

12-1-2020

Opportunities and challenges of geospatial analysis for promoting urban livability in the era of big data and machine learning

Anna Kovacs-Györi
Universität Salzburg

Alina Ristea
Northeastern University

Clemens Havas
Universität Salzburg

Michael Mehaffy
The Royal Institute of Technology (KTH)

Hartwig H. Hochmair
University of Florida Institute of Food and Agricultural Sciences

See next page for additional authors

Follow this and additional works at: https://scholarworks.sjsu.edu/faculty_rsca

Recommended Citation

Anna Kovacs-Györi, Alina Ristea, Clemens Havas, Michael Mehaffy, Hartwig H. Hochmair, Bernd Resch, Levente Juhasz, Arthur Lehner, Laxmi Ramasubramanian, and Thomas Blaschke. "Opportunities and challenges of geospatial analysis for promoting urban livability in the era of big data and machine learning" *ISPRS International Journal of Geo-Information* (2020). <https://doi.org/10.3390/ijgi9120752>








This Article is brought to you for free and open access by SJSU ScholarWorks. It has been accepted for inclusion in Faculty Research, Scholarly, and Creative Activity by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Authors

Anna Kovacs-Györi, Alina Ristea, Clemens Havas, Michael Mehaffy, Hartwig H. Hochmair, Bernd Resch, Levente Juhasz, Arthur Lehner, Laxmi Ramasubramanian, and Thomas Blaschke

Article

Opportunities and Challenges of Geospatial Analysis for Promoting Urban Livability in the Era of Big Data and Machine Learning

Anna Kovacs-Györi ^{1,*} , Alina Ristea ² , Clemens Havas ³ , Michael Mehaffy ⁴,
Hartwig H. Hochmair ⁵ , Bernd Resch ^{3,6} , Levente Juhasz ⁷ , Arthur Lehner ³,
Laxmi Ramasubramanian ⁸ and Thomas Blaschke ³ 

¹ IDA Lab, University of Salzburg, 5020 Salzburg, Austria

² Boston Area Research Initiative, School of Public Policy and Urban Affairs, Northeastern University, Boston, MA 02115, USA; a.ristea@northeastern.edu

³ Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria; clemensrudolf.havas@sbg.ac.at (C.H.); bernd.resch@sbg.ac.at (B.R.); arthur.lehner@stud.sbg.ac.at (A.L.); thomas.blaschke@sbg.ac.at (T.B.)

⁴ Centre for the Future of Places, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden; michael.mehaffy@gmail.com

⁵ Fort Lauderdale Research and Education Center, University of Florida, Davie, FL 33314, USA; hhhochmair@ufl.edu

⁶ Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA

⁷ GIS Center, Florida International University, Miami, FL 33199, USA; ljuhasz@fiu.edu

⁸ Department of Urban & Regional Planning, San José State University, San Jose, CA 95192-0185, USA; laxmi.ramasubramanian@sjsu.edu

* Correspondence: anna.gyori@sbg.ac.at; Tel.: +43-662-8044-5341

Received: 30 October 2020; Accepted: 10 December 2020; Published: 15 December 2020



Abstract: Urban systems involve a multitude of closely intertwined components, which are more measurable than before due to new sensors, data collection, and spatio-temporal analysis methods. Turning these data into knowledge to facilitate planning efforts in addressing current challenges of urban complex systems requires advanced interdisciplinary analysis methods, such as urban informatics or urban data science. Yet, by applying a purely data-driven approach, it is too easy to get lost in the ‘forest’ of data, and to miss the ‘trees’ of successful, livable cities that are the ultimate aim of urban planning. This paper assesses how geospatial data, and urban analysis, using a mixed methods approach, can help to better understand urban dynamics and human behavior, and how it can assist planning efforts to improve livability. Based on reviewing state-of-the-art research the paper goes one step further and also addresses the potential as well as limitations of new data sources in urban analytics to get a better overview of the whole ‘forest’ of these new data sources and analysis methods. The main discussion revolves around the reliability of using big data from social media platforms or sensors, and how information can be extracted from massive amounts of data through novel analysis methods, such as machine learning, for better-informed decision making aiming at urban livability improvement.

Keywords: spatial data science; livability; urban planning; big data; urban assessment; spatio-temporal analysis

1. Introduction

International landmark agreements such as the Sustainable Development Goals (SDG) or the New Urban Agenda (NUA) have focused attention on the improvement of the quality of urban life

in the past few years [1,2]. These agreements are addressing several natural and social challenges in cities including phenomena such as resilience to natural disasters, climate change, air pollution, access to clean water, public health, or equity in the access of public spaces. By tackling these issues, the overall goal is to provide better and safer circumstances for the growing urban population worldwide. Nowadays, also thanks to these international agreements, the existence and severity of these problems is getting more obvious and acknowledged, along with the importance of raising awareness amongst citizens and decision-makers. However—especially from the urban planners' point of view, by whom the majority of the local actions should be initiated—the NUA and SDG targets represent only the starting point of the long and challenging endeavor of actually solving the mentioned problems [3–6]. Although researchers and decision-makers are aware of what goals should be achieved and why, the actions should be adapted to the specific local circumstances and shorter timeframes, which is highly challenging mainly due to the complex nature of cities. In this paper, we intend to illustrate the opportunities and limitations of Geographic Information Systems (GIS)-based urban analysis techniques and novel data sources such as social media or sensor data to address these challenges of urban livability improvement, which includes big data analysis and machine learning approaches, among others.

The last decade has seen an increasing integration of GIS in everyday life and an almost ubiquitous production of big data across various contexts. These coupled technological and societal developments have resulted in the emergence of new fields, such as urban data science, and urban informatics [7–9]. These new intellectual domains enrich the study of urban systems and the practice of urban and regional planning or policy by making use of new datasets and novel methodological approaches [10,11]. Various disciplines are involved in shaping these interdisciplinary fields, including computer science, geography, geographic information science (GIScience), as well as urban studies. Such a revolutionary take on the investigation of cities brings along several advantages and can also benefit practical domains like urban planning.

However, understanding cities and their processes thoroughly, which is a key to improve urban livability, requires exploration beyond the merely technical and data-driven approaches to analysis techniques offered traditionally by methodological disciplines. This is what urban informatics and urban data science aim to achieve as a field “at the intersection of people, place and technology with a focus on cities” [7] (p. 1). Doing so will result in a better understanding of the theoretical foundations underlying human behavior in urban environments. This, in turn, is necessary to conceptualize and assess livability and to provide information for urban planners about the processes at different spatial (and often temporal) scales.

Although machine learning, urban data science and in general the use of big data in urban analytics have already received attention in the literature [10,12–14], the goal of this paper is to discuss some synergies and challenges of the interdisciplinary approach on GIS-based urban analysis in improving the livability of cities, by reviewing the key findings and open questions emerging from the state-of-the-art literature. The current work addresses three interrelated research questions concerning new data sources and novel analysis techniques in the context of cities and urban planning.

- RQ1: How can we assure the reliability of results based on social media data and Volunteered Geographic Information (VGI) for urban livability analysis and planning?
- RQ2: How can geospatial analysis aid urban livability assessment and planning that relies on machine learning methods?
- RQ3: How can relevant information be identified in urban big data to facilitate urban livability improvement?

Before investigating these questions in detail, Section 2 reviews urban theories related to cities and livability, as well as highlights use cases to demonstrate the potential of spatio-temporal analysis for urban planning using Twitter data. Section 3 provides a summary of methodological and data-related aspects of RQ1 and RQ2, whereas Section 4 describes a potential approach to extract information

(‘signal’) from the ‘noise’ of vast amounts of data focusing on the usage patterns and livability of cities instead of purely data-driven solutions, which is closely related to RQ3. Finally, Section 5 summarizes relevant findings concerning the potential and pitfalls of urban analysis and presents open questions for future research. We hope that this work will put forward the dialogue between methodological and academic research and benefits urban planning practitioners as well. Also, we hope to trigger a more intense debate about the values and challenges implied by making more extensive use of data in urban planning along with livability assessment and improvement.

2. Urban Theories and Assessment in the Light of Livability and BIG Data

In the 1930s, Lewis Mumford posed a simple question, “What is a City?” [15]. The simple answer, of course, is that cities are surprisingly complex. We argue that it is this complexity that has confounded legions of architects, planners, engineers, data scientists, and city managers who seek to impose a spatial (and social) order to make cities more ‘manageable’. Historically, social scientists have proposed theoretical frameworks that help to explain and predict the morphological structure of cities beginning in the early part of the 20th century such as Park and Burgess [16] to contemporary theorists like David Harvey [17] and Sassen [18]. For these social scientists, the city’s complexity is shaped by demographic, social, political, social, and technological changes. Collectively, these scientists take a process-oriented approach to city transformations. Yet, for the past 200 years or so, modern city development has also been shaped by positive visions of the future; many of which were products of individual imaginations [19,20] specifically intended to address the challenges of urban living. All these historic visions have addressed the challenges of density (some visions have advocated for more density than less) and spatial scale (large megalopolises versus small and intimate), alongside considerations of efficiency, access, and mobility. It is important to note that these are the same concerns that drive the conversations about ‘smart cities’ and the use of technology and data in urban planning.

The intention to assess cities and their livability has a long history, and the ubiquity of digital technology and big data opened up new horizons also in this regard [21]. The emergence of new aspects and methodologies such as big data [22], VGI [23,24], or citizen-generated data, coupled with social media analysis [25,26], and even remote sensing or earth observation [27] might offer relevant alternatives to traditional methodologies and concepts mentioned above. These approaches provide ways to extract useful information from data, also at finer spatial and/or temporal scales than national statistics, and if needed, even for individuals or various social groups. To a certain extent, it is now at least technologically possible to analyze how people interact with the urban environment. However, quite often this analysis focuses on only quantitative approaches [26,28–32] and neglects the subjective needs and perception of the residents. Therefore, it is the socio-spatial theories of urban planning that offer the most promise when trying to utilize advantages provided by big data to explain, predict, and manage future cities.

2.1. Urban Morphology Assessment

Urban planners, particularly scholar-practitioners have long championed bottom-up theory development, in other words, the development of explanatory frameworks about urban living through inductive reasoning. For instance, Kevin Lynch [33] proposed that urban environments that have clearly “imageable” (memorable) characteristics are more likely to be recognized, i.e., legible. Lynch put forward the notion that when cities or urban environments are memorable, that they create feelings of safety and security, and that they encourage movement and participation in the life of the city. Lynch proposed that humans created environmental images of their built environment by creating mental maps. These maps were easier to create for some cities/urban environments than others. Lynch also proposed a vocabulary to understand and describe urban environments—paths, edges, nodes, landmarks, and districts. Lynch used observational and survey data from everyday participants in three cities—Boston, Jersey City, and Los Angeles—to arrive at his seminal conclusions. The power of urban analytics nowadays provides the opportunity to better understand how individuals create

and store their mental maps and how these maps are deployed under different circumstances—for example, in times when they are under time constraints, or in other stressful situations. In addition, it facilitates to understand how people move spatially and temporally through urban environments, Lynch's theoretical frameworks can help identify and replicate successful urban morphologies—some street patterns work well and hold up to the test of time, others do not. Big data approaches help rapid prototyping and design of new environments in ways previously not feasible [34].

2.2. Urban Livability Assessment

In an environment of rapid transformation and locational competition on a global scale, research on the livability of cities has gained much attention during the last years. Research investigating the quality of life in cities has already started very early to deal with the spatial relationships of and within cities due to the complexity and diversity of livability standards, taking into account the global aspect of such standards and indices [35–37]. One declared important objective of such research is the objectification of economic, social, and structural-spatial factors of influence. In this context, Onnom et al. [38] discuss several indices that allow for measuring the quality, the development and the livability of cities. Others investigate the relationship between the livability of an urban area and its (sustainable) development [37,39].

Based on the research of Kovacs-Györi et al. [40] we argue that livability can serve as a useful conceptual and analytical framework to assess and improve the quality of urban life, also considering the personal aspects of the residents by focusing on the personal-environment relationship instead of merely statistical indices with no spatial relevance. Figure 1 shows the key elements and their relationship in livability to clarify what is meant conceptually when using this term throughout the article [40]. The lower part represents the urban environment (i.e., actual state), which consists, on the one hand of the built and natural environment along with the infrastructure, described as urban form. It also includes what this environment can provide for the citizens as urban functions, reflecting how the people use the urban space in the light of the functionality. This component of livability is more tangible, and spatial data for further analysis are available or acquirable, respectively. The upper part of the figure corresponds to the individual preferences (i.e., expectations) in livability and the person–environment relationship, consisting of human needs that the urban functions (or even urban form) need to satisfy and the personal values that define and influence the person's preferences. Due to the high level of subjectivity, individual preferences bring an inherent limitation to livability assessment, as means or data to extract these preferences are often not available in an intersubjective way. Based on this framework, urban livability describes how well the environment can satisfy citizen needs and expectations, i.e., to what degree the actual state matches the goal state.

However, as mentioned above, in the context of urban planning analysis, the difficulty exists in two main circumstances.

First, it is often challenging to derive recommendations for actions to urban planners, because they are often limited in the actions they can take. In private space, planners can only be active in supervision or demand for specifications in a restricted framework. In public space, they have more freedom of action, yet they are bound by various guidelines and social necessities and must take into account the private space. The translation of identified indicators and parameters into planning-relevant actions must therefore take place in such a way that it facilitates the identification of space for maneuver. Such methodologies target the simplification of the identification of the effects of measures and facilitate scenario formation when there are alternative courses of action.

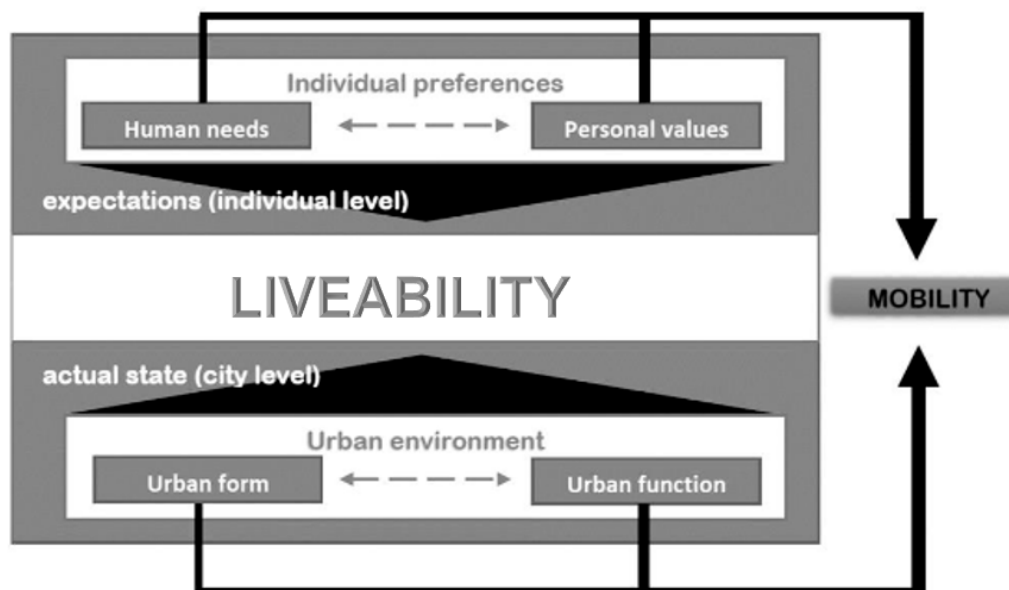


Figure 1. Conceptual framework for livability [40].

Furthermore, the mere accessibility depicted by minutes or meters cannot reflect the purpose or the quality of the satisfaction of needs at the target destination nor the actual quality on the way itself; the crucial question why a way or a means of transport has not been chosen; as such, this question cannot be directly mapped or only indirectly depicted. In times of resource scarcity and decreasing urban budgets, providing answers to such questions is essential for efficient urban planning. In this way, and as highlighted in Figure 1, mobility has a special role within livability assessment both in terms of conceptual and practical aspects. Collecting data on mobility is quite well-established and also various analysis techniques are available depending on the purpose of the analysis. Besides direct information about the paths or routes of residents, different approaches exist to provide information about mobility and its qualitative aspects. Pigliautile et al. [41] present wearable sensing techniques and automatized data-processing for urban microclimate mapping. Such mapping could provide information about the comfort of pedestrians and for other outside traffic participants that move within the urban area. A single trajectory can tell a lot in terms of subjective preferences and the satisfaction of needs that are both cornerstones of livability assessment as described above. This is because the destination usually reveals the need to be satisfied whereas the transportation mode and the route taken can provide insights into personal characteristics or routing preferences. Overall, we think that mobility analysis where livability aspects are considered (even only at a conceptual level) can combine the strong theoretical knowledge from urban planning with the new technological advancement from GIS and big data resulting in new approaches to improve urban quality of life.

2.3. Illustrations of GIS-Based Social Media Data Analysis to Support Urban Planning

In this section, we provide practical examples from previous research, to illustrate the potential of geospatial analysis for urban planning purposes using social media data. These examples provide only an insight into how geospatial analysis of Twitter data can facilitate urban planning but we do not provide an extensive review of the state-of-the-art approaches in this field. However, Section 3 reviews more in depth the potential of VGI in general for these purposes.

In Sections 2.3.1 and 2.3.2, it is demonstrated how the analysis of Twitter data can support urban planners by providing new insights about urban phenomena and challenges. Concept-wise, morphology (2.1) and livability (2.2) assessments are good examples of methodological solutions from the GIS domain to provide new insights for urban theories that have a long history. This often happens by applying mixed methods to grasp the complexity of urban phenomena. Mixed methods research originated in the social sciences and it is a third category of approaching an analysis, by combining

the first two categories, namely qualitative and quantitative. The main idea of this methodology of research is that such complex integration permits developing better synergies in using the data than by following separated collection and analysis [42]. More than two decades ago, urban researchers discussed the possible integration of mixed methods strategies in planning, with examples from New York City [43]. However, with the growth of data and technologies, the in-depth mixed methods got into development about a decade ago. Although these approaches are often difficult to evaluate, due to the data-rich environment they can still be informative, as it is illustrated throughout this section.

2.3.1. Towards Citizen-Contributed Urban Planning Using Twitter Data—Case Study for Planned Large Events

In a 2018 study, Kovacs-Györi et al. [44] investigated how social media data can be utilized to investigate planned large events in cities, using the Olympic Games in London, 2012. Similar studies have been performed also for the Rio Olympics [45], transportation during planned (and unplanned) events [46], or to detect events based on Twitter data [47–49].

Using over 12 million tweets, the authors in [44] identified potential residents and visitors of the Olympic Games in 2012 and analyzed their tweeting behavior in terms of spatial and temporal patterns, as well as the sentiments of their tweets. The analysis of the spatio-temporal and sentiment patterns revealed that it was possible to identify (spatial) hotspots for both groups during the Olympic Games, and these patterns were significantly different. The real potential of this result is to provide information for planners and decision-makers even during an event with a quite high spatial and especially temporal resolution, which is often a limitation with traditional surveys or other existing authoritative methods. However, it is important to emphasize that the usefulness of social media data (or VGI in general) highly depends on the purpose of the analysis. As different social media and VGI platforms are not used by each demographic groups equally, researchers must consider how this inherent selection bias might influence their results. In the case of a planned event or disasters the high spatial and temporal resolution of the data and their near-real time availability bring clearly more advantage and information even if it is not fully representative, compared to other scenarios, where representativeness might be essential such as the availability of specific urban services. Section 3 describes further considerations and analysis techniques regarding the reliability of different VGI data sources.

Another method demonstrated in the paper [44] was topic modeling to extract topics from the tweets of the previously identified groups of visitors and residents. This method can also be interesting for urban planners and other stakeholders to get indirect feedback regarding different aspects (e.g., transportation) from the two groups during an event. In a near-real-time scenario, it can be informative, especially in bigger amounts. The paper includes further assessment potentials both on micro and macro scale, usually by suggesting the combination of different data sources to dig even deeper into urban phenomena.

2.3.2. Classifying Parks and Their Visitors in London Based on Twitter Data Analysis

In another study, Kovacs-Györi et al. [50] investigated how social media data analysis can help the planners monitoring urban park visits and the perception of the visitors. Twitter data is often used for this purpose [51], for example to investigate seasonal variation in physical activities [52], or determine the emotional responses of people to urban green spaces [53]. Similar to the above-presented case study for the Olympic Games and other large events, the high spatial and temporal resolution of the tweets can provide new insights compared to other methods, such as in-situ assessments (surveys, counting, etc.). The study included the analysis of the distance between park tweets and the activity center of the user (geometric average of the coordinates where the given user tweeted from), followed by a sentiment analysis and emotion extraction of the tweets. By performing spatio-temporal analysis and clustering it was possible to identify different types of parks based on the time of the visit (during the day, week, and year) and also to identify parks where people tended to tweet more positively or

experiencing specific emotions. The research demonstrated, also in this case, the potential of using social media and GIS analysis for urban planning in this regard, as well as pointed out some of the limitations (most of all the lack of representativeness of the dataset). Overall, this method can serve as an input for more in-depth analysis performed by planners or other stakeholders, by revealing new insights or providing information about parks in greater amount and for a longer time than it is possible with traditional techniques, such as questionnaires.

3. Valuable Data Source Types and Methodologies for Analyzing Complex Urban Systems and Their Quality

Urban informatics can tap into a rapidly growing number of data sources providing a large variety of data sets of different spatial and temporal scales, completeness, and reliability in different combinations [54]. For example, metro card user data or social media check-ins allows analyzing and modeling people's individual movement, property appraisal data allow for an estimation of parcel values and determining factors using hedonic models, whereas traffic cameras or inductive loops provide information about traffic congestion at certain road segments. On the temporal scale, weather station data from fixed sensors provide both hourly information but can also be used for analyzing long term trends in climate change. Similarly, floating car data provide continuous information about speed, however at varying locations depending on the cars' locations. Other geo-data, such as household traffic surveys or on-board transit surveys, are collected less frequent at a smaller sample size which limits suitability for local analysis tasks. The different channels provide a large pool of data, which renders machine learning algorithms and artificial intelligence potentially useful for the identification of spatio-temporal activity patterns of various phenomena in urban environments at all spatial and temporal levels. Artificial intelligence, including machine learning and natural language processing as methods for urban data science, have, for example, been used for the early detection of disease outbreaks, such as COVID-19, Zika, or H1N1 Influenza, based on big data from health-wearable devices [55], or for urban flood risk mapping in urban environments [56].

3.1. User-Generated Data Sources

Quantitative urban research has traditionally relied on data from censuses, surveys, and specialized sensor systems. Low response rates of surveys and the high cost associated with surveys and replacing sensor systems, have, however, led to increasing interest in alternative ways of supplementing the urban data infrastructure [28]. The emergence of Web 2.0 has led to a plethora of VGI and social media data as part of big data that has been used for a wide range of urban analyses also addressing societal questions. This includes exploring urban dynamics from open and user-generated content posted on location-based social networks Foursquare, Google Places, Twitter, and Airbnb [57], assessing the effect of natural disasters on urban mobility patterns using tweets [58], or for monitoring the development of the COVID-19 epidemic through WeChat keyword analysis [59]. User-generated data sources provide different types of geo-data, including points of interest, event patterns, travel trajectories, traffic information, sentiments, or social activities. Such data form the building blocks for geo-applications and analysis tools in the context of urban planning. Consequently, knowledge about the quality of crowdsourced data is important to determine their fitness for purpose [60]. Whereas the abundance and multi-dimensionality of crowd-sourced big data lend itself to machine learning techniques in the spatial and temporal domain, machine learning has also been used to assess and improve the data quality of VGI, such as OpenStreetMap [61], and to remove automated bots in social media data, such as tweets [62].

The International Organization for Standardization Technical Committee (ISO/TC) 211 developed a set of international standards that define the measures of geographic information quality and provide a structure for describing digital geographic data. The ISO metadata standard 19,115 [63] defines the schema required for describing geographic information and services by means of metadata. It distinguishes between five quantitative measures (completeness, consistency, positional accuracy,

temporal accuracy, and thematic accuracy) and three qualitative quality indicators (purpose, usage, and lineage). At the emergence of crowdsourced data comparison to authoritative data (called extrinsic methods) was a common approach to determine data quality [64,65]. With the absence of available comparison data, additional qualitative indicators as a proxy for traditional ISO-defined data quality indicators [66] were developed, partially mitigating the need for reference data but rather relying on contribution history. Hence, data quality assessment began to utilize intrinsic methods recently as well [67]. These methods rely exclusively on the analyzed data source itself. Examples include the analysis of historical metadata as a means of inferring the inherent quality of the data [68], or the measurement of text metrics (e.g., number of words, structure, spelling) to determine the quality of crowd-sourced text content [69].

Alternatively, to determine their data quality, crowd-sourced data can also be compared to other non-authoritative data [70–72], to the surrounding environment [73], or data contributed by other volunteers [74]. Furthermore, a volunteer’s trustworthiness, credibility, experience, recognition, or reputation can be used as a proxy for the quality of contributed data [75,76]. Other measures to determine the quality of user-generated content include its localness [77,78], or its vagueness [79]. Vagueness can, for example, be caused by low-resolution imagery, or by capturing only a snapshot of a temporal phenomenon. Data quality of crowdsourced datasets vary between world regions [80], and the spatial distribution of such datasets is uneven [81]. Since urban environments are centers of human activities (jobs, social life, leisure, high-density housing), urban areas provide a rich source of geo-data, which are potentially suited for spatial and spatio-temporal analytics in different research domains including human mobility, health and environment, social interaction, food and nutrition, gentrification, land-use, and urban functionality. These geo-data are of different lineage and type, and therefore may be best described with different quality measures or indicators. A detailed overview of existing methods to assess the quality of map-based VGI (e.g., positional accuracy, completeness), image-based VGI (e.g., thematic accuracy of tags), and text-based VGI (e.g., credibility) is provided in the work of Senaratne et al. [82]. Table 1 summarizes commonly used VGI quality assessment methods (extrinsic vs. intrinsic) together with the type of used reference datasets and quality measures. An x indicates the primary use of these measures, whereas (x) indicates less common use for a given method.

Table 1. Different methods and measures for VGI quality assessment

Assessment Method	Extrinsic		Intrinsic
	Authoritative	Non-Authoritative	None
Comparison Reference Dataset			
Measures			
Quantitative measures (completeness, consistency, positional/temporal/thematic accuracy)	x	x	(x)
Qualitative indicators (purpose, usage, lineage)			x
Proxy quality indicators (trustworthiness, credibility, text content quality, vagueness, local knowledge, experience, recognition, reputation)	(x)	(x)	x

3.2. Improving Data Quality and Reliability in Urban Analysis by Combining Data from Several Crowd-Sourcing Platforms

A user’s credibility has traditionally been derived from contribution patterns to individual citizen science projects, mapping platforms, or social media apps—such as Ushahidi [76], Flickr [83], or Twitter [84]—which also led to the development of methods for detecting data vandalism [85]. However, lately, individual users began to become active in multiple crowdsourcing platforms [86], and to use multiple social media services simultaneously. Therefore, we propose to extend the assessment of a user’s credibility and trustworthiness through consideration of activities across multiple platforms. One obvious challenge in that aspect is to identify which users contribute to multiple platforms. This is not a trivial task and largely unaddressed in the GIScience literature. An online survey—conducted by Levente Juhász [87]—was performed to test this hypothesis on a small

sample. Participants were asked about their geospatial application usage and requested to share their online activity in 10 geospatial applications. Participants were recruited through university mailing lists, discussion forums and social media channels, and the study mainly focused on Europe and the United States. Figure 2 illustrates the results where 32 out of 53 responding users indicated that they use more than one geolocated social media or crowdsourcing platform. Considering the identification of individuals who are using multiple social media platforms, a common approach is to try to match user names between different platforms. More computationally complex and resource-intensive methods could apply concepts of space-time geography on user contribution patterns to identify contributions to different platforms from the same user [88]. The basic idea behind this would be that space-time prisms, which show an individual's reachable positions based on a person's capability constraints (e.g., travel speed) or authority constraints (e.g., rules and regulations), determine whether locations extracted from different data sources could have been visited by the same individual user or not. With millions of users contributing to different platforms, matching up contributions from different platforms and identifying them to be from the same user, if so, will require the use of advanced space-time indexing, data mining approaches, and machine learning techniques [89]. There are also ethical challenges associated with linking individuals and their activity between multiple online platforms, which stem from the right to privacy. Therefore, combining data from multiple platforms should also consider locational privacy. This is best addressed by an approach that combines legal, ethical, technological, and education aspects of the issue [90,91].



Figure 2. Social media and crowdsourcing platforms used by a set of 53 users.

Although the small sample size of the study mentioned above does not allow to draw generalized conclusions about the percentage of social media users who use multiple platforms, we assume that data from multiple data sources (e.g., Twitter and Instagram) can be combined to provide a more

complete spatio-temporal representation of an individual's activities compared to information extracted from a single platform (e.g., Twitter). An illustration of this is given in Figure 3, where maps a–d show locations and activities extracted from Web 2.0 platforms over the course of a day for one user in a region in southern Austria. Figure 3e overlays all geocoded activities of the same individual during the same day and therefore provides a more complete spatial representation of the user's activities than one single platform. It also provides information about transportation between the cities (i.e., Strava cycling activity). Although this example suggests that the approach leads to increased credibility and completeness, this step of data fusion from different sources needs to be repeated for a larger, more representative, and diverse user base to be able to draw conclusions about its usefulness and applicability for a more general population. Challenges, such as feature matching, which are inherent to data fusion from different sources [92], need to be addressed within this process.

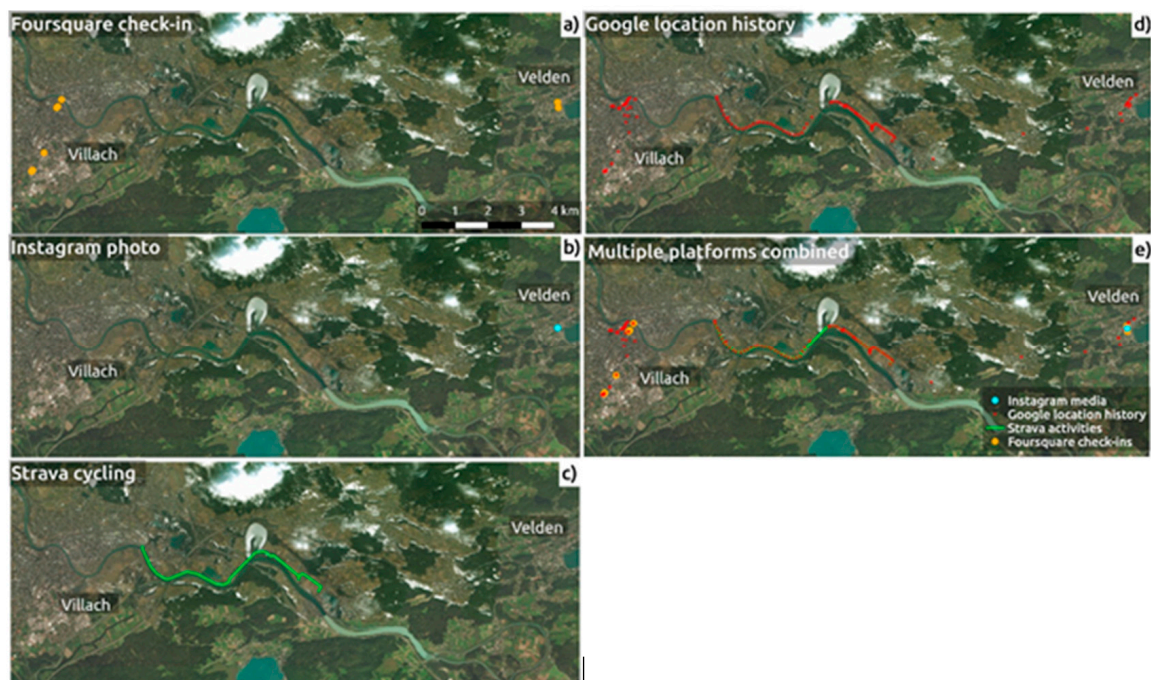


Figure 3. Spatial representation of user activities extracted from single platforms for a day (a–d), and data sources combined together providing a more complete picture of the user's activity locations (e).

Exploring the relationship between social media activity spaces of a user on different platforms and quantifying their similarity is a novel research direction [93]. Understanding the limitations of using data from only one platform and the potential benefits of combining user data on the individual level is relevant, since urban analytics already utilizes social media and crowdsourced datasets to extract valuable spatio-temporal information to understand human mobility and citizen opinions on certain topics [40,44].

A new trend in contribution patterns of user-generated content is also to cross-reference information between platforms, i.e., to obtain information from one data source and insert it into another, which can be referred to as “cross-tagging” [86]. An example is to use the hotel name found on a Mapillary image, which is then attached to an OSM feature as a name attribute value during the OSM editing procedure, based on which the Mapillary photo ID is linked to the map feature. Automated systems can then utilize this link to display an image of the linked map feature. In addition, contributors can also look up information from various data sources and platforms and use that information in other platforms, without explicitly documenting this process in the edited data. This process can be referred to as “cross-viewing” [94]. It is yet an open question if and how cross-tagging and cross-viewing affect data quality of the connected data sources. On the one side, it can be argued that cross-tagging

and cross-viewing add more eyes on the data, which should improve data quality, based on Linus' Law [95]. However, if the original data is not cross-checked, its potentially inferior data quality may propagate across platforms and negatively affect the data quality of the targeted data source, similar to how correct (or incorrect) news can propagate within the social media landscape [96] or how errors propagate in measurement-based GIS into the indirect values of the end product [97]. Joint use of datasets from different sources can provide better completeness, e.g., in feature change detection over time, which has been applied in urban environments [98]. However, different demographic groups do use crowd-sourcing tools, such as bicycle tracking apps [99] or social media apps [100] differently, which is important to consider in efforts of data fusion from different apps and in conclusions drawn from this. As far as cross-tagging goes, we propose to add another quality measure of user-generated content, namely the abundance of cross-linked data sources in the data set of interest.

In conclusion, it can be stated that the evolution of novel data sources from user-generated content, and their combination, necessitate a review and adaptation of quality metrics to appropriately capture and describe their validity and usefulness.

3.3. Potential of Machine Learning Algorithms in Geospatial Urban Analysis and Assessment

In a recent transformational development, the GIScience community has faced a transition from a data-scarce to a data-rich environment [101]. This advancement also reshapes the scientific discipline of geoinformatics through the more acknowledged investigation of data-driven approaches. Therefore, the development of new methods for data acquisition, storage, and analysis become inevitable, including unsupervised machine learning algorithms or semi-supervised learning systems. This also applies to the particular idea of assessing walkability, where researchers have used machine learning algorithms, for instance, on Google Street View imagery to generate three measures on visual enclosure [102]. The proposed methodology addresses the dimension of urban design that has usually been analyzed through urban characteristics like street network density or block size. The authors adopt an approach that uses Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to identify sky areas for measuring the proportion of visible sky from using Google Street View imagery, from which a measure for visual enclosure is derived.

Also, Resch et al. [25] developed a methodology that extracts emotion information from social media for improving urban planning processes. The authors present a semi-supervised machine learning approach that is based on similarities between social media posts in geographic space, temporal space, and semantic space. They conclude that social media are a valuable and reliable data source for urban planning related issues, but that traditional GIS methods are oftentimes not appropriate for analyzing social media data. These and other recent machine learning-based approaches show that the dichotomy between purely hypothesis-driven approaches (mostly developed by domain experts) and data-driven approaches (oftentimes pursued by computer scientists and statisticians) are currently closing up. In other words, urban planners tend to have a better understanding of the technical capabilities of cutting-edge technology, while computer science-oriented researchers become more and more familiar with the concepts developed in the urban planning community. This makes the field of urban analysis a highly interdisciplinary one by bringing together experts from a variety of scientific fields.

With respect to the analysis of human-generated data like social media to better understand urban contexts, the high degree of uncertainty—including textual ambiguities, positional and temporal inaccuracies, and semantic irregularities—has sparked the development of new machine learning-based analysis methods [12]. Various machine learning methods have recently been used and developed for analyzing urban spaces, including self-organizing maps to map recreational beneficiaries [103], to analyze urban traffic conditions [104], and to mine latent urban activity patterns [105], or methods of natural language processing (NLP) to extract urban perceptions [106]. These research efforts allow for generating new, unseen insights into urban processes through analyzing a new set of data sources that

previously posed severe challenges to scientific research due to their unstructured, unstandardized, and voluminous nature.

Concluding, these new algorithms seem to be a promising avenue for several reasons: First, they reduce or eliminate the need for a priori knowledge with respect to linguistic structures, geospatial correlations, or semantic meaning. Second, it is possible to incorporate the geospatial dimension into the analysis process, in many cases in a simple fashion without having to modify the original method. This step is essential because many machine learning algorithms have originally not been designed for handling and analyzing geospatial data. Third, machine learning methods have the power to deal with large amounts of data in that data-driven approaches can be applied to mine latent, unanticipated patterns in human-generated data [12].

4. Potential Methodological Problems for Urban Planning and Livability

4.1. Finding the Signal in the Noise

For urban planners, the era of large-scale geodata sets, and other forms of big data, is creating remarkable opportunities, but also raising a number of important problems. One of the most challenging problems is the sheer volume of the data available, which is often noisy in character. That is, it has inherent imprecision, ambiguity and inaccuracy, often obscuring both the full dynamics of urban challenges, and the path to their resolution. Researchers have an opportunity to help find strategies to clarify the relevant information—the ‘signal’—amongst an increasing deluge of information noise.

To be sure, many specific benefits of urban big data, and its related smart cities movement, have been described enthusiastically and in detail by proponents [107,108]. Among the benefits touted, information from sensors can be used to ease traffic jams, optimize energy demand, and give users real-time updates on transit and other systems. New mapping technologies can identify, and help to correct, urban problems that were previously much more difficult to identify and manage. Social media and crowdsourcing can target problems large and small, down to which potholes to fix. In this and other respects, it is claimed, the new technologies can even make governments more responsive to citizen needs [109].

There is, of course, a darker view of big data and smart cities. Both are criticized as fads, as technology marketing schemes, or as tools of covert manipulation and control, threatening to abridge political and personal rights (notably including the right to privacy). At best, they over-simplify the complex challenges of the city, with dubious results. For example, Hollands [110] asserts that with smart cities, researchers and decision-makers know too little about “what the label ideologically reveals as well as hides”. Söderström, Paasche, and Klauser [111] attack smart cities as “corporate storytelling” that can “obfuscate more urgent needs” such as “technology-poor affordable housing or sewage systems (that) are arguably more urgent in many of the world’s cities”. Kitchin [112] argues that “data-informed urbanism” is being replaced by “data-driven urbanism”, and its efforts too often “fail to recognize that cities are complex, multifaceted, contingent, relational systems, full of contestation and wicked problems that are not easily captured or steered, and that urban issues are often best solved through political/social solutions and citizen-centred deliberative democracy, rather than technocratic forms of governance”.

However, the vast streams of data are already here, and so are the challenges. The more pointed question is, therefore, how will planners, decision-makers and other stakeholders manage this already existing (and growing) technology in order to make progress? In what ways might it be engaged as a helpful resource to deal with the most pressing human challenges, also addressed by SDG and the NUA? To avoid “data-driven urbanism” planners and researchers can start by asking, what is it that they wish the data in a “data-informed urbanism” to reveal to them? They must be clear in the first instance what they are counting, and why.

The focus of this paper, the goal of livability, points in an important and helpful direction. If livability is defined as the ability to endure, survive, inhabit, and experience quality of life in an

environment, from that we can develop objective criteria to measure (for example) access to resources, economic opportunities, income, health, recreation, and freedom from dangers, threats, obstacles, pollution, and so on. It is also possible to develop subjective and intersubjective measurements of satisfaction and well-being, which help to indicate problem sources. From there, researchers and planners can identify both existing and new datasets that give us a picture of the current situation, and the specific ways that we might target improvements, most likely through an iterative process of action, measurement, and refinement.

Here is where another crucial point arises: Is it only the experts (researchers and planners), working as external agents who can contribute to the achievement of goals also defined by them? A number of researchers have demonstrated that a city is a partly self-organizing system, emerging from myriad incremental actions by distributed agents, i.e., the citizens [113,114]. In part, those actions create adaptations aimed at achieving a more livable environment with a higher quality of life. The data must therefore inform not only the experts working externally upon the city, but the actors themselves working within (and as part of) the city.

4.2. Distinguishing Information about the City, and Information within the City

To consider further opportunities and challenges, it might be important to draw a basic distinction between the information about a city, and the information within a city. The former is the customary focus of technologists and officials, who make measurements and draw conclusions for their own externally applied actions. The latter includes citizens, together with the structures of their environments. This addition is key: the interaction between information with the environment itself, and the people within it, generates new information and new outcomes, often quite apart from any external data process. In biology, this self-organizing phenomenon is known as stigmergy [115]. Applied to cities, it creates the capacity for urban self-organization [116]. This concept is also explored in actor network theory (ANT), developed in sociology [117]. Edelenbos et al. [118] argue that ANT points to an alternative working framework to the ‘data determinism’ that tends to dominate smart city discourse, and “accords specifically well with an understanding of cities as complex, adaptive, self-organizing systems”.

A question then arises concerning the relation between the information about a city, and the information within a city, including stigmergic information. These two forms of information do not exist in isolation, but rather, they occur together and in characteristic interactions. For example, planners and policymakers create infrastructure and regulatory frameworks, which then go on to produce stigmergic patterns by users, which can then be studied and acted upon externally, and so on. As a particular example, planners might create street networks with traffic signals, which then produce the stigmergic patterns of traffic as it self-organizes, which is then modified according to feedback from smart city sensors, and so on. A more distributed example is the stigmergic response of users to smartphone traffic congestion data.

All of these different forms of information and stages of process require clarity of understanding, lest the tasks of geodata application become commingled and confused. We, therefore, close by enumerating several critical ones.

First is the distinction between centralized information and distributed information—and especially, distributed information that is interacted with, acted upon, often created by, distributed agents who are acting at a range of scales. Particularly important are the distinct and often overlapping scales of democratic governance, in what the political scientist Elinor Ostrom referred to as “polycentric” governance [119]. Those who engage geodata in the pursuit of urban goals (including livability) must surely honor the need of these democratic institutions and individuals to be empowered by this distributed information, and not merely to be controlled by it.

Second is the distinction between three different kinds of information, which operate in distinct modes:

- Descriptive information about the city, used to guide actions usually by centralized agents. Examples of descriptive information include maps, measurement datasets, and user survey data.

- Prescriptive information in the form of rules that prescribe actions. They generally produce static configurations, e.g., “all cars stop at the line when a light is red”, and so on. Examples include zoning codes, traffic laws, and other regulations.
- Generative information. This is information within the city, operating iteratively between distributed agents and/or their environments (actor-networks). This kind of information is capable of generating emergent structures through self-organization, and more particularly, through the dynamic of stigmergy [116]. Examples include built-up environmental patterns (like the modest pathway changes in Figure 4), pop-up structures, neighborhood-scale cooperative projects, and other ‘bottom-up’ iterative changes by residents.



Figure 4. A modest example of stigmergy, where pedestrians build up pathway patterns over time that depart from what planners anticipate. Some planners revise their permanent structures based on these information patterns. Image: George Redgrave via Flickr.

It is important to note that descriptive, prescriptive, and generative information can all exist independently or in combination, and in either centralized or distributed forms—or indeed in combinations. There are of course cases in which distributed information would operate in tandem with centralized information, combining descriptive, prescriptive and/or generative kinds—say, in a Smartphone app. The key conclusion is that a geodata strategy that overlooks distributive and generative aspects is not necessarily wrong, but it is incomplete—perhaps dangerously so.

To be sure, the new era of geodata applied to urban livability offers exciting prospects. There is understandable excitement at the usefulness (and yes, the potential economic value) of vast new streams of information and new kinds of analysis about the city, with undeniably great capability and promise. However, as this discussion suggests, no less important is to consider how information moves within the city, within its distributed actors, and within its structures. The new technologies must surely work with both kinds of information, empowering residents and actors at all scales to identify their own signals to pursue within the noise, providing appropriate tools and strategies to do so—some new, and some no doubt yet to be developed.

5. Conclusions

Big data, coupled with machine learning algorithms and geospatial analysis bring without a doubt new insights into urban planning and livability improvement through the acquisition and extraction of vast amount of information about different urban systems. There are numerous disciplines contributing to this effort, therefore understanding the complexity and interrelationship of these urban systems clearly requires an interdisciplinary approach. In this paper, we highlighted some of the potential synergies of applying geospatial data sources and methodologies for urban issues based on reviewing the literature and discussing relevant issues and open questions, especially regarding the application of VGI and various geospatial analytics techniques for livability assessment and improvement. We want to acknowledge that this interdisciplinary take can be expected to strengthen

the collaboration between urban planners, urban scientists, and researchers from the GIS and other domain rather than substituting traditional urban theories and urban analysis.

Through our research questions, we addressed some key aspects of urban analysis and livability assessment in the era of big data, such as the reliability of these new data sources, the potential and need of applying machine learning algorithms for big data related to urban problems, and the extraction of truly meaningful information from the vast amount of data in the case of cities and planning. The issue of data ethics and geoprivacy are gaining more and more relevance due to the advanced usage, especially when different data sources are linked to analyze users [120]. Future research should address these aspects as well more in depth; however, it was beyond the scope of the current article.

This paper although investigated just a subset of potential approaches or limitations—the field of studying cities might be as (or probably even more) complex than cities themselves. Our goal was to demonstrate the strength of these new data types and methodologies in livability assessment and improvement as key elements in urban planning.

Also, we want to emphasize the importance of collaborative work, and diversity integration. One example can be the relationship between academics and stakeholders such as policymakers, governmental bodies, non-profit organizations, urban planning businesses. Framing a research problem and a possible solution can make the difference between being listened or not by a large crowd or a group of decision-making experts. Overall, we cannot emphasize enough the relevance of livability, as an overarching concept in the case of urban analysis. Cities are for and about people, the ultimate goal of any urban analysis or planning action is (or should be) to enhance the quality of life in cities. This requires the consideration of qualitative aspects (such as the perception or satisfaction of the dwellers) as well, beyond the merely data-driven and technology-focused analysis. However, the fact should not be neglected, that this qualitative approach has inherent limitations, which can only be partially conquered by having big data or machine learning methods. Further research should consider and address these qualitative aspects by deepening the cooperation in making cities more livable between traditional urban research domains and the newer, more technology-based fields such as urban data science and GIS.

Author Contributions: Conceptualization, Anna Kovacs-Györi, Alina Ristea, Clemens Havas, and Michael Mehaffy; Writing—original draft preparation, Anna Kovacs-Györi, Alina Ristea, Clemens Havas, Michael Mehaffy, Hartwig H. Hochmair, Bernd Resch, Levente Juhasz, Arthur Lehner, Laxmi Ramasubramanian, and Thomas Blaschke; Writing—review and editing, Anna Kovacs-Györi, Hartwig H. Hochmair, and Levente Juhasz. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience at the University of Salzburg (DK W 1237-N23). We would also like to express our gratitude to Land Salzburg for supporting the research at IDA Lab (Projekt-ID: 20102-F1901166-KZP and 20204-WISS/225/197-2019). The APC was funded by FWF.

Acknowledgments: Open Access Funding by the Austrian Science Fund (FWF).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. United Nations General Assembly. Transforming our World: The 2030 Agenda for Sustainable Development. 2015. Available online: <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication> (accessed on 1 October 2020).
2. United Nations General Assembly. *Habitat III New Urban Agenda: Quito Declaration on Sustainable Cities and Human Settlements for All*; United Nations General Assembly: Quito, Ecuador, 2016.
3. Caprotti, F.; Cowley, R.; Datta, A.; Broto, V.C.; Gao, E.; Georgeson, L.; Herrick, C.; Odendaal, N.; Joss, S. The New Urban Agenda: Key opportunities and challenges for policy and practice. *Urban Res. Pract.* **2017**, *10*, 367–378. [CrossRef]

4. Kabisch, S.; Finnveden, G.; Kratochvil, P.; Sendi, R.; Smagacz-Poziemska, M.; Matos, R.S.; Bylund, J. New Urban Transitions towards Sustainability: Addressing SDG challenges (Research and Implementation Tasks and Topics from the Perspective of the Scientific Advisory Board (SAB) of the Joint Programming Initiative (JPI) Urban Europe). *Sustainability* **2019**, *11*, 2242. [\[CrossRef\]](#)
5. Mehaffy, M.W.; Haas, T.; Elmlund, P. Public Space in the New Urban Agenda: Research into Implementation. *Urban Plan.* **2019**, *4*, 134–137. [\[CrossRef\]](#)
6. Elmlund, P.; Haas, T.; Mehaffy, M.W. Public Space in the New Urban Agenda. The Challenge of Implementation. *J. Public Space* **2018**, *3*, 165–170. [\[CrossRef\]](#)
7. Foth, M.; Choi, J.H.; Satchell, C. Urban informatics. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work—CSCW '11, Hangzhou, China, 19–23 March 2011.
8. Kang, W.; Oshan, T.; Wolf, L.J.; Boeing, G.; Frias-Martinez, V.; Gao, S.; Poorthuis, A.; Xu, W. A roundtable discussion: Defining urban data science. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1756–1768. [\[CrossRef\]](#)
9. Law, T.; Legewie, J. Urban Data Science. In *Emerging Trends in the Social and Behavioral Sciences*; Wiley: Hoboken, NJ, USA, 2018; pp. 1–12.
10. Batty, M. Big Data and the City. *Built Environ.* **2016**, *42*, 321–337. [\[CrossRef\]](#)
11. Singleton, A.D.; Spielman, S.; Folch, D. *Urban Analytics*; Sage: Thousand Oaks, CA, USA, 2017; ISBN 9781473958630.
12. Resch, B.; Szell, M. Human-Centric Data Science for Urban Studies. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 584. [\[CrossRef\]](#)
13. Balduini, M.; Brambilla, M.; Della Valle, E.; Marazzi, C.; Arabghalizi, T.; Rahdari, B.; Vescovi, M. Models and Practices in Urban Data Science at Scale. *Big Data Res.* **2019**, *17*, 66–84. [\[CrossRef\]](#)
14. Nosratabadi, S.; Mosavi, A.; Keivani, R.; Ardabili, S.F.; Aram, F. State of the Art Survey of Deep Learning and Machine Learning Models for Smart Cities and Urban Sustainability. In *Green Technology for Smart City and Society*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2020; pp. 228–238.
15. Mumford, L. What's a City? *Archit. Rec.* **1937**, *37*, 235–264.
16. Park, R.E.; Burgess, E.W. *The City: Suggestions for Investigation of Human Behavior in the Urban Environment*; Chicago Press: Chicago, IL, USA, 1925.
17. Harvey, D. *Justice, Nature and the Geography of Difference*; Wiley: Hoboken, NJ, USA, 1997; ISBN 9781557866813.
18. Sassen, S. *The Global City: New York, London, Tokyo*; Book Collections on Project MUSE; Princeton University Press: Princeton, NJ, USA, 2001; ISBN 9780691070636.
19. Howard, E. *Garden Cities of Tomorrow*; Swan Sonnenschein & Co.: London, UK, 1902; ISBN 0571061893 9780571061891.
20. le Corbusier; Etchells, F. *The City of To-Morrow and Its Planning*; Rodker, J., Ed.; Courier Corporation: Chelmsford, MA, USA, 1929.
21. Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013; ISBN 9780262019521.22.
22. Batty, M. Big data, smart cities and city planning. *Dialog Hum. Geogr.* **2013**, *3*, 274–279. [\[CrossRef\]](#)
23. Jokar Arsanjani, J.; Helbich, M.; Bakillah, M.; Hagenauer, J.; Zipf, A. Toward mapping land-use patterns from volunteered geographic information. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2264–2278. [\[CrossRef\]](#)
24. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 571–590. [\[CrossRef\]](#)
25. Resch, B.; Summa, A.; Zeile, P.; Strube, M. Citizen-Centric Urban Planning through Extracting Emotion Information from Twitter in an Interdisciplinary Space-Time-Linguistics Algorithm. *Urban Plan.* **2016**, *1*, 114–127. [\[CrossRef\]](#)
26. Crooks, A.; Pfoser, D.; Jenkins, A.; Croitoru, A.; Stefanidis, A.; Smith, D.; Karagiorgou, S.; Efentakis, A.; Lamprianidis, G. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [\[CrossRef\]](#)
27. Netzband, M.; Stefanov, W.L.; Redman, C. (Eds.) *Applied Remote Sensing for Urban Planning, Governance and Sustainability*; Springer: Berlin/Heidelberg, Germany, 2007.
28. Thakuria, P.; Tilahun, N.; Zellner, M. *Introduction to Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; pp. 1–9.
29. Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-In Data. *PLoS ONE* **2014**, *9*, e97010. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Sagl, G.; Resch, B.; Hawelka, B.; Beinath, E. *From Social Sensor Data to Collective Human Behaviour Patterns—Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments*; Jekel, T., Car, A., Strobl, J., Griesebner, G., Eds.; GI_Forum 2012 Geovisualisation; Herbert Wichmann Verlag, VDE VERLAG GMBH: Berlin/Offenbach, Germany, 2012; pp. 54–63. ISBN 978-3-87907-521-8.
31. Jacobs-Crisioni, C.; Rietveld, P.; Koomen, E.; Tranos, E. Evaluating the Impact of Land-Use Density and Mix on Spatiotemporal Urban Activity Patterns: An Exploratory Study Using Mobile Phone Data. *Environ. Plan. A Econ. Space* **2014**, *46*, 2769–2785. [[CrossRef](#)]
32. Calabrese, F.; Colonna, M.; Lovisolo, P.; Parata, D.; Ratti, C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Trans. Intell. Transp. Syst.* **2010**, *12*, 141–151. [[CrossRef](#)]
33. Lynch, K. *The Image of the City*; The MIT Press: Cambridge, MA, USA, 1960; ISBN 978-0-262-62001-7.
34. Boeing, G. Spatial information and the legibility of urban form: Big data in urban morphology. *Int. J. Inf. Manag.* **2019**, 102013. [[CrossRef](#)]
35. Giap, T.K.; Thye, W.W.; Aw, G. A new approach to measuring the liveability of cities: The Global Liveable Cities Index. *World Rev. Sci. Technol. Sustain. Dev.* **2014**, *11*, 176. [[CrossRef](#)]
36. Tan, K.G.; Woo, W.T.; Tan, K.Y.; Low, L.; Aw, G.E.L. *Ranking the Liveability of the World's Major Cities: The Global Liveable Cities Index (GLCI)*; World Scientific: Singapore, 2012; ISBN 981-4417-32-7.
37. Newton, P. Liveable and Sustainable? Socio-Technical Challenges for Twenty-First-Century Cities. *J. Urban Technol.* **2012**, *19*, 81–102. [[CrossRef](#)]
38. Onnom, W.; Tripathi, N.K.; Nitivattananon, V.; Ninsawat, S. Development of a Liveable City Index (LCI) Using Multi Criteria Geospatial Modelling for Medium Class Cities in Developing Countries. *Sustainability* **2018**, *10*, 520. [[CrossRef](#)]
39. Ley, A.; Newton, P.; Kallidaikurichi, S.; Yuen, B. Creating and Sustaining Liveable Cities. In *Developing Living Cities*; World Scientific Publisher Co Pte Lt.: Singapore, 2010; pp. 191–229.
40. Kovacs-Gyori, A.; Cabrera-Barona, P.; Resch, B.; Mehaffy, M.W.; Blaschke, T. Assessing and Representing Livability through the Analysis of Residential Preference. *Sustainability* **2019**, *11*, 4934. [[CrossRef](#)]
41. Pigliautile, I.; D'Eramo, S.; Pisello, A.L. Intra-urban microclimate mapping for citizens' wellbeing: Novel wearable sensing techniques and automatized data-processing. *J. Clean. Prod.* **2021**, *279*, 123748. [[CrossRef](#)]
42. Shorten, A.; Smith, J. Mixed methods research: Expanding the evidence base. *Evid. Based Nurs.* **2017**, *20*, 74–75. [[CrossRef](#)] [[PubMed](#)]
43. Gaber, J.; Gaber, S.L. Utilizing Mixed-Method Research Designs in Planning: The Case of 14th Street, New York City. *J. Plan. Educ. Res.* **1997**, *17*, 95–103. [[CrossRef](#)]
44. Kovacs-Gyori, A.; Ristea, A.; Havas, C.; Resch, B.; Cabrera-Barona, P. #London2012: Towards Citizen-Contributed Urban Planning through Sentiment Analysis of Twitter Data. *Urban Plan.* **2018**, *3*, 75–99. [[CrossRef](#)]
45. Condeço-Melhorado, A.; Mohino, I.; Moya-Gómez, B.; García-Palomares, J.C. The Rio Olympic Games: A Look into City Dynamics through the Lens of Twitter Data. *Sustainability* **2020**, *12*, 7003. [[CrossRef](#)]
46. Zhang, Z.; Ni, M.; He, Q.; Gao, J. Mining Transportation Information from Social Media for Planned and Unplanned Events, Buffalo, NY, USA, 2016. Available online: <https://rosap.nhtl.bts.gov/view/dot/30838> (accessed on 2 October 2020).
47. Lee, R.; Sumiya, K. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks—LBSN '10, New York, NY, USA, 2 November 2010*; Association for Computing Machinery (ACM): New York, NY, USA, 2010; pp. 1–10.
48. Li, R.; Lei, K.H.; Khadiwala, R.; Chang, K.C.-C. TEDAS: A Twitter-based Event Detection and Analysis System. In *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, 1–5 April 2012*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2012; pp. 1273–1276.
49. Weng, J.; Yao, Y.; Leonardi, E.; Lee, F. Event Detection in Twitter. *Development* **2011**, *2*, 179–182. [[CrossRef](#)]
50. Kovacs-Gyori, A.; Ristea, A.; Kolcsar, R.; Resch, B.; Crivellari, A.; Blaschke, T. Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 378. [[CrossRef](#)]
51. Roberts, H.V. Using Twitter data in urban green space research: A case study and critical evaluation. *Appl. Geogr.* **2017**, *81*, 13–20. [[CrossRef](#)]

52. Roberts, H.; Sadler, J.P.; Chapman, L. Using Twitter to investigate seasonal variation in physical activity in urban green space. *Geo. Geogr. Environ.* **2017**, *4*, e00041. [[CrossRef](#)]
53. Roberts, H.; Sadler, J.; Chapman, L. The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation. *Urban Stud.* **2019**, *56*, 818–835. [[CrossRef](#)]
54. Krishnamurthy, R.; Smith, K.L.; DeSouza, K.C. Urban Informatics: Critical Data and Technology Considerations. In *Springer Geography*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2017; pp. 163–188.
55. Allam, Z.; Dey, G.; Jone, D.S. Artificial Intelligence (AI) Provided Early Detection of the Coronavirus (COVID-19) in China and Will Influence Future Urban Health Policy Internationally. *AI* **2020**, *1*, 156–165. [[CrossRef](#)]
56. Darabi, H.; Choubin, B.; Rahmati, O.; Haghighi, A.T.; Pradhan, B.; Kløve, B. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *J. Hydrol.* **2019**, *569*, 142–154. [[CrossRef](#)]
57. Martí, P.; García-Mayor, C.; Serrano-Estrada, L. Taking the urban tourist activity pulse through digital footprints. *Curr. Issues Tour.* **2020**, 1–20. [[CrossRef](#)]
58. Han, S.Y.; Tsou, M.-H.; Knaap, E.; Rey, S.; Cao, G. How Do Cities Flow in an Emergency? Tracing Human Mobility Patterns during a Natural Disaster with Big Data and Geospatial Data Science. *Urban Sci.* **2019**, *3*, 51. [[CrossRef](#)]
59. Lu, Y.; Zhang, L. Social media WeChat infers the development trend of COVID-19. *J. Infect.* **2020**, *81*, e82–e83. [[CrossRef](#)]
60. Hochmair, H.H.; Juhász, L.; Cvetojevic, S. Data Quality of Points of Interest in Selected Mapping and Social Media Platforms. *Lect. Notes Geoinforma. Cartography* **2017**, 293–313. [[CrossRef](#)]
61. Vargas-Muñoz, J.E.; Tuia, D.; Falcão, A.X. Deploying machine learning to assist digital humanitarians: Making image annotation in OpenStreetMap more efficient. *Int. J. Geogr. Inf. Sci.* **2020**, 1–21. [[CrossRef](#)]
62. Yang, K.; Varol, O.; Davis, C.A.; Ferrara, E.; Flammini, A.; Menczer, F. Arming the public with artificial intelligence to counter social bots. *Hum. Behav. Emerg. Technol.* **2019**, *1*, 48–61. [[CrossRef](#)]
63. ISO 19115-1:2014 Geographic Information—Metadata—Part 1: Fundamentals. Available online: <https://www.iso.org/standard/53798.html?browse=tc> (accessed on 28 November 2020).
64. Haklay, M. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703. [[CrossRef](#)]
65. Zielstra, D.; Hochmair, H.H. Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transp. Res. Rec. J. Transp. Res. Board* **2011**, *2217*, 145–152. [[CrossRef](#)]
66. Antoniou, V.; Skopeliti, A. Measures and indicators of VGI quality: An overview. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *1*, 345–351. [[CrossRef](#)]
67. Degrossi, L.C.; de Albuquerque, J.P.; Rocha, R.D.S.; Zipf, A. A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Trans. GIS* **2018**, *22*, 542–560. [[CrossRef](#)]
68. Barron, C.; Neis, P.; Zipf, A. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Trans. GIS* **2014**, *18*, 877–895. [[CrossRef](#)]
69. Bordogna, G.; Carrara, P.; Criscuolo, L.; Pepe, M.; Rampini, A. A linguistic decision-making approach to assess the quality of volunteer geographic information for citizen science. *Inf. Sci.* **2014**, *258*, 312–327. [[CrossRef](#)]
70. Ballatore, A.; de Sabbata, S. Charting the Geographies of Crowdsourced Information in Greater London. *Lect. Notes Geoinform. Cartogr.* **2018**, 149–168. [[CrossRef](#)]
71. Alivand, M.; Hochmair, H.H. Spatiotemporal analysis of photo contribution patterns to Panoramio and Flickr. *Cartogr. Geogr. Inf. Sci.* **2016**, *44*, 170–184. [[CrossRef](#)]
72. Juhász, L.; Hochmair, H.H. User Contribution Patterns and Completeness Evaluation of Mapillary, a Crowdsourced Street Level Photo Service. *Trans. GIS* **2016**, *20*, 925–947. [[CrossRef](#)]
73. Zielstra, D.; Hochmair, H.H. Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *J. Spat. Sci.* **2013**, *58*, 251–273. [[CrossRef](#)]
74. Foody, G.M.; See, L.; Fritz, S.; van der Velde, M.; Perger, C.; Schill, C.; Boyd, D.S. Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Trans. GIS* **2013**, *17*, 847–860. [[CrossRef](#)]

75. Flanagan, A.J.; Metzger, M.J. The credibility of volunteered geographic information. *GeoJournal* **2008**, *72*, 137–148. [[CrossRef](#)]
76. Hung, K.-C.; Kalantari, M.; Rajabifard, A. Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia. *Appl. Geogr.* **2016**, *68*, 37–47. [[CrossRef](#)]
77. Hecht, B.J.; Gergle, D. On the “localness” of user-generated content. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work—CSCW '10, Savannah, GA, USA, 6–10 February 2010*; Association for Computing Machinery (ACM): New York, NY, USA, 2010; pp. 229–232.
78. Johnson, I.L.; Sengupta, S.; Schöning, J.; Hecht, B. The Geography and Importance of Localness in Geotagged Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016*; Association for Computing Machinery (ACM): New York, NY, USA, 2016; pp. 515–526.
79. de Longueville, B.; Ostländer, N.; Keskitalo, C. Addressing vagueness in Volunteered Geographic Information (VGI)-A case study. *Int. J. Spat. Data Infrastruct. Res.* **2010**, *5*, 463–1725.
80. Rice, M.T.; Paez, F.I.; Mulhollen, A.P.; Shore, B.M.; Caldwell, D.R. *Crowdsourced Geospatial Data: A Report on the Emerging Phenomena of Crowdsourced and User-Generated Geospatial Data*; Defense Technical Information Center (DTIC): Fort Belvoir, VA, USA, 2012.
81. Graham, M.; Zook, M. Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web. *Environ. Plan. A Econ. Space* **2013**, *45*, 77–99. [[CrossRef](#)]
82. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [[CrossRef](#)]
83. Senaratne, H.; Bröring, A.; Schreck, T. Using Reverse Viewshed Analysis to Assess the Location Correctness of Visually Generated VGI. *Trans. GIS* **2013**, *17*, 369–386. [[CrossRef](#)]
84. Zhao, B.; Sui, D.Z. True lies in geospatial big data: Detecting location spoofing in social media. *Ann. GIS* **2017**, *23*, 1–14. [[CrossRef](#)]
85. Neis, P.; Goetz, M.; Zipf, A. Towards Automatic Vandalism Detection in OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2012**, *1*, 315–332. [[CrossRef](#)]
86. Juhász, L.; Hochmair, H.H. Cross-Linkage Between Mapillary Street Level Photos and OSM Edits. *Lect. Notes Geoinf. Cartogr.* **2016**, *64*, 141–156. [[CrossRef](#)]
87. Juhász, L. Geo-Social Activity Research. Available online: <https://research.jlvente.com/tool/> (accessed on 28 November 2020), (unpublished work).
88. Miller, H.J. A Measurement Theory for Time Geography. *Geogr. Anal.* **2005**, *37*, 17–45. [[CrossRef](#)]
89. Huang, P.-C.; Lee, S.-S.; Kuo, Y.-H.; Lee, K.-R. A flexible sequence alignment approach on pattern mining and matching for human activity recognition. *Expert Syst. Appl.* **2010**, *37*, 298–306. [[CrossRef](#)]
90. Keßler, C.; McKenzie, G. A geoprivacy manifesto. *Trans. GIS* **2017**, *22*, 3–19. [[CrossRef](#)]
91. McKenzie, G.; Keßler, C.; Andris, C. Geospatial Privacy and Security. *J. Spat. Inf. Sci.* **2019**. [[CrossRef](#)]
92. McKenzie, G.; Janowicz, K.; Adams, B. A weighted multi-attribute method for matching user-generated Points of Interest. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 125–137. [[CrossRef](#)]
93. Juhász, L.; Hochmair, H.H. Cross-checking user activities in multiple geo-social media networks. In *Proceedings of the 21st AGILE Conference on Geo-Information Science, Lund, Sweden, 12–15 June 2018*.
94. Juhász, L.; Hochmair, H.H. How do volunteer mappers use crowdsourced Mapillary street level images to enrich OpenStreetMap? In *Proceedings of the 20th AGILE Conference on Geo-Information Science, Wageningen, The Netherlands, 9–12 May 2017*.
95. Haklay, M.; Basiouka, S.; Antoniou, V.; Ather, A. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus’ Law to Volunteered Geographic Information. *Cartogr. J.* **2010**, *47*, 315–322. [[CrossRef](#)]
96. Pei, S.; Muchnik, L.; Andrade, J.S., Jr.; Zheng, Z.; Makse, H.A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **2015**, *4*, srep05547. [[CrossRef](#)]
97. Leung, Y.; Ma, J.-H.; Goodchild, M.F. A general framework for error analysis in measurement-based GIS Part 1: The basic measurement-error model and related concepts. *J. Geogr. Syst.* **2004**, *6*, 325–354. [[CrossRef](#)]
98. Ahmouda, A.; Hochmair, H.H. Using Volunteered Geographic Information to measure name changes of artificial geographical features as a result of political changes: A Libya case study. *GeoJournal* **2017**, *83*, 237–255. [[CrossRef](#)]
99. Griffin, G.P.; Jiao, J. Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *J. Transp. Health* **2015**, *2*, 238–247. [[CrossRef](#)]

100. Smith, A.; Andersen, M. Social Media Use in 2018. *Pew Res. Cent.* **2018**. Available online: https://www.pewinternet.org/wp-content/uploads/sites/9/2018/02/PI_2018.03.01_Social-Media_FINAL.pdf (accessed on 28 November 2020).
101. Miller, H.J.; Goodchild, M.F. Data-driven geography. *GeoJournal* **2014**, *80*, 449–461. [[CrossRef](#)]
102. Yin, L.; Wang, Z. Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Appl. Geogr.* **2016**, *76*, 147–153. [[CrossRef](#)]
103. Gosal, A.; Geijzendorffer, I.R.; Václavík, T.; Poulin, B.; Václavík, T. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosyst. Serv.* **2019**, *38*, 100958. [[CrossRef](#)]
104. Steiger, E.; Resch, B.; de Albuquerque, J.P.; Zipf, A. Mining and correlating traffic events from human sensor observations with official transport data using self-organizing maps. *Transp. Res. Part C Emerg. Technol.* **2016**, *73*, 91–104. [[CrossRef](#)]
105. Steiger, E.; Resch, B.; Zipf, A. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1694–1716. [[CrossRef](#)]
106. Liu, Y.; Yuan, Y.; Zhang, F. Mining urban perceptions from social media data. *J. Spat. Inf. Sci.* **2020**. [[CrossRef](#)]
107. Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference on Digital Government Innovation in Challenging Times—dg.o '11*, College Park, MD, USA, 12–15 June 2011; Association for Computing Machinery (ACM): New York, NY, USA, 2011; pp. 282–291.
108. Albino, V.; Berardi, U.; Dangelico, R.M. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *J. Urban Technol.* **2015**, *22*, 3–21. [[CrossRef](#)]
109. Pereira, G.V.; Parycek, P.; Falco, E.; Kleinhans, R. Smart governance in the context of smart cities: A literature review. *Inf. Polity* **2018**, *23*, 143–162. [[CrossRef](#)]
110. Hollands, R.G. Will the real smart city please stand up? *City* **2008**, *12*, 303–320. [[CrossRef](#)]
111. Söderström, O.; Paasche, T.; Klauser, F. Smart cities as corporate storytelling. *City* **2014**, *18*, 307–320. [[CrossRef](#)]
112. Kitchin, R. The ethics of smart cities and urban science. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20160115. [[CrossRef](#)] [[PubMed](#)]
113. Portugali, J.; Haken, H.; Benenson, I.; Omer, I.; Alfasi, N. *Self-Organization and the City*; Springer Series in Synergetics; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2012; ISBN 9783662040997.
114. Bettencourt, L.M.A. The Kind of Problem a City Is: New Perspectives on the Nature of Cities from Complex Systems Theory. *Decod. City* **2014**. [[CrossRef](#)]
115. Theraulaz, G.; Bonabeau, E. A Brief History of Stigmergy. *Artif. Life* **1999**, *5*, 97–116. [[CrossRef](#)] [[PubMed](#)]
116. Mehaffy, M.; Elmlund, P. Smart cities: Missing the stigmergy? In *Spatial Knowledge as a Tool for Strategic and Data-Based Regional Policy*; Regional Science Academy and Université de Lyon: Lyon, France, 2020.
117. Latour, B. On actor-network theory: A few clarifications. *Soz. Welt* **1996**, *4*, 369–381.
118. Edelenbos, J.; Hirzalla, F.; van Zoonen, L.; van Dalen, J.; Bouma, G.; Slob, A.; Woestenburger, A. *Governing the Complexity of Smart Data Cities: Setting a Research Agenda*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2017; pp. 35–54.
119. Ostrom, E. Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *Am. Econ. Rev.* **2010**, *100*, 641–672. [[CrossRef](#)]
120. Kounadi, O.; Resch, B. A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *J. Empir. Res. Hum. Res. Ethic* **2018**, *13*, 203–222. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).