.3. The Data-Intensive Scientific Approach to Urban Sustainability Science and Related Wicked Problems

Cities are full of complex issues that are not easily captured or steered. The problems of cities are primarily about people and their environment and life. Physical, infrastructural, environmental, economic, and social issues in contemporary cities define what planners call 'wicked problems', a term

that has gained currency in urban planning and policy analysis, especially after the adoption of sustainable development within urban planning and development since the early 1990s. Cities are characterized by wicked problems [61,62], i.e., difficult to define, unpredictable, and defying standard principles of science and rational decision-making. When tackling wicked problems, they become worse due to the unanticipated effects and unforeseen consequences that were overlooked, because the systems in question were not approached from a holistic perspective, or were treated in too immediate and simplistic terms [1]. The essential character of wicked problems is that they, according to Rittel and Webber (1973) [62], cannot be solved in practice by a central planner. Bettencourt (2014) [21] reformulates some of their arguments in modern form in what is called the 'planner's problem,' which has two distinct facets: (1) the knowledge problem and (2) the calculation problem. The first problem refers to the planning data needed to map and understand the current state of the smart sustainable city in this context. It is conceivable that urban life and physical infrastructure could be adequately sensed in several million places at fine temporal rates, generating huge but manageable rates of information flow by the advanced forms of ICT. It is not impossible, albeit still implausible, to conceive and develop technologies that would enable a planner to have access to detailed information about every aspect of the infrastructure, services, social lives, and environmental states in a smart sustainable city. The second problem refers to the computational complexity to carry out the actual task of planning in terms of the number of steps necessary to identify and assess all possible scenarios and choose the best possible course of action. Unsurprisingly, the exhaustive approach of assessing all possible scenarios in such city is impractical due to the fact that it entails the consideration of impossibly large spaces of possibilities.

As a scientific discipline, urban sustainability science integrates urban sustainability, urban science, and sustainability science. The notion of sustainability has been applied to urban planning and development since the early 1990s. This was marked by the emergence of the notion of urban sustainability. Urban sustainability science has theoretical foundations and assumptions from which it has grown that have solidified into a defined science after the establishment of sustainability science, which emerged in the early 2000s [1]. Sustainability science focuses specifically on understanding the dynamic interactions of socio-ecological systems, of which cities represent perfect examples. As such, it can serve as a theoretical basis for urban planning and development under what is labeled 'sustainable urbanism' that can effectively engage with the wicked problems that are presented by cities and their sustainability. The objective of urban sustainability is to uphold the changing dynamics and hence reciprocal relationships (within and across levels and scales) that maintain the ability of the city to provide not just life-supporting, but also life-enhancing, conditions, as exhibited by its collective behavior. To achieve this, the city should work towards enhancing the underlying environmental, physical, social, and economic systems over the long run by means of sustainable interventions and programs using advanced technologies and their novel applications, with the primary purpose of maintaining predictable patterns of behavior, and hence stable reciprocal relationships that are responsible for generating such patterns. Typically, such relationships cycle to produce the behavioral patterns that the city exhibits as a result of its operational functioning, planning, design, and development in the context of sustainability. In particular, as the positive adaptation of the city depends upon how well it is adjusted with the environment, it needs to make changes to protect itself and grows to accomplish its goals in terms of achieving the ultimate goal of sustainability. One way of doing this is to self-correct itself based on reactions from the natural/environmental system with respect to climate change and related hazards and upheavals. This feature relates to the adaptive nature of complex systems in that they have the capacity to change and learn from experience. To put it differently, the objective of urban sustainability can be accomplished by rendering the city dynamic in its conception, scalable in its design, efficient in its operational functioning, and flexible in its planning, which is of crucial importance in dealing with population growth, environmental pressures, changes in socio-economic needs, global shifts, discontinuities, and societal transitions [1]. This involves maintaining the critical structures, key dependencies, functional integrity, resource availability,

well-being, and capacity for the regeneration and evolution of the city. What is important with respect to ensuring the persistence of structures and conditions necessary for keeping the city system within a preferred stability state is the need for continuous reflection as an effective way to learn from both failures and successes, as well as to achieve a deep understanding of how socio-ecological systems function to be able to work with, anticipate, and harness the dynamics within such systems.

The quest for finding an urban planning and development paradigm that can accommodate the wicked problems of cities and their sustainability and overcome the complexity and unpredictability introduced by social factors is increasingly inspiring scholars to combine urban sustainability and sustainability science under what has recently been termed as 'urban sustainability science'. This is, in turn, being informed by urban science and thus big data science and analytics. While the introduction of sustainability to the goals of urban planning and development added another layer of complexity brought about by the consideration of environmental externalities and social and economic concerns, the new urban science has opened new windows of opportunity to deal with such issues in cities on the basis of modern computation and data abundance. Indeed, sustainability as entailing complex dynamics of human-natural system interactions requires a decisive, radical change in the way that science is undertaken and developed. This change is what data-intensive scientific development is about—as enabled and driven by big data science and analytics.

The great innovation of big data science and analytics and the underlying technologies is that the urban problems should be approached in full knowledge, which supposes a new approach to scientific development that is based on massive-scale data. As an evolving, systematic enterprise building and organizing knowledge in the form of explanations and predictions about the world, data-intensive scientific development entails using data-driven inductive empiricism and data-driven science. These recent epistemological approaches are at the core of urban science [8], which informs urban sustainability science. This is due to their critical importance and relevance to urban practices, such as operational functioning, planning, design, and development in the context of sustainability.

There are various reasons that justify the adoption of data-intensive scientific development in urban sustainability science. It is imperative for urban sustainability science, a field that focuses on understanding the dynamic interactions of the social and ecological systems of the city, to develop and apply an advanced approach to scientific inquiry and exploration for dealing with the kind of wicked problems and intractable issues pertaining to urbanism as a set of multifaceted, contingent practices. Additionally, urban sustainability science should embrace data-intensive scientific development in order to be able to transform knowledge regarding how the natural and human systems in cities interact in terms of the underlying (changing) dynamics for the purpose of designing, developing, implementing, evaluating, and enhancing human engineered systems as practical solutions and interventions that support the idea of the socio-ecological system in balance. This embrace is additionally aimed at nurturing and sustaining the linkages between scientific research and technological innovation and policy and public administration processes in relevance to sustainability. To put it differently, the data-intensive approach to urban sustainability science is of high relevance in the cultivation, integration, and application of knowledge about natural systems gained, especially from the historical sciences, and its coordination with knowledge about human interrelationships gained from the social sciences and humanities. This is of crucial importance in evaluating, mitigating, and minimizing the intended and unintended consequences of anthropogenic influence on social and ecological systems across the globe and into the future. More to the appropriateness of data-intensive scientific development, urban sustainability science mixes and fuses disciplines across the natural sciences, social sciences, formal sciences, and applied sciences. The philosophical and analytical framework of urban sustainability science draws on and links with numerous disciplines and fields, and it is studied and examined in various contexts of environmental, social, economic, and cultural development and managed over many temporal and spatial scales. The focus ranges from macro levels starting from the (sustainability) of planet Earth to the sustainability of societies, regions, and cities, as well as economies, ecosystems, and communities, and to micro levels

that are encompassed in streets, buildings, and individual lifestyles [24]. In view of that, big data computing can perform more effectively with respect to achieving the desired outcomes that are expected from the application of interdisciplinarity and transdisciplinarity as scholarly enterprises due to the underlying analytical power, coupled with the data deluge available for scientific inquiry and exploration. This is particularly important in the context of urban science for gaining new interactional and unifiable knowledge necessary for exploring and exploiting the opportunity of using advanced technologies to solve real-world problems and challenges, especially those that are associated with sustainability and urbanization.

The solutions to the kind of wicked problems and intractable issues that are associated with urban sustainability are anchored in the recognition that the urban world has become integrated, complex, intricate, contingent, and uncertain. The data-intensive approach to urban sustainability science is primarily meant to facilitate the link of such problems and issues to the type of problems and issues explored and probed by sustainability science, as well as demonstrates how the understanding of cities, as instances of socio-ecological systems, provides a conceptual and analytical framework for addressing and overcoming some of the challenges that are characteristic to such problems and issues. There is a host of new practices that sustainability science could bring to urban sustainability under the umbrella or sphere of data-intensive science, an argument that needs to be further developed and to become part of mainstream debates in urban research and practice. This argument is being stimulated by the ongoing discussion and development of the new ideas regarding the untapped potential of big data computing and the underlying enabling technologies for advancing both sustainability science and urban sustainability as well as merging them into a holistic framework that is informed by the new urban science. This kind of integrated, holistic framework should focus on probing the complex mechanisms that are involved in the profound interactions between environmental, social, economic, and physical systems to understand their behavioral patterns and changing dynamics so as to develop upstream solutions for tackling the complex challenges that are associated with the systematic degradation of the natural and environmental systems and the concomitant perils to the human and social systems. Urban sustainability science as a research field seeks to give the broad-based and crossover approach of urban sustainability a solid scientific foundation. It also provides a critical and analytical framework for urban sustainability and, to draw on Reitan (2005) [63], it must encompass different magnitudes of scales (of time, space, and function), multiple balances (dynamics), multiple actors (interests), and multiple failures (systemic faults). In addition, it should be viewed as a field that is defined more by the kind of wicked problems and intractable issues it addresses rather than by the scientific and academic disciplines that it employs, thereby being neither basic nor applied research; it serves the need for advancing both knowledge discovery and actionable decisions by creating a dynamic bridge between the two thanks to new big data analytics techniques.

By and large, the link between sustainable urban development and urban data science stems from the idea that the former is an aspiration that should, as realized by many scholars over the past two decade or so, be achieved only on the basis of advanced scientific knowledge, and thus the approach to producing it, thereby the relevance of big data science and analytics. This has justified the establishment of a new branch of science—urban sustainability science—due to the fact that the city is confronted at an ever unprecedented rate and larger scale with the ramifications of its own success as a product of social revolution. The way that things have changed in recent years (and the attempts being undertaken to take this into account) calls for a novel approach to science for explaining, predicting, and understanding the underlying web of the ongoing, reciprocal relationships that are cycling to generate the patterns of behavior that the complex city system is exhibiting, and for figuring out the mechanisms that such a system is using to control itself. The point is that the complexities, uncertainties, and hazards of the human adventure are triggering drastic changes that increasingly require insights from all of the sciences to tackle them if there is a shred of seriousness about the aspiration to improve sustainability, resilience, and the quality of life, i.e., sustainable urban development.

The essential opportunities and challenges of the use of big data computing and the underlying enabling technologies in smart sustainable/sustainable smart urbanism as a practical application of urban sustainability science have, despite their appeal, not been sufficiently systemized and formally structured. In particular, the necessary conditions for the strategic application of big data science and analytics in such urbanism need to be spelled out, and their limitations must also be anticipated and elucidated [1]. There are different ways of addressing these and other important questions when considering the available interdisciplinary and transdisciplinary knowledge of such urbanism in the big data age (see Bibri 2019a [58] for an overview). In this line of thinking, Bettencourt (2014) [21] attempts to answer some of these questions by formalizing the use of big data analytics in urban planning and policy in light of the conceptual frameworks of engineering, and this shows that this formalization enables identifying the necessary conditions for the effective use of big data in urban planning and policy that address a large array of urban issues. This is intended to demonstrate that big data computing and the underlying enabling technologies are providing new opportunities for the application of advanced engineering solutions to smart sustainable/sustainable smart cities. It is conspicuous that big data science and analytics may offer radically novel solutions to the wicked problems of urban sustainability related to planning, design, and development. Big data computing is so fast in comparison to most physical, environmental, social, and economic phenomena that myriads of key urban planning and policy problems are falling within this window of opportunity [24]. In such circumstances, models of system response enabled by big data analytics can be very simple and crude, and typically be linearized (see [64]). Thus, the analytical engineering approach conveniently bypasses the complexity that can arise in the nested systems of smart sustainable/sustainable smart cities at longer temporal or larger spatial scales. Bibri (2019) [1] summarizes some of the key urban sustainability problems with their typical temporal and spatial scales and the nature of their operating outcomes. The unique potential of big data science and analytics in this regard lies in essentially advancing urban sustainability and solving related complex issues without coherent models, unified theories, or any mechanistic explanation at all, thanks to the data deluge that makes the scientific method obsolete, and within big data studies correlation supersedes causation. Many examples of planning, design, management, and policy practices in those cities that use data successfully can have this flavor, irrespective of whether their development and implementation involve organizations or computer algorithms [1].

5. Scholarly Shifts

5.1. *Building the New Urban Science and Establishing the Related Research Domain*

5.1.1. Research Status

Urban science is an interdisciplinary field within which data science is practiced to inform and sustain the core of data-driven urbanism. Positioned at the intersection of science and design, it seeks to exploit the development of modern computation and the growing abundance of data. As a research field, urban science is concerned with the study of diverse urban issues and problems, and it thereby aims to produce both theoretical and practical knowledge that contributes to understanding and solving them in contemporary society [1]. In this respect, it entails making sense of cities as they are by identifying relationships and urban laws, as well as predicting and simulating likely future scenarios under different conditions, potentially providing valuable insights for planning and development decision-making and policy formulation [65]. As such, it involves data-analytic thinking and computational modelling and simulation approaches to exploring, understanding, and explaining urban processes, and also addressing several challenges that are posed by urban data. The two fundamental ones are: (1) how to handle and make sense of billions of observations that are being generated on a dynamic basis [19] and (2) how to translate the insight that is derived into new urban theory (fundamental knowledge) and actionable outcomes (applied knowledge) [3,66,67]. The new urban science—which is underpinned by urban sustainability science, a transdisciplinary field that

fuses theories from urban sustainability and sustainability science, seeks to make cities more sustainable, resilient, efficient, livable, and equitable by rendering them more measurable, knowable, and tractable in terms of their operational functioning, management, planning, design, and development [1].

In view of the above, urban science is associated with scientific research as applied research: the search for solutions to practical problems using the knowledge and applied research as the aim of basic research, which is another form of scientific research. A great deal of our understanding comes from the curiosity-driven undertaking of basic research, although some scientific research is applied to specific problems. This leads to options for technological advancements that were not planned, or sometimes even imaginable, i.e., big data computing and the underlying enabling technologies. However, the new urban science, a field in which data science and big data analytics are practiced, aims to make cities more sustainable, resilient, liveable, and transparent by rendering them more measurable, knowable, and controllable in terms of their operational functioning, planning, design, and development. Indeed, cities are becoming highly responsive to a form of data-driven urbanism that is the key mode of production for what have widely been termed smart sustainable/sustainable smart cities, whose monitoring, understanding, and analysis are increasingly relying on the core enabling technologies of big data analytics. Research on such cities, the leading global paradigm of urbanism, is garnering growing attention and rapidly burgeoning, and its status is consolidating as one of the most enticing and fanciest areas of investigation today, which makes the relevance and rationale behind the smart sustainable/sustainable smart city debate of high significance and value with respect to the future form of urbanism. This area is typically concerned with addressing a large number and variety of issues that are related to both sustainable cities and smart cities in the context of sustainability, as well as to the amalgamation of these two classes of cities as landscapes and approaches in the context thereof [58]. A large part of research in this area focuses on exploiting the potentials and opportunities of advanced technologies and their novel applications, especially big data computing/analytics, as an effective way of mitigating or overcoming the issue of sustainable cities and smart cities being extremely fragmented as landscapes and weakly connected as approaches, particularly at the technical and policy levels [1,58].

Further, at the heart of data-driven smart sustainable/sustainable smart urbanism is a computational understanding of city systems that brings urban life to a set of logic, calculative, and algorithmic procedures, as well as an endeavor of drawing together and interlinking urban big data to provide synoptic city intelligence primarily directed for improving, advancing, and maintaining the contribution of both sustainable cities and smart cities to the goals of sustainable development in an increasingly urbanized world. This is underpinned by epistemological realism and instrumental reality, which sustain and are shaped by urban science, which in turn seeks to make cities more sustainable, resilient, and efficient. In view of that, urban science deeply pursues quantitative and computational approaches to understanding and dealing with cities. It involves a combination of two scientific inquiry methods: (1) a deductive scientific approach that was aimed at uncovering the common processes that influence the structure and dynamics of all cities (based on data-driven science) and (2) a descriptive approach that is based upon field work and surveys (or inductive empiricism using big data analytics techniques) aimed at understanding the specificity of cities. All in all, with the world in a phase of unprecedented urbanisation, coupled with the mounting challenges of sustainability, the new urban science is emerging as a coherent body of theory or a systematically organized body of knowledge driven by data-intensive science that can contribute to a more sustainable urban world.

Across the globe, many research institutions are, as supported by governments, exploring the rapidly expanding scholarly movement of urban science. As demonstrated in a series of scientific and academic publications, new waves of scientific interest in cities are emerging, and urban research is increasingly being colonized by other disciplines, i.e., appropriated as a domain for their own use, and the projected scale and extension of urban science research institutions is growing and predicted to grow to huge investment in the next decade. There is a quest for new research models that leverage digital platforms for the wide participation in urban research, and that mainstream

its benefits for sustainability in an increasingly urbanized, technologized, and computerized world. In this regard, there are many endeavors developing several visions of how urban science might unfold in the coming decades (see, e.g., [3,4,8,19,24,65,68,69]) for the purpose of informing policy, supporting research and innovation, and mainstreaming institutional research. Foresight is of crucial importance for informing strategic planning for effectively advancing urban science over the critical coming phase of evolution of this nascent movement, as pushed by big data computing, especially in the context of smart sustainable/sustainable smart cities of the future [24,58]. Such cities are indeed seen as the most important arena for sustainability transitions, and thus central to ensuring a sustainable future, because they constitute the hubs and sites of innovation within different, yet related, innovation systems, namely national, regional, sectoral, technological, and Quadruple Helix of university–industry–government–citizen relations. Besides, any advancement in the science of cities in support of sustainability requires long-term strategic planning, and foresight studies can serve as a basis for inspiration in the discussions and decision-making processes. In addition, in each of the three main pillars of sustainability—environmental, social, and economic—urbanization now plays a key role. In this respect, urban science is required to draw from the natural, engineering, and social sciences, as well as the arts and humanities, while directly linking into practice. This is crucial for urban science to be collectively greater than the sum of its parts. Especially, smart sustainable/sustainable smart cities are complex systems par excellence, more than the sum of their parts, and consequently, the underlying urbanism has become ever more complex, with the very technologies being used to make sense of and deal with it as involving special conundrums, wicked problems, intractable issues, and complex challenges that are associated with sustainability and urbanisation. This is well reflected in the operational functioning, planning, design, and development of smart sustainable/sustainable smart cities as a leading paradigm of urbanism. Consequently, to tackle such cities as complex systems and dynamically changing environments requires innovative solutions and sophisticated approaches as to the way that they can be monitored, understood, and analysed so as to be effectively operated, managed, planned, designed, developed, and governed in line with the long-term goals of sustainability, thereby strategically improving, advancing, and maintaining their contribution to the objectives of sustainable development. This can be accomplished, by developing and applying new urban intelligence functions as new conceptions of the way that such cities function and utilize complexity science, data science, and urban science in fashioning new powerful forms of urban simulation models and optimisation and prediction methods on the basis of big data analytics that generate urban forms and structures that improve the sustainability, efficiency, resilience, and quality of life.

5.1.2. Challenges and Prospects

There are numerous challenges and opportunities that pertain to the development of urban science that address and overcome the pressing issues of sustainability and urbanization and enable more effective science–policy interfaces [4,8,19,24,58,65,69]. In terms of science in the big data era, there are major shifts in the human condition (e.g., urbanization, sustainability, etc.) that require new changes in science, and vice versa (e.g., data science, pervasive computing, etc.). Today, the increasing rate, scale, and speed of urbanization, coupled with the mounting challenges of sustainability are again pushing the scientific frontier. In the meantime, policy bodies across governance scales, from the local to the multilateral, often struggle to gather an adequate response to the pace or specificity of urban change in terms of evidence base and the capacity to grasp it as a drastic shift from a holistic perspective. In light of such growing spotlights on cities and their sustainable transformation and other dynamic changes, developing more effective knowledge for actionable and successful interventions that are guided and supported by policy has become more imperative than ever. Smart sustainable/sustainable smart cities are increasingly seen as a place of high potential for scientific inquiry and professional training that will provide the raw material for generating adequate scientific knowledge for responding to the challenges of sustainability and urbanization. The ultimate aim is to fill the shortage that urban research and education are facing nowadays in key respects, especially data science, big data computing, urban

science, urban informatics, sustainability science, and data-driven urbanism in the context of such cities. It is of crucial importance and strategic value to invest in research activities and education programs that are focused on these domains. In this context, the emergence of such cities should be approached from the lens of hospitality, one through which we can align their affordances as well as embrace and realize their advantages and better understand their disadvantages. It is in research institutions and universities where this hospitality can best be achieved, and which can use and explore such cities as laboratories of innovation [24]. These research and educational entities are well positioned to readily host the emergence of such cities, and seek innovative solutions and sophisticated approaches that are of relevance to the challenges of sustainability and urbanization with regard to developing new metrics for measuring, and new methods for assessing, the urban progress towards achieving the long-term goals of sustainable development [24]. Another strategic value of investing in research and education lies in training and educating a new generation of interdisciplinary and transdisciplinary researchers, scholars, and practitioners within the domain of urban sustainability science, as well as in gaining new knowledge to explore and exploit the opportunity of using and applying advanced technologies to solve real-world problems and challenges in the realm of such cities.

Across academia, urban knowledge is out-dated and underfunded [70]. Much of what we know about cities, to date, has been gleaned from studies that are characterized by data scarcity [5], and they thus involve the use of traditional data collection and analysis methods with inherent limitations and constraints. In other words, urban research tends to rely on selective samples, so we still know very little regarding the majority of urban settlements and challenges around the world. As widely acknowledged in the domain of smart and sustainable urbanism in relation to academic and scientific research, 'small data' studies are associated with high cost, infrequent periodicity, quick obsolescence, incompleteness, inaccuracy, as well as subjectivity and biases, adding to capturing a relatively limited sample of data that is tightly focused, less representative, restricted in scope and scale, time and space specific, and relatively expensive to generate and analyze [2]. Moreover, urban research grapples with the shortcomings of studies today, as they only focus on selective categories of subjects, and neglect important questions of environmental sustainability, inequality, and vulnerability. There is an urgent need for scientific research endeavors for developing systematic frameworks for city analytics and 'big data' studies in relation to the domain of smart sustainable/sustainable smart urbanism [24]. This is in response to the emerging paradigm of big data computing and the increasing influence of big data analytics, and its application on enabling, operating, organizing, and planning the processes of smart sustainable/sustainable smart cities as a leading paradigm of urbanism. The intention is to utilize and apply well-informed, knowledge-driven decision-making and enhanced insights to improve and optimize the urban operations, functions, services, designs, strategies, and policies in line with the vision of sustainability.

Further, small pockets of well-funded research domains are often aligned to opportunistic themes that are driven by industry, policy, and market dynamics beyond academia, such as climate change, resilient cities, smart cities, and cyber-physical cities—rather than offering the wider coverage that is necessary for balanced interventions by practitioners from a variety of professional fields, especially in relation to sustainability dimensions. Overall, research outcomes remain patchy in the sense of being both inconsistent and not of the same quality as well as happening in small, isolated areas due to the complex nature of urban scholarship as being mismatched by the adequacy of such outcomes. Significant issues, like new urban intelligence functions and related simulations models, and prediction and optimization methods, which involve control, automation, management, efficiency, and enhancement as related to urban operations, functions, designs, services, strategies, and policies have made minimal headway into urban research in the context of sustainability [24], although the phenomenon that is associated with the escalation of scientific work on sustainability is set to accelerate some progress on urban science. Smart sustainable/sustainable smart cities should evolve new urban intelligence functions for monitoring, planning, and designing the operating and organizing processes of urban life, which relates to what has been termed laboratories for innovation [19,24,58]. Current

urban research regarding the pressing global issues pertaining to sustainability and urbanization is rudimentary and fragmented at a time when the window of urban transformation demands robust and sophisticated urban research, along with focused and strategic technological innovation and advancement. It is in large part trapped in the twentieth-century tradition of the systematic study of cities, particularly in terms of planning, engineering, and design. Consequently, this keeps us in a long distance from understanding the fabric of city systems that shape the way urban areas impact humanity with respect to sustainability dimensions, and vice versa.

There is a need for urban research to become a coherent urban science, and for urban scholarship to become well-informed in the ways it can convey the full spectrum of major urban changes—sustainability transitions. Urban research also needs to be adequately directed to real-problem applications that are associated with sustainability and urbanization beyond the need for a stronger interdisciplinary and interdisciplinary lens. At present, like many other city-related fields, urban science is segmented by disciplinary boundaries when the urban transformation related to sustainability and urbanization demands truly holistic urban research, and whereas solutions to related problems require integrated (cross-disciplinary) knowledge. The new urban science is multidisciplinary and it draws upon theoretical ideas from across the contributing disciplines and hence bring together architects, engineers, ecologists, social scientists, computer scientists, data scientists, and built and natural environment specialists. All of these actors are undertaking research and developing strategies and programs to tackle the challenging elements of urbanism, e.g., investigating the metabolism of cities as ecosystems characterized by stocks and flows of resources, including energy, material, water, capital, and information, and urban infrastructure systems as complex socio-technical systems that are composed not only of engineered structures, but also the people who make up the subsystems of urbanites and administrators. However, discrete research communities are unable to jointly advise urban planners on the complex multi-dimensional nature of urban sustainability problems, or indeed on the most appropriate prioritization of urban solutions. While smart sustainable development in which effective policy frameworks and measures and relevant institutional structures and practices are needed for integrating the research and innovation agenda of advanced ICT with the agenda of sustainable development, while aligning and mobilizing strategic urban planning and development actors in this direction [24], urban science requires institutional, political, and managerial expertise, alongside academic skills for its establishment, as well as injecting specialist expert knowledge into practice for the purpose of establishing data-driven smart sustainable/sustainable smart urbanism. There is a need for a close collaboration between urban scientists and urban planners as communities of research and practice, while converging on the places where knowledge is most needed to solve the sustainability and urbanization problems and challenges, and where the best scholarship has to be produced. In this respect, it is important to keep in mind that rectifying the parlous state of urban knowledge production and dissemination in the context of smart sustainable/sustainable smart cities is not just about scaling up and undertaking more research endeavours on more topics, but much urban knowledge that is related to the operational aspects of sustainability remains context specific, despite the likely universality of urban conditions. Urban scholars might not reveal their thoughts readily as to offering advice to policymakers and practitioners if they are unsure about how their research applies to context-specific problems that they have not directly dealt with.

5.1.3. The Need for Re-Casting and Reforms

Urban science needs to be re-cast in several ways, and also important reforms are required, which present particular challenges, to provide a step change in the way that urban knowledge is produced, enhanced, and advanced, and ultimately to secure sustainability transition as a major urban change. The need for re-casting and reforms pertains to the following:

- A re-orientation in how cities are conceived.
- A reconfiguration of the epistemology of urban science to openly recognize the contingent and relational nature of urban science as well as urban systems and processes.

- Effectively linking scholarly contributions with practice.
- Pushing new research agendas that are adequate to the scale of the challenges cities face today in relation to sustainability and urbanization.
- A fundamental shift of scientific and research paradigms in urban science, together with a reorganization of the institutional forms that bring scientists, experts, and researchers on common grounds.
- A restructuring of urban data systems, urban research, and the scientific research-to-practical applications continuum.
- A transformation of urban research in ways that respond to the requirements of sustainability and the demands of urbanization, which necessitates a fundamental restructuring of the scale, location, and mandate of the research institutions and universities that generate urban knowledge and sustain its production.
- Prompting the urban scientific inquiry and exploration that zooms in and focuses on specific urban dimensions or opportunistic themes that are driven by industry and market drivers, and concurrently strengthening the global assessments that enable a high level of granularity, which is necessary for understanding sustainability and urbanization issues.
- Developing an urban science that can grapple with issues of the policy relevance, prioritization, and co-determinants of sustainability transition in the context of the increasingly urbanized world.

It is deemed useful to elaborate on the first three of the above for clarification purposes. Concerning the re-orientation in the way that cities are conceived, instead of 'being cast as bounded, knowable and manageable systems that can be steered and controlled in mechanical, linear ways, cities need to be framed as fluid, open, complex, multi-level, contingent, and relational systems that are full of culture, politics, competing interests and wicked problems and often unfold in unpredictable ways. Reducing this complexity into models and then using the outcomes to drive urban management produces a reductionist and limiting understanding of cities and overly technocratic forms of governance. Rather, these models need to be complemented with other forms of knowledge ... In other words, city analytics and its instrumental rationality should not be allowed to simply trump reason and experience, or other sources of information and insight such as those based on 'small data' studies, in shaping and driving urban governance. Instead, they should be used contextually and in conjunction with each other' [8] (p. 11).

From a philosophical perspective, instrumental rationality, as a specific form of rationality focusing on the most efficient means to achieve a specific end, is not in itself reflecting on the value of that end. Indeed, while it is a crucial, and presumably indispensable, component of practical rationality, it is partly constitutive of intention, need, desire, or action. In particular, expressing a means to an end, it means doing whatever it takes to achieve a goal, so long as it aligns with certain ultimate objective. The idea here is that only the end really matters. As such, it can be highly problematic in terms of the increasing rationalization of the city and the unending drive for efficiency, and thus should be combined with values as to guiding our choices and decisions. In fact, we humans possess two independent rational capacities: operational reasoning to acquire technical knowledge regarding means that are functional (instrumental rationality) and ethical reasoning to acquire moral knowledge of ends that are right (value rationality). Nevertheless, the negative connotation of instrumental rationality might not be of relevance in this context, so long as the specific end is achieving the long-term goals of sustainability as a form of environmental and economic efficiency for improving societal outcomes. Originally, the philosophical concept of rationality emerged in the broader context of major developments in science, technology, philosophy, politics, and society when people started to focus more on themselves to guide their lives and make rational choices.

As regards the need for re-casting the epistemology of urban science, the main argument relates to the social shaping of science-based technology or the social construction of scientific knowledge and its practical application, which relate to the analytical and philosophical framework of STS (see Bibri and Krogstie 2016 [9] for a detailed discussion). In light of this, the re-casting in question involves

recognizing that the realist assumptions, which posit that urban science can reveal fundamental truths about the city, are flawed. Urban science can only produce a particular view through a specific lens, and cannot provide neutral, objective, God's eye views of the city [8]. On the one hand, the data that were used do not exist independently of the ideas, instruments, systems, practices, and knowledge employed—and embedded within a multidimensional context (e.g., local, national, social, political, cultural, organizational, regulatory, etc.)—to generate and process them [1,71]. To put it differently, the data are never raw, but they are always already cooked to a particular recipe for a particular purpose [72,73]. On the other hand, big data computing and the underlying enabling technologies are socio-technical in nature. As such, they are a not neutral, purely technical means of assembling and making sense of data; instead, philosophical ideas, socio-political frameworks, and ideological means shape them [8,9]. In particular, big data technology is 'cultural', since it can be conceptualized as a discourse prioritizing specific concepts, ideas, claims, assumptions, and visions about the nature and practice of S&T in society and the role of diverse actors in shaping them, to draw on Bibri (2015) [17]. There is potential for realizing that the big data-driven technologized nature of the city is neither apolitical nor inevitable. Furthermore, when engaging in a discursive-material analysis, the politics of this science-based technology does not become the result of the unconditioned agency of the involved actors, e.g., scholars, scientists, experts, engineers, and technologists. Rather, such technology can be conceived as specific techno-socio-political practice that depends on the agency of various actors promoting it and forming coalitions on particular technological innovations, and on the political regulation of S&T in society. In a nutshell, big data technology is the outcome of social processes that involve diverse intertwined factors and many stakeholders with a vested interest. Accordingly, urban science as a field in which data science and big data computing are practiced needs to recognize that it does not reflect the world as it actually is and to openly acknowledge its contingencies, limitations, and inherent politics, but rather actively frames and produces the world [8,74]. This is, though, not to say: 'the fundamental approach of analytics, modelling, and simulation is radically altered, but rather that how these approaches work in messy practice is detailed and grand claims as to their veracity or validity is tempered. This would include detailing how ethical issues were considered and the research design altered appropriately' ([8], p. 11).

The above two dimensions along which the respective re-casting needs to proceed relate to the numerous ethical issues that arise from the development and deployment of smart sustainable/sustainable smart city technologies and accompanying urban science. This has led to a number of critiques concerning the underlying concepts, ethos, and practices of urbanism, as underpinned by such technologies (e.g., [4,24,58,75–77]). 'One response to these critiques is to call for a fundamentally different approach to urban development and the practice of other forms of urban studies rather than urban science. Another is to argue that ... urban science needs to be re-imagined and re-cast' ([8], p. 11). It needs to be recognized that smart sustainable/sustainable smart city technologies do provide many benefits to urban planners, managers, administrators, and citizens. Similarly, urban science does provide novel and useful insights into cities, their systems, and citizens [8].

As to linking scholarly contributions with practice, it is argued that there is a need for strong collaboration between academics and practitioners. Especially, the contribution of science to shaping the future of smart sustainable/sustainable smart urbanism tends to be poor in terms of interlinking academia and practice, despite the critical juncture at which urban scholarship and practice is positioned. Focusing on the state of urban research and its science-policy nexus is one way to redress these shortcomings and develop more urban cross-disciplinary and policy-engaged research.

In addition, a renewed urban research agenda should be based on a stronger interdisciplinarity and transdisciplinarity character, with respect to all of the natural, social, and human sciences and practices—against the backdrop of the rising challenges of sustainability and urbanization and their complexity. This scholarly spirit signifies that urban research must embrace the diversity, integration, and fusion of many relevant disciplines, which allows for holistic and novel insights into various social, environmental, and economic issues [24,78], as well as for the prioritization of effective

advice to urban policymakers [79]. This is also of importance for building effective science-policy interfaces for addressing and overcoming the challenges of sustainability and urbanization in terms of both up-skilling scientists to speak to politics and policy and decision and policy makers to read science—rather than simply professionalizing urban science for the primary purpose of effectively operating, managing, planning, and designing cities in line with the vision of sustainability, thereby lacking key co-determinants in the social process shaping major societal transitions.

It is of equal importance not to hinder efforts towards reproducibility in urban research if we ultimately want to produce a balanced and comprehensive knowledge through the collection of diverse kinds of data within urban science [80]. Reproducibility measures whether an experiment/study can be reproduced in its entirety, either by the same researcher or by someone else working independently. Reproducibility is one of the main principles of the scientific method and it is often associated with threats. Indeed, if an experiment cannot be repeated to produce the same results, this implies that the original results might have been flawed or in error. Hence, it is common for an experiment to be performed multiple times. For significant results, some of the scientists are also inclined to replicate the results for themselves, especially if these results would be of particular importance to their own work. As illustrated in Figure 1, there are different potential threats to reproducible science, including:

- lack of replication;
- hypothesizing after the results are known;
- poor study design;
- low statistical power, analytical flexibility;
- *P*-hacking;
- publication bias; and,
- lack of data sharing.

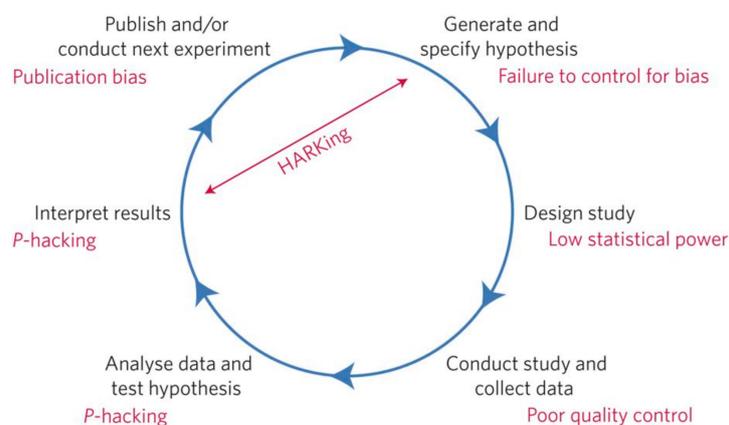


Figure 1. Threats to reproducible science. Source: Munafò et al. (2017) [80].

Figure 1 shows an idealized version of the hypothetico-deductive model of the scientific method.

Together, the above threats will serve to undermine the robustness and reliability of published research, and they may also impact on the ability of science to self-correct. Nevertheless, the data-intensive approach to science driven by big data science and analytics holds great potential to mitigate such threats, thanks to the underlying scientific systems, processes, and methods, as well as the systematic framework and analytical power pertaining to big data computing, coupled with the massive data that are involved in scientific inquiries and explorations. Indeed, reconsidering the foundation of building a focused urban science will need cross-topic, cross-scale, and cross-location studies that require novel analytical methods and vastly different competencies (especially data analysts, data scientists, computer scientists, urban scientists, social scientists, engineering scientists, big data developers and engineers, environmental scientists, technical planners, etc.), as well as abundant

data. However, the colossal amounts of the data being routinely sensed in real time, coupled with the emergence and adoption of the data-intensive approach to scientific and academic research, will require sophisticated tools, techniques, algorithms, and models for data management, processing, analysis, and interpretation as part of big data analytics processes, such as data mining and pattern recognition, statistical analysis, and simulation and prediction modeling. If we can fashion novel ways of handling the unfolding and soaring deluge of urban data for extracting useful knowledge and valuable insights for enhanced decision-making and scientific uses, which works very well in the new big data era where the numbers speak for themselves, we will move the field of urban science forward in great leaps.

5.2. Urban Knowledge Discovery/Data Mining and Big Data Studies and Related Issues

In contrast with urban knowledge that is derived from longer standing, more traditional urban studies, data science, as practiced within the field of urban science offers the potential for the kind of urban knowledge that is inherently longitudinal, and it has greater breadth, depth, scale, and timeliness [8,19] in the context of smart sustainable/sustainable smart urbanism. This is being enabled and afforded by the unfolding and soaring deluge of urban big data. With respect to the data-driven urban knowledge, emphasis has been on the development of new big data analytics that utilize sophisticated techniques and advanced mathematical models designed to process and analyse enormous datasets (e.g., [3–5,8,24]) that contained varied, real-time, exhaustive, fine-grained, indexical, flexible, evolvable, and relational type of data. This pertains to the process of knowledge discovery, which involves carefully choosing variable selection mechanisms, encoding schemes, preprocessing, reductions, and projections of the data, prior to discovering the intended patterns and building the relevant models, as well as their evaluation, interpretation, and visualization [24]. The pursuit of mastering the complexity of the process of knowledge discovery for smart sustainable/sustainable smart cities requires building an entirely new holistic system for big data analytics that involve their operational functioning, planning, design, and development in terms of applied intelligence functions that are primarily directed for improving, advancing, and maintaining their contribution to the goals of sustainable development through continuously optimizing and enhancing their operations, functions, services, designs, strategies, and policies in line with the vision of sustainability. The entire analytical process that is able to create the needed knowledge services or associated with extracting useful knowledge and valuable insights in the form of such functions pertaining to decision-making processes should be expressible within a system that supports the following:

- the acquisition of data from multiple distributed sources;
- the management of data streams;
- the integration of heterogeneous data into coherent databases;
- the definition of observables to extract relevant information from available datasets;
- data transformation and preparation;
- methods for distributed data mining and network analytics;
- the organisation and composition of the extracted models and patterns as well as the evaluation of their quality;
- tools for visual analytics to study the behavioural patterns and models;
- the availability of visualizations to planners, strategists, and decision makers;
- methods for the simulation and prediction of the mined patterns and models; and,
- mining strategies for overcoming the scalability issues associated with big data in distributed environments.

Recent years have witnessed remarkable progress within computer science, information science, and data science with regard to handling and extracting valuable knowledge and deep insights from large masses of data, and these have been utilized in urban science, to reiterate. In parallel, several projects of knowledge discovery across the globe as precursors in mining data related to different

urban domains have developed various analytical and mining methods for spatiotemporal and spatial data [19,24]. They have shown to support the complex knowledge discovery process from the raw urban data, being capable of supporting the decisions of different urban planners, administrators, and managers, thereby revealing the striking analytical power of big data. As an advanced form of decision support, the complex process of knowledge discovery/data mining is, by far, the most applied big data analytics technique or widely used framework for automatically extracting useful knowledge and valuable insights from large masses of data for enhanced decision-making in the domain of smart sustainable/sustainable smart urbanism (see, e.g., [19,24]). Data mining (also known as knowledge discovery) is the computational process of probing colossal datasets in order to find frequent, hidden, and previously unsuspected and unknown patterns and subtle relationships; to make useful, meaningful, and valid correlations from these discoveries; and, to summarize the results in novel ways and then visualize them in understandable formats prior to their deployment for decision-making purposes [2]. According to several codifications of the process of data mining, this process consists of well-defined stages, namely problem understanding, data understanding, data preparation, model building, result evaluation, and result deployment [24].

Bibri and Krogstie (2018) [2] propose, illustrate, and discuss a systematic framework (Figure 2) for city analytics and 'big data' studies in relation to the domain of smart sustainable urbanism based on cross-industry standard process for data mining. It consists of six steps, namely:

1. understanding and specifying urban sustainability problems;
2. understanding urban data;
3. preparing and combining urban data from diverse sources;
4. building models and generating patterns as true regularities;
5. evaluating and interpreting the obtained results; and,
6. deploying the results for urban operations, functions, services, strategies, and policies.

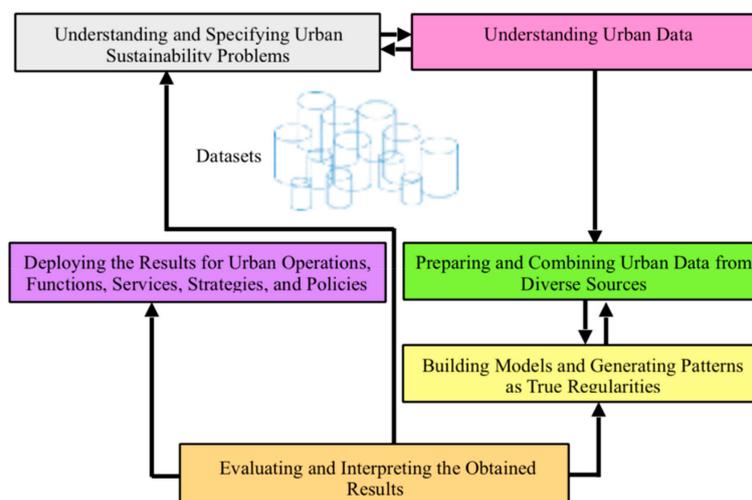


Figure 2. A data mining framework for urban analytics and big data studies. Source: Bibri and Krogstie (2018) [2].

Figure 2 illustrates a systematic framework for urban analytics and big data studies, which places a structure on the problems pertaining to urban sustainability or sustainable urbanism, allowing for reasonable consistency, repeatability, and objectiveness. The derivation of this data mining framework is based on cross-industry standard process for data mining, as supported by several sources, as referenced in.

Using the process of knowledge discovery/data mining is increasingly gaining traction and foothold in many academic and scientific research fields, taking over the scientific method that has

prevailed for centuries. Here, the use of this process is seen as an important and effective way to, in addition to conducting scientific exploration and discovery that are based on big data, solve complex problems within a wide number and variety of domains, including smart sustainable/sustainable smart urbanism. By mining urban data, it is possible to discover laws and principles of sustainable development that pertain to environmental and socio-economic aspects of the city. This development will allow an inference of the varied city stakeholders' responses to operations, functions, services, designs, strategies, and policies in relation to multiple urban domains in the context of sustainability. Indeed, data-analytic and sustainable thinking and practice, as an integrated approach into urbanism, connect the best elements of data science technologies and urban sustainability practices.

Data science has brought a novel approach to the way problems can be conceived of, understood, and tackled within a wide variety of domains. Accordingly, big data computing is changing the paradigm of scientific development, shifting from mainly formulating and testing hypotheses, as well as manually collecting data and examining and reflecting on them to relying more and more on data generation, organization, processing, analysis, modeling, simulation, and verification [24]. This paradigm shift obviously spans many major academic and scientific research domains. In this context, it will help to make decisions easier to judge, knowledge-driven, and strategic, and hence support and enhance existing, and create new, practices, strategies, and policies. For instance, big data analytics and related simulation models and optimisation and prediction methods hold great potential for completely redefining urban problems, as well as offering entirely innovative opportunities to tackle them on the basis of new urban intelligence and planning functions, thereby doing more than merely enhancing existing urban practices. Further, experiences have shown that traditional scientific and academic research paradigms lead to questionable and challengeable assumptions regarding the evolution of social practices. Therefore, it is more beneficial and effective to search for new practices by rather using data-driven research approaches (as part of data-driven science and inductive empiricism), and thus the wider application of big data analytics techniques in the domain of smart sustainable/sustainable smart urbanism. In this sense, new practices can develop around big data technology, which can, in turn, be adapted and integrated into these practices, thereby further advancing its use in a way that fits into a wider strategy or formula that makes this technology more meaningful and relevant at the practical level [24]. All in all, big data analytics is increasingly becoming a salient factor for academic and scientific innovation with regard to addressing complex challenges, wicked problems, and pressing issues, i.e., responding to major environmental concerns and socio-economic needs, mitigating the risks posed by ICT itself to environmental and social sustainability, and containing the potential effects of urbanization.

There has recently been much enthusiasm in the domain of smart sustainable/sustainable smart urbanism about the immense possibilities and fascinating opportunities that are created by the deluge of urban data and its extensive sources with regard to improving urban operational functioning, management, planning, and design in line with the goals of sustainable development as a result of thinking about and understanding sustainability and urbanization and their relationships in a data-analytic fashion for the purpose of generating and applying knowledge-driven, fact-based, strategic decisions in relation to urban domains, such as transport, traffic, mobility, energy, environment, education, healthcare, public safety, public services, governance, economy, and science and innovation [1,58]. This emanates from the ability of big data computing and underlying technologies to effectively monitor, understand, and analyze smart sustainable/sustainable smart cities to improve and maintain their sustainability performance through continuously optimizing their operations, functions, services, designs, strategies, and policies across multiple urban domains, in line with the vision of sustainability. The use of big data computing and the underlying enabling technologies offers the prospect of cities in which natural resources can be sustainably and efficiently managed to enhance societal and economic outcomes by means of data-driven methods. This, in fact, epitomizes what smart sustainable/sustainable smart cities of the future entail and aim for: a set of transformative, innovative urban processes, and approaches that amalgamate technological

capabilities and strategic, data-driven decisions for boosting the performance of urban systems on the basis of a quest for promoting the health of individual citizens, communities, and natural ecosystems; conserving resources; and, fostering economic development [24]. In light of this, the prospect of developing and implementing such cities based on big data analytics and its novel applications is fast becoming the new reality, as manifested in the ever-growing embeddedness of advanced data sensing, data processing platforms, cloud and fog computing infrastructures, and wireless communication technologies as core enabling technologies of big data analytics into the fabric of urban environments for the purpose of solving the challenges of sustainability (e.g., [1,19,24,81]).

Unsurprisingly, big data have become the fundamental ingredient for the next wave of urban analytics and big data studies [1,58]. As a result, many governments have started to exploit and harness urban data to reap their numerous benefits to support the development of their cities regarding to sustainability, efficiency, resilience, equity, and the quality of life. In the meantime, it has become of critical importance for urban professionals, analysts, and researchers to understand the fundamental concepts of data science, and thus data mining/knowledge discovery, even if they never intend to approach urban sustainability or sustainable urbanism problems from a data-analytical perspective, merely because data analysis has now become so critical to urbanism practices. Sustainable cities and smart cities are both increasingly driven by big data analytics (e.g., [1,4,8,19,20,24,58,65,68,81,82]), so there is a great professional and academic advantage in terms of interacting data-analytically with smart sustainable/sustainable smart cities as an emerging holistic approach to urbanism in an efficient and capable way. Understanding the fundamental concepts of data science and making an effective use of the available frameworks for organizing data-analytic thinking, especially the process of data mining/knowledge discovery, not only will allow urban professionals, analysts, and researchers to competently interact with such cities, but it will help to envision tremendous opportunities for improving urbanism practices in the context of sustainability based on data-driven decision-making in the context of smart sustainable/sustainable smart urbanism. The emerging cities badging or regenerating themselves as smart sustainable/sustainable smart are exploiting new and existing data resources for environmental and socio-economic gains and benefits. They gather data science teams and urban scholars and practitioners on common ground to bring big data computing and the underlying enabling technologies, as well as sustainability practices to bear to increase the contribution of such cities to the goals of sustainable development. Increasingly, urban administrators need to oversee analytics teams and analysis endeavors across multiple urban domains, and local city governments must be able to invest wisely in urban projects and initiatives with substantial data assets being directed for improving the different aspects of sustainability, and urban strategists and policy makers must be able to devise plans and design regulatory policies that exploit and leverage data in the needed transition towards sustainability and its advancement [1].

6. Discussion and Conclusions

As a new area of S&T, big data science and analytics possesses unparalleled potential to revolutionize society in a way that no one is able to predict in terms of the dramatic change that it will have on our lives. In this respect, it is dramatically changing the rules and procedures by which smart cities and sustainable cities function in the face of urbanization.

This paper examined the unprecedented paradigmatic and scholarly shifts that the sciences underlying smart sustainable urbanism are undergoing in light of big data science and analytics and the underlying enabling technologies, as well as discussed how these shifts intertwine with and affect one another in the context of sustainability. With respect to the paradigmatic and scientific shifts, it shed light on the old and new way of doing science while providing insights into how big data science and analytics is reshaping the approach to scientific discovery and development. It also elucidated why and how data-intensive science makes a paradigmatic shift from, or an epistemological break with, the current scientific paradigm. Additionally, the dilemma of wicked problems in planning was elaborated, and the relevance of data-intensive approach to urban sustainability science was

highlighted and discussed. However, a worldview that dominates science for a period of time during which that worldview as a reigning single paradigm is extended is characteristic to science. It is worth noting in this regard that the acceptance or rejection of a scientific paradigm is a complex process that involves societal structures. Furthermore, as with scientific facts that are never really more than opinions whose dominance is transitory and far from conclusive, theories should be properly evaluated not as solutions but as beginnings. Therefore, theories cannot tell us how things should be because they are not—yet continue to strive for—a value-free view of reality [17].

With regard to the scholarly shifts, the focus was on the issues, challenges, prospects, and reforms pertaining to the development of the new urban science and the establishment of the related research domain, in addition to big data studies and the process of knowledge discovery/data mining as a new approach of data collection and analysis, which is being increasingly adopted in scholarly research practice within the domain of smart sustainable/sustainable smart urbanism. The new urban science as informed by big data science and analytics heavily depends upon advanced technologies, which provide not only previously inconceivable analytical power and thus answers challenging analytical questions, but also access to huge amounts of data ripe for exploitation with respect to scientific explorations and scholarly investigations. Indeed, many of the questions that are related to sustainability and urbanization and how they interrelate at several levels that urban science attempts to answer have only become tractable because of the increasing availability of urban data. What will be exciting to witness in the near future is how data science will evolve and affect urban science and sustainability science; what new techniques will be invented that would not have come into existence if not for the amalgamation of the parental disciplines of data science, as well as the extent to which they will radically change urban sustainability science; and, what new kinds of urban problems will urban sustainability science, using more advanced big data computing and the underlying technologies, be able to solve. With respect to the latter, there is currently a need for sophisticated digital laboratories, which can be equipped with state-of-the-art tools for modelling multi-dimensional data and simulating urban phenomenon that pertain to sustainability and urbanization.

In the past few years, we have seen the emergence of concepts, such as urban science, urban informatics, urban analytics, urban intelligence functions, and urban simulation modeling (see Bibri 2019 [1] for a detailed account and discussion of these concepts), which are seen to hold great promise for improving the functioning and planning of both smart cities and sustainable cities, as well as for enabling their integration at the technical and policy levels. This integrated approach can be facilitated by the ability to analyse the ever-increasing deluge of the data that are generated by the ubiquitous sensorisation, datafication, and computerization of the built environment, among other things. The opportunities are vast, but so are the challenges of establishing the foundations of the sciences underlying smart sustainable/sustainable smart urbanism. Both empirical work and theoretical advances are needed to cope with the new challenges that are raised by scientific urbanism. Equally challenging is integrating data and models into urban processes and practices: operational functioning, planning, management, development, and governance, which are still largely based on qualitative considerations. There is a sense that state-of-the-art urban intelligence functions and related simulations models and predictive and optimization methods are immature with respect to operational and strategic use as well as institutional integration [1]. There is a need to figure out how to share and harness data in a form that fluidly intersects collective and participatory decision-making processes pertaining to smart sustainable/sustainable smart urbanism. Recent advances in areas, such as agent-based computational modelling, group decision theory, and network theory, as well as the intrinsically holistic approach that was advocated by complexity science, appear as a suitable framework for the development of the sciences underlying smart sustainable/sustainable smart urbanism. This can, in turn, lead to new advances in the way cities are planned, designed, operated, managed, and governed, allowing us to address the enormous challenges that are related to sustainability and urbanization in the 21st century.

Author Contributions: The author read and approved the final manuscript.

Funding: The study is an integral part of a Ph.D. research endeavor being undertaken at NTNU.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bibri, S.E. *Big Data Science and Analytics for Smart Sustainable Urbanism: Unprecedented Paradigmatic Shifts and Practical Advancements*; Springer: Berlin, Germany, 2019.
2. Bibri, S.E.; Krogstie, J. The big data deluge for transforming the knowledge of smart sustainable cities: A data mining framework for urban analytics. In Proceedings of the 3rd Annual International Conference on Smart City Applications, ACM, Tetouan, Morocco, 11–12 October 2018.
3. Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013.
4. Kitchin, R. The real-time city? Big data and smart urbanism. *Geo J.* **2014**, *79*, 1–14. [[CrossRef](#)]
5. Miller, H.J. The data avalanche is here. Shouldn't we be digging? *J. Region Sci.* **2010**, *50*, 181–201. [[CrossRef](#)]
6. Barlow, M. *The Culture of Big Data*; O'Reilly Media, Inc: Sebastopol, CA, USA, 2013.
7. Donoho, D. 50 Years of Data Science (PDF). Based on a talk at Tukey Centennial workshop, Princeton NJ. 18 September 2015.
8. Kitchin, R. The ethics of smart cities and urban science. *Philos. Trans. R. Soc. A* **2016**, *374*, 20160115. [[CrossRef](#)]
9. Bibri, S.E.; Krogstie, J. On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society. *Sustain. Cities Soc.* **2016**, *29*, 219–246. [[CrossRef](#)]
10. Gauch, H.G., Jr. *Scientific Method in Practice*; Cambridge University Press: Cambridge, UK, 2003.
11. Peirce, C.S. A Neglected Argument for the Reality of God; Wikisource. pp. 90–112. with added notes. Reprinted with previously unpublished part, Collected Papers v. 6, paragraphs 452–85, The Essential Peirce v. 2, pp. 434–50, and elsewhere, 1908.
12. Stuart, A.; Ord, K.; Arnold, S. Kendall's advanced theory of statistics: Volume 2A—classical inference & the linear model (Arnold). In *Statistics in Medicine*; Wiley Library: Hoboken, NJ, USA, 1999.
13. Robert, B.J. Thought experiments. In *Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2014.
14. Yeates, L.B. Thought Experimentation: A Cognitive Approach. Master's Thesis, University of New South Wales, Sydney, NSW, Australia, 2004.
15. Wilson, E.O. *The Natural Sciences. Consilience: The Unity of Knowledge*; Vintage: New York, NY, USA, 1999; pp. 49–71.
16. Kuhn, T.S. *The Structure of Scientific Revolutions*. University of Chicago Press: Chicago, IL, USA, 1962/1996.
17. Bibri, S.E. *The Shaping of Ambient Intelligence and the Internet of Things: Historico-Epistemic, Socio-Cultural, Politico-Institutional and Eco-Environmental Dimensions*; Springer: Berlin/Heidelberg, Germany, 2015.
18. Kuhn, T.S. Logic of Discovery or Psychology of Research. In *Criticism and the Growth of Knowledge*; Lakatos, I., Musgrave, A., Eds.; Cambridge University Press: Cambridge, UK, 1972.
19. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J.* **2012**, *214*, 481–518. [[CrossRef](#)]
20. Bibri, S.E.; 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*, 449–474. [[CrossRef](#)]
21. Bettencourt, L.M.A. *The Uses of Big Data in Cities*; Santa Fe Institute: Santa Fe, NM, USA, 2014.
22. Jabareen, Y.R. Sustainable urban forms: Their typologies, models, and concepts. *J. Plan. Educ. Res.* **2006**, *26*, 38–52. [[CrossRef](#)]
23. Neuman, M. The compact city fallacy. *J. Plan. Educ. Res.* **2005**, *25*, 11–26. [[CrossRef](#)]
24. Bibri, S.E. *Smart Sustainable Cities of the Future: The Untapped Potential of Big Data Analytics and Context Aware Computing for Advancing Sustainability*; Springer: Berlin, Germany, 2018.
25. Khan, M.A.; Islam, M.Z.; Hafeez, M. Evaluating the performance of several data mining methods for predicting irrigation water requirement. In Proceedings of the Tenth Australasian Data Mining Conference, Sydney, NSW, Australia, 5–7 December 2012; Volume 134, pp. 199–207.
26. Milovic, B.; Milovic, M. Prediction and decision making in healthcare using data mining. *Intern. J. Public Health Sci.* **2012**, *1*, 69–78.
27. Benevolo, C.; Dameri, R.P.; D'Auria, B. Smart Mobility in Smart City. In *Empowering Organizations*; Springer International Publishing: Cham, Switzerland, 2016; pp. 13–28.

28. Sin, K.; Muthu, L. Application of big data in education data mining and learning analytics—a literature review. *ICTAT J. Soft Comput.* **2015**, *5*, 1035–1049. [[CrossRef](#)]
29. Bachelard, G. *The Formation of the Scientific Mind: A Contribution to a Psychoanalysis of Objective Knowledge*; Beacon Press: Boston, MA, USA, 1986.
30. Foucault, M. *The Order of Things: An Archaeology of the Human Sciences*; Random House: New York, NY, USA, 1970.
31. Lehmann, E.L.; Romano, J.P. *Testing Statistical Hypotheses*, 3rd ed.; Springer: New York, NY, USA, 2005.
32. Lehmann, E.L. Introduction to Neyman and Pearson (1933) on the problem of the most efficient tests of statistical hypotheses. In *Breakthroughs in Statistics, Volume 1*; Kotz, S., Johnson, N.L., Eds.; Springer-Verlag: Berlin, Germany, 1992.
33. Hubbard, R.; Parsa, A.R.; Luthy, M.R. The spread of statistical significance testing in psychology: The case of the journal of applied psychology. *Theory Psychol.* **1997**, *7*, 545–554. [[CrossRef](#)]
34. Lenhard, J. Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson. *Br. J. Philos. Sci.* **2006**, *57*, 69–91. [[CrossRef](#)]
35. Mayo, D.G.; Spanos, A. Severe testing as a basic concept in a neyman–pearson philosophy of induction. *Br. J. Philos. Sci.* **2006**, *57*, 323–357. [[CrossRef](#)]
36. Chow, S.L. *Statistical Significance: Rationale, Validity and Utility*; Sage Publications: Thousand Oaks, CA, USA, 1997.
37. Harlow, L.L.; Mulaik, S.A.; Steiger, J.H. *What If There Were No Significance Tests?* Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1997.
38. Kline, R. *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*; American Psychological Association: Washington, DC, USA, 2004.
39. McCloskey, D.N.; Ziliak, S. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*; University of Michigan Press: Ann Arbor, MI, USA, 2008.
40. Morrison, D.; Henkel, R. (Eds.) *The Significance Test. Controversy*; AldineTransaction: Piscataway, NJ, USA, 2006.
41. Oakes, M. *Statistical Inference: A Commentary for the Social and Behavioural Sciences*; Wiley: Chichester, UK, 1986.
42. Pease, C. Deliberate bias: Conflict creates bad science. In *Science for Business, Law and Journalism*; Vermont Law School: South Royalton, VT, USA, 2006; Chapter 23.
43. Van Gelder, T. *Heads I Win, Tails You Lose: A Foray Into the Psychology of Philosophy*; University of Melbourne: Melbourne, VIC, Australia, 1999.
44. Krinsky, S. *Science in the Private Interest: Has the Lure of Profits Corrupted the Virtue of Biomedical Research*; Rowman & Littlefield: Lanham, MA, USA, 2003; ISBN 978-0-7425-1479-9. OCLC 185926306.
45. Shatz, D. *Peer Review: A Critical Inquiry*; Rowman & Littlefield: Lanham, MA, USA, 2004.
46. Bulger, R.E.; Heitman, E.; Reiser, S.J. *The Ethical Dimensions of the Biological and Health Sciences*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2002.
47. Backer, P.R. *What is the Scientific Method?* San Jose State University: San Jose, CA, USA, 2004; Retrieved 28 March 2008.
48. Bell, G.; Hey, T.; Szalay, A. Computer science: Beyond the data Deluge. *Science* **2009**, *323*, 1297–1298. [[CrossRef](#)]
49. Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*, 23 June 2012. Available online: <https://www.wired.com/2008/06/pb-theory/> (accessed on 12 October 2012).
50. Kelling, S.; Hochachka, W.; Fink, D.; Riedewald, M.; Caruana, R.; Ballard, G.; Hooker, G. Data-intensive science a new paradigm for biodiversity studies. *BioScience* **2009**, *59*, 613–620. [[CrossRef](#)]
51. Bird, A.; Zalta, E.N. (Eds.) Thomas Kuhn. In *Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2013.
52. Blei, D.; Smyth, P. Science and data science. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8689–8692. [[CrossRef](#)]
53. Dhar, V. Data science and prediction. *Commun. ACM* **2013**, *56*, 64. [[CrossRef](#)]
54. Cleveland, W.S. Data science: An action plan for expanding the technical areas of the field of statistics. *Int. Stat. Rev.* **2001**, *69*, 21–26. [[CrossRef](#)]
55. Collins, F.S.; Tabak, L.A. NIH plans to enhance reproducibility. *Nature* **2014**, *505*, 612–613. [[CrossRef](#)]
56. McNutt, M. Reproducibility. *Science* **2014**, *343*, 229–229. [[CrossRef](#)] [[PubMed](#)]
57. Peng, R.D. Reproducible research and biostatistics. *Biostatistics* **2009**, *10*, 405–408. [[CrossRef](#)] [[PubMed](#)]
58. Bibri, S.E. On the sustainability of smart and smarter cities in the era of big data: An interdisciplinary and transdisciplinary review and synthesis. *J. Big Data* **2019**, *6*, 1–64. [[CrossRef](#)]

59. Bibri, S.E.; Krogstie, J. The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: A review and synthesis. *J. Big Data* **2017**, *4*, 38. [CrossRef]
60. West, G.F. Big Data Needs a Big Theory to Go with It. *Scientific American*, 1 May 2013. Available online: <https://www.scientificamerican.com/article/big-data-needs-big-theory/> (accessed on 16 November 2013).
61. Rittel, H.W.J. Panel on policy sciences. *Am. Assoc. Adv. Sci.* **1969**, *4*, 155.
62. Rittel, H.W.J.; Webber, M.M. Dilemmas in a general theory of planning. *Policy Sci.* **1973**, *4*, 155–169. [CrossRef]
63. Reitan, P. Sustainability science—And what’s needed beyond science. *Sustain. Sci. Pract. Policy* **2005**, *1*, 77–80. [CrossRef]
64. Astrom, K.J.; Murray, R.M. *Feedback Systems: An Introduction For Scientists and Engineers*; Princeton University Press: Princeton, NJ, USA, 2008.
65. Kitchin, R. *Data-Driven, Networked Urbanism. A Theory of Good City Form*; MIT Press: Cambridge, MA, USA, 2015.
66. Foth, M. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*; IGI Global: Hershey, PA, USA, 2008.
67. Ratti, C.; Offenhuber, D. *Decoding the City: How Big Data Can Change Urbanism*; Birkhauser Verlag AG: Basel, Switzerland, 2014.
68. Bibri, S.E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustain. Cities Soc.* **2018**, *38*, 230–253. [CrossRef]
69. Kitchin, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*; Sage: London, UK, 2014.
70. McPhearson, T.; Pickett, S.T.; Grimm, N.B.; Niemelä, J.; Alberti, M.; Elmqvist, T.; Qureshi, S. Advancing urban ecology towards a science of cities. *Bioscience* **2016**, *66*, 198–212. [CrossRef]
71. Ribes, D.; Jackson, S.J. Data bite man: The work of sustaining long-term study. In *'Raw Data' is an Oxymoron*; Gitelman, L., Ed.; MIT Press: Cambridge, MA, USA, 2013; pp. 147–166.
72. Bowker, G. *Memory Practices in the Sciences*; MIT Press: Cambridge, MA, USA, 2005.
73. Gitelman, L. *Raw Data is an Oxymoron*; MIT Press: Cambridge, UK, 2013.
74. Kitchin, R.; Lauriault, T.P.; McArdle, G. Knowing and governing cities through urban indicators, city benchmarking & real-time dashboards. *Reg. Stud. Reg. Sci.* **2015**, *2*, 1–28.
75. Greenfield, A. *Against the Smart City*; Do Publications: New York, NY, USA, 2013.
76. Mattern, S. 2013 Methodolatry and the Art of Measure: The New Wave of Urban Data Science. *Design Observer: Places*, 5 November 2013. Available online: <http://designobserver.com/places/feature/0/38174/> (accessed on 15 November 2013).
77. Marvin, S.; Luque-Ayala, A.; McFarlane, C. (Eds.) *Smart Urbanism: Utopian Vision or false Dawn?* Routledge: London, UK, 2016.
78. Bibri, S.E. A foundational framework for smart sustainable city development: Theoretical, disciplinary, and discursive dimensions and their synergies. *Sustain. Cities Soc.* **2018**, *38*, 758–794. [CrossRef]
79. Robinson, J.D.; Parnell, S. The global urban: Difference and complexity in urban studies and the science of cities. In *The SAGE Handbook of the 21st Century City*; Hall, S., Burdett, R., Eds.; SAGE: London, UK, 2017; pp. 13–31.
80. Munafò, M.R.; Nosek, B.A.; Bishop, D.V.; Button, K.S.; Chambers, C.D.; Du Sert, N.P.; Ioannidis, J.P. A manifesto for reproducible science. *Nat. Hum. Behav.* **2017**, *1*, 1–9.
81. Al Nuaimi, E.; Al Neyadi, H.; Nader, M.; Al-Jaroodi, J. Applications of big data to smart cities. *J. Internet Serv. Appl.* **2015**, *6*, 1–15. [CrossRef]
82. Bibri, S.E.; Krogstie, J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustain. Cities Soc.* **2017**, *31*, 183–212. [CrossRef]

