Received: 19 September 2020 | Revised: 4 October 2021 | Accepted: 2 November 2021
DOI: 10.1002/bse.2942

SPECIAL ISSUE ARTICLE

Business Strategy and the Environment WILEY

Citizens and cities: Leveraging citizen science and big data for sustainable urban development

Francesco Cappa[1,2] | Stefano Franco[2] | Federica Rosso[3,4]

[1]Department of Engineering, Campus Bio-Medico University, Rome, Italy
[2]Luiss Business School, Luiss University, Rome, Italy
[3]Department of Civil, Construction and Environmental Engineering, Sapienza University of Rome, Rome, Italy
[4]Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

Correspondence
Francesco Cappa, Department of Engineering, Campus Bio-Medico University, 00128 Rome, Italy.
Email: francesco.cappa@unicampus.it; fcappa@luiss.it

Abstract

Citizen science (CS), that is, the involvement of citizens in data collection or analysis for research projects, is becoming more widespread. This is due to the increasing digitalization of the general public and due to the increasing number of grand challenges that society is facing. Thanks to the contributions of common citizens in data collection and data analysis conducted through technology-mediated interactions, CS can produce a number of benefits for researchers, public organizations, policymakers, citizens, and society as a whole. Given the high density of socio-economic activities in cities, CS can be implemented in a particularly effective way in urban environments to help tackle many "grand challenges", namely, the pressing environmental and social issues that societies are facing at present. However, CS still has untapped potential to be explored. Indeed, we contend that even though CS involves citizens for precisely defined scientific objectives, the interaction that occurs can also be leveraged to collect data beyond the original aim, thereby producing big data (BD). Through a multiple case studies analysis, we highlight how CS can be used to collect BD as well, which can be a valuable resource for researchers, public organizations, and policymakers. With this aim in mind, this study proposes the definition of a *citizen-sourcing* framework that jointly employs CS and BD, and it highlights which processes can be implemented to favor the sustainable development of urban environments. Moreover, we also discuss the looming dangers associated with *citizen-sourcing* as a result of technology-mediated interactions and the use of digital technologies, and we highlight possible future developments.

KEYWORDS
big data, citizen science, citizen-sourcing, crowd, digitalization, grand challenge, sustainability, urban areas

1 | INTRODUCTION

Citizen science (CS), that is, the involvement of citizens in data collection or analysis for research projects, is a phenomenon that dates back to the early 19th century, although in recent years it has spread quickly thanks to advancements in information technology (IT) (Sauermann et al., 2020; Young et al., 2019). CS involves crowds from outside organizational boundaries in the creation of new knowledge (Franzoni & Sauermann, 2014). CS has become popular thanks to the ease with which organizations can involve external individuals, that is, common citizens, through web-based platforms, and thanks to a growing focus on environmental and social sustainability (Cappa et al., 2018). In particular, CS involves common citizens without a specific background requirement in data collection or analysis to advance research projects and improve decision-making processes at the organizational level. Examples are monitoring birds, preventing deforestation, and planning a better alignment of energy production and consumption (Arcanjo et al., 2015; Cappa et al., 2020; Land-Zandstra et al., 2015; Pocock et al., 2017; Wildschut, 2017).

Research bodies and public organizations are looking for methods to gather large amounts of data in order to tackle "grand challenges," that is, the pressing environmental and social issues that we are facing at present (Guida & Carpentieri, 2021; Sauermann et al., 2020) and CS is a promising way to achieve these goals thanks to the involvement of citizen volunteers through technology-mediated interactions (Riesch et al., 2013; Rowbotham et al., 2019). Indeed, CS makes it is possible to collect and analyze data for socially and environmentally oriented projects in a rapid and convenient way. At the same time, CS increases the literacy of citizens, and thus, it fulfills social and educational purposes as well. Through CS, all three pillars of sustainability can benefit. Indeed, thanks to CS, research projects are conducted faster and cheaper (economic pillar), the well-being and literacy of citizens can improve (social pillar), and environmental impacts can be mitigated (environmental pillar) (Cappa et al., 2020; Hansmann et al., 2012). Due to its inclusive nature, as no special previous knowledge is required, CS is particularly useful in urban contexts, where most of the world population lives (United Nations, 2018). These city dwellers are conceivably more eager to contribute to problems affecting their immediate surroundings. Cities are also the places where the majority of social, economic, and environmental issues are concentrated (Acuto et al., 2018; Bai et al., 2018; Hallin et al., 2021).

Given the numerous benefits generated by CS, there is increasing scientific, managerial, and policymaking attention towards this phenomenon. Previous studies have thus far focused on how to involve higher numbers of citizens in order to collect a greater quantity of contributions during CS campaigns, thus increasing the quality of the outcomes, and on highlighting the benefits of CS in terms of scientific results and sustainability achievements (Aristeidou et al., 2017; Cappa et al., 2018; Mueller et al., 2018; Roman et al., 2017; Tinati et al., 2017). However, beyond the data collected or analyzed for other scientific aims, there are additional potential benefits that CS can bring to organizations that have often been overlooked. In this study, we argue that while organizations collect data for CS projects, they could ask for additional information from these citizens, since they have already become engaged, and this could yield a valuable resource to use in the sustainable development of urban areas.

In this manner, CS can function as a source of big data (BD), which is another recent phenomenon. Indeed, CS has expanded in recent years thanks to the increasing digitalization of the general public and the emergence of new digital technologies (Ardito et al., 2019; Bohnsack et al., 2019; Cappa et al., 2021; Elia et al., 2019; Hanelt et al., 2020; Visconti & Morea, 2019). These digital advancements led to the creation of 41 zettabytes of data every day in 2019, that is, billions of terabytes, and it is estimated that society will create 175 zettabyte of data by the end of 2021 (Wang & Wang, 2020). In fact, BD is increasingly being recognized as "the new oil"—which demonstrates the great potential it can have for organizations (Oyer, 2019).

There has been much debate through the years regarding a definition for BD. It is sometimes incorrectly related to simply having high volumes of data. Recently, instead, there has been increasing scholarly consensus that BD can be identified as information characterized not only by high volume but also by high levels of variety, high velocity, high veracity, and high value (the so-called 5 Vs) (Jin et al., 2015). Recent studies have shown how BD can be crucial in nurturing innovations, making better decisions, and delivering better outcomes (Batistič & van der Laken, 2019; Cappa et al., 2021; Huang et al., 2018; Jin et al., 2015; Johnson et al., 2017; Marshall et al., 2015; Müller et al., 2018). BD has also proven beneficial for urban environments, transport management, and planning (Boeing, 2021; Creutzig et al., 2019). In fact, BD may assist sustainable development by exploiting predictive analysis that can solve social and environmental problems (Dubey et al., 2019). However, since collecting and handling BD is not an easy task due to the difficulties of reaching a broad audience willing to provide information (Philip Chen & Zhang, 2014), CS can contribute to solving this challenge by facilitating contact with a broad audience of citizens in a trusted way, thanks to the scientific community leading activities.

Nevertheless, the benefits of CS and of BD for urban areas have mainly been analyzed individually. Previous research has highlighted the benefits that urban environments can gain from properly implementing CS strategies, such as better urban planning, improved environmental monitoring, and an enhanced quality of life (Craglia & Eds, 2014; Mueller et al., 2018; Paulos et al., 2008) while other scholars have discussed the role of BD in such contexts (Ahvenniemi et al., 2017). However, further consideration is needed regarding the possible benefits that may arise from the simultaneous implementation of CS and BD. In particular, the research question we address is the following: How can CS and BD be jointly implemented for the sustainable development of the urban environment?

By performing a qualitative, multiple case studies analysis, we provide evidence that even though CS involves citizens in a precisely defined manner for a specific scientific aim, the online interaction that takes place during CS can also be leveraged to collect additional data that provide BD to be used later. Our analysis focuses on two extreme cases (Pratt, 2009) that illustrate examples of the joint application of CS and BD: "New York Street Tree Map" in New York City, USA ("New York City Street Tree Map,", 2021) and "B.E. (Building Energy) Smart" in Rome, Italy ("B.E. Smart,", 2021). These two projects are very relevant case studies, as further detailed in the Methodology section, primarily because they were set in two globally renowned but heterogeneous urban contexts and thus they contribute external validity to the study (Flick et al., 2004).

Based on the insights gathered from the above-mentioned cases, using both primary and secondary data, we define the boundaries of a new framework that is at the intersection of CS and BD, that is, *citizen-sourcing*, which can give researchers and policymakers the means to further involve citizens and favor sustainable urban development. While organizations have so far been focused on one aspect at a time, that is, CS or BD collection, *citizen-sourcing* provides a framework through which organizations can collect additional personal information from citizens, producing BD for future initiatives, while they contribute to the scientific purpose at hand, by collecting or analyzing data that can sustain CS projects.

In addition, we identify the main processes that characterize this framework, that is, *citizen profiling*, *involvement of technological capabilities*, and *data protection*. We thereby advance scientific knowledge of the CS and BD phenomena by laying the foundations for *citizen-sourcing* and by highlighting the processes through which it is possible to simultaneously implement them. The joint implementation of CS and BD may favor sustainable urban development, that is, leading cities to become more responsive to citizen needs and promoting higher quality of life by nurturing economic development, social well-being, and environmental protection (Angelidou et al., 2018; Guzmán et al., 2017; Simwanda & Murayama, 2018). We also provide empirical evidence for researchers and policymakers on how to collect valuable information from citizen scientists to better manage upcoming CS projects and to implement future target actions. Moreover, we also address some perils of *citizen-sourcing* as a result of technology-mediated interaction and the use of digital technologies, and we highlight possible future developments.

2 | THEORETICAL BACKGROUND

In this section we provide a literature review of the CS and BD phenomena, with particular focus on their application in the urban context, and we outline the reasoning and theoretical foundation for our *citizen-sourcing* framework.

2.1 | CS in the urban context

The past decade has been characterized by a rapid evolution in the extent to which people outside of organization boundaries have been involved in the generation and dissemination of new knowledge. Indeed, the digitalization of the business environment and the emergence of new digital technologies make it easier for organizations to reach dispersed individuals (Acciarini et al., 2021). This increases the popularity of different forms of crowd involvement, like crowdsourcing and crowdfunding; the former involves dispersed individuals in the collection of innovative ideas and in the latter individuals contribute funds towards knowledge creation thanks to new ventures (e.g., Block et al., 2018; Messeni Petruzzelli et al., 2019; Natalicchio et al., 2017). Moreover, the way in which research projects are conducted has also changed thanks to the inclusion of crowds in data collection and analysis, which has led to the consolidation of CS (Arcanjo et al., 2015; Franzoni & Sauermann, 2014; Wildschut, 2017). CS is the term commonly used for technology-mediated collaboration between citizens, or members of the public, and scientific organizations in order to produce or analyze scientific data (Haklay, 2013). In order to be effective, however, a pressing need in CS is to get motivated participants involved, since CS projects often fail due to low participation (Cappa et al., 2016; Druschke & Seltze, 2012). Moreover, there is a close link between the quantity of the contributions collected and the quality of the scientific outcomes, and this also applies to the increased literacy of the population as a whole (Cappa et al., 2018, 2020; Nov et al., 2014). For example, higher participation in the Brooklyn Atlantis CS project resulted in more precise and reliable mapping of water quality (Cappa et al., 2018). Furthermore, thanks to participation in CS projects, citizens increase their technological and environmental awareness (Cappa et al., 2020). Thus, understanding how to motivate citizens to participate has been the focus of the majority of studies on CS so far (Cappa et al., 2018, 2020; Laut et al., 2017).

Currently, there are numerous collaborations between well-known research institutes and common citizens, such as the CS projects carried out by NASA (NASA, 2021). Thanks to CS, it is indeed possible to conduct research that would not otherwise have been possible if relying solely on researchers' resources. The benefits generated by CS projects are numerous. For researchers, having a greater number of people available to collect and analyze data reduces costs and the time to complete scientific projects (Haklay, 2013; Nov et al., 2014). Moreover, citizens may increase their scientific knowledge, while being socially useful and thus intrinsically satisfied, and they also connect with the scientific world that had previously been distant from their problems (Paul et al., 2014). Finally, society as a whole benefits from CS as well, thanks to better and faster scientific outcomes, which are mainly environmental and social in content, and thanks to the increased literacy gained by citizens (Cappa et al., 2020). In particular, CS may contribute to addressing the "grand challenges" that society is currently facing (Foray et al., 2012; Kuhlmann, 2014). Such pressing environmental and social issues have a profound effect on urban areas, where population density is high (Cappa et al., 2020; Centobelli et al., 2020; Pinkse & Dommisse, 2009; Sakshi et al., 2020), and CS projects can help address these issues at lower cost, and faster, thanks to the voluntary involvement of a large number of citizens already locally available in the urban area.

McCaffrey (2005) was among the first researchers to show the benefits that CS can bring to the ecosystems of cities by describing how CS enabled extensive data collection for bird-monitoring projects, providing scientific outcomes that might otherwise have been impossible to achieve. Later, other studies showed that CS could help reduce electricity consumption (Cappa et al., 2020), clean bodies of water (Cappa et al., 2018), find clean water spots in the urban environment (Young et al., 2019), monitor air pollution (Sauermann et al., 2020; Silvertown, 2009; Young et al., 2019), prevent deforestation (Arcanjo et al., 2015), map invasive plant species (Luigi Nimis et al., 2019), and enhance the well-being of citizens (Ferster et al., 2017). In these situations, volunteers can help solve pressing environmental or social problems by either providing data or analyzing data on a scale that would have otherwise been impossible for researchers to achieve. In fact, thanks to technology-mediated interactions and digital technologies, citizens can connect from all over the world and contribute during their spare time. Moreover, previous research has shown how urban administrations can engender effective actions that integrate citizen contributions in a participatory decision-making process (Åström, 2020; Mueller et al., 2018; Paulos et al., 2008) and simultaneously increase the literacy of citizens by raising their awareness of significant issues affecting the socio-economic context where they live. Indeed, CS leads to "raising awareness, building capacity, and strengthening communities" (Craglia & Eds, 2014, p. 48), rather than merely collecting data.

Table 1 summarizes the main findings of the above-mentioned studies, tracing the environmental, social, and economic benefits that CS produces for the sustainable development of urban areas.

In CS, it is crucial to reach the greatest number of individuals to maximize benefits for organizations. This provides a chance to also collect BD, which is another emerging phenomenon that has benefited from the digitalization of the general public and advancements in digital technologies. In the next subsection we explore the

TABLE 1 Previous studies on CS benefits for sustainable urban development

Benefits for urban development	Authors	Main findings	Geographical context	Methodology	Case
Environmental	McCaffrey (2005) and Young et al. (2019)	CS enhances the ecology of cities and bird preservation in urban centers	USA; USA and Canada	Qualitative; quantitative	Bird monitoring; natural heritage protection
	Silvertown (2009) and Sauermann et al. (2020)	CS helps monitor the quality of air, soil, and water	Global; global	Conceptual; conceptual	Ecology; multiple cases
	Arcanjo et al. (2015)	CS improves the effectiveness of monitoring green areas	Brazil	Quantitative	Deforestation
	Cappa et al. (2018)	Social and monetary rewards may help CS in cleaning up bodies of waters	USA	Quantitative	Water monitoring
	Luigi Nimis et al. (2019)	CS helps map invasive species in the city	Italy	Qualitative	Invasive plants monitoring
	Cappa et al. (2020)	CS reduces energy consumptions and related emissions	Italy	Quantitative	Energy management
Social	Paulos et al. (2008)	CS leads to the empowerment of citizens in directing policymakers' decisions	Global	Conceptual	Participatory urbanism
	Silvertown (2009)	CS raises people's awareness of social and environmental issues	Global	Conceptual	Citizen awareness
	Craglia and Eds (2014)	CS helps cities raise awareness of social problems among citizens and strengthen communities	European Union	Qualitative	Smart cities
	Ferster et al. (2017)	CS allows citizens to report insights on bike accidents and cycling behaviors	Canada	Qualitative	Cycling safety
	Mueller et al. (2018)	CS is a strategy for cities to integrate citizen ideas and wishes in the urban planning process	Global	Qualitative	Urban design
	Åström (2020)	Interactions between citizens and administrations lead to the design of participatory urban plans	Sweden	Quantitative	Participatory urban planning
Economic	Haklay (2013), Nov et al. (2014), and Cappa et al. (2020)	CS reduces costs and timelines in the collection of data from a multitude of citizens	Global; USA; Italy	Conceptual; quantitative; quantitative	Geographic information; citizen participation; nudging

Abbreviation: CS, citizen science.

phenomenon of BD and develop our reasoning regarding the joint application of CS and BD, based on their similarities.

2.2 | BD in the urban context

The proliferation of new technologies and the digitalization of the general public have contributed significantly to the generation of enormous quantities of data available in real time on product–customer interactions, market needs, product life cycles, opinions, preferences, and consumer attitudes (Ardito et al., 2018; Bharadwaj & Noble, 2017; Trabucchi et al., 2018). The unprecedented growth in the volume, variety, and velocity with which data are generated and transferred over the past decade has led to the birth of the concept of BD. Hashem et al. (2015, p. 100) define BD as "*a set of techniques and technologies that require new forms of integration to discover great values hidden within different, complex and large data sets*." There is a growing general consensus on identifying BD as data characterized by high values of volume, velocity, variety, veracity and value, constituting the so called 5 Vs (Cappa et al., 2021; Jin et al., 2015).

Recent studies have highlighted the positive impact that the use of BD can have on organizations in different sectors (Cappa et al., 2021; Corte-Real et al., 2017). For example, BD analysis in the healthcare sector can reduce costs and improve the quality of services provided (Tiwari et al., 2018), and in the social sphere the large amount of information contained in BD is useful in various political and economic contexts (Jin et al., 2015). More generally, BD, thanks to the insights collected, can lead to better decisions, increase efficiency by improving the services and support provided, improve customer retention rates, and facilitate product and service innovation (Batistič & van der Laken, 2019; Henke et al., 2016; Marshall et al., 2015; McAfee & Brynjolfsson, 2012). Whereas in CS information is collected through voluntary contributions for a current scientific aim, in BD information is collected both through voluntary and automatized ways, for example, cookies from websites and data from mobile applications, to be later analyzed for any further insights. Examples include the data collected by Amazon and Target through browsing histories, which was revealed to be helpful in improving dynamic pricing strategy and which is also then used to forecast buying trends in the long run (Gandomi & Haider, 2015; Wamba et al., 2017). Other examples have come from Alibaba, which uses data collected to build a credit scoring system and deliver a loan business (Nonninger, 2018), and Ford, which has exploited the data on customer vehicle use to direct product innovation practices (Erevelles et al., 2016).

However, recently, it has also been argued that there are several drawbacks associated with BD. These problems are mainly related to high costs for database management and analysis (CoolaData, 2014), reputational risks in case of data breaches (Brook, 2019), and privacy issues, which need to be considered carefully (Choe et al., 2013; Thurm & Kane, 2010). Nevertheless, it has been contended that, if properly managed, the benefits of BD for organizations can outweigh the costs associated with it (Cappa et al., 2021).

Depending on the way BD is owned and managed, it could be public BD, private BD, or open BD. Public BD refers to data generated by individuals and used for research purposes by public entities, while open BD offers accessibility to everyone interested (George et al., 2014; Pantelis & Aija, 2013). Private BD is, in contrast, created and owned by a private firm that tries to extract value from this distinctive resource to gain entrepreneurial rents. This taxonomy makes it clear that BD can be collected both by public and private organizations and be made partially or totally available to the general public. In this study we focus on the type termed public BD, which we refer to as simply BD in the rest of the manuscript.

BD can be extremely useful for public organizations, researchers, and policymakers. The increasing focus on BD is shown by the numerous grants that have recently been launched in the USA and China to encourage the use of BD for public organizations (Weiss & Zgorski, 2012; Wu et al., 2014). The benefits of BD have been also studied in the urban environment (Athey, 2017). An example is provided by the European ESPON project in Estonia (European Regional Development Fund, 2020), where citizens' GPS positions, collected by mobile devices, provide precise, rich, and quick information that allows policymakers to make better transportation plans compared with conventional methods. In addition, there is increasing evidence that cities can benefit from the collection of BD, e.g. in mitigating climate change, improving air quality, reducing traffic, and implementing shared mobility programs (Allam & Dhunny, 2019; Bai et al., 2018; Creutzig et al., 2019; Ford et al., 2016; Rathore et al., 2018). Furthermore, BD has been shown to contribute to increasing citizens' quality of life by providing insights that help tackle environmental and social issues such as industrial production and waste management, health systems, and mobility (Hashem et al., 2016). Other examples of the impact that BD may have on citizen health is using restaurant reviews to guide inspections and reduce food related health risks and costs (Glaeser et al., 2018). More generally, BD is starting to be at the core of city administration and planning because it makes it possible to simulate, evaluate, and monitor interventions in the urban environment (Hao et al., 2015).

Table 2 shows the main findings of the above-mentioned studies, tracing the benefits that BD can contribute to the environmentally, socially, and economically sustainable development of urban areas.

However, the value of BD also depends on the number of different sources integrated into database construction (Blazquez & Domenech, 2018; George et al., 2014). As a result, volunteer citizen scientists may be considered an additional and effective source of BD that can be exploited, as argued in detail in the following subsection.

2.3 | Joint implementation of CS and BD in urban contexts

In the previous sections we reported the benefits of CS and BD, which have so far been viewed as independent phenomena, whereas their simultaneous implementation has been overlooked. More specifically, a critical analysis of previous studies and cases of

TABLE 2 Previous studies on BD benefits for sustainable urban development

Benefits for urban development	Authors	Main findings	Geographical context	Methodology	Case
Environmental	Ford et al. (2016)	BD can help to manage risks of climate change	USA	Conceptual	Early warning and monitoring climate change
	Bai et al. (2018)	BD used to connect and control water, power, communication, and transport systems	Global	Conceptual	Reduce carbon emissions and support digital transformation
	Allam and Dhunny (2019)	BD used to reduce pollution in urban environment	Global	Conceptual	Artificial intelligence and smart city.
	Creutzig et al. (2019)	BD can generate environmental benefits in urban transport	Global	Conceptual	Sustainable urban transport
Social	Hao et al. (2015)	BD used to simulate, evaluate and monitor interventions in the urban environment	China	Review	Planning practices
	Hashem et al. (2016)	BD used in monitoring waste management, health systems, traffic and mobility	European Union	Qualitative	Smart city
	Athey (2017)	BD may aid in directing policy interventions	USA	Conceptual	Planning practices
	Rathore et al. (2018)	BD collected from sensors may improve quality of life, for example, by providing traffic alerts	Global	Qualitative	Smart city
Economic	Hashem et al. (2016)	BD may boost industry productivity and the economic development of cities	European Union	Qualitative	Smart city
	Glaeser et al. (2018)	BD can result in the ability to augment city government functions by lowering the costs of government services	USA	Quantitative	Quality of city services

Abbreviation: BD, big data.

implementation suggests a number of reasons why it is worth considering a joint application of CS and BD: (1) both CS and BD can be tools for sustainable development in cities; (2) in order for CS to work properly, large amounts of data need to be collected; (3) BD needs to integrate numerous different sources in order to be effective, and CS can contribute to this; (4) CS offers a fast and trusted means of collecting BD from citizens; and (5) by leveraging CS it is possible to collect BD from citizen volunteers in an inexpensive way.

Based on the above, we argue that by collecting further data from citizen scientists, it is possible to benefit sustainable urban development to a greater extent. Specifically, considering the fact that any given project focuses on collecting a certain quantity of scientific data for a precise scientific aim, we contend that, through CS, it may also be possible to collect BD from citizens. In particular, it should be possible to further exploit this interaction with citizens for targeted actions that can be used in the long term. In this sense, CS is an underestimated opportunity to collect BD that allows specific clusters of citizens to be identified so that targeted actions can be implemented to better involve them. Indeed, through the lens of a resource-based view (Barney, 1991; Wernerfelt, 1984), BD collected via CS may constitute a valuable resource that can be stored and later used to support sustainability-oriented studies and urban planning decisions.

This study argues that CS constitutes an opportunity to collect additional information from citizens, in the form of BD, which can further aid in the sustainable development of urban areas. Indeed, citizens that voluntarily participate in CS projects are highly engaged and motivated and more likely to allow the collection of BD for such aims. If we consider the numerous privacy concerns that individuals increasingly have regarding mobile devices (Acquier et al., 2017; Hayes et al., 2020; Moreno et al., 2016), citizens are more likely to provide information if asked by recognized institutions and researchers with a scientific aim. In this case citizens know that the information collected will be protected properly and that it will be used solely for scientific and public interest purposes, in order to increase sustainability. However, privacy can still be one of the perils in the joint implementation of CS and BD, and specific attention should be devoted to managing this personal information and avoiding dangerous data breaches. In addition to efforts to properly secure data collected from citizens, it might also be useful to consider recently developed insurance policies to protect against any such data breaches (Brook, 2019). Other dangers associated with both CS and BD, and in general with all processes based on digital technologies and technology-mediated interactions, are those related to the obsolescence of data and the risk of job losses. Regarding the timeliness of data, organizations should seek to use recent data and continuously collect new data. Both updated and older contributions from citizens might be useful for the current scientific aim as well as for future insights, but strategies for setting aside outdated data should be studied. Nonetheless, the fear of job loss due to digital transformation (Balsmeier & Woerter, 2019; Krutova et al., 2021) is not relevant in this instance. Indeed, data collected for a CS project aim would not have been collected in other ways, and so digitalization would not be taking anyone's job. On the contrary, the data collected as BD will require additional human resources to be effectively processed for future insights.

Notwithstanding the perils associated with CS and BD, which call for further scrutiny in future studies, the above considerations show that the way organizations can benefit from the simultaneous application of BD and CS has so far been overlooked, even though it is worth investigating. A qualitative approach to comprehend how public organizations can benefit from the joint adoption of CS and BD, by focusing on relevant case studies, has been adopted in this study. In the following section we specify the methodology and empirical outcomes of the case studies considered.

3 | METHODOLOGY

Due to the novelty of the phenomenon under investigation, that is, the joint application of CS and BD, we performed a qualitative analysis of multiple case studies. This method allowed us to explore in-depth and provide substantial support for the development of preliminary theoretical concepts (Eisenhardt, 1989; Leonard-Barton, 1990). To this end, we considered two relevant case studies (Esteves et al., 2021; Hallin et al., 2021) pertaining to highly populated and iconic urban environments in North America and Europe, that is New York and Rome. We selected these two cases for three main reasons: (i) they are among the few cases that show initial evidence of the joint applications of CS and BD, (ii) they are representative of different geographical and cultural contexts, and (iii) they provide a number of differentiated data sources, which is crucial for data triangulation and for the external validity of results (Flick et al., 2004). Consequently, the nature of these cases is consistent with the aim of analyzing and discussing the phenomenon under investigation (Siggelkow, 2007). We therefore employed maximum variation sampling (Soutaris & Zerbinati, 2014) and analyzed cases that cover a spectrum of lenses regarding the phenomenon investigated (Eisenhardt & Graebner, 2007; Miles & Huberman, 1984).

3.1 | Data collection

The data collected comprised both primary and secondary data and included the following sources: archival material, documentary information, surveys, and a semistructured interview. Table 3 illustrates the different data sources we used for each case.

For both cases, we collected data between February and April 2021, totaling 135 pages of transcribed data. By collecting a differentiated set of data sources, we were able to reach triangulation with both primary and secondary data covering different perspectives (i.e., citizens, organizations, managers, and media), which allowed us to garner a comprehensive viewpoint on CS and BD phenomena (Denzin, 2012; Jick, 1979).

3.1.1 | Documentary information

These data consisted of information collected from official sources for the two cases, that is, their websites and other materials directly produced by the organizations in charge of developing the projects, for example, YouTube videos. Collecting documentary information allowed the authors to gather basic information that could explain the main mechanisms behind the cases (Salvato & Corbetta, 2013), providing in-depth information and technicalities that codified the main processes implemented.

TABLE 3 Data source for each case analyzed

Case	Primary versus secondary data	Data source	Material
New York Street Tree Map (New York, USA)	Secondary	Documentary information	• Official website • YouTube videos
		Archival records	• Newspaper articles • Magazine articles • External reports
B.E. Smart (Rome, Italy)	Primary	Semi-structured interview	B.E. Smart software engineer
	Secondary	Documentary information	Official website
		Archival records	Magazine articles
		Surveys	Survey responses from citizens through the web-based platform

Abbreviation: B.E., building energy.

3.1.2 | Archival records

Archival materials consisted of a collection of external sources of data like specialized newspaper and magazine articles and reports on the cases, useful to confirm and verify the data gathered from official and primary sources (Layder, 1993).

3.1.3 | Surveys

Surveys were collected through questionnaires administered to citizens in the B.E. Smart project. These data also made it possible to record respondents' characteristics, which enriched the information provided by other sources of data with the perspective of the respondents themselves (Groves et al., 2011).

3.1.4 | Semistructured interviews

We conducted a semistructured interview with a key informant from the B.E. Smart project, that is, the software engineer (SE) Gianluca Squarcia. The interview was conducted in March 2021. The authors jointly interviewed him to provide an organizational perspective (Eisenhardt & Graebner, 2007). The interview was presented by introducing the general aim of the study, without making any presumptions on the topic, and it lasted 80 min.

3.2 | Data analysis

The analysis of data followed an inductive and iterative process (Miles & Huberman, 1984; Strauss & Corbin, 1997). We conducted within-case analysis for each case, extracting the main concepts from the data. Then we conducted a benchmark study on the two cases in order to find and compare possible differences and similarities. The first step in the analysis was to read and select all relevant quotes while focusing on the processes arising from each data source. After that, each researcher coded each quote independently in order to identify consistency between the data collected and the concepts found (Yin, 1994) and to develop a comprehensive understanding of the cases. After refining the concepts identified by merging the analyses of the researchers involved, we started to compare codes found in the two different cases. We found a great deal of consistency between them, despite the different data sources used and the different contexts.

Through this procedure, we identified a set of codes for each process. In some cases, recoding was needed because some of the processes identified were in conflict with some data items. Finally, we found patterns and relationships among concepts and built an analytical framework to address our research question. In the end, we conducted a further series of iterations between our primary and secondary data, and the literature on CS and BD, in order to refine the emerging framework of *citizen-sourcing* and better identify the theoretical foundations of our arguments.

4 | CASE STUDIES

In this section we will report the outcomes of our study for each one of the cases considered, that is, New York City Street Tree Map, in New York, USA, and B.E. Smart, in Rome, Italy. Then we will provide results of the cross-case analysis. All the sources consulted for the analyses of these case studies are reported in the Appendix.

4.1 | New York City Street Tree Map, New York, USA

The New York City Street Tree Map ("New York City Street Tree Map,", 2021), which was organized by the New York City Department of Parks and Recreation (NYC Parks) to conduct a census of all the trees in NYC, is the continuation of the CS project "TreesCount!". TreesCount! was launched in 2015 and lasted until 2016. The tree census is now regularly updated by NYC Parks (NYC Parks) and citizen volunteers can still provide updates on the condition of trees and take care of them by logging into the Tree Map web tool and registering their activities. New York City Street Tree Map made it possible to map 666,134 street trees and 2200 volunteer citizen scientists contributed their time and effort towards the Street Tree Map (Figure 1), with 12,000 h of volunteer work, saving the city about $100,112 in tax dollars. The data set comprises all the boroughs of New York, and citizen scientists contribute with data on tree species; state of health; trunk width, latitude, and longitude; and GPS coordinates. As reported on the project website, trees can indeed produce many benefits for their surrounding urban area, like filtering air pollution, releasing oxygen, reducing mental stress, providing shade, supporting wildlife, managing the flow of rainwater, and reducing traffic noise.

In her *New York Times* interview in 2019, the director of a New York-based nongovernmental organization involved in natural areas conservancy for New York, Ms. Charlop-Powers, said that over the years the project allowed them to "*collect data on native plants and the soil*." This extensive collection process made it possible to identify more than 200 hundred species of trees in New York City. In addition to producing a census of trees in order to better plan their management and care, it was also possible to recognize invaders, that is, trees that are not native to the area and that are spreading in particular areas of the city, threatening local biodiversity.

Such volunteer activities increase awareness of the trees present in the urban environment, which otherwise would not generally be known by citizens, and it also provides a way to inform citizens of the benefits of urban trees, such as the energy conserved in them and the amount of pollutant emission reduction and rainwater runoff reduction. These details were provided to citizens on the website and were publicly available to everyone, which could further increase interest in greenery and its care. Moreover, the data on shading, runoff, and CO_2 reductions also allowed policymakers and urban planners to implement future actions with greater awareness and consideration of the green assets that already exist in the urban

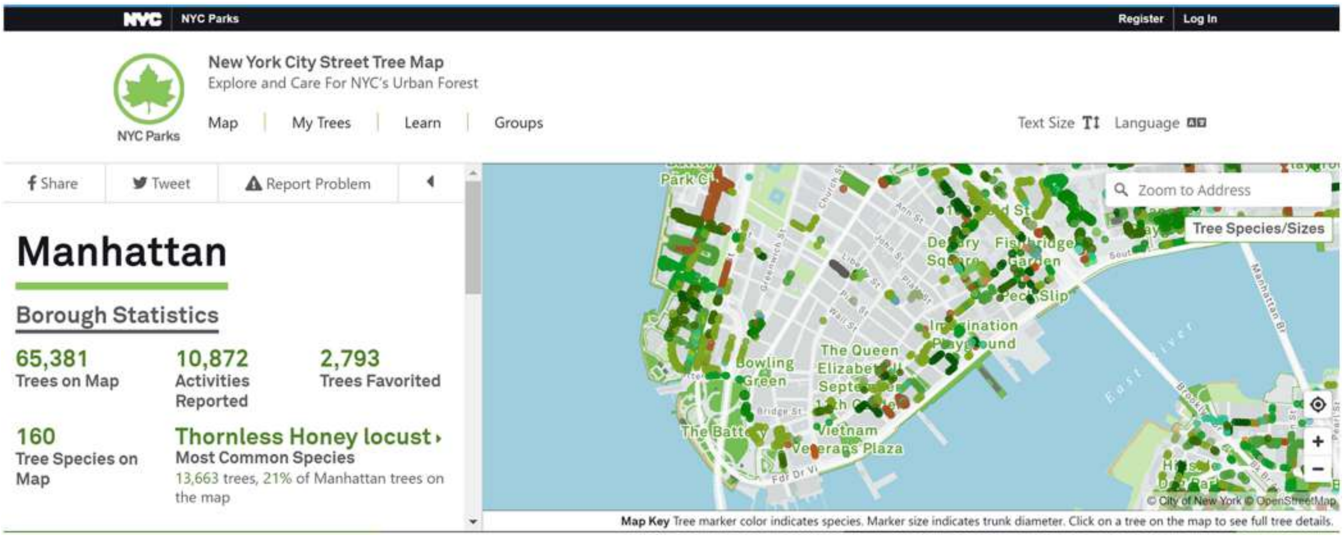

FIGURE 1 The web-based platform for the New York City Street Tree Map citizen science (CS) project [Colour figure can be viewed at wileyonlinelibrary.com]

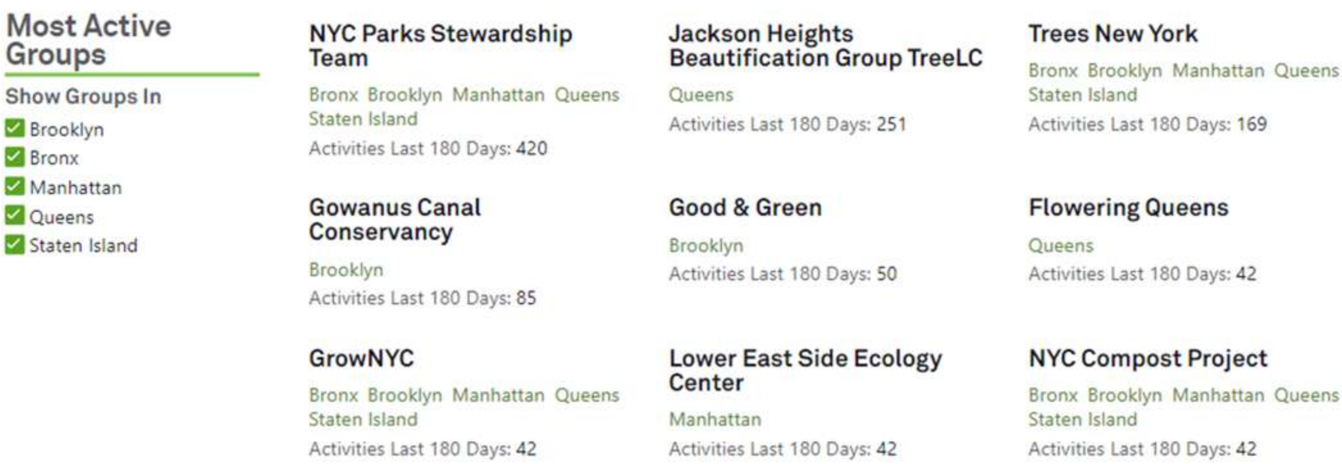

FIGURE 2 The New York City Street Tree Map's most active citizen groups [Colour figure can be viewed at wileyonlinelibrary.com]

environment. In addition, as shown in Figure 2, the website shows which groups of citizens and from which area of the city are the most active in mapping trees.

The kind of information collected through CS allows administrations to carry out targeted actions in specific areas of the city by profiling clusters of citizens and designing granular plans and strategies that take a number of additional data items into account. Indeed, the Forest Service report on the New York City Street Tree Map project states that: "*The better we understand the composition of who's engaged and who's not, the better we can serve those members of the public, we can craft messages and recruitment strategies that could be more effective in engaging more people.*"

As the New York City Park Department's Assistant Commissioner declared in an interview with Bloomberg, "*We have got excellent data at the core of this project*" and this extremely fine-grained information allowed them to better direct maintenance work in the city. This was only possible through the implementation of adequate technology, capable of collecting and managing such volumes of data. As the official website for the project ("New York City Street Tree Map,", 2021) reports, there are three technological aspects to consider: "*geospatial data collection, methods for mapping trees, and techniques for mobilizing volunteers and tracking progress.*" Such complex technologies need to be managed by specialized professionals: "*The simple and intuitive mapping method was designed by a local nonprofit, TreeKIT. The mapping technique leveraged a municipal geospatial dataset of curb edges to solve urban locational accuracy issues. The data collection method was integrated by the software company Azavea into a web application featuring online training modules, event management and community engagement tools to provide a seamless volunteer experience.*" A third aspect—mobilizing volunteers and tracking progress—regards the engagement of citizens and is described by the official website in these words: "*The user experience was designed to scale for thousands* *of nontechnical volunteers to collect standardized and consistent data with minimal training. The platform supported individual volunteers as well as local organizations who committed to recruiting volunteers and hosting events to map specific sections of the City."* This excerpt demonstrates how the project was able to collect large amounts of data while pursuing its CS goals, thanks to the inclusion of specialized professionals capable of managing the technologies needed to run different parts of the project. In other words, one of the key aspects is the need to extend the CS project to the general population through a system of engagement that also makes it possible to collect BD.

When implementing technological capabilities for profiling citizens, a major caveat may also emerge. As seen in the promotional video of New York City Street Tree Map on YouTube, individuals can contribute to the "urban forest" of New York from their "*desktop computer or from mobile applications*" and they are encouraged to share their activities on social media. It is evident that the information they share can therefore be easily collected. As a result, protecting citizens' privacy becomes fundamental for CS projects that aim to collect BD. There is no immediately available information available for the New York City Street Tree Map in this respect, as no privacy issues were reported. In the different steps in the relationship between the platform and citizens, no immediate information is given about data protection or the use of any sensitive information collected. This is relevant if we consider the different technologies, such as cookies and backend data gathering, involved in the project, as they can collect private data from users for citizen profiling (Dai et al., 2017; Schermer, 2011).

4.2 | B.E. Smart, Rome, Italy

B.E. Smart ("B.E. Smart," 2021), which stands for "Building Energy Smart," is a citizen science project that deals with electrical appliance usage in the context of residential energy-demand management. It suggests people "Be Smart," as the project's aim is to implement interventions in electricity production and usage that can lead to savings in pollutant emissions and costs. This project, run by an Italian university, seeks to achieve a better alignment between energy production and consumption, that is, energy-demand management. Typically, the demand for electricity peaks in the middle of the day, which forces carriers to switch on inefficient production plants for a few hours, like those using oil and coal, which causes higher electricity costs and increased emission of pollutants (Fagerberg, 2018). Better alignment of electricity production and usage could contribute to diminishing these drawbacks (Zhou et al., 2016). Citizen scientists are asked to provide data concerning the time at which they use their appliances. This information documenting their habits regarding appliance usage can be used to better plan electricity production and usage levels.

This ongoing CS project runs on an in-house web-based platform that is accessible on any device (Figure 3). Data collected from citizen scientists consist of their appliance usage habits, that is, the hours of their electricity usage for appliances, and the type and energy class of the appliances themselves. In addition, the website also asks participants for additional information in an anonymized form. These data include personal identification information, answers to surveys, and optional feedback that may be requested.

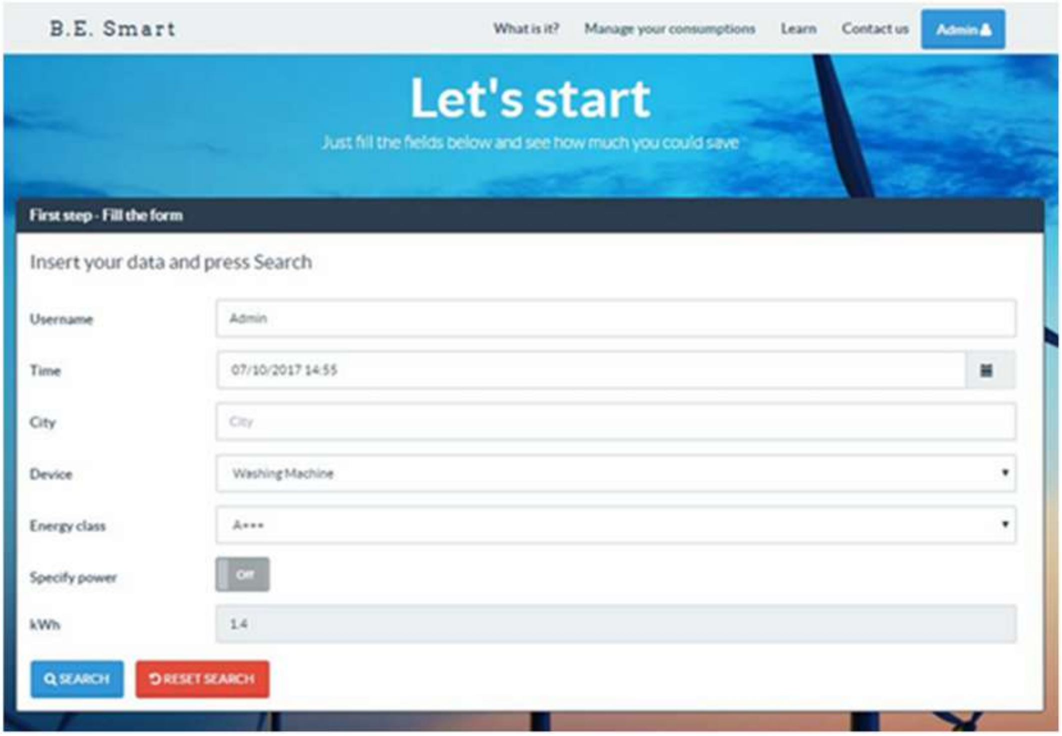

FIGURE 3 The building energy (B.E.) Smart citizen science (CS) project's web-based platform [Colour figure can be viewed at wileyonlinelibrary.com]

For this case we collected documentary information provided by the official website and archival materials retrieved from magazine articles. In addition, we conducted a semistructured interview with the project's SE and had access to the surveys conducted through the website (Table 3).

An article published in the magazine *Luiss Open* reported that *"Through B.E. Smart it was possible to provide personalized feedback for each citizen involved in the analysis to direct them towards more virtuous energy behaviors and to motivate them to provide data on their consumption habits."* As a matter of fact, one of the aims of the project was to profile citizens by matching their demographic and motivational characteristics with their appliance usage. In addition to the basic information collected, citizens could share other information on the website. This BD included the age, gender, and level of education of each participant; their level of engagement, enjoyment, and learning; as well as an open box for feedback and suggestions. Therefore, it was possible, ex-post, to extract additional valuable insights regarding the motivations of citizen scientists. Anonymized demographic data items are also collected through the web-based survey box administered online when users log onto the system.

Findings have shown that among citizen scientists, males are on average less engaged and derive less enjoyment in such activities. This may indicate that males could be motivated to engage in CS in different ways, for example, through gamification, and it may also provide evidence that females are the best target segment for attempts to increase the number of participants. Moreover, level of education has been found to have a positive effect on enjoyment level and a negative one on learning. This evidence suggests that individuals with higher levels of education might be those who should be targeted in the initial phases of a project because they may be more enthusiastic about CS. On the other hand, those with lower education levels may be the ones who stand to learn the most through direct experience with a CS project, and therefore CS would be most advantageous to them. As concerns the effect of age, younger individuals experience a higher level of learning and higher levels of enjoyment and engagement. In other words, elderly citizens should be involved at the beginning to increase the numerosity of the crowd, but younger citizens are those that learn the most through this experience.

In addition, a considerable amount of feedback has been provided through the online "open box." Citizen scientists suggested extending the CS approach to educational programs, in order to further motivate participants and expand its adoption ("*I would like to learn more about such research/learning projects*"; "*Traditional courses should incorporate such practical applications*"; "*I suggest devoting greater space in classes to such programs at university and in high school*"). This feedback is a valuable starting point for the policymakers and the researchers managing the B.E. Smart project when planning future actions to undertake. It is evident that, thanks to the BD collected, it is possible to implement targeted interventions to better leverage CS projects.

IT experts are needed to properly develop and maintain the interfaces used to interact with citizens when collecting data, as well as to manage the BD collected. Additionally, in the case of B.E. Smart, the BD collected through the web platform need to be stored and managed through advanced technological capabilities. As the SE of the project said in his interview: *"Our system allowed us to collect a large quantity of significant additional data not merely related to the project. At the beginning, we had some trouble managing this amount of information, as you cannot do that without the right technologies and people that know how to manage them."* Indeed, although CS projects do not always require deep technological competences to manage information, their joint implementation with BD collection cannot be accomplished without the presence of extensive technical know-how.

Finally, B.E. Smart clearly underscores the pertinence of privacy protection by declaring it a major point on its official website. A dedicated tab on the web-based platform ("B.E. Smart,", 2021) states that: *"While data constitutes a precious form of help in the sustainable development of our community, we respect citizen privacy and all the information collected is anonymized since the initial collection. In addition, we have personnel dedicated to the management of the data collected to prevent data breaches"* When collecting BD, in addition to the information needed for the primary CS aim, the project must also clearly emphasize attention to data protection issues. As previously described, B.E. Smart does indeed collect a number of sensitive information items in order to profile citizens and their appliance usage. Collecting this information helps public and urban administrations enhance their policy strategies on sustainable development, that is, energy efficiency, while pursuing ethical goals. In the same ethical vein, B.E. Smart stresses its commitment to data protection because it is conscious of its importance in a mutually respectful relationship between citizens and organizations (Hayes et al., 2020; Srinivasan, 2019). As also confirmed by the SE in our interview, *"Citizen respect and data protection are the keys to collecting BD. Not only to comply with regulations. This reflects our social commitment."*

4.3 | The processes for jointly implementing CS and BD

After the description of each of the cases considered, we conducted a cross-case analysis to compare and provide evidence of common processes that emerge, through which it is possible to jointly implement CS and BD. Following the coding scheme provided by the Gioia methodology (Gioia et al., 2013), we generated a number of theoretical themes from the data. The process was then refined, and from these theoretical themes, we developed further coding that was validated by comparing it with the first order concepts and the second-order themes (Figure 4). We ended up finding three processes (i.e. aggregate dimensions) that characterize the joint implementation of CS and BD, that is, *citizen profiling*, *data protection*, and *involvement of technological capabilities*.

We grouped the quotes, reported in detail in the previous subsections containing the case descriptions, according to the emerging theoretical themes as *second-order themes*. The first theme is the

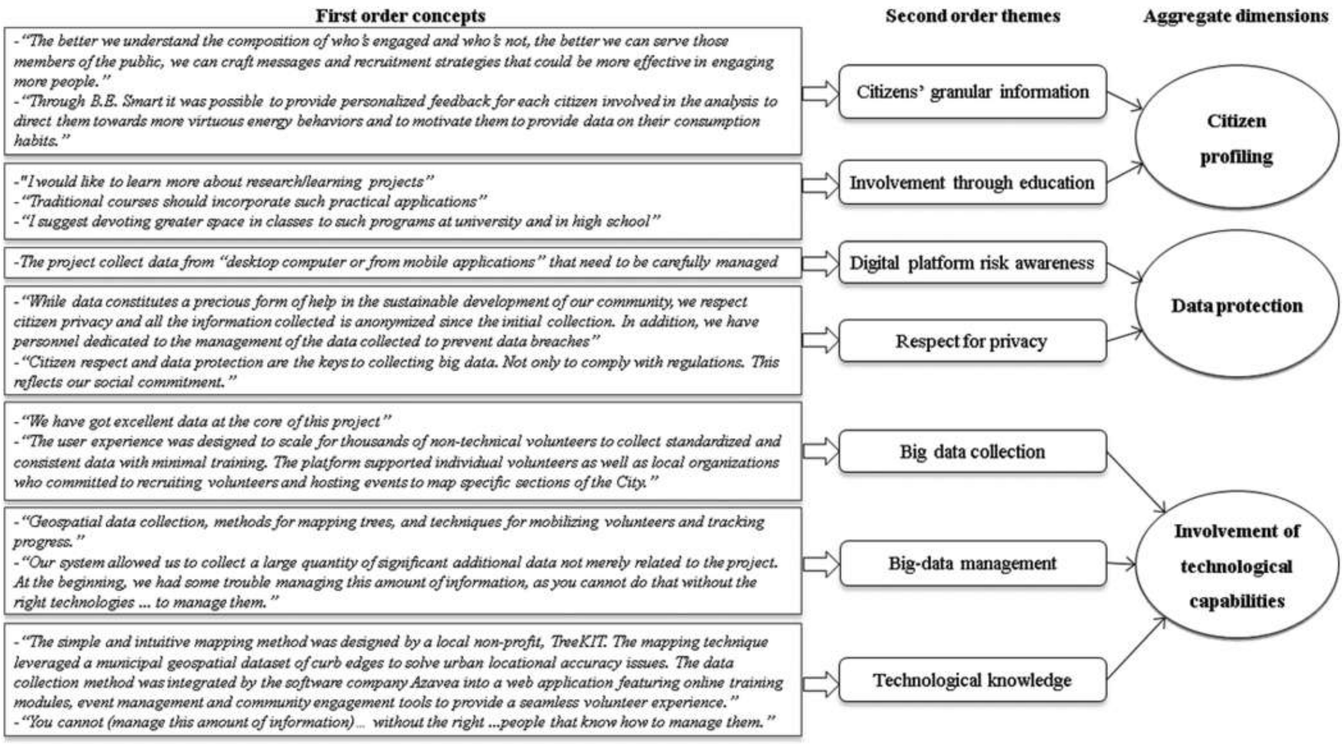

FIGURE 4 Data structure and coding scheme based on Gioia et al. (2013)

citizens' granular information, as CS makes it possible to collect personalized information about citizens. In this way it is possible to gather BD that may be used for future purposes. Another related theme is citizen *involvement through education*. Indeed, thanks to its social and environmental focus, CS can include the whole population in the collection of data for immediate scientific purposes and personal information for future usage, while also increasing scientific and environmental knowledge. Therefore, citizens also voluntarily participate through a willingness to improve their literacy. These two themes drive the collection of BD in CS projects because they compel more citizens to get involved, while collecting precise data. The dimension aggregating these two themes, that is, a third-order category theme, is called *citizen profiling*. Thanks to *citizen profiling*, it will be possible to collect granular information, so that researchers and urban administrations can understand how to better involve people, based on their characteristics, in other CS projects or in other participatory activities. This information could make it possible to implement tailored interventions that might increase the motivation level of groups of people that seldom participate; they could be the targeted recipients of one of the state-of-the-art techniques used to further motivate citizen scientists. Interventions like face-to-face interactions with project managers or social acknowledgement have been found to be effective in motivating citizens, although they can be costly to organize and implement. Thus, identifying the most suitable target audience can be extremely helpful in order to optimize costs and increase the effectiveness of these interventions. Besides getting less active citizen scientists more involved, knowledge of these data can also be useful for urban developers who could organize focus groups with the most active citizens in the neighborhood in order to implement better decision-making processes.

Another emerging second-order theme is *digital platform risk awareness*. It regards the fact that data from CS projects are collected through digital technologies such as mobile applications and computers that are subject to IT risks. These platforms provide access to large amounts of sensitive data, thus representing a possible source of power for digital hackers, including in public organizations. Entities should thus be aware of such risks and take appropriate precautions. Moreover, organizations are increasingly sensitive to respecting the privacy of their contributors as evidence of their social commitment to them. *Respect for privacy* is thus an additional second-order theme that refers to efforts to protect contributors' privacy through, for example, anonymization or encryption of data, as well as the technological infrastructure of defensive countermeasures to prevent data breaches. These two second-order themes converge in the *data protection* aggregate third-order dimension. This process is needed to guarantee proper functioning in terms of protection of personal information, which is crucial for the joint implementation of CS and BD, as citizens tend to consent to data collection for organizations that are mindful of privacy issues.

Finally, three additional second-order themes have emerged. The first one relates to *BD collection*. Under the circumstances described, CS projects can collect large amounts of data thanks to the digital platforms used by organizations. Proper mechanisms for collecting personal information from citizens should be put in place. At the same time, there is a need to properly store and administer the BD collected, an aspect that should not be overlooked. Efforts in this direction reveal the second-order theme of *BD management*, as the amount and nature of the information collected require complex storage and management systems. Finally, to complement efforts in *BD collection* and *BD management*, the need arises for *technological knowledge* that can properly implement CS and BD simultaneously. Such knowledge could be sourced both inside and outside the organization. In the New York case, for example, external parties are involved in the management of BD, and in the Roman case, the technological knowledge is internalized. Therefore, in order to properly satisfy the three themes just described and exploit the maximum value of a CS project, *involving technological capabilities* is needed, and this is our last aggregated dimension. Technology is indeed a fundamental aspect of CS and a conduit for the collection of BD.

In conclusion, *citizen profiling*, *data protection*, and the *involvement of technological capabilities* emerge as the three processes that characterize the joint implementation of CS and BD, as reported in Figure 4.

5 | DISCUSSION

CS is increasingly recognized as a valuable way to tackle the "grand challenges" we face today and to nurture sustainable development, thanks to the contributions made by numerous volunteer citizens to research projects (Craglia & Eds, 2014). In this study, we focused on the application of CS in urban environments due to the impact of "grand challenges" in this context (Guida & Carpentieri, 2021; Whiteman et al., 2011) and due to the opportunity cities offer to reach a large number of individuals who directly encounter such problems in their everyday lives. As a result, there is a huge amount of scientific and practical ferment around CS in urban areas, with studies and projects growing continuously. However, we have highlighted the fact that the benefits arising from CS are still far from having been thoroughly tapped.

In this study, we have provided evidence of how CS projects allow many kinds of data to be collected from volunteer citizens, in addition to the information currently needed for scientific aims. Indeed, the two projects considered in this study have provided evidence of data collected from citizens regarding the area in which they live, their gender, level of education, and interest in learning, and this is collected in addition to the original CS objective. These extra pieces of information allowed policymakers and researchers to identify target audiences for whom similar kinds of interaction requests could be better received in the future. At the same time, they have also discovered which proportion of the population may instead need targeted actions to increase their willingness to contribute to CS activities that improve the urban environment in which they live. The collection of such information constitutes BD, that is, a valuable resource that can be used for future purposes. Indeed, data that meet the 5 Vs can generate valuable insights (Cappa et al., 2021; Jin et al., 2015). In the context of urban environments the availability of BD collected from citizens may help public administrators and researchers to better direct and manage sustainability activities in future.

Based on the above, the possibility to jointly implement CS and BD may be fundamental in tackling current issues while also building a priceless resource that can produce valuable prospects in the future. The outcomes of this research contribute to revealing of how public organizations can do this in the urban context. In particular, our explorative analysis has identified three processes through which the joint implementation of CS and BD activities can be valuable and effective for public organizations, as summarized in Table 4.

The first process, *citizen profiling*, highlights the value of the joint implementation of CS and BD. Indeed, it may be possible to implement targeted actions in favor of sustainability with segments of profiled citizens. Thanks to the analysis of BD collected through CS projects, it is possible to profile citizens or clusters of citizens and identify emerging needs for further CS projects that directly involve the end users of these urban services, that is, citizens, in generating fast and precise feedback from their experiences. In addition, the information collected allows administrations to improve the quality of public decisions and interventions, for example, by making decisions that are better suited for specific neighborhoods or groups of citizens (Åström, 2020).

While adopting CS has thus far mainly involved researchers and policymakers, our analysis has demonstrated the need to *involve technological capabilities*, and therefore IT experts, as an additional process required in the joint implementation of CS and BD. IT experts are needed to craft interactions with the volunteers in the best possible way, as well as to properly store, manage, and analyze the BD

TABLE 4 The processes characterizing the joint implementation of CS and BD collection

Process	New York City Street Tree Map	B.E. Smart
Citizen profiling	Identification of clusters of active citizens that can drive granular urban policies	Identification of active citizens' demographic characteristics, which can drive targeted interventions
Involvement of technological capabilities	Use of three different technologies to collect, analyze and profile data	Need to manage a huge quantity of diverse data, including personal data from the CS project team collected by the web-based platform
Data protection	Information collected from users calls for a clear and immediate indication on data treatment and protection	Explicit commitment to protect citizen privacy in compliance with regulations and for ethical reasons

Abbreviations: B.E., building energy; BD, big data; CS, citizen science.

collected. This aspect is of great importance because the various technologies involved in these projects constitute the exact point of connection between citizens and public organizations.

Finally, *data protection* is the third process needed in joint CS-BD implementation. Due to the vast amount of data collected, which includes both general and sensitive information, even if anonymized, regarding citizens, there is a growing need for organizations to protect citizens' personal information (Dai et al., 2017; Hayes et al., 2020). The case of the New York City Street Tree Map illustrates the fact that some limitations still remain, at least in terms of immediately communicating their commitment to protecting citizens' data and privacy. In contrast, and as clearly implemented in the B.E. Smart project, citizens' immediate awareness of virtuous practices in this respect is fundamental in order to gain their trust in the project, thus favoring their commitment and collaboration.

Our findings suggest that cities should rely on data from citizens as important strategic resources to design and implement urban development activities (Larty et al., 2017). We have argued that the joint application of CS and BD constitutes an additional valuable resource because it might allow both short-term and future benefits to be generated. Based on the above, we have laid the foundations for a new framework named *citizen-sourcing*, which

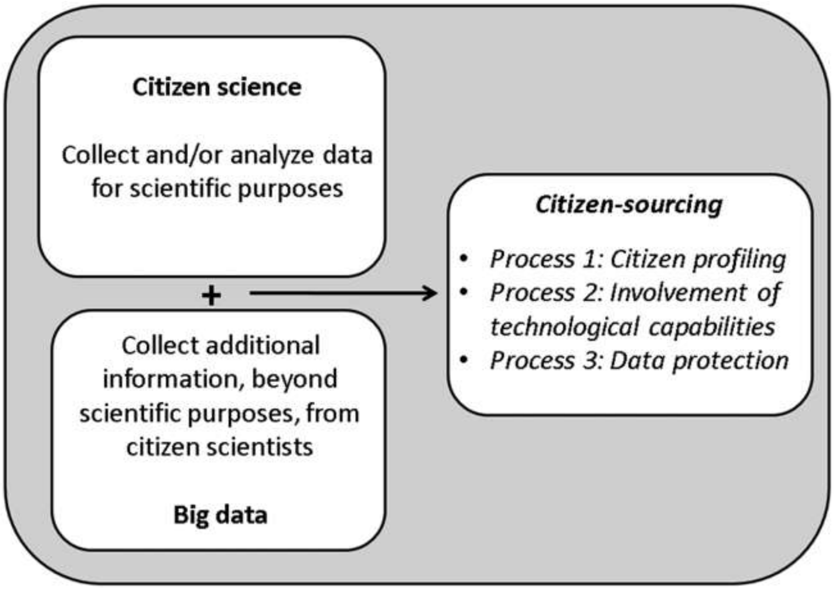

FIGURE 5 The joint implementation of citizen science (CS) and big data (BD) in citizen-sourcing

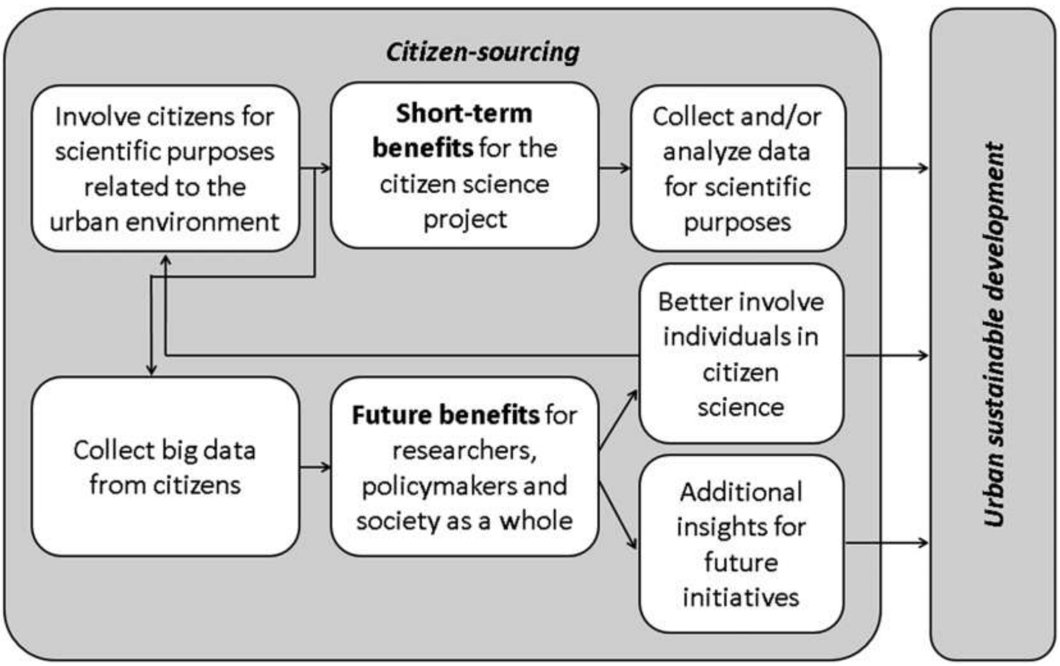

FIGURE 6 How to implement the citizen-sourcing framework and the benefits that can arise for the urban environment

makes it possible to collect additional personal information from citizens for future use while they are contributing to a current scientific purpose by collecting or analyzing primary data. *Citizen-sourcing*, whose boundaries are reported in Figure 5, synthesizes the joint implementation of CS and BD towards the goal of sustainable urban development.

In *citizen-sourcing*, rather than collecting information that is only instrumental to the precise aims of the CS project at hand, it would be possible to use CS to collect other types of information that could constitute BD. Having high volumes thanks to the numerous citizens involved, high variety thanks to the additional information collected, and high velocity thanks to the use of digital instruments for participation, it is possible to create extremely valuable and veracious BD. These data may prove to be valuable not only for the current needs of the CS project but also for future uses. Indeed, it may be possible to acquire additional insights about individuals who have participated in CS projects, in order to profile them and carry out targeted actions to better involve citizens in urban sustainability. In addition, researchers and policymakers may also collect feedback based on suggestions that come directly from the community to improve decision-making, as well as other information that can be used in future activities related to participatory planning and policy decisions. By outlining the benefits that can spring from the simultaneous application of CS and BD, we aim to focus the attention of researchers, policymakers, and citizens on this framework and further stimulate consideration of the benefits generated by *citizen-sourcing*.

The way a *citizen-sourcing* framework can be implemented in the urban context and the outcomes that can be obtained are summarized in Figure 6. In this manner it will be possible to achieve immediate benefits for CS projects while simultaneously collecting BD that could produce future benefits; both of these favor the sustainable development of urban environments.

6 | CONCLUSIONS

In this research we provide a number of contributions for theory and practice. First, while CS has already proven to be an effective tool for promoting scientific research, we have argued that it is also possible to add an additional step to common CS practices, that is, collecting BD in addition to the information needed for the primary aim of the scientific project. We have extended the resource-based view in this context to show how the joint application of CS and BD, that is, the *citizen-sourcing* framework, may be a valuable resource in the urban context. In this fashion, *citizen-sourcing* can be implemented to further benefit all the parties involved, that is, researchers, policymakers, and citizens collectively. We contribute to this by defining the boundaries of the new framework of *citizen-sourcing* and outlining the benefits that may arise from its implementation. Moreover, we have also outlined the position that *citizens-sourcing* may make it be possible to achieve both present and future benefits in sustainable urban development.

In addition to advancing scientific knowledge of the joint implementation of CS and BD, we have also provided practical indications on how to properly implement and benefit from *citizen-sourcing*. Indeed, we identify the three processes that characterize the functioning of *citizen-sourcing*: (1) *citizen profiling*, (2) *involvement of technological capabilities*, and (3) *data protection*. While *citizen profiling* emphasizes the potential benefit of *citizen-sourcing*, the *involvement of technological capabilities* and *data protection* highlights the way public organizations should carefully manage *citizen-sourcing* to achieve maximum benefits. In fact, the intended outcome of introducing digital technologies was to facilitate interactions and allow access to a broader set of data, as highlighted by the first two processes indicated in this study. However, digital technologies also produce unintended consequences, such as those connected to privacy concerns, which is the third process identified in this research and which indeed demands particular care from project organizers.

Above all, this study aims to shed light on *citizen-sourcing* in order to attract the attention of scholars, public organizations, and policymakers to the usefulness of this framework in tackling grand challenges in cities.

This research is not exempt from limitations, but these also provide directions for future development. First, while this study has illustrated the processes that need to be followed to better manage current CS projects and the additional benefits that can derive from BD collection through *citizen-sourcing*, the actual impact of the framework may be further validated by empirical studies. Moreover, the joint application of CS and BD has been explored in this study mainly in relation to the sustainable development of the urban environment. Indeed, the majority of grand challenges that can be addressed through CS are concentrated in cities, and there are numerous individuals that would be likely to participate voluntarily in addressing them. This is conducive to the collection of BD from the surrounding audience, that is, citizens. However, considering the several benefits brought about by *citizen-sourcing*, future studies could evaluate how this framework can also be applied in other contexts where the match between CS contributors and participants' data is not strictly limited to a target audience. For example *citizen-sourcing* can be applied to address other social and environmental issues like the discovery and study of vaccines, which began in a promising way in the USA in 2021 (Dockser Marcus, 2021, January). In such cases, indeed, citizen scientists and their data represent a broader geographical area compared with CS projects focused on urban issues. The application of *citizen-sourcing* in other contexts is promising but requires further refinement and studies, which we encourage scholars to pursue. In addition, the drawbacks associated with *citizen-sourcing* should be more clearly addressed from a theoretical and empirical standpoint, and with a multidisciplinary perspective, in future studies. Indeed, considerations regarding the perils of data protection, namely, data management and privacy issues and of the obsolescence of data should be further explored. Furthermore, the data considered in this study deal with public BD but it may also be possible to implement open BD strategies, rather than public BD alone; this would give citizens themselves the opportunity to fully consult the collected data, too, which may improve transparency and knowledge sharing. Future studies may consider this further development and enrich our understanding of the benefits produced by *citizen-sourcing*.

ACKNOWLEDGEMENT

Open Access Funding provided by Universita Campus Bio-Medico di Roma within the CRUI-CARE Agreement. [Correction added on 20 May 2022, after first online publication: CRUI funding statement has been added.]

FUNDING INFORMATION

Francesco Cappa and Federica Rosso would also like to gratefully acknowledge Ermenegildo Zegna that supported their research thanks to the EZ Founder's Scholarship 2017–2018, 2018–2019, 2019-2020 and 2020–2021. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID

Francesco Cappa https://orcid.org/0000-0001-5628-731X
Stefano Franco https://orcid.org/0000-0001-7341-8318
Federica Rosso https://orcid.org/0000-0003-2151-3780

REFERENCES

Acciarini, C., Borelli, F., Capo, F., Cappa, F., & Sarrocco, C. (2021). Can digitalization favour the emergence of innovative and sustainable business models? A qualitative exploration in the automotive sector. *Journal of Strategy and Management*. In Press, 1–18. https://doi.org/10.1108/JSMA-02-2021-0033, ahead-of-print

Acquier, A., Daudigeos, T., & Pinkse, J. (2017). Promises and paradoxes of the sharing economy. *Technological Forecasting and Social Change*, *125*-(July), 1–10. https://doi.org/10.1016/j.techfore.2017.07.006

Acuto, M., Parnell, S., & Seto, K. C. (2018). Building a global urban science. *Nature Sustainability*, *1*(1), 2–4. https://doi.org/10.1038/s41893-017-0013-9

Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., & Airaksinen, M. (2017). What are the differences between sustainable and smart cities? *Cities*, *60*, 234–245. https://doi.org/10.1016/j.cities.2016.09.009

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80–91. https://doi.org/10.1016/j.cities.2019.01.032

Angelidou, M., Psaltoglou, A., Komninos, N., Kakderi, C., Tsarchopoulos, P., & Panori, A. (2018). Enhancing sustainable urban development through smart city applications. *Journal of Science and Technology Policy Management*, *9*(2), 146–169. https://doi.org/10.1108/JSTPM-05-2017-0016

Arcanjo, J. S., Luz, E. F. P., Fazenda, Á. L., & Ramos, F. M. (2015). Methods for evaluating volunteers' contributions in a deforestation detection citizen science project. *Future Generation Computer Systems*, *56*, 550–557. https://doi.org/10.1016/j.future.2015.07.005

Ardito, L., Cerchione, R., Del Vecchio, P., & Raguseo, E. (2019). Big data in smart tourism: Challenges, issues and opportunities. *Current Issues in Tourism.*, *22*, 1805–1809. https://doi.org/10.1080/13683500.2019.1612860

Ardito, L., Scuotto, V., Del Giudice, M., & Messeni, A. (2018). A bibliometric analysis of research on Big Data analytics for business and management. *Management Decision*, *57*(8), 1993–2009. https://doi.org/10.1108/MD-07-2018-0754

Aristeidou, M., Scanlon, E., & Sharples, M. (2017). Profiles of engagement in online communities of citizen science participation. *Computers in Human Behavior*, *74*, 246–256. https://doi.org/10.1016/j.chb.2017.04.044

Åström, J. (2020). Participatory urban planning: What would make planners trust the citizens? *Urban Planning*, *5*(2), 84–93. https://doi.org/10.17645/UP.V5I2.3021

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *485*(February), 483–485. https://doi.org/10.1126/science.aal4321

Bai, X., Dawson, R. J., Ürge-Vorsatz, D., Delgado, G. C., Barau, A. S., Dhakal, S., … Schultz, S. (2018). Six research priorities for cities. *Nature*, *555*, 23–25. https://doi.org/10.1038/d41586-018-02409-z

Balsmeier, B., & Woerter, M. (2019). Is this time different? How digitalization influences job creation and destruction. *Research Policy*, *48*(8), 103765. https://doi.org/10.1016/j.respol.2019.03.010

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Batistič, S., & Van der Laken, P. (2019). History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. *British Journal of Management*, *30*(2), 229–251. https://doi.org/10.1111/1467-8551.12340

Bharadwaj, N., & Noble, C. (2017). Finding innovation in data rich environments. *Journal of Product Innovation Management*, *34*(5), 560–564. https://doi.org/10.1111/jpim.12407

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*(July 2017), 99–113. https://doi.org/10.1016/j.techfore.2017.07.027

Block, J. H., Colombo, M. G., Cumming, D. J., & Vismara, S. (2018). New players in entrepreneurial finance and why they are there. *Small Business Economics*, *50*(2), 239–250. https://doi.org/10.1007/s11187-016-9826-6

Boeing, G. (2021). Spatial information and the legibility of urban form: Big data in urban morphology. *International Journal of Information Management*, *56*(February 2019), 102013. https://doi.org/10.1016/j.ijinfomgt.2019.09.009

Bohnsack, R., Pinkse, J., & Bidmon, C. (2019). Special issue "Sustainability in the digital age.". *Business Strategy and the Environment*, *28*(6), 1283–1284. https://doi.org/10.1002/bse.2365

Brook, C. (2019). What is data breach or cyber security insurance? Retrieved May 20, 2019, from https://digitalguardian.com/blog/what-data-breach-or-cyber-security-insurance

Cappa, F., Laut, J., Nov, O., Giustiniano, L., & Porfiri, M. (2016). Activating social strategies: Face-to-face interaction in technology-mediated citizen science. *Journal of Environmental Management*, *182*, 374–384. https://doi.org/10.1016/j.jenvman.2016.07.092

Cappa, F., Laut, J., Porfiri, M., & Giustiniano, L. (2018). Bring them aboard: Rewarding participation in technology-mediated citizen science projects. *Computers in Human Behavior*, *89*, 246–257. https://doi.org/10.1016/j.chb.2018.08.017

Cappa, F., Oriani, R., Peruffo, E., & McCarthy, I. P. (2021). Big data for creating and capturing value in the digitalized environment: Unpacking the effects of volume, variety and veracity on firm performance. *The Journal of Product Innovation Management*, *38*(1), 49–67. https://doi.org/10.1111/jpim.12545

Cappa, F., Rosso, F., Giustiniano, L., & Porfiri, M. (2020). Nudging and citizen science: The effectiveness of feedback in energy-demand management. *Journal of Environmental Management*, *269*, 110759. https://doi.org/10.1016/j.jenvman.2020.110759

Centobelli, P., Cerchione, R., Chiaroni, D., Del Vecchio, P., & Urbinati, A. (2020). Designing business models in circular economy: A systematic literature review and research agenda. *Business Strategy and the Environment*, *29*(4), 1734–1749. https://doi.org/10.1002/bse.2466

Choe, E. K., Jung, J., Lee, B., & Fisher, K. (2013). Nudging people away from privacy-invasive mobile apps through visual framing. In *Lecture Notes in Computer Science* (Vol. 8119) (pp. 74–91). Springer. https://doi.org/10.1007/978-3-642-40477-1_5

CoolaData. (2014). The true cost of building a big data solution. Retrieved August 19, 2019, from https://www.cooladata.com/blog/true-cost

Corte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in European firms. *Journal of Business Research*, *70*, 379–390. https://doi.org/10.1016/j.jbusres.2016.08.011

Craglia, M., & Eds, C. G. (2014). Citizen science and Smart cities. European Commission Citizen science and Smart cities. *European Commission*, 1-57. https://doi.org/10.2788/80461

Creutzig, F., Franzen, M., Moeckel, R., Heinrichs, D., Nagel, K., Nieland, S., & Weisz, H. (2019). Leveraging digitalization for sustainability in urban transport. *Global Sustainability*, *2*, 1–6. https://doi.org/10.1017/sus.2019.11

Dai, W., Qiu, M., Qiu, L., Chen, L., & Wu, A. (2017). Who moved my data? Privacy protection in smartphones. *IEEE Communications Magazine*, *55*(1), 20–25. https://doi.org/10.1109/MCOM.2017.1600349CM

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, *6*(2), 80–88. https://doi.org/10.1177/1558689812437186

Dockser Marcus, A. (2021, January). Covid-19 patients are doing their own research. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/covid-19-patients-are-doing-their-own-research-11611982860

Druschke, C. G., & Seltze, C. E. (2012). Failures of engagement: Lessons learned from a citizen science pilot study. *Applied Environmental Education and Communication*, *11*(3–4), 78–188. https://doi.org/10.1080/1533015X.2012.777224

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., & Roubaud, D. (2019). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change*, *144*(June 2017), 534–545. https://doi.org/10.1016/j.techfore.2017.06.020

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/AMJ.2007.24160888

Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2019). A multi-dimension framework for value creation through big data. *Industrial Marketing Management, in Press.*, *90*, 508–522. https://doi.org/10.1016/j.indmarman.2019.08.004

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001

Esteves, A. M., Genus, A., Henfrey, T., Penha-Lopes, G., & East, M. (2021). Sustainable entrepreneurship and the sustainable development goals: Community-led initiatives, the social solidarity economy and commons ecologies. *Business Strategy and the Environment*, *30*(3), 1432–1435. https://doi.org/10.1002/bse.2706

European Regional Development Fund. (2020). Using big data for policy-making in the public interest|ESPON. Retrieved September 17, 2020, from https://www.espon.eu/using-big-data-policy-making-public-interest

Fagerberg, J. (2018). Mobilizing innovation for sustainability transitions: A comment on transformative innovation policy. *Research Policy*, *47*(9), 1568–1576. https://doi.org/10.1016/j.respol.2018.08.012

Ferster, C., Nelson, T., Laberee, K., Vanlaar, W., & Winters, M. (2017). Promoting crowdsourcing for urban research: Cycling safety citizen science in four cities. *Urban Science*, *1*(4), 21. https://doi.org/10.3390/urbansci1020021

Flick, U., Kardorff, E. V., & Steinke, I. (2004). *A Companion to Qualitative Research. A Companion to Qualitative Research (First edit)*. SAGE Publications.

Foray, D., Mowery, D., & Nelson, R. (2012). Public R&D and social challenges: What lessons from mission R&D programs? *Research Policy*, *41*(10), 1697–1792. https://doi.org/10.1016/j.respol.2012.07.011

Ford, J. D., Tilleard, S. E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A. C., MacDonald, G. K., Hsu, A., Chen, C., & Bizikova, L. (2016). Big data has big potential for applications to climate change adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10729–10732. https://doi.org/10.1073/pnas.1614023113

Franzoni, C., & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, *43*(1), 1–20. https://doi.org/10.1016/j.respol.2013.07.005

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

George, G., Haas, M., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326. https://doi.org/10.5465/amj.2014.4002

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, *56*(1), 114–137. https://doi.org/10.1111/ecin.12364

Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., & Tourangeau, R. (2011). *Survey Methodology*. John Wiley & Sons.

Guida, C., & Carpentieri, G. (2021). Quality of life in the urban environment and primary health services for the elderly during the Covid-19 pandemic: An application to the city of Milan (Italy). *Cities*, *110*(April 2020), 103038. https://doi.org/10.1016/j.cities.2020.103038

Guzmán, P. C., Roders, A. R. P., & Colenbrander, B. J. F. (2017). Measuring links between cultural heritage management and sustainable urban development: An overview of global monitoring tools. *Cities*, *60*(A), 192–201. https://doi.org/10.1016/j.cities.2016.09.005

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing Geographic Knowledge* (pp. 105–122). Springer. https://doi.org/10.1007/978-94-007-4587-2_7

Hallin, A., Karrbom-Gustavsson, T., & Dobers, P. (2021). Transition towards and of sustainability—Understanding sustainability as performative. *Business Strategy and the Environment*, (June 2020), 1–10. https://doi.org/10.1002/bse.2726, *30*, 1948, 1957

Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2020). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies, In press*, *58*, 1159–1197. https://doi.org/10.1111/joms.12639

Hansmann, R., Mieg, H. A., & Frischknecht, P. (2012). Principal sustainability components: Empirical analysis of synergies between the three pillars of sustainability. *International Journal of Sustainable Development and World Ecology*, *19*(5), 451–459. https://doi.org/10.1080/13504509.2012.696220

Hao, J., Zhu, J., & Zhong, R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management*, *4*(2), 92–124. https://doi.org/10.1016/j.jum.2015.11.002

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city.

International Journal of Information Management, *36*(5), 748–758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems.*, *47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006

Hayes, D., Cappa, F., & Le-Khac, N. A. (2020). An effective approach to Mobile device management: Security and privacy issues associated with Mobile applications. *Digital Business*, *1*(1), 100001. https://doi.org/10.1016/j.digbus.2020.100001

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016). *The Age of Analytics: Competing in a Data-Driven World* (p. 1). McKinsey Global Institute.

Huang, C., Wang, T., & Huang, T. (2018). Initial evidence on the impact of big data implementation on firm performance. *Information Systems Frontiers*, *22*, 1–13. https://doi.org/10.1007/s10796-018-9872-5

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*, 602. https://doi.org/10.2307/2392366

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, (Vol. 2, pp. 59–64). Elsevier. https://doi.org/10.1016/j.bdr.2015.01.006

Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. *Journal of Product Innovation Management*, *34*(5), 640–658. https://doi.org/10.1111/jpim.12397

Krutova, O., Koistinen, P., Turja, T., Melin, H., & Särkikoski, T. (2021). Two sides, but not of the same coin: Digitalization, productivity and unemployment. *International Journal of Productivity and Performance Management*, 1–27. https://doi.org/10.1108/IJPPM-05-2020-0233, ahead-of-print

Kuhlmann, S. (2014). The Challenge of Addressing Grand Challenges. European Union Framework Programme Horizon 2020 (1–11).

Land-Zandstra, A. M., Devilee, J. L. A., Snik, F., Buurmeijer, F., & van den Broek, J. M. (2015). Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science*, *25*(1), 45–60. https://doi.org/10.1177/0963662515602406

Larty, J., Jack, S., & Lockett, N. (2017). Building regions: A resource-based view of a policy-led knowledge exchange network. *Regional Studies*, *51*(7), 994–1007. https://doi.org/10.1080/00343404.2016.1143093

Laut, J., Cappa, F., Nov, O., & Porfiri, M. (2017). Increasing citizen science contribution using a virtual peer. *Journal of the Association for Information Science and Technology*, *68*(3), 583–593. https://doi.org/10.1002/asi.23685

Layder, D. (1993). New strategies in social research: An introduction and guide. *Polity Press.*, *31*, 31-1232. https://doi.org/10.5860/choice.31-1232

Leonard-Barton, D. (1990). A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, *1*(3), 148–266. https://doi.org/10.1287/orsc.1.3.248

Luigi Nimis, P., Pittao, E., Altobelli, A., De Pascalis, F., Laganis, J., & Martellos, S. (2019). Mapping invasive plants with citizen science. A case study from Trieste (NE Italy). *Plant Biosystems*, *153*(5), 700–709. https://doi.org/10.1080/11263504.2018.1536085

Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. *Strategy and Leadership*, *43*(5), 32–39. https://doi.org/10.1108/SL-06-2015-0054

McAfee, A., & Brynjolfsson, E. (2012). Big data. The management revolution. *Harvard Buiness Review*, *90*(10), 61–68. https://doi.org/10.1007/s12599-013-0249-5

McCaffrey, R. (2005). Using citizen science in urban bird studies. *Urban Habitats*, *3*(1), 70–86. ISSN 1541-7115

Messeni Petruzzelli, A., Natalicchio, A., Panniello, U., & Roma, P. (2019). Understanding the crowdfunding phenomenon and its implications for sustainability. *Technological Forecasting and Social Change*, *141*, 138–148. https://doi.org/10.1016/j.techfore.2018.10.002

Miles, M. B., & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, *13*(5), 20–30. https://doi.org/10.3102/0013189X013005020

Moreno, J., Serrano, M. A., & Fernández-Medina, E. (2016). Main issues in big data security. *Future Internet*, *8*(3), 1–16. https://doi.org/10.3390/fi8030044

Mueller, J., Lu, H., Chirkin, A., Klein, B., & Schmitt, G. (2018). Citizen design science: A strategy for crowd-creative urban design. *Cities*, *72*(April 2017), 181–188. https://doi.org/10.1016/j.cities.2017.08.018

Müller, O., Fay, M., & vom Brocke J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, *35*(2), 488–509. https://doi.org/10.1080/07421222.2018.1451955

NASA. (2021). Citizen science. Retrieved April 28, 2021, from https://science.nasa.gov/citizenscience

Natalicchio, A., Messeni Petruzzelli, A., & Garavelli, A. C. (2017). Innovation problems and search for solutions in crowdsourcing platforms—A simulation approach. *Technovation*, *64*, 28–42. https://doi.org/10.1016/j.technovation.2017.05.002

New York City Street Tree Map. (2021). Retrieved from https://tree-map.nycgovparks.org/

Nonninger, L. (2018). China Tencent and Alibaba's new credit scoring solution—Business insider. Retrieved December 20, 2018, from https://www.businessinsider.com/china-tencent-and-alibabas-new-credit-scoring-solution-2018-2?IR=T

Nov, O., Arazy, O., & Anderson, D. (2014). Scientists@ home: What drives the quantity and quality of online citizen science participation? *PLoS ONE*, *9*(4), e90375. https://doi.org/10.1371/journal.pone.0090375

Oyer, P. (2019). How to make data and AI add up. Retrieved June 5, 2019, from https://www.ft.com/content/8395f842-7712-11e9-be7d-6d846537acab

Pantelis, K., & Aija, L. (2013). Understanding the value of (big) data. In *2013 IEEE International Conference on Big Data* (pp. 38–42). IEEE. https://doi.org/10.1109/BigData.2013.6691691

Paul, K., Quinn, M. S., Huijser, M. P., Graham, J., & Broberg, L. (2014). An evaluation of a citizen science data collection program for recording wildlife observations along a highway. *Journal of Environmental Management*, *139*, 180–187. https://doi.org/10.1016/j.jenvman.2014.02.018

Paulos, E., Honicky, R. J., & Hooker, B. (2008). Citizen science: Enabling participatory urbanism. In *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City* (pp. 414–436). IGI Global. https://doi.org/10.4018/978-1-60566-152-0.ch028

Philip Chen, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, *275*, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Pinkse, J., & Dommisse, M. (2009). Overcoming barriers to sustainability: An explanation of residential builders' reluctance to adopt clean technologies. *Business Strategy and the Environment*, *18*(8), 515–527. https://doi.org/10.1002/bse.615

Pocock, M. J. O., Tweddle, J. C., Savage, J., Robinson, L. D., & Roy, H. E. (2017). The diversity and evolution of ecological and environmental citizen science. *PLoS ONE*, *12*(4), e0172579. https://doi.org/10.1371/journal.pone.0172579

Pratt, M. (2009). From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal.*, *52*, 856–862. https://doi.org/10.5465/AMJ.2009.44632557

Rathore, M. M., Paul, A., Hong, W. H., Seo, H. C., Awan, I., & Saeed, S. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. *Sustainable Cities and Society*, *40*, 600–610. https://doi.org/10.1016/j.scs.2017.12.022

Riesch, H., Potter, C., & Davies, L. (2013). Combining citizen science and public engagement: The open Air Laboratories Programme. *Journal of Science Communication*, *12*(3), 1–19. https://doi.org/10.22323/2.12030203

Roman, L. A., Scharenbroch, B. C., Östberg, J. P. A., Mueller, L. S., Henning, J. G., Koeser, A. K., Sanders, J. R., Betz, D. R., & Jordan, R. C. (2017). Data quality in citizen science urban tree inventories. *Urban Forestry and Urban Greening*, *22*, 124–135. https://doi.org/10.1016/j.ufug.2017.02.001

Rowbotham, S., McKinnon, M., Leach, J., Lamberts, R., & Hawe, P. (2019). Does citizen science have the capacity to transform population health science? *Critical Public Health*, *29*(1), 118–128. https://doi.org/10.1080/09581596.2017.1395393

Sakshi, S., Cerchione, R., & Bansal, H. (2020). Measuring the impact of sustainability policy and practices in tourism and hospitality industry. *Business Strategy and the Environment*, *29*(3), 1109–1126. https://doi.org/10.1002/bse.2420

Salvato, C., & Corbetta, G. (2013). Transitional leadership of advisors as a facilitator of Successors' leadership construction. *Family Business Review.*, *26*, 235–255. https://doi.org/10.1177/0894486513490796

Sauermann, H., Vohland, K., Antoniou, V., Balázs, B., Göbel, C., Karatzas, K., Mooney, P., Perelló, J., Ponti, M., Samson, R., & Winter, S. (2020). Citizen science and sustainability transitions. *Research Policy*, *49*(5), 103978. https://doi.org/10.1016/j.respol.2020.103978

Schermer, B. W. (2011). The limits of privacy in automated profiling and data mining. *Computer Law and Security Review*, *27*(1), 45–52. https://doi.org/10.1016/j.clsr.2010.11.009

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, *50*(1), 20–24. https://doi.org/10.5465/amj.2007.24160882

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Simwanda, M., & Murayama, Y. (2018). Spatiotemporal patterns of urban land use change in the rapidly growing city of Lusaka, Zambia: Implications for sustainable urban development. *Sustainable Cities and Society*, *39*, 262–274. https://doi.org/10.1016/j.scs.2018.01.039

B.E. Smart. (2021). Retrieved from http://besmart.scienceontheweb.net/

Soutaris, V., & Zerbinati, S. (2014). How do corporate venture capitalists do deals? An exploration of corporate investment practices. *Strategic Entrepreneurship Journal*, *8*, 321–348. https://doi.org/10.1002/sej

Srinivasan, D. (2019). The antitrust case against Facebook: A monopolist's journey towards pervasive surveillance in spite of consumers' preference for privacy. *Berkeley Business Law Journal*, *16*(1), 1–39.

Strauss, A., & Corbin, J. (1997). *Grounded Theory in Practice*. SAGE Publications.

Thurm, S., & Kane, Y. I. (2010). Your Apps Are Watching You. https://www.wsj.com/articles/SB10001424052748704694004576020083703574602

Tinati, R., Luczak-Roesch, M., Simperl, E., & Hall, W. (2017). An investigation of player motivations in eyewire, a gamified citizen science project. *Computers in Human Behavior*, *73*, 527–540. https://doi.org/10.1016/j.chb.2016.12.074

Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers and Industrial Engineering*, *115*, 319–330. https://doi.org/10.1016/j.cie.2017.11.017

Trabucchi, D., Buganza, T., Dell'Era, C., & Pellizzoni, E. (2018). Exploring the inbound and outbound strategies enabled by user generated big data: Evidence from leading smartphone applications. *Creativity and Innovation Management*, *27*(1), 42–55. https://doi.org/10.1111/caim.12241

United Nations (2018). World Urbanization Prospect. https://population.un.org/wup/

Visconti, R. M., & Morea, D. (2019). Big data for the sustainability of healthcare project financing. *Sustainability*, *11*(13), 3748. https://doi.org/10.3390/su11133748

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S., Subey, R., & Childe, S. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, W. Y. C., & Wang, Y. (2020). Analytics in the era of big data: The digital transformations and value creation in industrial marketing. *Industrial Marketing Management*, *86*, 12–15. https://doi.org/10.1016/j.indmarman.2020.01.005

Weiss, R., & Zgorski, L. (2012). Obama administration unveils "big data" initiative: Announces $200 million in new R&D investments. Office of science and technology policy, Washington. Retrieved from http://www.whitehouse.gov/

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180. https://doi.org/10.1002/smj.4250050207

Whiteman, G., de Vos, D. R., Chapin, F. S., Yli-Pelkonen, V., Niemelä, J., & Forbes, B. C. (2011). Business strategies and the transition to low-carbon cities. *Business Strategy and the Environment*, *20*(4), 251–265. https://doi.org/10.1002/bse.691

Wildschut, D. (2017). The need for citizen science in the transition to a sustainable peer-to-peer-society. *Futures*, *91*, 46–52. https://doi.org/10.1016/j.futures.2016.11.010

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Yin, R. K. (1994). Case study research. In *Design and Methods*. Sage.

Young, B. E., Dodge, N., Hunt, P. D., Ormes, M., Schlesinger, M. D., & Shaw, H. Y. (2019). Using citizen science data to support conservation in environmental regulatory contexts. *Biological Conservation*, *237*, 57–62. https://doi.org/10.1016/j.biocon.2019.06.016

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews.*, *56*, 215–225. https://doi.org/10.1016/j.rser.2015.11.050

How to cite this article: Cappa, F., Franco, S., & Rosso, F. (2022). Citizens and cities: Leveraging citizen science and big data for sustainable urban development. *Business Strategy and the Environment*, *31*(2), 648–667. https://doi.org/10.1002/bse.2942

APPENDIX A

Sources consulted for the analyses of these case studies:

https://tree-map.nycgovparks.org/

https://www.nytimes.com/2019/03/29/nyregion/newyorktoday/nyc-news-invasive-species-forests.html

https://www.bloomberg.com/news/articles/2016-11-04/explore-new-york-city-s-street-tree-map

https://www.openculture.com/2019/10/the-new-york-city-street-tree-map.html

https://secretnyc.co/map-every-single-tree-nyc-streets/

https://www.nycgovparks.org/trees/treescount https://www.green.it/nyc-street-tree-map-quanto-ci-risparmiare-un-albero/

https://www.fs.usda.gov/features/volunteers-count-every-street-tree-new-york-city-0

https://www.wired.com/2015/04/vivid-map-pinpoints-every-street-tree-nyc/

https://www.wsj.com/articles/new-tool-lets-cities-see-where-trees-are-needed-1492395000

https://www.fs.fed.us/nrs/pubs/jrnl/2017/nrs_2017_johnson_001.pdf

http://senseable.mit.edu/treepedia

http://media.nycgovparks.org/images/web/TreesCount/Index.html

https://www.youtube.com/watch?v=Jb3IK6Mo71c

https://www.youtube.com/watch?v=YrR8OMfemd8

Why count trees? Volunteer motivations and experiences with tree monitoring in New York City. Arboriculture & Urban Forestry, 2018. 44 (2): 59–72.

http://besmart.scienceontheweb.net/

Interview with the Software Engineer of B.E. Smart project, Dr. Gianluca Squarcia.

Surveys collected from the B.E. Smart project and made available by the researchers of the Citizen Science project.

https://open.luiss.it/2020/06/01/nudging-e-la-citizen-science-per-la-ricerca-di-nuove-soluzioni/