Journal of Remote Sensing
A SCIENCE PARTNER JOURNAL

REVIEW ARTICLE

Crowdsourcing Geospatial Data for Earth and Human Observations: A Review

Xiao Huang[1*], Siqin Wang[2*], Di Yang[3], Tao Hu[4], Meixu Chen[5], Mengxi Zhang[6], Guiming Zhang[7], Filip Biljecki[8], Tianjun Lu[9], Lei Zou[10], Connor Y. H. Wu[11], Yoo Min Park[12], Xiao Li[13], Yunzhe Liu[14], Hongchao Fan[15], Jessica Mitchell[16], Zhenlong Li[17], and Alexander Hohl[18]

[1]Department of Environmental Sciences, Emory University, Atlanta, GA, USA. [2]Spatial Sciences Institute, University of Southern California, Los Angeles, CA, USA. [3]Wyoming Geographic Information Science Center, University of Wyoming, Laramie, WY, USA. [4]Department of Geography, Oklahoma State University, Stillwater, OK, USA. [5]Department of Geography and Planning, University of Liverpool, Liverpool, UK. [6]Carilion School of Medicine,Virginia Tech, Blacksburg, VA, USA. [7]Department of Geography & the Environment, University of Denver, Denver, CO, USA. [8]Department of Architecture, National University of Singapore, Singapore, Singapore. [9]Department of Epidemiology and Environmental Health, College of Public Health, University of Kentucky, Lexington, KY, USA. [10]Department of Geography, Texas A&M University, College Station, TX, USA. [11]Department of Management Science and Information Systems, Oklahoma State University, Stillwater, OK, USA. [12]Department of Geography, University of Connecticut, Storrs, CT, USA. [13]Transport Studies Unit, University of Oxford, Oxford, UK. [14]The MRC Centre for Environment and Health, Imperial College London, London, UK. [15]Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway. [16]Spatial Analysis Lab, University of Montana, Missoula, MT, USA. [17]Geoinformation and Big Data Research Laboratory, Department of Geography, The Pennsylvania State University, University Park, PA, USA. [18]Department of Geography, The University of Utah, Salt Lake City, UT, USA.

*Address correspondence to: xiao.huang2@emory.edu (X.H.); siqinwan@usc.edu (S.W.)

Citation: Huang X, Wang S, Yang D, Hu T, Chen M, Zhang M, Zhang G, Biljecki F, Lu T, Zou L, et al. Crowdsourcing Geospatial Data for Earth and Human Observations: A Review. *J. Remote Sens.* 2024;4:Article 0105. https://doi.org/10.34133/remotesensing.0105

Submitted 12 September 2023
Accepted 14 December 2023
Published 19 January 2024

The transformation from authoritative to user-generated data landscapes has garnered considerable attention, notably with the proliferation of crowdsourced geospatial data. Facilitated by advancements in digital technology and high-speed communication, this paradigm shift has democratized data collection, obliterating traditional barriers between data producers and users. While previous literature has compartmentalized this subject into distinct platforms and application domains, this review offers a holistic examination of crowdsourced geospatial data. Employing a narrative review approach due to the interdisciplinary nature of the topic, we investigate both human and Earth observations through crowdsourced initiatives. This review categorizes the diverse applications of these data and rigorously examines specific platforms and paradigms pertinent to data collection. Furthermore, it addresses salient challenges, encompassing data quality, inherent biases, and ethical dimensions. We contend that this thorough analysis will serve as an invaluable scholarly resource, encapsulating the current state-of-the-art in crowdsourced geospatial data, and offering strategic directions for future interdisciplinary research and applications across various sectors.

Introduction

Over the past several decades, human and Earth observations have been overwhelmingly dictated by traditional, authoritative data sources, such as population censuses, surveys, satellite imagery, and other physical sensors. However, the landscape of data creation and analysis has undergone a seismic shift in recent times, fueled primarily by the advent of revolutionary paradigms such as Web 2.0 [1] and Big Data [2]. This paradigm shift was precipitated by several key factors, including widespread internet access, the ubiquity of smartphones, and a general surge in participatory culture [3]. The impact of this transition has been profound across various industries. In sectors like urban planning, transportation, and environmental monitoring, user-generated data have provided unprecedented real-time insights and community-driven perspectives, often leading to more responsive and adaptive decision-making processes [4]. In the commercial sector, businesses harness user-generated content for enhanced market research, customer engagement, and trend analysis, leading to more customer-centric product development and marketing strategies. The importance of this shift lies in its empowerment of ordinary individuals to contribute to and influence fields traditionally dominated by experts and authorities. This democratization has not only diversified the types of data available but also led to richer, more multifaceted insights into human behavior and environmental changes.

It is crucial to recognize that this democratization and the ensuing influx of user-generated content have crucially linked human experiences with environmental monitoring and analysis. For instance, the real-time tracking of human mobility patterns using smartphone data can greatly enhance our understanding of urban dynamics, traffic management, and even disaster response, effectively bridging the gap between human behaviors and environmental impacts. Similarly, public engagement in reporting environmental changes, like air quality or weather conditions, through mobile applications or social media platforms, brings a unique and valuable human dimension to Earth observations. These interconnected contributions emphasize a vital, yet previously underexplored synergy between human and Earth data sources, illustrating a more cohesive narrative of how human activities and Earth systems are inextricably linked.

The term "crowdsourcing (crowdsourced) geospatial data," which is used extensively throughout this paper, encapsulates the data acquisition process undertaken by large, diverse groups of individuals who often lack professional training [3]. The term "NeoGeography," introduced by Turner [5], conveys a broader contextual understanding through the sharing of location data, enabled by an ever-expanding array of freely accessible tools. In the same vein, "Volunteered Geographic Information (VGI)" [6] signifies the rising trend of ordinary citizens playing an active role in the creation of geographic information. VGI is characterized as "the employment of tools to create, assemble, and distribute geographic data provided voluntarily by individuals." Another pertinent term, "Citizen Science," refers to the active participation of the public in scientific research, monitoring, and action research, which often culminates in scientific progress and a broader public understanding of scientific principles [7,8]. Despite the varying focuses of these definitions, they all emphasize the growing importance and impact of non-authoritative data sources. The simultaneous advancement of rapid, accurate positioning technology, the prevalence of digital devices, the accessibility of high-speed communication links, and the progression in data management techniques have expedited the conceptual, methodological, and practical evolution of crowdsourced geospatial data. Contrasting with the traditional human and Earth observation methods, which are primarily coordinated by governmental and large institutional entities, the data collection process has been increasingly democratized, incorporating the participation of everyday users. This development effectively diminishes the erstwhile barrier between data producers and users. Such an innovative, decentralized approach to data collection, bolstered by an extensive global user base, potentially facilitates high-resolution spatio-temporal observations that were previously unattainable.

Crowdsourcing geospatial data has been analyzed through a multitude of lenses in recent studies. Numerous scholarly reviews have adopted a categorical approach to structure their analyses, centering on data source types. These include various platforms such as social media [9,10], OpenStreetMap (OSM) [11], and an array of other participatory datasets [12,13]. On the other hand, some reviews have taken a domain-focused approach, examining the practical applications of crowdsourcing geospatial data in fields like disaster mitigation [14], public health [15], remote sensing [16], and urban sciences [17]. While these studies offer valuable insights, a common limitation is the absence of a comprehensive, overarching perspective that ties together the various data sources and application domains. The primary motivation behind this review is to bridge this existing gap by offering a holistic perspective of crowdsourcing geospatial data. This will aid in fostering an improved interdisciplinary dialogue, supporting the development of innovative strategies and facilitating more efficient utilization of crowdsourced geospatial data across different sectors.

In addressing the interdisciplinary nature of this topic, we opted for a narrative review over traditional systematic or meta-analytical methods. Systematic reviews and meta-analyses, reliant on keyword-based database queries and quantitative data aggregation, often fail to adequately capture the nuanced and theoretical aspects crucial for such a multifaceted topic, necessitating substantial post-selection refinement [10]. Our narrative approach, emphasizing seminal works identified by subject-matter experts, offers a more targeted and insightful exploration. This method ensures comprehensive coverage and a contextual understanding, circumventing the limitations of keyword-dependent searches and the quantitative constraints of meta-analyses in addressing the complex dimensions of our interdisciplinary study. The articles assessed in this undertaking were intentionally chosen by the authors, who possess substantial expertise in crowdsourcing studies and have conducted interdisciplinary inquiries using crowdsourcing data. This targeted selection process ensures that the review encapsulates diverse perspectives and a rich array of experiences in this complex and evolving field.

In this study, we conduct an exhaustive analysis of the current efforts, possibilities, and obstacles associated with crowdsourced geospatial data across two fundamental perspectives: human observations ("Crowdsourcing Earth Observations" section) and Earth observations ("Crowdsourcing Human Observations" section). We group the applications of crowdsourcing geospatial data into varying domains, dissect the traits of data and contributors for widely recognized crowdsourced geospatial platforms, and investigate their data collection paradigms and applicable potential in detail. Furthermore, we discuss the intrinsic challenges ("Challenges in Crowdsourcing Earth and Human Observations" section) connected to crowdsourced geospatial data, considering facets such as data quality and accuracy, data bias, privacy concerns, legal and ethical dimensions, the sustainability of data collection, training and validation requirements, and issues surrounding data interpretation. This section is followed by a forward-looking discussion on prospective directions and pathways ("Future Directions and Pathways" section). The organizational layout of this review is illustrated in Fig. 1.

We believe that this comprehensive review will serve as an invaluable touchstone, encapsulating the concerted efforts in human and Earth observations utilizing crowdsourced geospatial data and providing future direction for effectively utilizing information gathered from crowdsourcing platforms to address extant and future challenges.

Crowdsourcing Earth Observations

In the era of rapidly evolving technology and increasing environmental concerns, crowdsourcing Earth observations has

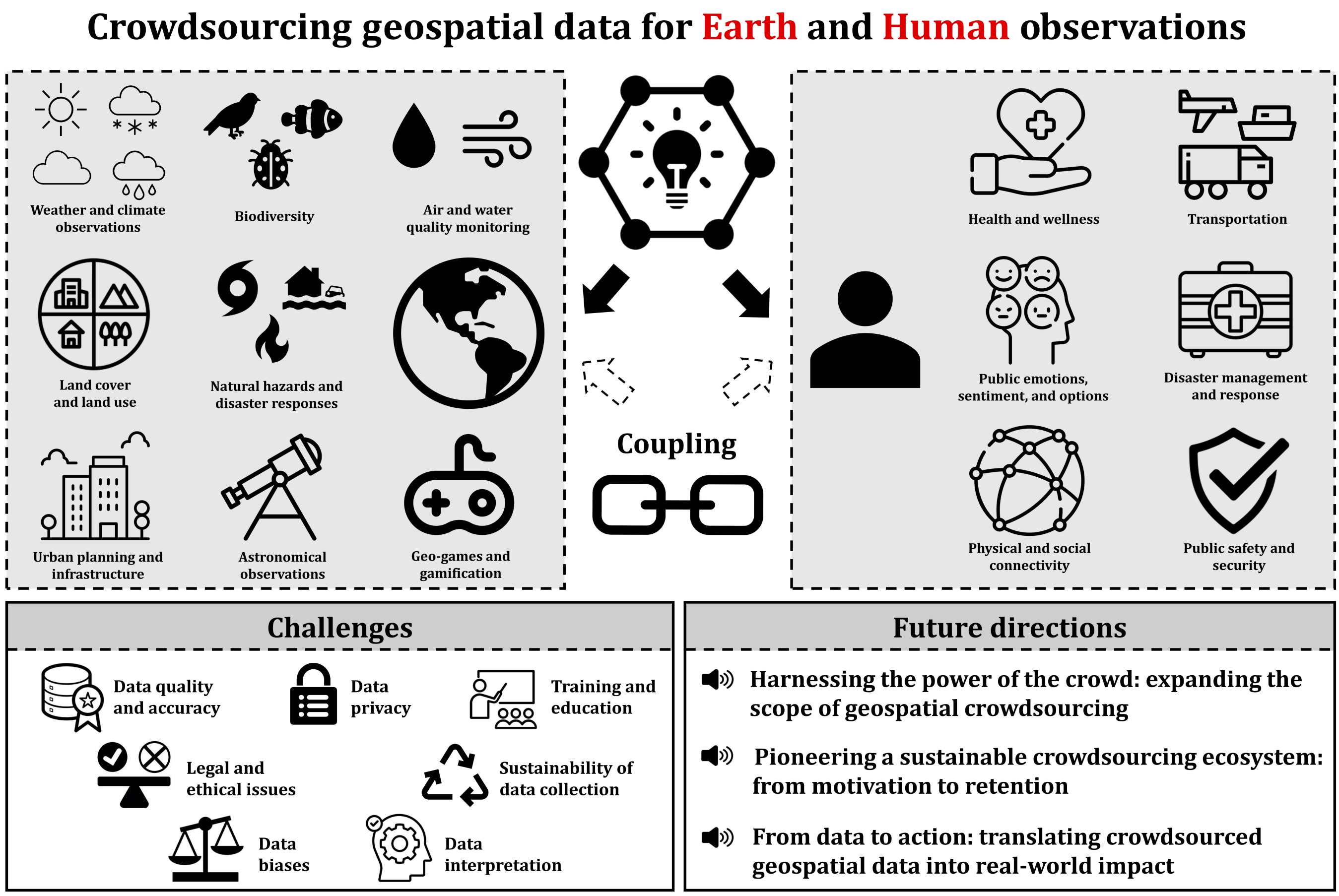

Fig. 1. The structure of this review.

emerged as an innovative and participatory approach to data gathering and analysis. Harnessing the collective intelligence and technological resources of citizens around the globe, this democratized methodology has fundamentally transformed the way we understand our planet. The approach transcends traditional barriers, involving the public in scientific research and decision-making processes. It has a wide array of applications that benefit both science and society. In the following sessions, we will delve deep into various facets of crowdsourcing Earth observations, examining its impact and efficacy in areas such as weather and climate observations, biodiversity assessment, air and water quality monitoring, and natural hazards and disaster response. We will also explore its role in understanding land cover and land use changes, aiding urban planning and infrastructure development, contributing to astronomical observations, and even creating geo-games and gamification strategies for educational and engagement purposes. The structure of our review on crowdsourcing Earth observation is presented in Fig. 2.

Weather and climate observations

The progression of technology has significantly advanced crowdsourcing methods for geospatial data collection for weather and climate observations. These methodologies can be effectively segmented into four principal categories, i.e., citizen science, social media, in situ sensors, and smart devices, with each offering unique benefits, facing specific challenges, and bringing distinct values.

Citizen Science stands out as a transformative force, especially in the domains of historical data retrieval and ongoing environmental monitoring. Endeavors such as Old Weather [18] and Cyclone Centre [19] highlight the tremendous capability of leveraging the general populace in data transcription and categorization. The Global Learning and Observations to Benefit the Environment Programme (GLOBE) further integrates students and educators in environmental measurements that adhere to stringent scientific standards. Moreover, ventures such as CoCoRaHS [20] and We Sense It (http://www.wesenseit.com/web/guest/home) underscore the significant contributions of community-driven networks, ensuring superior data collection quality. These data profoundly influence areas like climate change assessments [21], meteorological forecasting [22], and advancing knowledge about extreme weather events [23].

Social media platforms including Twitter and several mobile apps are progressively utilized as real-time data repositories. Initiatives like the UK snow map (https://uksnowmap.com/) and Twitcident (http://twitcident.com/) [24] emphasize the importance of user-generated content in monitoring events, from snow patterns to storm occurrences. Moreover, applications such as Metwit (https://metwit.com/) and Weddar (http://www.weddar.com/) procure localized meteorological data, forming a nexus between personal experiences and analytical data. Yet, the intrinsic attributes of social media, which can occasionally propagate misinformation, demand the adoption of meticulous filtering mechanisms.

In situ sensors have become pivotal in the data collection landscape. The incorporation of internet-capable, cost-effective sensors, integrated into individual weather stations or larger

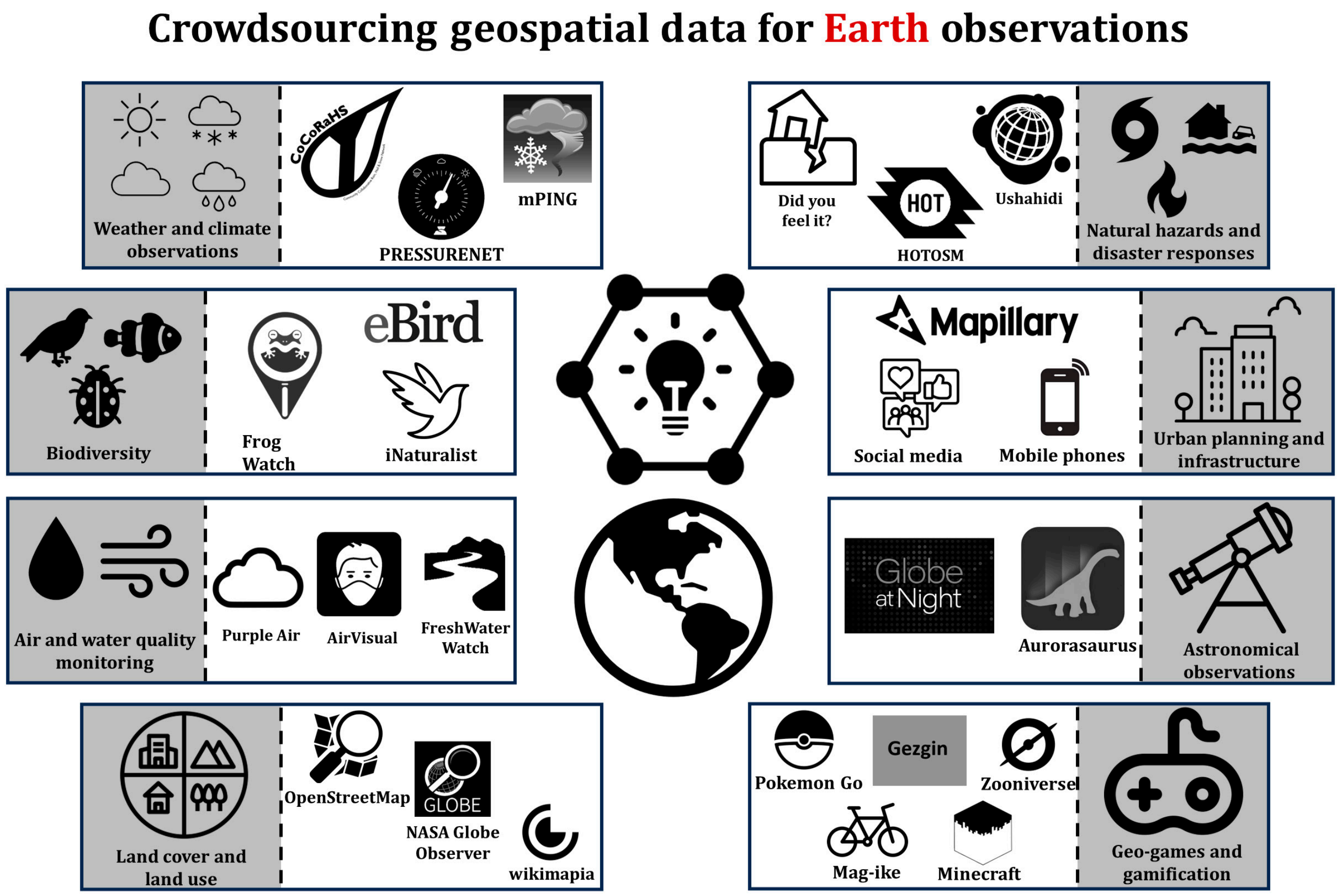

Fig. 2. Crowdsourcing geospatial data for Earth observations.

networks, has magnified both the quantity and detail of data accrued. Networks like Air Quality Egg (https://airqualityegg.com/home), a community-led air quality-sensing network, and Weather Underground (https://www.wunderground.com/) expedite real-time data acquisition from diverse origins. Although these devices are cost-effective, they necessitate rigorous calibration to validate data precision.

Smart devices are increasingly designed to connect with a variety of sensors, such as the BlutolTemp Thermometer [25], iCelsius thermistor (https://www.icelsius.com/), and iSPEX aerosol measuring sensor (www.ispex.nl), greatly facilitating dense data acquisition, predominantly in metropolitan settings. Notable initiatives, like the N-Smarts pollution project [26], leverage these sensor-equipped smartphones to understand urban air pollution's effects on individuals and communities. This plethora of sensors facilitates a crowdsourcing approach termed "human-in-the-loop sensing," enabling the collection and analysis of real-world data. Apps like OpenSignal (https://www.opensignal.com/apps) and PressureNet (http://pressurenet.cumulonimbus.ca) harness smartphone sensors to collect real-time weather data. However, there are challenges in using such data, primarily due to the potential variations in local weather conditions, which raises concerns about data accuracy and consistency.

In conclusion, the convergence of technological innovation and community collaboration has significantly transformed the realm of geospatial data collection, especially in weather and climate studies. Citizen science initiatives democratize the scientific process and ensure a consistent influx of essential data. Although social media introduces certain complexities, it provides unparalleled real-time insights. In situ sensors and smart devices enhance the precision and depth of data collection, facilitating more sophisticated interpretations. When integrated with rigorous validation and calibration methodologies, these tools are poised to drive future breakthroughs in environmental and atmospheric research.

Biodiversity

A plethora of crowdsourcing geospatial data has been accumulated through biodiversity citizen science projects for documenting and monitoring plants, animals, and other species on Earth [27]. Biodiversity citizen science projects provide infrastructure and platforms (e.g., website and app) to communities of volunteers (e.g., nature observers) who are interested in any aspects of biodiversity to contribute species observations [28,29]. A variety of biodiversity citizen science projects are in operation, attracting millions of contributors to report species observations. These projects may differ in species specialization and/or geographic scope, but together, they provide valuable data on species distribution, helping scientists monitor biodiversity. Three representative biodiversity citizen science projects are introduced here for illustration.

iNaturalist (https://www.inaturalist.org/), launched in 2008, enables anyone to share species observations of all taxa around the world by uploading species photos [29,30]. Each species observation is referenced with a geographic location and time (e.g., extracted from photos), and a species identification vetted by the community. Information of the observer and identifier is also retained. iNaturalist is arguably the largest biodiversity citizen science project in the world, having compiled over 148.5 million observations on more than 431,900 species based on contributions from over 2.7 million observers and 317,000 identifiers. eBird (https://ebird.org/) is for birdwatchers around the world to share bird sightings by submitting their birding checklists [28]. Essential information in a checklist includes location and time of the birding event, list of bird species observed, bird activity (e.g., breeding and behavior codes), bird group size, and information of the observer. A total of 10,715 bird species have been reported to eBird based on 82.3 million complete checklists contributed by 899,200 birders since the launch of eBird in 2002. FrogWatch USA (https://www.akronzoo.org/frogwatch), established in 1998, is a citizen science program for volunteers at a network of chapters to collect and submit data on local frog and toad populations in the United States. Volunteers are trained to listen to and identify frogs and toad calls. Frog and toad observations are recorded with location, time, and descriptions of habitat characteristics. A total of 178,795 observations have been submitted to this project by 15,641 volunteers.

Datasets resulted from biodiversity citizen science projects contain, at the minimum, spatially and temporally referenced records of the observed species and, in some cases (e.g., eBird and iNaturalist), even information of the underlying human volunteers who carry out the observations [31]. Species records can be analyzed to reveal the spatiotemporal dynamics of species distributions [32,33] to help inform conservation strategies, while information regarding data contributors allows examining volunteers' data contribution behavior patterns [34,35]. It should be noted that, although crowdsourcing offers an effective means for compiling timely biodiversity data at very large scales, the volunteers behind biodiversity data production are of varied levels of expertise [36] and their observation efforts are highly variable across space, time, and observation targets [34], leading to potential biases in the data (e.g., more bird observations are made during migration seasons on common species in accessible geographic areas). Such biases must be assessed and mitigated where necessary in order to make robust inferences from the data [37,38,39].

Air and water quality monitoring

Crowdsourcing emerges as a cost-effective tool for the data collection on air quality, especially enabling a broader spatial coverage and increased temporal resolutions as compared to traditional regulatory-grade monitoring. PurpleAir, a very affordable, accessible, and easy-to-use low-cost air sensor network, has become one of the most popular and largest crowdsourced networks worldwide [40]. The dense network is built with the help of stakeholders such as residents, environmental and public health agencies, and university researchers to measure real-time particulate matter (PM) at residential areas, industrial facilities, schools, and various other places of interest [41,42]. For example, the PurpleAir sensor data are also integrated into AirNow, a one-stop source of air quality platform developed by major agencies, including the US Environmental Protection Agency (EPA), to provide historical, current, and future air quality data, especially during wildfire seasons [43]. Similarly, "AirVisual," developed by IQAir, also engages citizen scientists to collect various pollutants (e.g., PM, ozone, and nitrogen dioxide) mainly through indoor (e.g., home, office, and hospital) and mobile (e.g., travel) applications [44]. SmartCitizen offers a kit with open-source files and schematics, enabling users to customize their air measurement needs through citizen science [45]. Clarity monitoring network enhances traditional air monitors by incorporating solar panels to sustain internal batteries and enable the measurement of black carbon [46]. Additionally, Air Quality Egg aims to empower K-12 students to become citizen scientists by measuring multiple pollutants, including carbon monoxide and sulfur dioxide [47]. Overall, crowdsourced platforms emerged as integrated systems for short-term and long-term community-based air quality monitoring networks, enabling the public to measure and report real-time crowdsourced data. Importantly, the crowdsourced air quality data could be combined with satellite remote sensing data, meteorological data, noise data, and other smart-phone applications to revolutionize environmental exposure assessment, disaster resilience, urban planning, environmental justice, and epidemiological studies [48,49].

Crowdsourcing has also become increasingly valuable in water quality monitoring, engaging citizen scientists to collect data on a large scale. For example, the Secchi disk, a plain white, circular disk with a diameter of 30 cm, is commonly used for measuring water transparency or turbidity. Citizens, scientists, nongovernment organizations, and other stakeholders have employed this tool to increase networks of in situ measurements mainly in oceans [50] and lakes [51,52], supplementing traditional monitoring networks. Particularly, the North American Lake Management Society has held annual crowdsourcing events—"Secchi Dip-In"—since 1994 to engage lake enthusiasts and volunteers to contribute to a comprehensive database of water quality [53]. State agencies also encourages the use of crowdsourced air quality monitoring to supplement routine water monitoring activities, such as the Citizens Statewide Lake Assessment Program [54] and the Clean Water Team under the Surface Water Ambient Monitoring Program [55]. Volunteers are trained to collect samples and measure various parameters, such as dissolved oxygen, pH, and nutrient levels to characterize water quality. Additionally, environmental charity such as Earthwatch Europe has launched a global citizen science project—FreshWater Watch in 2012. This initiative engages volunteers in using the same standardized research method to monitor various water bodies on a broader scale and to identify regional and global trend, including rivers, lakes, streams, ponds, and wetlands [56]. The Surfrider Foundation's Blue Water Task Force is a crowd-sourcing initiative focused on monitoring water quality at beaches and coastal areas to protect public health and advocate for clean water policies [57]. These example programs commonly value the advantages of crowd-sourcing data in water quality monitoring. By leveraging citizen science and engaging the public to collect data on parameters essential to understanding water quality, the crowdsourced database contributes to enhance spatial and temporal coverage, expand types of waters, support decision-making, and inform efforts in water resource management and conservation.

Natural hazards and disaster response

Natural hazards like flash floods or earthquakes usually take place in a short period, while their impacts vary by communities and individuals and depend on local context, e.g., hazard conditions, natural and built environment, socioeconomic characteristics, management strategies, and responding behaviors. Therefore, timely and hyperlocal observations of natural hazards and affected communities are necessary to understand the impacts of natural hazards and formulate immediate, actionable disaster response that can minimize loss of lives, property damages, and social and environmental disruptions. Harnessing crowdsourced geospatial data from social media platforms, smartphone apps, or crowdsourcing websites offers a novel avenue for observing short-term, localized events with exceptional spatial and temporal resolutions [58]. Several initiatives have adeptly harnessed crowdsourced data to observe natural hazards and their impacts, as well as supporting disaster responses.

For example, the United States Geological Survey (USGS) developed the "Did You Feel It?" (DYFI) website in 1999 to gather information about the impacts and effects of earthquakes from people who have experienced them. When an earthquake occurs, individuals in the affected area can visit the DYFI website and provide details about their experience, including their location, the level of shaking they felt, and any observed effects, such as swaying of buildings, rattling objects, or other impacts. This information is then aggregated and displayed on an interactive map, allowing users to see how the shaking was perceived across different areas. Crowdsourced DYFI data have been demonstrated valuable in better understanding the spatial distribution of shaking intensity, which can be used to refine earthquake hazard maps [59], evaluate the performance of buildings and infrastructure, and improve earthquake engineering practices [60].

Another notable example is Ushahidi (meaning "testimony" in Swahili), an open-source platform that facilitates crowdsourcing, mapping, and data visualization for hazardous events or crises. This platform empowers both individuals and organizations to collect reports that elucidate local damages and identify individuals needing help during natural hazards. These reports are derived from diverse sources like text messages, social media posts, emails, and web forms [61]. The reports are then aggregated and displayed on an interactive web map to coordinate disaster response missions. Since its inception in 2007, Ushahidi has been employed in over 90,000 cases, received more than 6.5 million reports from 160 countries, and played a pivotal role in several disaster response tasks, such as searching and rescuing victims during the 2010 Haiti Earthquake [62].

CrowdSource Rescue (CRS) is a similar platform that uses crowdsourced geospatial data to bolster disaster response. It was established in 2017, prompted by the unprecedented flooding caused by Hurricane Harvey in Houston and the surrounding areas. During Harvey, many people failed to evacuate on time while the 911 system was overloaded [63]. As a result, flood victims turned to online platforms (e.g., Twitter and Facebook) and volunteer groups (e.g., Cajun Navy) to seek help [64]. CRS emerged as a solution to aggregate these help requests through a crowdsourcing methodology and present them via an interactive WebGIS platform. This platform enables victims to submit help requests and allows certified volunteers to access the geographical locations of individuals seeking help and extends assistance accordingly. As of 2023, CRS has engaged over 13,513 rescuers and volunteers, facilitating aid to more than 94,036 survivors across 28 hazardous incidents.

In addition, disaster-affected regions with limited geospatial data and experts can benefit from geospatial data crowdsourced by cartography professionals or volunteers. The Humanitarian OSM Team (HOT) is a pioneering and collaborative effort that leverages crowdsourced geospatial data to co-produce mapping products for decision-making in disaster response. Using OSM, cartography professionals or volunteers can contribute to satellite image digitization and mapping roads, buildings, rivers, and essential features in disaster-affected regions with limited mapping resources. This collective endeavor yields comprehensive and up-to-date geographical data that significantly enhance disaster response efforts by supporting the assessment of fundamental infrastructure (e.g., transportation and shelters), identification of resource availability, delineation of impacted zones, evaluation of damages, and estimation of affected populations. HOT's establishment was prompted by the 2010 Haiti Earthquake, and its contributions have been demonstrated in various natural hazard events, including the 2015 Nepal Earthquake [65].

Land cover and land use

Citizen science has emerged as a powerful tool for collecting and analyzing data across a wide range of disciplines, including land cover and land use studies. The integration of citizen science with Earth observation has the potential to provide valuable calibration and validation data, covering a diverse set of fields from disaster response to environmental monitoring. This integration has yet to be fully exploited, and there is a significant opportunity for citizen science to contribute to the achievement of the United Nations Sustainable Development Goals, including those related to land use and land cover [66].

Several projects have demonstrated the potential of citizen science in this domain. For instance, the Geo-Wiki project has leveraged a global network of volunteers to improve the quality of geographic data by validating and correcting existing land cover maps [67]. This project has harnessed the collective efforts of a global network of volunteers to enhance the quality of geographic data, primarily through the validation and correction of existing land cover maps [67,68]. The tool has found particular utility in Central Europe, where it has been extensively employed to improve land cover maps [69]. Similarly, the LUCID project has engaged participants in identifying types of land cover in satellite images to monitor land use changes over time [70]. The Missing Maps project has mobilized volunteers to use satellite images to map areas that are home to vulnerable populations but are poorly covered in existing maps. These projects, among others, have shown that citizen science can provide high-quality data for land use and land cover classification tasks, and that local knowledge and professional background have a minimal impact on volunteer performance in these tasks [66]. Additionally, these ground-based observations frequently provide essential datasets for training machine learning algorithms or for developing mapping rules, thereby enhancing the interpretation of satellite imagery and improving the accuracy and dependability of land cover classifications obtained from satellite data [71].

The integration of citizen and community science into land cover, land use, and land change detection processes has been explored in various contexts. For instance, Olteanu-Raimond et al. [72] proposed an experimental framework for integrating citizen science into a national mapping agency. Theobald [73] put forth a general-purpose spatial survey design for collaborative science and monitoring of global environmental change. In addition, Kolstoe et al. [74] leveraged citizen science to study the differential impacts of climate and land cover on bird populations in the Pacific Northwest. In a similar vein, Whitehorn et al. [75] utilized a decade of citizen science observations to investigate the effects of climate and land use on British bumblebees.

These and other projects demonstrate the diverse applications of citizen science in monitoring and understanding changes in land cover and land use.

Urban planning and infrastructure

Data on the infrastructure in the built environment, such as buildings, roads, amenities, and public open spaces, is crucial for urban planning and supporting livable and smart cities. The acquisition and management of such data has traditionally been in the realm of governments, but the crowdsourcing route has been gaining momentum in the past decade with increasing completeness and application of such data in academia and practice. Further, the surge of sensors and the volume of user-generated geographic information in cities introduced new means to collect information on the built environment. For example, recent crowdsourced datasets that have been gaining attention in the built environment are real estate ads and accommodation reviews to collect information on building characteristics [76,77,78].

OSM, the open and collaborative map of the world that is built by a community of mappers, is perhaps the most relevant instance of crowdsourcing in this context [79]. The community contributes and maintains data about roads, trails, cafes, railway stations, etc., with the content reaching high levels of quality globally. For example, currently, there are more than half billion buildings mapped around the world, reaching full completeness, high quality, and rich semantic information in many urban areas around the world [80,81]. Taking advantage of this trend, data on a variety of features sourced from OSM have been used for myriads of purposes in urban planning, supporting, e.g., modeling urban change and characterizing the urban form, real estate analyses, and population studies [82,83,84,85]. Considering the continuous popularity of OSM and the growing role of corporate editors [86], i.e., companies that contribute data to OSM, the platform is expected to remain very relevant and instrumental in the crowdsourced mapping of the urban infrastructure and its applications in urban planning.

Mapillary, a platform that manages crowdsourced imagery to create a visual representation of the world for improving maps [87], is another increasingly popular instance of crowdsourcing spatial data in the built environment as street-level imagery plays an important role in urban planning and infrastructure management. Despite being a relatively new type of spatial data, because of offering a new perspective and other advantages such as dense coverage, street-level imagery has rapidly gained attention for a variety of use cases in the built environment. Initially, use cases have been dominated by data provided by commercial services such as Google Street View and Baidu Maps, with many demonstrated uses of it for mapping infrastructure and supporting urban planning, such as collecting information on buildings, assessing perception of streetscapes, and mapping greenery [88,89,90]. However, recent years have seen a growing use of Mapillary for similar purposes, offering an alternative at a liberal license and with various advantages such as imagery taken from bicycles and in open public spaces, owing to the heterogeneity of contributors, and offering dense coverage in urban areas that are often not available in commercial counterparts. For example, Mapillary has been used to extract detailed information of buildings to generate three-dimensional (3D) building models at high level of detail, map networks of bicycle paths, and measure greenery along roads [91,92,93,94]. While Mapillary remains the most popular crowdsourced platform for street-level imagery, there are alternative services that are more popular in particular regions, e.g., KartaView, which is particularly focused on Southeast Asia.

Astronomical observations

Astronomical observations, as part of the emerging concept of "Citizens as Sensors," leverage the potential of crowdsourcing for capturing celestial phenomena and contributing to a global dataset. This active data collection, engaged by voluntary contributors worldwide, provides a unique opportunity to gather scientific data on a scale that would be otherwise impossible with traditional methods. The "Globe at Night" and "Aurorasaurus" projects, two remarkable examples of this burgeoning field, harness the enthusiasm and curiosity of citizen scientists to contribute to our understanding of the cosmos.

"Globe at Night" is an international citizen science project inviting global participants to measure and submit observations of their night sky's brightness. It aims to construct a global dataset of light pollution, a growing concern that affects astronomical observations and the natural world. This initiative provides a cost-effective method for obtaining comprehensive geospatial data on light pollution, with citizen scientists worldwide acting as "sensors" in their locales. The protocol is simple: Participants engage in a "star hunt" during moonless nights, recording the faintest star visible to them. These data are then submitted along with the date, time, and location, contributing to a worldwide light pollution map. From 2006 to 2019, this project accumulated over 190,000 data points with more than 200,000 measurements from 180 countries, providing a rich dataset utilized in numerous domains, including city ordinances, school science projects, and monitoring conditions near observatory sites. The participatory nature of the project encourages public awareness about light pollution, turning citizen scientists into advocates for darker skies [95].

Similarly innovative, the "Aurorasaurus" project crowdsources observations of the aurora (both positive and negative), providing real-time data that can assist in forecasting these extraordinary events. The utility of this initiative was demonstrated during a significant geomagnetic storm event when an unprecedented number of sightings were reported, illustrating the platform's potential for large-scale, real-time data collection on auroral activity. Like the "Globe at Night," this project serves dual purposes: collecting observations and fostering public understanding of auroras and space weather. The process of submitting observations involves detailing auroral activity, color, and height in the sky, often accompanied by a photograph. The platform further enriches its dataset by combing social media (e.g., Twitter) for likely aurora sightings [96].

Both "Globe at Night" and "Aurorasaurus" are emblematic of the potential of crowdsourcing in astronomical observations. By capitalizing on the enthusiasm of the public and the ubiquity of smart devices, these projects collect data on a scale that would be unfeasible through traditional means. Moreover, they turn every participant into an advocate for scientific understanding, fostering a deeper appreciation for the natural world. In this sense, they embody the essence of gamification described by Ahlqvist and Schlieder [97], transforming a mundane data collection exercise into an engaging and enjoyable activity. While this session has mainly focused on these two projects, the vast potential and applicability of crowdsourced astronomical observations is worth noting. The combination of citizen science and geospatial data collection represents a powerful tool for scientific discovery that is only beginning to be fully realized.

Geo-games and gamification

Gamification is an emerging strategy used by cities to promote, persuade, invite, engage, and educate people through game-based approaches (e.g., geo-games and geoplay) that include a geospatial element for data collection, validation, and analysis of geo-information. With minimal cost, the crowdsourcing strategy utilizes gamification approaches to carry out activities by incorporating game elements that provide meaningful results for further analysis. Ahlqvist and Schlieder [97] provide detailed elaboration on spatial gamification from both within and outside the realm of geo-information science, highlighting its applications in education, spatial planning, tourism, product marketing, and other areas. Augmented reality games or gamified apps have been developed to encourage environmental exploration or reward users for documenting specific environmental features. Major geospatial organizations, such as OSM, WikiMapia, DigitalGlobe-tomnod, GeoWiki, and Zooniverse, have reported the gamification of their work related to managing innovation processes over the past decade [97]. Several important features define gamification: first, the focus on fun as the primary element to ensure an enjoyable game experience; second, the emphasis on collective intelligence to avoid information imbalances, where players generate information explicitly through commenting or rating, as well as implicitly without realizing it; and third, the utilization of stable architectures and gamification systems that obtain geographic information through people's activities, such as check-ins, place registrations, and message postings.

Building on the three aforementioned features, several typical examples of geo-games have been developed and are increasingly popular in the field. For instance, SocialVenue [98] is a location sharing app designed to facilitate communication among users through location sharing features. Mag-ike (magic bike) is a biking game developed by a research team in Spain [99] with the aim of gathering crowdsourced commuting data. It employs a multi-cache approach, providing daily reports with accumulated scores and game status to help players improve their results. Gezgin is a geo-game application [100] developed to evaluate the benefits, values, and skills related to the global connections learning area of the social studies curriculum in Turkey. It is based on the four-component instructional design (4C/ID) model and incorporates expert opinions. NavApps, designed by Geotech, a research team in Germany [101], aims to raise awareness among high school students about their surrounding locations and educate citizens about existing services in smart cities, such as traffic conditions. Additionally, some games have been involved in evaluating the cultural and historical significance of cities. For example, Pokemon GO is primarily used to identify tangible attributes and values from textual descriptions, while Minecraft is a 3D block-building geogame developed for (re) designing buildings, cities, and landscapes [102]. These games have gained popularity and attracted a large online community of players who contribute to the creation and adaptation of worlds, fostering autonomy, 3D and spatial awareness, creativity, and social interactions. They serve as emerging approaches for crowdsourced data generation and collection.

Crowdsourcing Human Observations

In the intricate landscape of human experiences and societal complexities, the methodology of crowdsourcing human observations has emerged as a pivotal instrument for achieving unparalleled granularity and scalability in data collection. By facilitating the contribution of real-time information from individuals, this model is effecting transformative changes across diverse sectors, ranging from healthcare to public safety. This participatory framework extends beyond mere data accumulation to actively involve communities, thereby amplifying voices that might otherwise remain marginalized. In the forthcoming sessions, we will engage in a comprehensive exploration of the various dimensions of crowdsourcing human observations, examining its seminal impact on areas such as health and wellness, the optimization of transportation systems, and the real-time capture of public sentiments and opinions. Furthermore, we will discuss the utility of this approach in the arenas of disaster management and response, the enhancement of physical and social connectivity, as well as the fortification of public safety and security protocols. Each specialized session aims to furnish attendees with nuanced understandings of the interplay among technological, ethical, and societal considerations, thereby offering a balanced view of both the benefits and limitations inherent in leveraging collective intelligence. The structure of our review on crowdsourcing human observation is presented in Fig. 3.

Health and wellness

Crowdsourcing technology has played a pivotal role in advancing health and medical research, providing enormous opportunities to transcend geographical and organizational barriers faced by traditional research processes. One of the most notable benefits of this technology in addressing complex medical and public health issues is the ability to accelerate the collection of health-related data from a large number of individuals across various geographic locations and demographic groups [103]. Powered by electronic/mobile health (e/m-health) and wearable sensor technologies, a wealth of individual-level geospatial data has contributed to making groundbreaking discoveries that would otherwise be impossible due to the large number of research participants required for data collection [104,105]. For instance, Mappiness—a mobile application developed to collect large-scale geo-referenced data on subjective well-being—has allowed researchers to understand the effect of environmental aesthetics on happiness at a population level by capturing the spatiotemporal variability in happiness experienced in a wide range of environments [106]. The real-time nature of these data has been proven to significantly reduce recall bias since researchers no longer have to rely on people's recollections of their feelings and locations. Similarly, crowdsourced disease surveillance platforms that enable real-time geospatial data collection, such as Mosquito Alert and Flu Near You, have become an essential means for open collaboration between public health professionals and the general public to identify geographic hot spots of infectious disease and implement timely interventions [107,108].

Due to the principle of the self-selected sample in the crowdsourcing health data, however, particular attention should be paid to the characteristics of the population generating the data and the possibility of under- or over-representation of certain population groups [109]. Some researchers thus suggest that crowdsourcing approaches are best suited for studies in which rapid data collection from a large group of people is crucial but a representative sample is not necessary [110], such as Qin et al.'s [111] geo-crowdsourcing application developed to offer people with mobility or visual impairments real-time information on the locations of navigation obstacles. However, when executed carefully, crowdsourcing

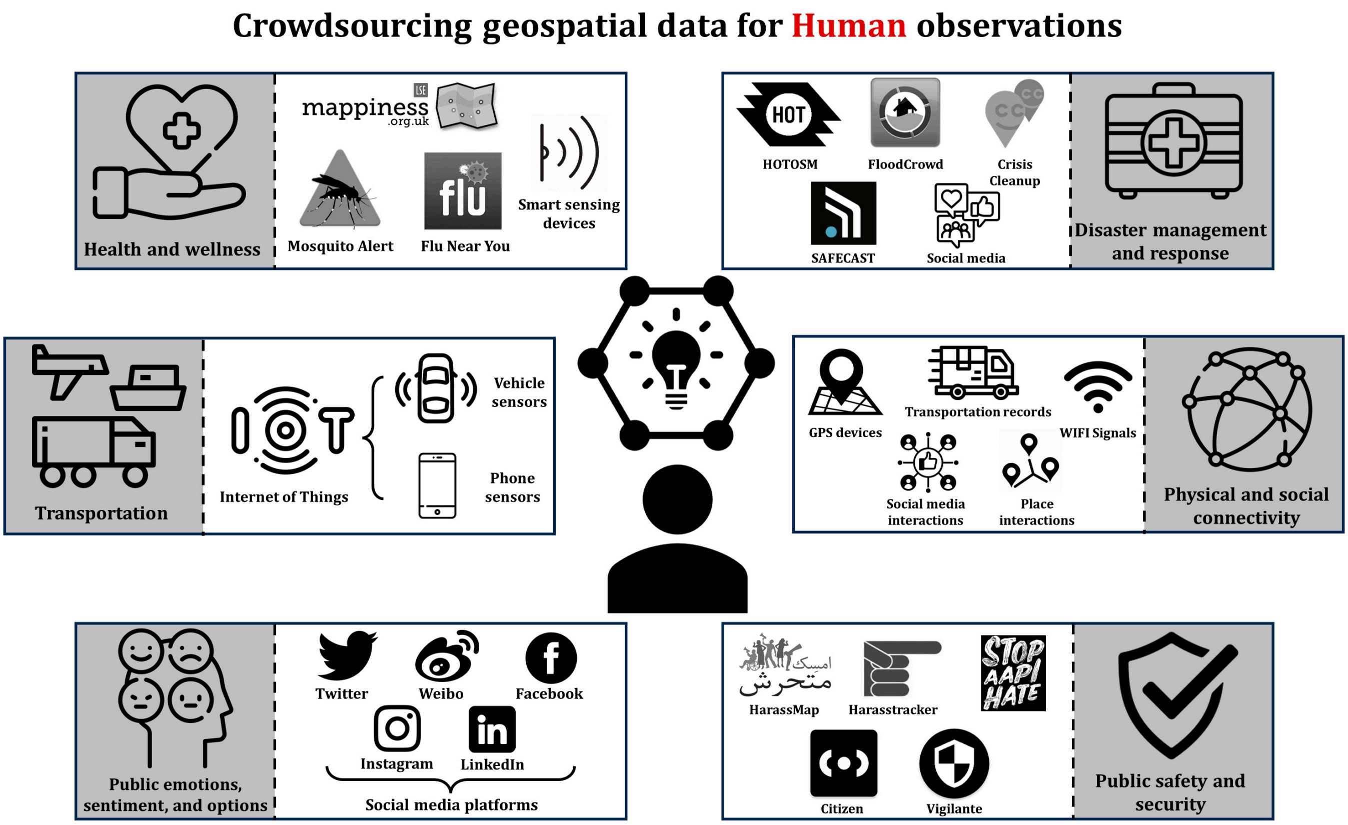

Fig. 3. Crowdsourcing geospatial data for human observations.

approaches hold significant potential for addressing health disparities by facilitating engagement with underrepresented communities. Park et al. [112] utilized GeoAir2, a portable air sensor that does not require technical proficiency in users or a local Wi-Fi network for data collection—which often hinder underserved groups' participation [113]—to ensure the inclusion of low-income immigrant communities in real-time air quality monitoring and empower them to take data-driven action. These examples demonstrate that crowdsourcing will remain increasingly a powerful tool for addressing public health challenges as e/m-health and participatory sensing technologies continue to evolve.

Transportation

Transportation management is rapidly evolving into a data-driven discipline. The integration and analysis of data have become essential for enhancing efficiency, safety, and decision-making in transportation systems [114,115]. The advancement of Internet of Things (IoT) technologies and the widespread use of mobile devices have facilitated the active contribution and passive collection of vast amounts of traffic information from various road users [116,117]. This has opened new possibilities for cost-efficient solutions in transportation monitoring and management. Crowdsourced observations have gained widespread adoption in different transportation management tasks [118,119,120]. This review will primarily introduce two key applications of crowdsourced data in transportation studies: traffic volume estimation and road safety assessment.

Traffic volume is a critical component of transportation planning and management. Traditionally, traffic volume data were collected through stationary sensors or manual surveys, which are not only expensive but also limited in their spatiotemporal coverage [121]. Many studies have shown that speed patterns on roadways are closely related to volume patterns, making them valuable for estimating the volume of road segments not covered by traffic sensors. Researchers can potentially estimate traffic volume for extended areas by leveraging speed patterns from crowdsourced Floating Car Data (FCD) [122,123]. FCD refers to real-time or near-real-time driving information collected from individual vehicles moving through the road network [122]. Common sources of FCD include phone-based navigation apps (e.g., Google Maps, Waze, and HERE WeGo), connected vehicles, probe vehicles, and car-sharing and ride-hailing platforms (e.g., Uber), providing valuable insights into vehicle movements. Apart from vehicle volumes, the number of active transportation users, such as pedestrians and cyclists, also plays a critical role in urban and transportation planning. However, there is often a lack of counts for nonmotorized traffic modes [124]. To address this, many mobile apps have been created, allowing users to track, report, and share their walking or cycling activities. One prominent data source in this regard is Strava, which anonymizes and aggregates data to help urban planners understand the volume of active activities, such as walking, running, and cycling, obtained from their registered users. By linking crowdsourced activity/user counts with additional features such as road inventory, built-environment characteristics, and sociodemographic data, researchers have proposed various models achieving satisfactory results in estimating volumes of different active transportation users [124,125].

Crowdsourced observations are also widely used in road safety studies. Traditionally, official crash records have been the primary data source for such studies. However, analyzing crash data is a retrospective approach, often requiring 3 to

5 years of crash data to obtain statistically reliable results. Moreover, using crash data alone may lead to an underestimation of traffic risk as many unreported crashes, incidents, and near-miss events are not captured [126]. To address these limitations, an increasing number of studies are exploring the potential of surrogate safety measures, such as traffic conflicts and abnormal driving behaviors, in road safety assessments (e.g., identifying traffic blackspots) instead of relying solely on crash data [122,127]. With the advent of mobile sensing techniques, a vast volume of hazardous driving behaviors (e.g., hard braking, fast acceleration, and frequent lane changes) can be detected using crowdsensing solutions through phone- or vehicle-based sensors. Different crowdsourced driving behaviors, such as hard braking [128], driving jerks [129], and speed variations [130], have been proven to be strongly correlated with crash risks. Additionally, people can also actively report traffic incidents through mobile apps. For instance, Waze, a leading crowdsourcing platform, can efficiently collect traffic incidents reported by its registered users, proving to be a valuable data source for road safety assessment. By combining Waze-captured incidents with historical crash records, a more comprehensive understanding of traffic risks can be achieved [131].

In addition to applications in traffic volume estimation and road safety assessment, crowdsourced data have also found successful use in various transportation management tasks, including traffic congestion detection [121], road surface assessment [119,132], transport asset management [133], and transportation planning [134]. These applications demonstrate the great potential of crowdsourced information in supporting the establishment of intelligent transport systems.

Public emotions, sentiments, and opinions

In comparison to traditional techniques such as surveys and questionnaires, crowdsourcing is particularly effective in studies that necessitate broad spatiotemporal coverage and involvement of large population sizes. One of the major arenas where crowdsourcing shines is in the extraction and interpretation of public emotions, sentiments, and opinions. The rapid advancements in natural language processing and deep learning techniques have substantially enhanced the application of crowdsourced data in sentiment analysis [135,136]. Social media, offering an extensive, continually updated, and diverse array of user-generated content, serves as a rich data source for sentiment analysis. Platforms like Twitter, Flickr, Weibo, and Facebook are particularly significant, as sentiment analysis techniques applied to data from these platforms have found extensive usage across various disciplines, providing invaluable insights that shape strategic decision-making [137,138,139,140].

The application of sentiment analysis using social media data spans across multiple areas. In healthcare, it helps in monitoring public attitudes toward healthcare policies, tracking disease outbreaks, and understanding the social and psychological impacts of various health conditions [141,142,143]. In politics, it is used for predicting election outcomes by assessing public sentiment toward general elections and users' reaction to political campaign [144,145,146]. Furthermore, the technique has been employed to analyze public perceptions of the urban built environment, thereby informing urban landscape planning [147]. In the field of marketing, sentiment analysis plays a crucial role in understanding consumers' or employers' sentiments toward products and brands, informing strategic marketing decisions [148,149].

Within the realm of public health, the application of sentiment analysis using crowdsourced data, particularly from social media, has been particularly impactful. For instance, Broniatowski et al. [150] leveraged Twitter data as a surveillance tool to concurrently monitor influenza cases and the related public reactions. Similarly, Ahmed et al. [151] adopted a similar approach during the 2009 H1N1 pandemic, using Twitter data to dissect public sentiment and responses. Further enhancing such approach, Müller and Salathé [152] introduced "Crowdbreaks," an open-source platform that streamlines this process with crowdsourced labeling, enabling efficient sentiment analysis of health trends in real time, thereby accelerating pace of research in the public health domain.

The recent worldwide coronavirus disease 2019 (COVID-19) pandemic saw a significant application of a similar approach in understanding public sentiment related to the disease and its respective vaccines. Several researchers, including but not limited to Ibrahim et al. [153], Hussain et al. [154], and Hussain et al. [155], have demonstrated the value of sentiment analysis using social media data. Ibrahim et al. [153] built a Hierarchical Twitter Sentiment Model to discern sentiment polarities within COVID-19-related tweets, while Hussain et al. [154] extended the use of artificial intelligence (AI) techniques to analyze over 300,000 social media posts from Facebook and Twitter about COVID-19 vaccines in the UK and the US. Complementing this, Hussain et al. [155] conducted a parallel analysis of posts discussing adverse effects following immunization, effectively underscoring the utility of social media analysis as a supportive mechanism in traditional pharmacovigilance. This broadened usage of crowdsourced data in sentiment analysis signifies its critical role in comprehending and navigating public sentiment during health-related crises.

Disaster management and response

The real-time nature of crowdsourced data enables rapid response to dynamic situations, mitigating the impacts of disasters. It leverages the collective intelligence and capabilities of diverse individuals, tapping into a wide range of knowledge and experiences to offer a holistic view of disaster situations. Importantly, crowdsourcing democratizes disaster management by empowering communities to contribute to response efforts, fostering resilience at a grassroots level.

A number of applications and tools have been developed to harness the power of crowdsourcing in disaster management and response. For example, Ushahidi is an open-source platform designed to crowdsource crisis information, visualizing these data on a map to provide situational awareness during disaster events. Similarly, HOT uses crowdsourcing to generate geographic data, aiding relief organizations in their operations. During the 2010 earthquake in Haiti, Ushahidi was used to collect and map incidents of collapsed buildings, trapped individuals, and medical emergencies. This allowed emergency services to prioritize their resources effectively [156]. Meanwhile, there was an immediate need for high-quality maps of the affected areas to aid in rescue and relief operations. The HOT used available satellite imagery to map the affected areas and a virtually blank map was transformed into a detailed spatial dataset in a short time. FloodCrowd is a UK-based citizen science project that crowdsources flood reports from the public. Using a simple online form, anyone can report flooding incidents, providing key details, such as the location, timing, and impact of the flood. The collected data are then used to improve the understanding and modeling of flood risks. Many other tools, such as Crisis

Cleanup and Safecast, also illustrate the transformative role of crowdsourcing in optimizing disaster management strategies and actions.

In addition to the tools above, researchers often use data from social media platforms (e.g., Twitter, Weibo, and Facebook) to gain situational awareness and improve disaster response [157]. The gathered data are analyzed using Geographical Information System (GIS) and natural language processing techniques to extract useful information such as sentiments, needs, and locations of affected individuals. For example, Ashktorab et al. [158] developed Tweedr, an architecture for collecting and analyzing Twitter data to identify actionable information for disaster response. Tweedr was shown to effectively mine disaster-related information from the vast amount of data on Twitter. De Albuquerque et al. [159] proposed an approach for integrating social media data with authoritative data for improved disaster management. The authors demonstrated that social media could provide timely, geographically diverse information that complements traditional data sources. Wang et al. [160] examined the use of social media in managing flood emergencies in urban areas, focusing specifically on the 2012 Beijing rainstorm. They used data from Weibo, a popular Chinese microblogging site, to highlight the potential of social media in real-time information dissemination and public participation in flood management. To improve the quality of data available to emergency responders, Tien Nguyen et al. [161] developed a deep learning model to automatically filter out irrelevant images from social media during crises. These studies illustrate the valuable insights that can be gleaned from social media data during disasters, from improving situational awareness to understanding public sentiment and aiding in disaster response coordination. However, these researchers also note the challenges, such as data validity and privacy concerns, that must be addressed when using such data.

Physical and social connectivity

The emerging "Web 2.0" and "Citizens as Sensors" represent a forward-thinking concept for leveraging the potential of crowdsourcing in the accumulation of digital imprints left by users of digital devices, capitalizing on the burgeoning trend of geopositioning technologies. Both passive and active data collection means have greatly facilitated our understanding of physical connectivity, quantified by human moving patterns. Passive data collection, for instance, comprises information acquired from sources such as mobile phone GPS [162,163], smart card transactions [164,165], and wireless networks [166,167]. The spatial connections originating from these passive traces often exhibit high degrees of representation, due to their broad data penetration ratios [168]. This prevalence, nonetheless, prompts significant apprehensions regarding privacy and confidentiality. An alternative that offers reduced intrusion and ameliorates privacy concerns incorporates spatial data gathered from social media platforms [169,170,171]. Given the active sharing characteristic inherent in these platforms, data extracted from social media sources are typically less abundant compared to passively collected GPS locations from mobile devices. The physical interlinkages derived from the abovementioned sources have been deciphered and employed for a myriad of purposes. These include transportation planning [172,173], disease modeling [174,175,176], identification of urban functional zones [177,178], disaster management [179,180], and marketing and business development [181,182].

Transitioning to social connectivity, crowdsourcing data present a rich reservoir for understanding and analyzing social interactions and patterns. One principal source of social connectivity data is online social media platforms, where user-generated content can provide significant insights into social behavior and community dynamics. These platforms inherently encourage interaction, engagement, and social sharing, resulting in a plethora of records that, when analyzed, reveal a complex network of social relationships [183]. An in-depth examination of engagement indicators, encompassing likes, shares, comments, and even the subtle nuances of language usage, offers a robust methodology for gauging sentiment [184], identifying social affiliations [185], and discovering shared interests [186]. Furthermore, the applicability of social media transcends the realm of direct social engagement. It serves as an invaluable tool in tracing the spread of information [187,188], monitoring societal attitudes [189,190], and understanding online communities [191]. The advent of location tagging introduces a new spatial facet to social connectivity, further enriching the depth and breadth of analytical possibilities.

Public safety and security

Traditional data that measure public safety and security are typically collected through government agencies, law enforcement organizations, research institutions, and other formal sources [192]. For example, crime reports that compile crimes documented various criminal activities, incidences requiring immediate assistance collected from emergency call centers, and injuries resulting from incidents related to public safety from hospital records [192]. Those types of data are typically well-established and reliable. However, the data may not be able to document dynamic changes in the cases and may not be able to be shared with the public promptly. The development of crowdsourcing data complements these limitations through a rapid and cost-effective data collection and sharing process.

Mobile phone applications are popular tools to collect and share crowdsourcing data, which can affect public safety and security. For example, Citizen app, an application that keeps users updated about nearby crimes, accidents, and emergencies in real time, sources information from police scanners and application user reports. Based on a recent online survey, 87% of the participants expressed their willingness to use the Citizen app to report accidents, crime, and corruption [193]. Researchers also find that the Citizen app generates earlier notifications in traumatic cardiac arrest compared with standard Emergency Medical Service radio communications [194]. The out-of-hospital information provided by the app may create a complementary source for the emergency department to make rapid resuscitative decisions for upcoming patients [194].

Crowdsourcing data can also be used to serve specific populations, for example, Stop AAPI Hate, a website that operates the nation's largest reporting center tracking acts of hate against Asian Americans and Pacific Islanders (AAPI). The website was initiated in 2020 due to the rising of AAPI hate during the COVID-19 pandemic. From 2020 March 19 to 2021 December 31, a total of 10,905 hate incidents against the AAPI community were reported through this website [195]. The data collected have been used in research and reports measuring the experience related to AAPI hate incidences [196,197,198]. The collection of crowdsourcing data combating public safety and security is a global effort. In Egypt, HarassMap was initiated to encourage people to report instances of sexual harassment via texting or internet reporting [199]. The reports are then plotted on a map, highlighting hotspots of such activities. This initiation has inspired people in other countries to establish similar systems, for example, Safe City in India, Harasstracker in Lebanon, and Biyoya in Bangladesh.

Although crowdsourcing has the benefit of collecting timely data through novel and cost-effective approaches, it is still subject to numerous concerns and challenges. For example, the information collected might not be reliable as some of the information comes from unverified sources. Additionally, those systems might increase people's anxiety about public safety [200]. Participants who reported using neighborhood apps perceived local crime rates as higher than those who do not use the apps, independent of actual crime rates [201]. For example, the Citizen app mentioned above was originally released as Vigilante, a banned application in 2017, which encouraged users to develop a vigilante-style network to protect themselves from potential offenders before the police needed to intervene. The app has the potential to incite violence and put innocent people in danger. Citizens did not seem to have learned from previous experience but rather created a dangerous effort to seek an arsonist of a wildfire in Los Angeles' Pacific Palisades neighborhood through app users [202]. Thus, regulation is needed to use crowdsourcing efforts legally and ethically to protect public safety and security.

Challenges in Crowdsourcing Earth and Human Observations

Data quality and accuracy

Crowdsourced data can vary significantly in quality and accuracy. Contributors might have different levels of expertise, commitment, and access to high-quality recording devices. Data validation and quality control processes are essential but can be complex and resource-intensive [203].

Data quality is of paramount importance when discussing crowdsourced geoinformation data. This is because contributors vary in their levels of expertise, dedication, and access to high-quality recording devices. Furthermore, due to the anonymity of crowdsourcing platforms, there is an inherent risk of vandalism [204]. Prior to utilizing crowdsourced data in experiments, applications, or projects, stakeholders typically seek to understand the quality of the data to a certain degree. Nevertheless, there are persistent challenges from various perspectives regarding data quality.

On the one hand, establishing appropriate criteria for quality assessment is challenging due to the evolving nature of crowdsourced data and the different application-specific requirements in terms of input data quality. These data might introduce novel types of information and geographic features that are not present in authoritative databases, rendering the evaluation of such new data difficult. Existing research has proposed quality criteria for crowdsourced data, which often encompass geometric, temporal, and positional accuracy, data completeness, logical consistency, and fitness for purpose [205]. However, delineating precise thresholds to categorize quality—such as distinguishing between high, medium, or low quality—is problematic, given the varying perceptions of quality across different domains.

On the other hand, executing quality assessments presents its own set of challenges. In many instances, reference data (typically sourced from authoritative or commercial databases) may not be readily available due to the prohibitive costs associated with acquisition, especially at a large scale. Even when such reference data are accessible, its utility in quality assessments is often limited due to its slower update frequency relative to crowdsourced data. While some might advocate for intrinsic quality assessment methods [206], the absence of standardized quality criteria and associated methodologies introduces ambiguity into the assessment process.

In addition, new kinds of crowdsourced data may have their own set of particularities that render currently established data quality assessment procedures and standards insufficient, presenting another challenge. For example, researchers have identified that crowdsourced street-level imagery (see the "Urban planning and infrastructure" section) has a number of quality aspects that have not been foreseen in existing approaches to gauge the quality of crowdsourced geographic information and have been working on establishing a quality assessment framework that is tailored for such form of data [207].

Data bias

Crowdsourced geospatial data inherently contain biases as it relies on voluntary contributions, leading to less credible inferences compared to conclusions drawn from a randomly sampled population [208]. Biases can affect the reliability, representativeness, and usability of the data derived from various factors and sources. Common categories of biases encompass spatial biases, temporal biases, demographic biases, cognitive biases, and systematic biases [208,209,210,211,212]. Spatial biases in crowdsourced geospatial data occur as a result of the unequal geographical distribution of contributors or specific local characteristics of crowdsourcing tasks, causing certain regions or places to receive higher contributions due to factors such as popularity, population density, internet accessibility, and task localization [213,214,215,216]. Temporal biases can arise when crowdsourcing tasks are limited to specific events or time frames, potentially distorting the comprehensive understanding of individuals' characteristics, behaviors, or opinions over time and resulting in biased insights or conclusions [217,218].

Beyond spatial and temporal considerations, demographic biases play a substantial role in influencing the representativeness of crowdsourcing participants concerning age, gender, education, socioeconomic status, culture, and other demographic attributes [212,219,220,221]. These biases—often overrepresentation—tend to be more prevalent among young, male, well-educated, technologically literate, and affluent segments of the population [208,212,213,217,222,223], resulting in a lack of diversity and inclusivity in the participant pool and subsequently leading to an imbalance in perspectives and experiences. Additionally, other inherent biases, such as cognitive biases, derive from limitations in human cognitive processes, and systematic biases arise due to flaws in the data collection process, study design, or analysis methods. Both biases require attention when assessing the quality and representation of crowdsourcing geospatial data [224,225]. They can manifest in varying ways, impacting the responses of participants and the decision-making of researchers and decision-makers.

Different strategies have been explored to effectively mitigate and address these data bias challenges in crowdsourcing geospatial data: first, performing data preprocessing, reweighting, or sensitivity analysis to reduce the impact of biases on downstream analysis or modelling [211,219]; second, combining crowdsourced data with authoritative geospatial sources to gain a more balanced view of the data [168,208,226]; third, designing tasks with clear guidelines and varying perspective considerations, such as relying on long-term trends and unbiased statements or questions [211,224]; fourth, encouraging a broad range of contributors with different demographic backgrounds from different regions for inclusivity and diversity [220,227]; finally, implementing quality control measures during the crowdsourcing process can help identify and filter out biased responses [225,228]. This can involve prescreening workers, incorporating validation questions, or using redundancy to compare multiple worker responses.

Data privacy

Apart from inherent data biases, the use of crowdsourced data, particularly from social media or mobile devices, can raise data privacy concerns. One of the primary data privacy issues is location privacy. Geospatial information often reveals precise details about individuals' whereabouts, activities, behaviors, or even their home addresses. Without adequate safeguards, such data could be exploited by malicious actors to track or identify individuals, leading to potential risks related to personal safety and security [229,230]. Furthermore, crowdsourced geospatial data might inadvertently contain personal identifiers, such as usernames or profile information, which could lead to the re-identification of individuals. Studies have shown that it is still possible to re-identify individuals through cross-referencing with external datasets [230,231,232]. This poses a significant challenge as it compromises the anonymity of contributors and exposes them to potential privacy breaches. Additionally, third-party access and sharing are also data privacy challenges of crowdsourcing platforms, such as research or commercial use. This raises concerns about how these entities handle the data and whether they adhere to privacy regulations. The lack of explicit consent for data sharing may lead to unexpected data usage, emphasizing the need for transparent data sharing policies and stringent agreements with third-party partners [233,234].

It is crucial for organizations and researchers to implement robust privacy measures, uphold ethical standards, and comply with relevant data protection regulations to ensure the confidentiality and security of crowdsourced geospatial data while maximizing its potential for beneficial insights. Various techniques have been proposed to avoid violating user privacy, aiming to strike a balance between data utility and individual privacy, ensuring responsible data usage. Anonymization, aggregation, and privacy-preserving methodologies are essential strategies to mitigate location privacy risks [231,235,236], specifically anonymizing the geospatial data by removing direct identifiers, such as names and direct details, aggregating data at higher spatial or temporal resolutions to enhance privacy by obscuring specific locations, and implementing clear and transparent data sharing policies by explicating the consent mechanism for contributors. To help protect the identity of participants and minimize the risk of re-identification, techniques like pseudonymization and data encryption can be employed by enabling secure computation on encrypted geospatial data [229,233]. Secure multi-party computation is also a promising approach that allows multiple parties to jointly analyze data without sharing raw information [237].

Legal and ethical issues

Using crowdsourced data inevitably raises legal and ethical issues due to the nature of the data, which is typically contributed by a diverse group of individuals. These issues include concerns about data ownership, intellectual property rights, and liability, which have been extensively discussed in the literature on citizen science [238]. Unlike traditional data sources where ownership is often more straightforward, determining the ownership rights of crowdsourced data can be complex since there are usually no clear guidelines or agreements regarding ownership and usage rights. This ambiguity gives rise to concerns regarding privacy, intellectual property, and the potential exploitation of contributors' data. Data scientists must navigate this ethical and legal landscape by establishing transparent protocols, consent mechanisms, and fair compensation approaches to ensure that the rights of contributors are protected while harnessing the full potential of crowdsourced data for valuable insights and innovation.

More specifically, crowdsourced Earth observation data, such as OSM, require users to attribute the source of the data and share any derivative works under the same license [239]. It is crucial to understand and comply with the licensing terms when using OSM data to avoid legal repercussions. OSM has a strong community that follows specific guidelines and norms. Ethical considerations involve respecting the principles of the OSM community, such as refraining from vandalizing or misrepresenting data, giving proper credit to contributors, and collaborating with the community to improve the dataset. In contrast, the legal and ethical issues surrounding crowdsourced human observation data, such as tweets, are even more significant. There is a growing need for universal guidelines addressing the ethics of social media research, particularly concerning the privacy and anonymity of social media users. Although social media data are often claimed to be anonymized, sharing such data via public repositories and platforms should involve discussions on obtaining consent and/or ethical approval for research purposes [240]. This is especially crucial for datasets containing user profile information, as these datasets can be potentially identifiable through cross-referencing data attributes [10]. While adhering to data sharing regulations and the principles of reproducibility, it is important to approach the sharing of processed social media data via public repositories and platforms with caution and establish reproducible workflows that can be utilized by end-users without a coding background.

Sustainability of data collection

The sustainability of data collection in crowdsourcing initiatives can be a significant challenge, because it heavily relies on the active participation and engagement of volunteers. As noted by Newman et al. [241], the sustainability of such efforts can be compromised when volunteers lose interest, leading to a drop in data input, and consequently affecting the efficacy and validity of the gathered data.

Moreover, sustainability is also influenced by the nature of the community driving the project. The absence of ongoing community involvement can contribute to this dwindling interest, leading to sporadic data collection that lacks consistency and continuity. According to Starbird and Palen [242], the effectiveness and sustainability of crowdsourcing efforts, especially during crisis situations, can be significantly improved with the presence of dedicated coordinators who can motivate volunteers, manage and direct efforts, and ensure that data collection continues in an organized and systematic manner.

Institutional support can also play a crucial role in the long-term sustainability of crowdsourcing initiatives. With the necessary resources and funding, institutions can maintain motivation and engagement among volunteers through incentives, training, and recognition of efforts. This can ensure the continued flow of data and enhance the sustainability of the project over time [243].

Therefore, sustainable crowdsourcing efforts, especially in terms of data collection, need strategic planning, community engagement, and strong institutional support. These factors can ensure the continuation of volunteer participation and data collection in prolonged periods, thus ensuring the effectiveness of crowdsourcing initiatives.

Data interpretation

The process of interpreting crowdsourced data, especially when employed for scientific research, is fraught with complexities attributable to the diverse nature of the data and the potential dearth of metadata. The task of extracting consequential insights from crowdsourced data, as applied to Earth and human observations, is underscored by the substantial challenge of data interpretation. Despite the surge in accessible information (e.g., OSM) and the evolution of sophisticated tools designed to manage these data (e.g., OSM Analytics Tool), the endeavor of unraveling the salient meaning and implicit subtleties within the data poses a demanding task.

The process of interpreting crowdsourced data necessitates a meticulous traverse through a multifarious landscape of informational noise [244,245]. Data acquisition, a composite process entailing the collection from an extensive range of sources, each varying in their level of expertise, precision, and consistency, often culminates in datasets marked by heightened complexity and diversity. This inherent heterogeneity, while advantageous to crowdsourcing, amplifies the task of isolating accurate, germane signals amidst an expanse of potentially discordant or erroneous data. The task of data interpretation is further intensified by the intrinsic subjectivity associated with human observations. This set of data, commonly influenced by personal biases [246], perceptual variations [247], and undulating levels of comprehension and expressive proficiency among contributors [66], can exert considerable influence over the final output. The dearth of a robust system to temper these variables could elevate the likelihood of data misinterpretation, potentially leading to skewed deductions and misplaced strategic decisions, thereby emphasizing the need for rigorous analytical approaches in the scientific processing and interpretation of crowdsourced data.

Addressing spatial and temporal variations is a critical aspect in the interpretation of crowdsourced data, particularly for earth and human observations. There can be notable fluctuations in the quality and frequency of data across distinct geographical areas and over varying time periods, thereby presenting significant hurdles in synthesizing a holistic and globally representative interpretation [248,249]. These inconsistencies mandate thorough attention and the employment of advanced analytical methodologies to enable trustworthy interpretations. Moreover, the absence of uniform protocols for data validation and verification intensifies the complexities involved in data interpretation [104]. Yet, the formulation and execution of such protocols pose significant challenges, especially considering the characteristically decentralized and often anonymized nature of crowdsourcing initiatives.

Responding to these challenges necessitates the adoption of inventive and rigorous methodologies for data management, analysis, and interpretation. We argue that emphasis should be placed on the evolution of more advanced machine learning algorithms, capable of filtering and standardizing crowdsourced data. This should occur in tandem with the application of robust statistical approaches designed to address inherent biases and discrepancies within the data, aiming to rectify any embedded biases and discrepancies within the data, thereby ensuring that subsequent interpretations of the data retain their validity and accuracy.

Training and education

In the realm of crowdsourced data collection, it is crucial to provide proper training and guidelines to volunteers to ensure the quality and consistency of the collected data. Data collectors, who are often volunteers, play a significant role in crowdsourcing initiatives by contributing their time and efforts to gather valuable information. However, without adequate training, the data collected may vary widely in terms of accuracy, completeness, and adherence to predefined standards. Data scientists should establish comprehensive training programs that equip volunteers with the necessary skills, knowledge, and understanding of the data collection process. This includes educating them about specific data requirements, providing clear instructions on data collection techniques, and familiarizing them with any relevant tools or technologies. An emerging trend in this domain is the use of robots to cope with data processing, such as the development of Roboturk, a crowdsourcing platform for robotic skill learning through imitation [250]. However, it should be noted that training robots involves different requirements and infrastructure compared to training human workers. Regardless of the subjects involved in data collection and manipulation, data scientists should be prepared for unexpected outcomes, as design choices in data collection can have a significant impact on the quality of crowdsourced user-generated content [251].

To achieve better results of training and education in data collection and manipulation, several key steps can be considered to ensure their effectiveness and the quality of the collected information. First, we need to begin by clearly defining the objectives and requirements of the data collection project. This includes specifying the type of data needed, the desired format, and any specific guidelines or standards to be followed. Second, there is a need to create comprehensive training materials that cover all aspects of the data collection process. These materials should be accessible, be easy to understand, and provide step-by-step instructions, including visual aids, examples, and real-world scenarios to facilitate learning. Third, it is important to offer opportunities for volunteers to gain hands-on experience by conducting practice data collection exercises. This can be done through simulated scenarios or by providing sample datasets for practice. Volunteers are encouraged to seek feedback and address any questions or concerns they may have during this practice phase. It is crucial to organize training sessions where volunteers can learn directly from data experts or experienced team members. These sessions can be conducted in person, through webinars, or using online platforms. Fourth, the introduction of data quality control measures is necessary to ensure the reliability and consistency of the collected data. This can involve periodic reviews, validation checks, or random audits of the data collected by volunteers. Meanwhile, we should provide feedback and constructive suggestions to help volunteers improve their data collection techniques. Fifth, it needs to have a supportive and collaborative environment where volunteers can share their experiences, ask questions, and learn from one another, through establishing communication channels, such as discussion forums or chat groups, where volunteers can interact and seek guidance from data experts or project coordinators. Finally, offering regular training and support throughout the data collection process could ensure that volunteers receive regular updates and refresh sessions, and address any issues or challenges that arise during the process. By following these steps, data scientists can effectively train volunteers in collecting crowdsourced data, ensuring a high level of quality, consistency, and adherence to project requirements.

Future Directions and Pathways

Harnessing the power of the crowd: Expanding the scope of geospatial crowdsourcing

Navigating the evolving landscape of crowdsourcing geospatial data collection and analysis reveals transformative perspectives. These include harnessing the temporal dimension, leveraging advanced AI and machine learning, integrating IoT technologies with crowdsourcing, and prioritizing inclusivity, particularly from underrepresented regions such as the Global South. We believe that the amalgamation of these insights is poised to significantly reshape our methodologies, enriching our understanding of the world through a comprehensive and representative approach to geospatial crowdsourcing. We illustrate these four perspectives in detail below.

Embracing the fourth dimension

Presently, a significant portion of crowdsourcing initiatives in geospatial data accumulation predominantly concentrates on static data. Nevertheless, prospective endeavors possess the capability to transcend conventional limitations by integrating the fourth dimension: time. By assimilating this temporal aspect more proficiently, geospatial crowdsourcing can expedite real-time or near-real-time data collation and evaluation. This dynamic strategy has the potential to enhance our competency in cultivating a more exhaustive and nuanced comprehension of our environmental milieu. Furthermore, it capacitates timely reactions to emerging circumstances and challenges, thereby fostering more informed decision-making processes and proactive initiatives. The incorporation of this temporal facet into geospatial crowdsourcing broadens the spectrum of potentialities and empowers us to harness the collective intelligence of the masses to stimulate consequential and impactful results.

Deepening the wisdom of crowds

The intensification of collective intelligence in geospatial crowdsourcing signifies a compelling venture to exploit avant-garde AI and machine learning methodologies. Utilization of these state-of-the-art technologies empowers the extraction of more intricate and sophisticated insights from the amassed data, thereby augmenting traditional analytical frameworks. AI and machine learning algorithms harbor the capacity to reveal latent patterns, associations, and tendencies inherent in geospatial data, facilitating an enhanced comprehension of our environment. These methodologies can supplement human potentialities by processing extensive quantities of data with expedience and efficiency, discerning intricate spatiotemporal patterns, and offering predictive analytics. By amplifying collective intelligence through AI and machine learning, we can unfetter unprecedented layers of comprehension and catalyze innovative solutions in geospatial analysis. Ultimately, this contributes to the refinement of decision-making processes and promotes sustainable development.

Seamless integration of IoT and crowdsourcing

The advent of the IoT offers an extraordinary opportunity for seamless integration with crowdsourcing initiatives. There lies tremendous potential in amalgamating sensor data emanating from diverse sources with crowdsourced information to furnish a more enriched, comprehensive depiction of our planet and human perceptions. This multifaceted integration not only optimizes the capacity to acquire extensive datasets but also enhances the depth of analysis by incorporating the vastness of sensor-based IoT data. This convergence of technologies empowers us to derive a finer granularity of insights and, ultimately, a more robust understanding of the patterns and processes shaping our world. Consequently, the synthesis of IoT and crowdsourcing technologies signifies an innovative stride toward more comprehensive and informed decision-making, fostering a proactive approach in our interactions with the environment.

Encouraging citizen science in the global south

Momentous efforts need to be marshaled to invigorate participation from areas that are currently underrepresented, particularly the Global South, within the sphere of crowdsourcing sciences. The adoption of an inclusive strategy for data collection propagates the cultivation of a more balanced, representative, and comprehensive database. This approach ensures the capture of diverse perspectives, thereby enriching our comprehension of multifarious geospatial phenomena. The proactive integration of these regions provides a crucial conduit to bridge extant data voids while fostering knowledge sharing and capacity development. Moreover, it engenders a sense of communal responsibility and global collaboration directed toward understanding and mitigating shared challenges. Hence, we believe that the advancement of Citizen Science in the Global South marks a vital stride toward shaping a more equitable and insightful scientific terrain, profoundly contributing to the enhancement and inclusivity of our global data reservoir.

Pioneering a sustainable crowdsourcing ecosystem: From motivation to retention

In the contemporary digital landscape, the opportunity has emerged for citizens to significantly contribute to scientific advancements via crowdsourcing. For this potent instrument to realize its full potential and to make it sustainable, it is imperative to fortify several foundational elements. This entails constructing a unified community of dedicated citizen scientists and crafting incentives that optimally balance motivation with genuine engagement. Equally important is the commitment to inclusivity, ensuring that technological progress does not inadvertently result in disparities or omit specific groups. Central to this endeavor is comprehensive education, which guarantees that participants are not only adept at their tasks but also cognizant of the wider ramifications of their input. We illustrate these four perspectives in detail below.

Building a robust community of citizen scientists

Developing strong communities around these efforts can improve the long-term sustainability of data. Cultivating a robust community of engaged citizen scientists is imperative for the longevity of crowdsourcing initiatives [247]. Projects that foster a sense of collective purpose and belonging can promote prolonged contributions from volunteers [252]. For instance, eBird's passionate birder community and discussion forums create social incentives that sustain participation [253]. Effective community-building entails establishing open communication channels, providing mentorship opportunities, and encouraging a participatory culture where volunteers feel valued in the scientific process [254]. Decentralizing leadership and facilitating collaborations via workshops and events also strengthens communal bonds [255]. Modular and personalized training resources further enhance sustainability by enabling volunteers to develop relevant skills while recognizing their contributions' significance [256]. For example, CitSci.org's adaptive courses on gathering field data provide tailored learning pathways based on needs and schedules, ensuring broad accessibility.

Incentivizing participation

Besides intrinsic motivations, crowdsourcing projects should explore supplementary incentives for attracting and retaining contributors [257]. These could include reputational rewards like leaderboards, milestone badges, and opportunities for public recognition [258]. More tangible benefits may include discounts on project merchandise, premium account features, or prize give-aways for active participants [259]. However, caution is necessary to avoid over-gamifying participation or introducing disproportionate incentives that skew data [253]. The SciStarter Project Finder illustrates how participants can be incentivized via different benefit categories (e.g., career development and social engagement), displayed transparently alongside each project [260].

Bridging the digital divide

Bridging digital divides is also critical for pioneering an inclusive crowdsourcing ecosystem, as technological and socioeconomic barriers can perpetuate representation gaps [261]. For example, community-driven monitoring of local air quality using low-cost sensors revealed participation discrepancies along socioeconomic lines [262]. Targeted outreach, infrastructure development, and offline participation options can help engage marginalized communities [263]. LOCALE facilitates neighborhood-level data collection by providing local access to equipment and training [262]. Text and telephone reporting systems also expand access, as exemplified by Mosquito Alert's multichannel disease surveillance [264]. Ensuring wide accessibility promotes representative data inputs unconstrained by demographic factors.

Education and training initiatives

Lastly, comprehensive education and training initiatives raise awareness of crowdsourcing's significance while empowering quality contributions [265]. Interactive workshops with field components enhance skills and data literacy for diverse audiences from students to policymakers [266]. For example, Public Lab's community events build capacity for using low-cost tools for environmental monitoring through hands-on learning [267]. Online resources like tutorial videos, customized teaching modules, and webinars enable self-paced learning. Knowledge exchange forums allow participants to learn from each other [268], as exemplified by the Cornell Bird Academy fostering an educative birder community [265]. By imparting skills and communicating larger purposes, robust education sustains crowdsourcing participation while benefiting society.

From data to action: Translating crowdsourced geospatial data into real-world impact

There are several pathways toward translating the analytical results generated by crowdsourced geospatial data into real-world impact. First, crowdsourced data play a vital role in informing policy decisions and driving policy changes, particularly in the domains of environmental [262], health [203], and urban planning policies [269]. This role has become increasingly important, especially following the outbreak of the COVID-19 pandemic [10]. Additionally, crowdsourced data have immense potential in advancing scientific research and enabling scientists to gather data at scales and resolutions that were previously unattainable. The abundance of crowdsourced earth observation data (e.g., OSM and Mapillary) and human observation data (e.g., sentiment measures derived from social media) with extensive temporal and spatial coverage facilitates the availability of global or nationwide time-series studies [190,221,248]. The analytical results, encompassing large spatial and temporal coverage, provide evidence that can be compared across countries and regions, offering policy implications for governments at various levels and international organizations.

Second, crowdsourced data contributed by individuals represent the intentions, ideas, and behavioral tendencies of the general public, often referred to as the "silent voice," aiming to raise public awareness and encourage participation in citizen science. In this sense, crowdsourced data promote a broader understanding of people's awareness regarding public health crises (e.g., COVID-19 and vaccination), environmental changes (e.g., natural hazards), and post-pandemic economic recovery [190,270,271]. Specifically, prior to the occurrence of these events and crises, crowdsourced data can enhance emergency preparedness and response through early warning systems, improved resource allocation, and better coordination on the ground. After these disasters and crises, crowdsourced data have the potential to facilitate real-time action through data mapping and monitoring, such as crisis mapping during disasters [272] or real-time air quality monitoring for public health advisories [273]. Furthermore, industries and businesses also incorporate crowdsourced geospatial data into their strategies to gain business insights and support decisions related to market analysis [274], product development [275], and logistics planning [276].

Third, due to the aforementioned advantages of crowdsourced data, it has wide-ranging benefits and implications in empowering communities and individuals to advocate for their needs and protect their rights [277]. It also facilitates global collaboration to address global challenges, such as climate change or pandemic tracking, which cannot be effectively tackled by traditional survey data or other types of small data. In the realm of urban planning and governance, crowdsourced data enable urban planners and government officials to make informed decisions regarding city development, transportation networks, and public infrastructure. It also enables the timely collection of feedback and suggestions from the general public through e-participation and e-governance channels [278]. The aforementioned benefits associated with the use of crowdsourced data can be further extended through the development of tools and platforms that not only facilitate data collection but also make the data accessible and usable for decision-makers, communities, and individuals. This, in turn, helps bridge the gap between data and action, aligning with the goals of smart city initiatives and citizen science, which aim to create inclusive cities with an improved quality of life and increased socioeconomic performance through data-driven approaches, intelligent resource management, and participatory governance [279].

Conclusion

In this comprehensive review, we have dissected the multifaceted realm of crowdsourced geospatial data, illuminating its myriad applications, inherent challenges, and expansive potential in both human and Earth observations. Our exploration traverses the diverse domains of application, analyzes the nature and contributions of the data, and examines current data collection paradigms. In doing so, we map the present landscape of this burgeoning field and chart strategic directions essential for steering future research and applications across varied sectors.

The integration of time-sensitive data collection, AI, and IoT within geospatial crowdsourcing, coupled with an inclusive approach that encompasses underrepresented communities, fosters a detailed, real-time understanding of Earth's dynamics and human experiences, supported by a strong network of contributors. The emphasis on the collection of time-sensitive data allows for the attainment of enhanced, real-time socioenvironmental insights. Furthermore, the integration of AI and machine learning technologies holds the promise of revealing more intricate patterns and understandings within these accumulated data. The incorporation of IoT innovations in conjunction with crowdsourcing methodologies yields a more detailed and holistic understanding of environments and societal interactions. It is critically important to include a broad range of perspectives, particularly from typically underrepresented communities, in these initiatives. This inclusivity not only broadens the scope and depth of the data gathered but also guarantees a representation that is truly global in scale. To maintain the viability and effectiveness of geospatial crowdsourcing, it is vital to cultivate a strong network of citizen scientists, incentivize participation effectively, and address technological disparities. This endeavor requires comprehensive educational initiatives and training programs that adequately prepare participants, thereby equipping them with the necessary skills and knowledge.

The exceptional possibilities offered by crowdsourced geospatial data in reshaping information environments are simultaneously promising and complex. This calls for our focus not only on its extraordinary potential but also on addressing its inherent, multifaceted challenges, necessitating a collaborative and interdisciplinary strategy for effective solutions. With an eye on real-world applicability, we aspire for this review to serve as a foundational reference, guiding both scholarly and pragmatic pathways in upcoming explorations and applications within this evolving field.

Acknowledgments

Funding: This research is supported by the Faculty Startup Fund of the College of Arts and Sciences at Emory University.
Author contributions: X.H. and S.W. conceptualized this work and designed the article structure. X.H., S.W., D.Y., T.H., M.C., M.Z., G.Z., F.B., T.L., L.Z., C.Y.W., Y.M.P., X.L., Y.L., and H.F. contributed to the writing of the manuscript. J.M., Z.L., and A.H. helped review and revise the manuscript.
Competing interests: The authors declare that they have no competing interests.

Data Availability

This is a review article. In its preparation, we did not generate or analyze any new datasets. Instead, this article synthesizes and discusses information from previously published literature. Therefore, there are no data repositories or datasets associated with this work. For further information, references cited in the text are provided to guide readers to the original sources of the reviewed materials.

References

1. O'reilly T. What is web 2.0: Design patterns and business models for the next generation of software. *Commun Strateg*. 2007;1:17.
2. McAfee A, Brynjolfsson E. Big data: The management revolution. *Harv Bus Rev*. 2012;90(10):60–66.
3. Heipke C. Crowdsourcing geospatial data. *ISPRS J Photogramm Remote Sens*. 2010;65(6):550–557.
4. Li G, Zheng Y, Fan J, Wang J, Cheng R. Crowdsourced data management: Overview and challenges. Paper presented at: Proceedings of the 2017 ACM international conference on Management of Data; 2017 May 14–19; Illinois, Chicago, USA.
5. Turner A. *Introduction to neogeography.* Sebastopol (CA): O'Reilly Media Publisher; 2006. p. 54.
6. Goodchild MF. Citizens as sensors: The world of volunteered geography. *GeoJournal*. 2007;69(4):211–221.
7. Strasser B, Baudry J, Mahr D, Sanchez G, Tancoigne E. "Citizen Science"? Rethinking science and public participation. *Sci Technol Stud*. 2019;32(2):52–76.
8. Doyle C, David R, Li Y, Luczak-Roesch M, Anderson D, Pierson CM. Using the web for science in the classroom: Online citizen science participation in teaching and learning. Paper presented at: Proceedings of the 10th ACM conference on web science; 2019 Jun 30–Jul 3; Massachusetts, Boston, USA.
9. Owuor I, Hochmair HH. An overview of social media apps and their potential role in geospatial research. *ISPRS Int J Geo Inf*. 2020;9(9):526.
10. Huang X, Wang S, Zhang M, Hu T, Hohl A, She B, Gong X, Li J, Liu X, Gruebner O, et al. Social media mining under the COVID-19 context: Progress, challenges, and opportunities. *Int J Appl Earth Obs Geoinf*. 2022;113:102967.
11. Mooney P, Minghini M. A review of OpenStreetMap data. In: *Mapping and the citizen sensor*. London (UK): Ubiquity Press; 2017. p. 37–59.
12. Zhao Y, Han Q. Spatial crowdsourcing: Current state and future directions. *IEEE Commun Mag*. 2016;54(7):102–107.
13. Kamel Boulos MN, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, Chuang KYS. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. *Int J Health Geogr*. 2011;10:67.
14. Poblet, M., García-Cuesta, E., Casanovas, P. Crowdsourcing tools for disaster management: A review of platforms and methods. In: *International Workshop on AI Approaches to the Complexity of Legal Systems*. Berlin, Heidelberg: Springer (Berlin Heidelberg); 2013.
15. Wazny K. Applications of crowdsourcing in health: An overview. *J Glob Health*. 2018;8(1):010502.
16. Saralioglu E, Gungor O. Crowdsourcing in remote sensing: A review of applications and future directions. *IEEE Geosci Remote Sens Mag*. 2020;8(4):89–110.
17. Niu H, Silva EA. Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. *J Urban Plan Dev*. 2020;146(2):04020007.
18. Blaser L. Old weather: Approaching collections from a different angle. In: *Crowdsourcing our cultural heritage*. New York (NY): Routledge; 2016. p. 45–56.

19. Knapp KR, Matthews JL, Kossin JP, Hennon CC. Identification of tropical cyclone storm types using crowdsourcing. *Mon Weather Rev*. 2016;144(10):3783–3798.
20. Reges HW, Doesken N, Turner J, Newman N, Bergantino A, Schwalbe Z. CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bull Am Meteorol Soc*. 2016;97(10):1831–1846.
21. Crimmins TM, Crimmins MA. Large-scale citizen science programs can support ecological and climate change assessments. *Environ Res Lett*. 2022;17(6):065011.
22. Liu BF, Seate AA, Iles I, Herovic E. Eyes of the storm: How citizen scientists contribute to government forecasting and risk communication. *Weather Clim Soc*. 2020;12(2):263–277.
23. Martín Y, Cívica M, Pham E. Constructing a supercell database in Spain using publicly available two-dimensional radar images and citizen science. *Ann Am Assoc Geogr*. 2021;111(5):1346–1366.
24. Abel F, Hauff C, Houben GJ, Stronkman R, Tao K. Twitcident: Fighting fire with information from social web streams. Paper presented at: Proceedings of the 21st International Conference on World Wide Web; 2012 Apr 16–20; Lyan, France.
25. EDN. Bluetooth temperature sensor works with smart phones. [accessed 13 December 2013] http://www.edn.com/electronics-products/other/4412268/Bluetooth-temperature-sensor-works-with-smart-phones.
26. Honicky R, Brewer EA, Paulos E, White R. N-smarts: Networked suite of mobile atmospheric real-time sensors. Paper presenterd at: Proceedings of the Second ACM SIGCOMM Workshop on Networked Systems for Developing Regions; 2008 Aug 18; Seattle, WA, USA.
27. Silvertown J. A new dawn for citizen science. *Trends Ecol Evol*. 2009;24(9):467–471.
28. Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S. eBird: A citizen-based bird observation network in the biological sciences. *Biol Conserv*. 2009;142(10):2282–2292.
29. Unger S, Rollins M, Tietz A, Dumais H. iNaturalist as an engaging tool for identifying organisms in outdoor activities. *J Biol Educ*. 2021;55(5):537–547.
30. Mesaglio T, Callaghan CT. An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildl Res*. 2021;48(4):289–303.
31. Zhang G, Zhu A-X. The representativeness and spatial bias of volunteered geographic information: A review. *Ann GIS*. 2018;24(3):151–162.
32. Fink D, Auer T, Johnston A, Ruiz-Gutierrez V, Hochachka WM, Kelling S. Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecol Appl*. 2020;30(3):e02056.
33. Fink D, Hochachka WM, Zuckerberg B, Winkler DW, Shaby B, Munson MA, Hooker G, Riedewald M, Sheldon D, et al. Spatiotemporal exploratory models for broad-scale survey data. *Ecol Appl*. 2010;20(8):2131–2147.
34. Zhang G. Spatial and temporal patterns in volunteer data contribution activities: A case study of eBird. *ISPRS Int J Geo Inf*. 2020;9(10):597.
35. Zhang G. Detecting and visualizing observation hot-spots in massive volunteer-contributed geographic data across spatial scales using GPU-accelerated kernel density estimation. *ISPRS Int J Geo Inf*. 11(1):55.
36. Kelling S, Fink D, La Sorte FA, Johnston A, Bruns NE, Hochachka WM. Taking a 'big Data'approach to data quality in a citizen science project. *Ambio*. 2015;44(4):601–611.
37. Johnston A, Hochachka WM, Strimas-Mackey ME, Ruiz Gutierrez V, Robinson OJ, Miller ET, Auer T, Kelling ST, Fink D. Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Divers Distrib*. 2021;27(7):1265–1277.
38. Zhang G. Mitigating spatial bias in volunteered geographic information for spatial modeling and prediction. In: Li B, Shi X, Zhu A-X, Wang C, Lin H, editors. *New thinking in GIScience*. Singapore: Springer Nature; 2022. p. 179–190.
39. Zhang G, Zhu A-X. A representativeness directed approach to spatial bias mitigation in VGI for predictive mapping. *Int J Geogr Inf Sci*. 2019;33(9):1873–1893.
40. PurpleAir. PurpleAir: Real time air quality monitoring. [6 June 2023] https://www2.purpleair.com.
41. Lu T, Liu Y, Garcia A, Wang M, Li Y, Bravo-Villasenor G, Han B. Leveraging citizen science and low-cost sensors to characterize air pollution exposure of disadvantaged communities in Southern California. *Int J Environ Res Public Health*. 2022;19(14):8777.
42. Lu T, Bechle MJ, Wan Y, Presto AA, Hankey S. Using crowd-sourced low-cost sensors in a land use regression of PM2. 5 in 6 US cities. *Air Qual Atmos Health*. 2022;15(4):667–678.
43. AirNow. Get air quality data where you live. [8 June 2023] Retrieved from https://www.airnow.gov.
44. AirVisual. Air quality API—The most trusted historical, real-time and forecast air quality data. [6 June 2023] https://www.iqair.com/us/air-pollution-data-api.
45. SmartCitizen. Order your smart citizen kit from Seeed Studio. [7 June 2023] https://smartcitizen.me.
46. Clarity. We make air quality measurement easy. [7 June 2023] https://www.clarity.io.
47. Air Quality Egg. The egg air quality learning system. [7 June 2023] https://airqualityegg.com/home.
48. Thompson JE. Crowd-sourced air quality studies: A review of the literature & portable sensors. *Trends Environ Anal Chem*. 2016;11:23–34.
49. Miskell G, Salmond J, Williams DE. Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments. *Sci Total Environ*. 2017;575:1119–1129.
50. Kirby RR, Beaugrand G, Kleparski L, Goodall S, Lavender S. Citizens and scientists collect comparable oceanographic data: Measurements of ocean transparency from the Secchi disk study and science programmes. *Sci Rep*. 2021;11(1):15499.
51. George G, Menon NN, Abdulaziz A, Brewin RJ, Pranav P, Gopalakrishnan A, Platt T. Citizen scientists contribute to real-time monitoring of lake water quality using 3D printed mini Secchi disks. *Front Water*. 2021;3:662142.
52. Cakmak EK, Ugurlu A, Anbaroglu B. Adopting citizen science approach for water quality monitoring in Uzungöl, Turkey. *Environ Monit Assess*. 2021;193(9):604.
53. North American Lake Management Society. The Secchi Dip-In. [12 June 2023] https://www.nalms.org/secchidipin.
54. New York State Department of Environmental Conservation. Citizens statewide lake assessment program. [12 June 2023] https://www.dec.ny.gov/chemical/81576.html.
55. California State Water Resources Control Board. SWAMP – Clean Water Team (CWT) – Citizen Monitoring. [12 June 2023] https://www.waterboards.ca.gov/water_issues/programs/swamp/clean_water_team.
56. FreshWater Watch. Investigating the health of global freshwater ecosystems. [20 June 2023] https://www.freshwaterwatch.org.

57. Surfrider Foundation. Blue Water Task Force. [20 June 2023] https://bwtf.surfrider.org.
58. Zou L, Lam NS, Cai H, Qiang Y. Mining twitter data for improved understanding of disaster resilience. *Ann Am Assoc Geogr*. 2018;108(5):1422–1441.
59. Atkinson GM, Wald DJ. "Did you feel it?" intensity data: A surprisingly good measure of earthquake ground motion. *Seismol Res Lett*. 2007;78(3):362–368.
60. Quitoriano V, Wald DJ. USGS "did you feel it?"—Science and lessons from 20 years of citizen science-based macroseismology. *Front Earth Sci*. 2020;8:120.
61. Rotich J. Ushahidi: Empowering citizens through crowdsourcing and digital data collection. *Field Actions Sci Rep*. 2017;16(17):36–38.
62. Heinzelman J, Waters C. *Crowdsourcing crisis information in disaster-affected Haiti*. Washington (DC): US Institute of Peace; 2010.
63. Zou L, Liao D, Lam NS, Meyer MA, Gharaibeh NG, Cai H, Li D. Social media for emergency rescue: An analysis of rescue requests on twitter during hurricane Harvey. *Int J Disaster Risk Reduct*. 2023;85:103513.
64. Zhou B, Zou L, Mostafavi A, Lin B, Yang M, Gharaibeh N, Mandal D. VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Comput Environ Urban Syst*. 2022;95:101824.
65. Poiani, T. H., Rocha, R. D. S., Degrossi, L. C., De Albuquerque, J. P. Potential of collaborative mapping for disaster relief: A case study of OpenStreetMap in the Nepal earthquake 2015. Paper presented at: 2016 49th Hawaii International Conference on System Sciences (HICSS); 2016; Koloa, HI, USA. p. 188–197.
66. Geldmann J, Heilmann-Clausen J, Holm TE, Levinsky I, Markussen BO, Olsen K, Tøttrup AP. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers Distrib*. 2016;22(11):1139–1149.
67. Fritz S, McCallum I, Schill C, Perger C, Grillmayer R, Achard F, Obersteiner M. Geo-wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sens*. 2009;1(3):345–354.
68. Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, van der Velde M, Kraxner F, Obersteiner M. Geo-Wiki: An online platform for improving global land cover. *Environ Model Softw*. 2019;31:110–123.
69. Laso Bayas JC, Lesiv M, Waldner F, Schucknecht A, Duerauer M, See L, Fritz S, Fraisl D, Moorthy I, McCallum I, et al. A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform. *Sci Data*. 2017;4(1):1–10.
70. Moltchanova E, Lesiv M, See L, Mugford J, Fritz S. Optimizing crowdsourced land use and land cover data collection: A two-stage approach. *Land*. 2022;11(7):958.
71. Ding Q, Shao Z, Huang X, Altan O, Hu B. Time-series land cover mapping and urban expansion analysis using OpenStreetMap data and remote sensing big data: A case study of Guangdong-Hong Kong-Macao Greater Bay Area, China. *Int J Appl Earth Obs Geoinf*. 2022;113:103001.
72. Olteanu-Raimond A-M, Jolivet L, Van Damme M-D, Royer T, Fraval L, See L, Sturn T, Karner M, Moorthy I, Fritz S. An experimental framework for integrating citizen and community science into land cover. *Land*. 2018;7(3):103.
73. Theobald DM. A general-purpose spatial survey design for collaborative science and monitoring of global environmental change: The global grid. *Remote Sens*. 2016;8(10):813.
74. Kolstoe S, Cameron TA, Wilsey C. Climate, land cover, and bird populations: Differential impacts on the future welfare of birders across the Pacific Northwest. *Agric Resour Econ Rev*. 2018;47(2):272–310.
75. Whitehorn PR, Seo B, Comont RF, Rounsevell M, Brown C. The effects of climate and land use on Britishbumblebees: Findings from a decade of citizen–science observations. *J. Appl. Ecol*. 2022;58(7):1837–1851.
76. Ma N, Zhang Q, Murai F, Braham WW, Samuelson HW. Learning building occupants' indoor environmental quality complaints and dissatisfaction from text-mining Booking.com reviews in the United States. *Build Environ*. 2023;237:110319.
77. Chen X, Biljecki F. Mining real estate ads and property transactions for building and amenity data acquisition. *Urban Informatics*. 2022;1:12.
78. Wang J, Chow YS, Biljecki F. Insights in a city through the eyes of Airbnb reviews: Sensing urban characteristics from homestay guest experiences. *Cities*. 2023;140:104399.
79. Yan Y, Feng C-C, Huang W, Fan H, Wang Y-C, Zipf A. Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. *Int J Geogr Inf Sci*. 2020;34(9):1756–1791.
80. Biljecki F, Chow YS, Lee K. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Build Environ*. 2023;237:110295.
81. Herfort B, Lautenbach S, Van Albuquerque JP, Anderson J, Zipf A. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nat Commun*. 2023;14(1):3985.
82. Zhang L, Pfoser D. Using OpenStreetMap point-of-interest data to model urban change—A feasibility study. *PLOS ONE*. 2019;14(2):e0212606.
83. Venerandi A, Feliciotti A, Fleischmann M, Kourtit K, Porta S. Urban form character and Airbnb in Amsterdam (NL): A morphometric approach. *Environ Plan B Urban Anal City Sci*. 2023;50(2):386–400.
84. Kang Y, Zhang F, Peng W, Gao S, Rao J, Duarte F, Ratti C. Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*. 2021;111:104919.
85. Thomson DR, Stevens FR, Chen R, Yetman G, Sorichetta A, Gaughan AE. Improving the accuracy of gridded population estimates in cities and slums to monitor SDG 11: Evidence from a simulation study in Namibia. *Land Use Policy*. 2022;123:106392.
86. Sarkar D, Anderson JT. Corporate editors in OpenStreetMap: Investigating co-editing patterns. *Trans GIS*. 2022;26(4):1879–1897.
87. Ma D, Fan H, Li W, Ding X. The state of Mapillary: An exploratory analysis. *ISPRS Int Geo inf*. 2019;9(1):10.
88. Kang J, Körner M, Wang Y, Taubenböck H, Zhu XX. Building instance classification using street view images. *ISPRS J Photogramm*. 2018;145:44–59.
89. Qiu W, Li W, Liu X, Zhang Z, Li X, Huang X. Subjective and objective measures of streetscape perceptions: Relationships with property value in Shanghai. *Cities*. 2023;132:104037.
90. Liu D, Jiang Y, Wang R, Lu Y. Establishing a citywide street tree inventory with street view images and computer vision techniques. *Comput Environ Urban Syst*. 2023;100:101924.

91. Zhang C, Fan H, Kong G. VGI3D: An interactive and low-cost solution for 3D building modelling from street-level VGI images. *J Geovies Spat Anal*. 2021;5:18.
92. Fan H, Kong G, Zhang C. An interactive platform for low-cost 3D building modeling from VGI data using convolutional neural network. *Big Earth Data*. 2021;5(1):49–65.
93. Ding X, Fan H, Gong J. Towards generating network of bikeways from Mapillary data. *Comput Environ Urban Syst*. 2021;88:101632.
94. Yap W, Stouffs R, Biljecki F. Urbanity: Automated modelling and analysis of multidimensional networks in cities. *npj Urban Sustain*. 2023;3:45.
95. Walker CE, Kisiel E, Newhouse M. Globe at night: Turning awareness into action through citizen science. (2020) https://ui.adsabs.harvard.edu/abs/2020AAS...23540102W.
96. Case NA, MacDonald EA. Aurorasaurus and the St Patrick's Day storm. *Astron Geophys*. 2015;56(3):3.13–3.14.
97. Ahlqvist O, Schlieder C. Introducing geogames and geoplay: Characterizing an emerging research field. In: *Geogames and geoplay: Game-based approaches to the analysis of geo-information*. New York (NY): Springer; 2018. p. 1–18.
98. Choi J, Kim G, Sung K. Strategies for transforming location sharing apps into location based games. *J Future*. 2012.
99. Pajarito D, Gould M, Miralles I, Frias D, Monfort A. *A biking geo-game to gather commuting data.* 2016. https://www.geogames-team.org/agile2016/submissions/Pajarito_et_al_Biking.pdf.
100. Ineç ZF. Developing a geo-game application for global connections in social studies teaching: Gezgin Case, Romanian. *Rev Geogr Educ*. 2021;10(1):36–55.
101. Ramos F, Miralles N. *Smart Beetles: Towards a geogame for smart citizens*. Paper presented at: 20th AGILE International Conference on Geographic Information Science; 2017; Wageningen, Netherlands.
102. McNally B, de Andrade B. Altered spaces: New ways of seeing and envisioning nature with Minecraft. *Vis Stud*. 2022;37(3):175–182.
103. Tucker JD, Day S, Tang W, Bayus B. Crowdsourcing in medical research: Concepts and applications. *PeerJ*. 2019;7:e6762.
104. Afshinnekoo E, Ahsanuddin S, Mason CE. Globalizing and crowdsourcing biomedical research. *Br Med Bull*. 2016;120(1):27–33.
105. Tonne C, Basagaña X, Chaix B, Huynen M, Hystad P, Nawrot TS, Slama R, Vermeulen R, Weuve J, Nieuwenhuijsen M. New frontiers for environmental epidemiology in a changing world. *Environ Int*. 2017;104:155–162.
106. Seresinhe CI, Preis T, MacKerron G, Moat HS. Happiness is greater in more scenic locations. *Sci Rep*. 2019;9(1):4498.
107. Južnič-Zonta Ž, Sanpera-Calbet I, Eritja R, Palmer JRB, Escobar A, Garriga J, Oltra A, Richter-Boix A, Schaffner F, et al. Mosquito alert: Leveraging citizen science to create a GBIF mosquito occurrence dataset. *GigaByte*. 2022;2022:gigabyte54.
108. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, Santillana M, Nguyen A, Brownstein JS. Flu near you: Crowdsourced symptom reporting spanning 2 influenza seasons. *Am J Public Health*. 2015;105(10):2124–2130.
109. Yank V, Agarwal S, Loftus P, Asch S, Rehkopf D. Crowdsourced health data: Comparability to a US National Survey, 2013-2015. *Am J Public Health*. 2017;107(8):1283–1289.
110. Chan Y-FY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, Genes N, Scott ER, Krock E, Badgeley M, et al. The asthma Mobile health study, a large-scale clinical observational study using ResearchKit. *Nat Biotechnol*. 2017;35:354–362.
111. Qin H, Rice RM, Fuhrmann S, Rice MT, Curtin KM, Ong E. Geocrowdsourcing and accessibility for dynamic environments. *GeoJournal*. 2016;81(5):699–716.
112. Park YM, Chavez D, Sousan S, Figueroa-Bernal N, Alvarez JR, Rocha-Peralta J. Personal exposure monitoring using GPS-enabled portable air pollution sensors: A strategy to promote citizen awareness and behavioral changes regarding indoor and outdoor air pollution. *J Expo Sci Environ Epidemiol*. 2023;33(3):347–357.
113. Park YM, Sousan S, Streuber D, Zhao K. GeoAir—A novel portable, GPS-enabled, low-cost air-pollution sensor: Design strategies to facilitate citizen science research and geospatial assessments of personal exposure. *Sensors*. 2021;21(11):3761.
114. Huang X, Li X, Yang D, Zou L. Crowdsourced geospatial data in human and earth observations: Opportunities and challenges. *Geoinformatics Geosci*. 2023;109–129.
115. Zhu L, Yu FR, Wang Y, Ning B, Tang T. Big data analytics in intelligent transportation systems: A survey. *IEEE Trans Intell Transp Syst*. 2019;20(1):383–398.
116. Macias E, Suarez A, Lloret J. Mobile sensing systems. *Sensors*. 2013;13(12):17292–17321.
117. Panichpapiboon S, Leakkaw P. Traffic density estimation: A mobile sensing approach. *IEEE Commun Mag*. 2017;55(12):126–131.
118. Fazeen M, Gozick B, Dantu R, Bhukhiya M, González MC. Safe driving Using Mobile Phones. *IEEE Trans Intell Transp Syst*. 2012;13(3):1462–1468.
119. Li X, Goldberg DW. Toward a mobile crowdsensing system for road surface assessment. *Comput Environ Urban Syst*. 2018;69:51–62.
120. Sairam N, Nagarajan S, Ornitz S. Development of Mobile mapping system for 3D road asset inventory. *Sensors*. 2016;16(3):367.
121. Jalilifar E, Li X, Martin M, Huang X. Toward a crowdsourcing solution to estimate border crossing times using market-available connected vehicle data. Paper presented at: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications; 2022 Nov 1; Seattle, Washington, USA.
122. Li J, Boonaert J, Doniec A, Lozenguez G. Multi-models machine learning methods for traffic flow estimation from floating car data. *Trans Res Part C: Emerg Technol*. 2021;132:103389.
123. Zhang Z, Li M, Lin X, Wang Y. Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data. *Trans Res Part C: Emerg Technol*. 2020;121:102870.
124. Lee K, Sener IN. Strava metro data for bicycle monitoring: A literature review. *Transp Rev*. 2021;41(1):27–47.
125. Dadashova B, Griffin GP, Das S, Turner S, Sherman B. Estimation of average annual daily bicycle counts using crowdsourced Strava data. *Transp Res Rec*. 2020;2674(11):390–402.
126. Li X, Dadashova B, Yu S, Zhang Z. Rethinking highway safety analysis by leveraging crowdsourced waze data. *Sustainability*. 2020;12(23):10127.
127. Tarko AP. Surrogate measures of safety. In: *Transport and sustainability*. Leeds (UK): Emerald Group Publishing Ltd.; 2018, vol. 11, p. 383–405.

128. Hunter M, Saldivar-Carranza E, Desai J, Mathew JK, Li H, Bullock DM. A proactive approach to evaluating intersection safety using hard-braking data. *J. Big Data Anal Transp*. 2021;3(2):81–94.
129. Mousavi SM, Zhang Z, Parr SA, Pande A, Wolshon B. Identifying high crash risk highway segments using jerk-cluster analysis. In: *International Conference on Transportation and Development 2019: Smarter and Safer Mobility and Cities—Selected Papers from the International Conference on Transportation and Development 2019*. Reston (VA): American Society of Civil Engineers; 2019.
130. Wang X, Zhou Q, Quddus M, Fan T, Fang S. Speed, speed variation and crash relationships for urban arterials. *Accid Anal Prev*. 2018;113:236–243.
131. Goodall N, Lee E. Comparison of Waze crash and disabled vehicle records with video ground truth. *Transp Res Interdiscip Perspect*. 2019;1:100019.
132. Chen K, Lu M, Tan G, Wu J. CRSM: Crowdsourcing based road surface monitoring. Paper presented at: Proceedings of the 2013 IEEE international Conference on High Performance Computing and Communications, HPCC 2013 and 2013 IEEE International Conference on Embedded and Ubiquitous Computing, EUC; 2013 Nov 13–15; Zhangjiajie, China.
133. Strain T, Wilson RE, Littleworth R. Computer vision for rapid updating of the highway asset inventory. *Transp Res Rec*. 2020;2674(9):245–255.
134. Le Dantec CA, Asad M, Misra A, Watkins KE. Planning with crowdsourced data: Rhetoric and representation in transportation planning. Paper presented at: Proceedings of the 18th ACM conference on Computer Supported Cooperative Work & Social Computing; 2015 Mar 14–18; BC, Vancouver, Canada.
135. Rodríguez-Ibánez M, Casánez-Ventura A, Castejón-Mateos F, Cuenca-Jiménez P-M. A review on sentiment analysis from social media platforms. *Expert Syst Appl*. 2023;223:119862.
136. Xu QA, Chang V, Jayne C. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decis Anal J*. 2022;3:100073.
137. Klimiuk K, Czoska A, Biernacka K, Balwicki Ł. Vaccine misinformation on social media–topic-based content and sentiment analysis of polish vaccine-deniers' comments on Facebook. *Hum Vaccin Immunother*. 2021;17(7):2026–2035.
138. Kumar A, Jaiswal A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurr Comput Pract Exp. 2020;32(1):e5107.
139. Yang X, Xu S, Wu H, Bie R. Sentiment analysis of Weibo comment texts based on extended vocabulary and convolutional neural network. *Procedia Comput Sci*. 2019;147:361–368.
140. Ye J, Peng X, Qiao Y, Xing H, Li J, Ji R. Visual-textual sentiment analysis in product reviews. Paper presented at: Proceedings of the International Conference on Image Processing, ICIP; 2019 Sep 22–25; Taipei, Taiwan.
141. Abd-Alrazaq A, Alhuwail D, Househ M, Hai M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. *J Med Internet Res*. 2020;22(4):e19016.
142. Manguri H, K., N. Ramadhan, R., & R. Mohammed Amin, P. Twitter sentiment analysis on worldwide COVID-19 outbreaks. *Kurd J Appl Res*. 2020;5(3):54–65.
143. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for Health Research: A systematic review. *Am J Public Health*. 2017;107(1):e1–e8.
144. Jaidka K, Ahmed S, Skoric M, Hilbert M. Predicting elections from social media: A three-country, three-method comparative study. *Asian J Commun*. 2019;29(3):252–273.
145. Sandoval-Almazan R, Valle-Cruz D. Sentiment analysis of Facebook users reacting to political campaign posts. *Digit Gov Res Pract*. 2020;1(2):1–13.
146. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time twitter sentiment analysis of 2012 U.S. presidential election cycle. Paper presented at: Proceedings of the Annual Meeting of the Association for Computational Linguistics; 2012; Jeju Island, Korea.
147. Wang Y, Gao S, Li N, Yu S. Crowdsourcing the perceived urban built environment via social media: The case of underutilized land. *Adv Eng Inform*. 2021;50:101371.
148. Kausar S, Huahu X, Shabir MY, Ahmad W. A sentiment polarity categorization technique for online product reviews. *IEEE Access*. 2020;8:3594–3605.
149. Sajid S, Volkova N, Wilson JA, Opoku-Asante E. Using text mining and crowdsourcing platforms to build employer brand in the US banking industry. *Glob Bus Organ Excell*. 2022;41(4):6–27.
150. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PLOS ONE*. 2013;8(12):e83672.
151. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 pandemic: A thematic analysis of twitter data. *Health Inf Libr J*. 2019;36(1):60–72.
152. Müller MM, Salathé M. Crowdbreaks: Tracking health trends using public social media data and crowdsourcing. *Front Public Health*. 2019;7:81.
153. Ibrahim AF, Hassaballah M, Ali AA, Nam Y, Ibrahim IA. COVID19 outbreak: A hierarchical framework for user sentiment analysis. *Comput Mater Contin*. 2022;70:2507–2524.
154. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, Sheikh A. Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study. *J Med Internet Res*. 2021;23(4):e26627.
155. Hussain Z, Sheikh Z, Tahir A, Dashtipour K, Gogate M, Sheikh A, Hussain A. Artificial intelligence–enabled social media analysis for pharmacovigilance of COVID-19 vaccinations in the United Kingdom: Observational study. *JMIR Public Health Surveill*. 2022;8(5):e32543.
156. Ushahidi. Crisis mapping Haiti: Some final reflections. July 13, 2023 [accessed 2010] https://wwwushahidicom/about/blog/crisis-mapping-haiti-some-final-reflections/.
157. Wang Z, Ye X. Social media analytics for natural disaster management. *Int J Geogr Inf Sci*. 2018;32(1):49–72.
158. Ashktorab Z, Brown C, Nandi M, Culotta A. Tweedr: Mining Twitter to inform disaster response. Paper presented at: International Conference on Information Systems for Crisis Response and Management (ISCRAM); 2014; University Park, PA, USA. p. 269–272.
159. De Albuquerque JP, Herfort B, Brenning A, Zipf A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int J Geogr Inf Sci*. 2015;29(4):667–689.
160. Wang Y, Wang T, Ye X, Zhu J, Lee J. Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. *Sustainability*. 2015;8(1):25.

161. Tien Nguyen D, Alam F, Ofli F, Imran M. Automatic image filtering on social networks using deep learning and perceptual hashing during crises. arXiv. 2017. https://arxiv.org/abs/1704.02602.
162. Amini A, Kung K, Kang C, Sobolevsky S, Ratti C. The impact of social segregation on human mobility in developing and industrialized regions. *EPJ Data Sci*. 2014;3:6.
163. Huang X, Zhao Y, Wang S, Li X, Yang D, Feng Y, Chen B. Unfolding community homophily in US metropolitans via human mobility. *Cities*. 2022b;129:103929.
164. Pelletier MP, Trépanier M, Morency C. Smart card data use in public transit: A literature review. *Transp Res Part C: Emerg Technol*. 2011;19(4):557–568.
165. Seaborn C, Attanucci J, Wilson NH. Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transp Res Rec*. 2009;2121(1):55–62.
166. Zhao M, Mason L, Wang W. Empirical study on human mobility for mobile wireless networks. In: *MILCOM 2008-2008 IEEE Military Communications Conference*. San Diego (CA): IEEE; 2008. p. 1–7.
167. Su W, Lee SJ, Gerla M. Mobility prediction in wireless networks. In: *MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No. 00CH37155)*. Los Angeles (CA): IEEE; 2000, vol. 1, p. 491–495.
168. Huang X, Wang C, Li Z. A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. *Ann GIS*. 2018;24(2):113–123.
169. Li Z, Huang X, Ye X, Jiang Y, Martin Y, Ning H, Li X. Measuring global multi-scale place connectivity using geotagged social media data. *Sci Rep*. 2021;11:14964.
170. Shen Y, Karimi K. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*. 2016;55:9–21.
171. Chen T, Hui EC, Wu J, Lang W, Li X. Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data. *Habitat Int*. 2019;89:102005.
172. Tang J, Liu F, Wang Y, Wang H. Uncovering urban human mobility from large scale taxi GPS data. *Phys A Stat Mech Appl*. 2015;438:140–153.
173. Jiang S, Guan W, Zhang W, Chen X, Yang L. Human mobility in space from three modes of public transportation. *Phys A Stat Mech Appl*. 2017;483:227–238.
174. Bisanzio D, Kraemer MU, Bogoch II, Brewer T, Brownstein JS, Reithinger R. Use of twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. *Geospat Health*. 2020;15(1).
175. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLOS ONE*. 2020;15(11):e0241957.
176. Xiong C, Hu S, Yang M, Luo W, Zhang L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc Natl Acad Sci*. 2020;117(44):27087–27089.
177. Yuan NJ, Zheng Y, Xie X, Wang Y, Zheng K, Xiong H. Discovering urban functional zones using latent activity trajectories. *IEEE Trans Knowl Data Eng*. 2014;27(3):712–725.
178. Du Z, Zhang X, Li W, Zhang F, Liu R. A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou City, China. *Trans GIS*. 2020;24(1):123–141.
179. Jiang Y, Yuan F, Farahmand H, Acharya K, Zhang J, Mostafavi A. Data-driven tracking of the bounce-back path after disasters: Critical milestones of population activity recovery and their spatial inequality. *Int J Disaster Risk Reduct*. 2023;92:103693.
180. Guadagno L. Human mobility in the Sendai framework for disaster risk reduction. *Int J Disaster Risk Reduct*. 2016;7:30–40.
181. Sunny EE, Anael O. Mobile marketing in a digital age: Application, challenges & opportunities. *Br J Econom Manage Trade*. 2016;11(1):1–13.
182. Gao L, Liu X, Zou H. The role of human mobility in promoting Chinese outward FDI: A neglected factor? *Int Bus Rev*. 2013;22(2):437–449.
183. Zeng D, Chen H, Lusch R, Li SH. Social media analytics and intelligence. *IEEE Intell Syst*. 2010;25(6):13–16.
184. Nemes L, Kiss A. Social media sentiment analysis based on COVID-19. *J Inf Telecommun*. 2021;5(1):1–15.
185. Weaver IS, Williams H, Cioroianu I, Williams M, Coan T, Banducci S. Dynamic social media affiliations among UK politicians. *Soc Networks*. 2018;54:132–144.
186. van Schalkwyk F, Dudek J, Costas R. Communities of shared interests and cognitive bridges: The case of the anti-vaccination movement on twitter. *Scientometrics*. 2020;125(2):1499–1516.
187. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. *Soc Sci Med*. 2019;240:112552.
188. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Scala A. The COVID-19 social media infodemic. *Sci Rep*. 2020;10(1):16598.
189. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: Social bots, emerging topics, and trends. *JMIR Public Health Surveill*. 2017;3(4):e8641.
190. Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, Li Z. Revealing public opinion towards COVID-19 vaccines with twitter data in the United States: Spatiotemporal perspective. *J Med Internet Res*. 2021;23(9):e30854.
191. Finin T, Joshi A, Kolari P, Java A, Kale A, Karandikar A. The information ecology of social media and online communities. *AI Mag*. 2008;29(3):77–77.
192. Department of Homeland Security, *Public safety analytics terminal: Technology scouting research summary*; Washington (DC): Department of Homeland Security; 2019.
193. Mitchell T, Krulicky T. Big data-driven urban geopolitics, interconnected sensor networks, and spatial cognition algorithms in smart city software systems. *Geopolit Hist Int Relat*. 2021;13(2):9–22.
194. Kelly GS, Clare D. Improving out-of-hospital notification in traumatic cardiac arrests with novel usage of smartphone application. *J Am Coll Emerg Physicians Open*. 2020;1(4):618–623.
195. Horse AJY, Jeung R, Lim R, Tang B, Im M, Higashiyama L, Chen M. *Stop AAPI hate national report.* San Francisco (CA): Stop AAPI Hate; 2021.
196. Jeung R, Horse AJY, Lau A, Kong P, Shen K, Cayanan C, Canayan C, Xiong M, Lim R. *Stop AAPI hate youth report*. Stop AAPI Hate; 2020. https://stopaapihate.org/wp-content/uploads/2021/04/Stop-AAPI-Hate-Report-Youth-Incidents-200917.pdf.
197. Nham K, Huỳnh J. Contagious heathens: Exploring racialization of COVID-19 and Asians through stop AAPI hate incident reports. *AAPI Nexus Policy Pract Commun*. 2020;17(1-2).

198. Saw A, Yellowhorse A, Jeung R. *Stop AAPI hate mental health report*. San Francisco (CA): Stop AAPI Hate Coalition; 2021.
199. Young C. HarassMap: Using crowdsourced data to map sexual harassment in Egypt. *Technol Innov Manag Rev*. 2014;4(3).
200. Chordia I, Tran L-P, Tayebi TJ, Parrish E, Erete S, Yip J, Hiniker A. Deceptive design patterns in safety technologies: A case study of the citizen app. Paper presented at the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; 2023 Apr 23–28; Hamburg, Germany.
201. Fetterman AK, Baker CD, Meier BP. Crime in your area: Use of neighborhood apps is associated with inaccurate perceptions of higher local crime rates. *Psychol Pop Media*. 2023.
202. VICE (Producer). 'FIND THIS FUCK': Inside Citizen's dangerous effort to cash in on vigilantism; 2021.
203. Wang S, Zhang M, Huang X, Hu T, Li Z, Sun QC, Liu Y. Urban-regional disparities in mental health signals in Australia during the COVID-19 pandemic: A study via twitter data and machine learning models. *Camb J Reg Econ Soc*. 2022;15(3):663–682.
204. Neis P, Goetz M, Zipf A. Towards automatic vandalism detection in OpenStreetMap. *ISPRS Int J Geo-Inf*. 2012;1:315–332.
205. Costa Fonte C, Fritz S, Olteanu-Raimond AM, Antoniou V, Foody G, Mooney P, See L. Mapping and the citizen sensor. London (UK): Ubiquity Press; 2017. p. 398.
206. Ballatore A, Zipf A. A conceptual quality framework for volunteered geographic information. In: *Spatial Information Theory: 12th International Conference, COSIT 2015, Santa Fe, NM, USA, October 12–16, 2015, Proceedings 12*. New York (NY): Springer International Publishing; 2015. p. 89–107.
207. Hou Y, Biljecki F. A comprehensive framework for evaluating the quality of street view imagery. *Int J Appl Earth Obs*. 2022;115:103094.
208. Basiri A, Haklay M, Foody G, Mooney P. Crowdsourced geospatial data quality: Challenges and future directions. *Int J Geogr Inf Sci*. 2019;33(9):1588–1593.
209. Comber A, Mooney P, Purves RS, Rocchini D, Walz A. Crowdsourcing: It matters who the crowd are. The impacts of between group variations in recording land cover. *PLOS ONE*. 2016;11(7):e0158329.
210. Haklay M. Why is participation inequality important? In: *European handbook of crowdsourced geographic information*. London (UK): Ubiquity Press; 2016.
211. Hettiachchi D, Sanderson M, Goncalves J, Hosio S, Kazai G, Lease M, Schaekermann M, Yilmaz E. Investigating and mitigating biases in crowdsourced data. Paper presented at: Proceedings of the ACM Conference on Computer Supported Cooperative Work. CSCW; 2021 Oct 23–27; Virtual Event, USA.
212. Mullen WF, Jackson SP, Croitoru A, Crooks A, Stefanidis A, Agouris P. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal*. 2015;80:587–605.
213. Chen M, Arribas-Bel D, Singleton A. Understanding the dynamics of urban areas of interest through volunteered geographic information. *J Geogr Syst*. 2019;21:89–109.
214. Goodchild MF, Li L. Assuring the quality of volunteered geographic information. *Spat Stat*. 2012;1:110–120.
215. Haklay M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ Plan B Urban Anal City Sci*. 2010.
216. Huang X, Li Z, Jiang Y, Ye X, Deng C, Zhang J, Li X. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. *Int J Digit Earth*. 2021;14(4):424–442.
217. Ballatore A, Arsanjani JJ. Placing Wikimapia: An exploratory analysis. *Int J Geogr Inf Sci*. 2019;33(8):1633–1650.
218. Yang A, Fan H, Jing N, Sun Y, Zipf A. Temporal analysis on contribution inequality in OpenStreetMap: A comparative study for four countries. *ISPRS Int J Geo Inf*. 2016;5(1):5.
219. Antoniou V, Skopeliti A. Measures and indicators of vgi quality: An overview, in: ISPRS annals of the photogrammetry. Paper presented at: ISPRS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences; 2015 Sep; La Grand Motte, France.
220. Millar EE, Hazell EC, Melles SJ. The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *Int J Geogr Inf Sci*. 2019;33(8):1612–1632.
221. Wang S, Huang X, Hu T, She B, Zhang M, Wang R, Bao S. A global portrait of expressed mental health signals towards COVID-19 in social media space. *Int J Appl Earth Obs Geoinf*. 2023;116:103160.
222. Duggan M, Brenner J. *The demographics of social media users—2012.* Washington (DC): Pew Research Center; 2013.
223. Longley PA, Adnan M, Lansley G. The geotemporal demographics of twitter usage. *Environ Plan A*. 2015;47(2).
224. Draws T, Rieger A, Inel O, Gadiraju U, Tintarev N. A checklist to combat cognitive biases in crowdsourcing. Paper presented at: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing. 2021 Nov; Virtual.
225. Eickhoff C. Cognitive biases in crowdsourcing. Paper presented at: WSDM 2018—Proceedings of the 11th ACM International Conference on Web Search and Data Mining; 2018 Feb 5–9; Marian Del Rey, CA, USA.
226. Dorn H, Törnros T, Zipf A. Quality evaluation of VGI using authoritative data-a comparison with land use data in southern Germany. *ISPRS Int J Geo Inf*. 2015;4(3): 1657–1671.
227. Bubalo M, van Zanten BT, Verburg PH. Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landsc Urban Plan*. 2019;184:101–111.
228. Hara K, Le V, Froehlich JE. Combining crowdsourcing and Google street view to identify street-level accessibility problems. Paper presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2013 Apr–27–May 2; Paris, France.
229. Song F, Ma T. A location privacy protection method in spatial crowdsourcing. *J Inf Secur Appl*. 2022;65:103095.
230. To H, Ghinita G, Shahabi C. Framework for protecting worker location privacy in spatial crowdsourcing. Paper presented at: Proceedings of the VLDB Endowment. 2014 Sep 1–5; Hangzhou, China. p. 919–930.
231. Gruteser M, Liu X. Protecting privacy in continuous location-tracking applications. *IEEE Secur Priv*. 2004;2(2):28–34.
232. Shokri R, Theodorakopoulos G, Troncoso C, Hubaux JP, Le Boudec JY. Protecting location privacy: Optimal strategy against localization attacks. Paper presented at: Proceedings of the ACM Conference on Computer and Communications Security; 2012; Raleigh, NC, USA.
233. Batty M, Crooks A, Hudson-Smith A, Milton R, Anand S, Jackson M, Morley J. Data mash-ups and the future of mapping. *JISC, Bristol*. 2010.
234. Meftah L, Rouvoy R, Chrisment I. Empowering mobile crowdsourcing apps with user privacy control. *J Parallel Distrib Comput*. 2021;147:1–15.

235. Jin F, Hua W, Francia M, Chao P, Orlowska ME, Zhou X. A survey and experimental study on privacy-preserving trajectory data publishing. *IEEE Trans Knowl Data Eng*. 2023;35(6):5577–5596.
236. Pournajaf L, Garcia-Ulloa DA, Xiong L, Sunderam V. Participant privacy in mobile crowd sensing task management: A survey of methods and challenges. *SIGMOD Rec*. 2015;44(4):23–34.
237. Zhu X, Hu D, Hou Z, Ding L. A location privacy preserving solution to resist passive and active attacks in VANET. *China Commun*. 2014;11(9):60–67.
238. Zheng F, Tao R, Maier HR, See L, Savic D, Zhang T, Popescu I. Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Rev Geophys*. 2018;56(4):698–740.
239. Gray JS, Hwang JT, Martins JR, Moore KT, Naylor BA. OpenMDAO: An open-source framework for multidisciplinary design, analysis, and optimization. *Struct Multidiscip Optim*. 2019;59:1075–1104.
240. da Silva M, Viterbo J, Bernardini F, Maciel C. Identifying privacy functional requirements for crowdsourcing applications in smart cities. In: *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*. Miami (FL): IEEE; 2018. p. 106–111.
241. Newman G, Wiggins A, Crall A, Graham E, Newman S, Crowston K. The future of citizen science: Emerging technologies and shifting paradigms. *Front Ecol Environ*. 2012;10(6):298–304.
242. Starbird K, Palen L. Voluntweeters self-organizing by digital volunteers in times of crisis. Paper presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 May 7–12; Vancouver, BC, Canada. p. 1071–1080.
243. See L, Comber A, Salk C, Fritz S, Van Der Velde M, Perger C, Obersteiner M. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLOS ONE*. 2013;8(7):e69958.
244. Zhang J, Sheng VS, Li T, Wu X. Improving crowdsourced label quality using noise correction. *IEEE Trans Neural Netw Learn Syst*. 2017;29(5):1675–1688.
245. Li C, Jiang L, Xu W. Noise correction to improve data and model quality for crowdsourcing. *Eng Appl Artif Intell*. 2019;82:184–191.
246. Courter JR, Johnson RJ, Stuyck CM, Lang BA, Kaiser EW. Weekend bias in citizen science data reporting: Implications for phenology studies. *Int J Biometeorol*. 2013;57:715–720.
247. Wiggins A, Crowston K. From conservation to crowdsourcing: A typology of citizen science. In: *2011 44th Hawaii International Conference on System Sciences*. Kiloa (HI): IEEE; 2011. p. 1–10.
248. Girres JF, Touya G. Quality assessment of the French OpenStreetMap dataset. *Trans GIS*. 2010;14(4):435–459.
249. Kaur J, Singh J, Sehra SS, Rai HS. Systematic literature review of data quality within openstreetmap. In: *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*. Jammu (India): IEEE; 2017. p. 177–182.
250. Mandlekar A, Zhu Y, Garg A, Booher J, Spero M, Tung A, Fei-Fei L. Roboturk: A crowdsourcing platform for robotic skill learning through imitation. In: *Conference on Robot Learning*. Zürich (Switzerland): PMLR; 2018. p. 879–893.
251. Lukyanenko R, Parsons J, Wiersma YF, Maddah M. Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Q*. 2019;43(2):623–648.
252. Rotman D, Preece J, Hammock J, Procita K, Hansen D, Parr C, Lewis D, Jacobs D. Dynamic changes in motivation in collaborative citizen-science projects. Paper presented at: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work; 2012 Feb 11–15; Seattle, Washington, USA.
253. Kelling S, Johnston A, Hochachka WM, Iliff M, Fink D, Gerbracht J, Lagoze C, Sorte F, Moore T, Wiggins A, et al. Can observation skills of citizen scientists be estimated using species accumulation curves? *PLOS ONE*. 2015;10(10):e0139600.
254. Freitag A, Pfeffer MJ. Process, not product: Investigating recommendations for improving citizen science "success". *PLOS ONE*. 2013;8(5):e64079.
255. Wald DM, Longo J, Dobell AR. Design principles for engaging and retaining virtual citizen scientists. *Conserv Biol*. 2016;30(3):562–570.
256. Kobori H, Dickinson JL, Washitani I, Sakurai R, Amano T, Komatsu N, Kitamura W, Takagawa S, Koyama K, Ogawara T, et al. Citizen science: A new approach to advance ecology, education, and conservation. *Ecol Res*. 2016;31(1):1–19.
257. Eveleigh A, Jennett C, Blandford A, Brohan P, Cox AL. Designing for dabblers and deterring drop-outs in citizen science. Paper presented at: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems; 2014 Apr 26–May 1; Toronto, Ontario, Canada.
258. Iacovides I, Jennett C, Cornish-Trestrail C, Cox AL. Do games attract or sustain engagement in citizen science? A study of volunteer motivations. Paper presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; 2013 Apr 27–May 2; Paris, France.
259. Simpson R, Page KR, De Roure D. Zooniverse: Observing the world's largest citizen science platform. Paper presented at: Proceedings of the 23rd International Conference on World Wide Web; 2014 Apr 7–11; Seoul, Korea.
260. Cavalier D, Kennedy EB. *The rightful place of citizen science*. Tempe (AZ): Consortium for Science, Policy & Outcomes; 2016.
261. Pandya RE. A framework for engaging diverse communities in citizen science in the US. *Front Ecol Environ*. 2012;10(6): 314–317.
262. English PB, Richardson MJ, Garzón-Galvis C. From crowdsourcing to extreme citizen science: Participatory research for environmental health. *Annu Rev Public Health*. 2018;39:335–350.
263. Noorashid N, Chin WL. Coping with COVID-19: The resilience and transformation of community-based tourism in Brunei Darussalam. *Sustainability*. 2021;13(15):8618.
264. Palmer JR, Oltra A, Collantes F, Delgado JA, Lucientes J, Delacour S, Bengoa M, Eritja R, Bartumeus F. Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes. *Nat Commun*. 2017;8:916.
265. Wiggins A. Free as in puppies: Compensating for ICT constraints in citizen science. Paper presented at: Proceedings of the 2013 Conference on Computer Supported Cooperative Work; 2013 Feb 23–27; Texas, USA.
266. Ballard HL, Dixon CGH, Harris EM. Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. *Biol Conserv*. 2017;208:65–75.
267. Davies R. Civic crowdfunding: Participatory communities, entrepreneurs and the political economy of place. Paper

presented at: Entrepreneurs and the Political Economy of Place; 2014 May 18–23.Davies R. Civic Crowdfunding: Participatory communities, entrepreneurs and the political economy of place. 2014. https://ssrn.com/abstract=2434615 or http://dx.doi.org/10.2139/ssrn.2434615.
268. Prestopnik NR, Crowston K. Gaming for (citizen) science: Exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. In: *2011 IEEE Seventh International Conference on e-Science Workshops*. Stockholm (Sweden): IEEE; 2011. p. 28–33.
269. Gao Z, Wang S, Gu J. Public participation in smart-city governance: A qualitative content analysis of public comments in urban China. *Sustainability*. 2020;12(20):8605.
270. Wang S, Huang X, She B, Li Z. Diverged landscape of restaurant recovery from the COVID-19 pandemic in the United States. *iScience*. 2023;26(6):106811.
271. Takahashi B, Tandoc EC Jr, Carmichael C. Communicating on twitter during a disaster: An analysis of tweets during typhoon Haiyan in the Philippines. *Comput Hum Behav*. 2015;50: 392–398.
272. Middleton SE, Middleton L, Modafferi S. Real-time crisis mapping of natural disasters using social media. *IEEE Intell Syst*. 2013;29(2):9–17.
273. Devarakonda S, Sevusu P, Liu H, Liu R, Iftode L, Nath B. Real-time air quality monitoring through mobile sensing in metropolitan areas. Paper presented at: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing; 2013 Aug 11; Chicago, Illinois.
274. Patino A, Pitta DA, Quinones R. Social media's emerging importance in market research. *J Consum Mark*. 2012;29(3):233–237.
275. Carr J, Decreton L, Qin W, Rojas B, Rossochacki T, wen Yang, Y. Social media in product development. *Food Qual Prefer*. 2015;40:354–364.
276. Gal-Tzur A, Grant-Muller SM, Kuflik T, Minkov E, Nocera S, Shoor I. The potential of social media in delivering transport policy goals. *Transp Policy*. 2014;32:115–123.
277. Blackwell L, Hardy J, Ammari T, Veinot T, Lampe C, Schoenebeck S. LGBT parents and social media: Advocacy, privacy, and disclosure during shifting social movements. Paper presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; 2016 May 7–12; San Jose, California, USA.
278. Rawat P, Yusuf JEW. Participatory mapping, E-participation, and E-governance: Applications in environmental policy. In: *Leveraging digital innovation for governance, public administration, and citizen services: Emerging research and opportunities*. Hershey (PA): IGI Global; 2020. p. 147–175.
279. Dameri RP. Searching for smart city definition: A comprehensive proposal. *Int J Comput Technol*. 2013;11(5):2544–2551.