[DC-02-029] Volunteered Geographic Information

Abstract

Volunteered geographic information (VGI) refers to geo-referenced data created by citizen volunteers. VGI has proliferated in recent years due to the advancement of technologies that enable the public to contribute geographic data. VGI is not only an innovative mechanism for geographic data production and sharing, but also may greatly influence GIScience and geography and its relationship to society. Despite the advantages of VGI, VGI data quality is under constant scrutiny as quality assessment is the basis for users to evaluate its fitness for using it in applications. Several general approaches have been proposed to assure VGI data quality but only a few methods have been developed to tackle VGI biases. Analytical methods that can accommodate the imperfect representativeness and biases in VGI are much needed for inferential use where the underlying phenomena of interest are inferred from a sample of VGI observations. VGI use for inference and modeling adds much value to VGI. Therefore, addressing the issue of representativeness and VGI biases is important to fulfill VGI's potential. Privacy and security are also important issues. Although VGI has been used in many domains, more research is desirable to address the fundamental intellectual and scholarly needs that persist in the field.

Keywords: citizen science, data quality, geographic data, geospatial big data, privacy, security, spatial bias, volunteered geographic information (VGI)

Author & citation

Zhang, G. (2021). Volunteered Geographic Information. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2021 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2021.1.1

Explanation

1. Definitions
2. Volunteered Geographic Information
3. Types of VGI
4. Advantages of VGI
5. Data Quality of VGI
6. Privacy and Security
7. Data Licenses and Copyright
8. Challenges and Outlook

1. Definitions

Web 2.0: A collection of technologies that harnesses the Web in a more interactive and collaborative manner, emphasizing social interaction and collective intelligence. Web 2.0 allows users both access content from Web sites and contribute to them.

User-generated content: Any form of content that has been posted by users on online platforms, for example, images, videos, text, and audio contributed by social media users

and wikis contributors.

Geo-referencing: Specifying the geographic location of an object, entity, phenomenon, image, concept, data, or information with universal parameters, code, or place.

Geo-tagging: The process of adding geospatial identification metadata to various types of media.

Crowdsourcing: The practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community.

Citizen science: Scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions.

Neogeography: The use of geographical techniques and tools for personal and community activities or by a non-expert group of users.

Participatory mapping: Approaches and techniques that combines the tools of modern cartography with participatory methods to record and represent the spatial knowledge of local communities.

Public participation GIS: The use of GIS to broaden public involvement in policymaking as well as to the value of GIS to promote the goals of nongovernmental organizations, grassroots groups, and community-based organizations.

2. Volunteered Geographic Information

Volunteered geographic information (VGI) is an umbrella term referring to geo-referenced data created by citizen volunteers (Goodchild, 2007). VGI broadly encompasses geographic data contributed by non-professional volunteers such as participants in citizen science, crowdsourcing, neogeography, participatory mapping, public participation GIS, etc. and social media users (Figure 1), as long as they share the characteristics of voluntary and non-expert geographic data contribution. Nonetheless, VGI is only a loose generalization of geographic data resulting from these sources. Each of the terms representing the sources has slightly different but important connotative differences (Sieber, 2006). In many cases, it can be problematic to treat the data as “volunteered”, for example, user-generated content with surveillance systems, web page trackers and cookies, etc. In such circumstances, VGI emphasizes more on the non-expert (instead of voluntary) nature of data contribution; the term “volunteered” has always sat uneasily alongside the actual practices people associate with it.

Examples of VGI are road networks of the world complied by OpenStreetMap (OSM) contributors (Haklay & Weber, 2008), species occurrence records across the globe contributed by eBirders (Sullivan et al., 2014), and geo-tagged social media posts. VGI has been used in a variety of applications such as environmental monitoring (e.g., species sightings, phenological observations), land management, land cover map validation, location-based services (e.g., routing and navigation), disaster response and humanitarian action (see Yan, Feng, Huang, Fan, & Wang, 2020 and references therein), human mobility research (Jurdak et al., 2015), public health (Goranson, Thihalolipavan, & di Tada, 2013),

crime analysis and community policing (Jelokhani-Niaraki, Bastami Mofrad, Yazdanpanah Dero, Hajiloo, & Sadeghi-Niaraki, 2019; White & Roth, 2010), sharing spatial data regarding business reviews (Rahimi, Mottahedi, & Liu, 2018), speeding cameras, traffic accidents, infrastructure closures (www.waze.com/livemap), etc.

VGI has proliferated mainly because of the advancement of technologies that enable the public to contribute geographic data. With the empowerment of Web 2.0 and ubiquitous access to the Internet and positioning services, ordinary citizens acting as “human sensors” using mobile smartphones and other location-aware portable devices can now easily contribute geo-referenced observations regarding social and natural environments of the world, as a specific form of user-generated content. VGI represents a paradigm shift in geographic data production and sharing and its content and characteristics (Elwood, Goodchild, & Sui, 2012). It may greatly influence GIScience and geography and its relationship to society (Goodchild, 2007). VGI is also an important source of big geospatial data that may propel geographic research towards a “data-driven” approach (Miller & Goodchild, 2014).

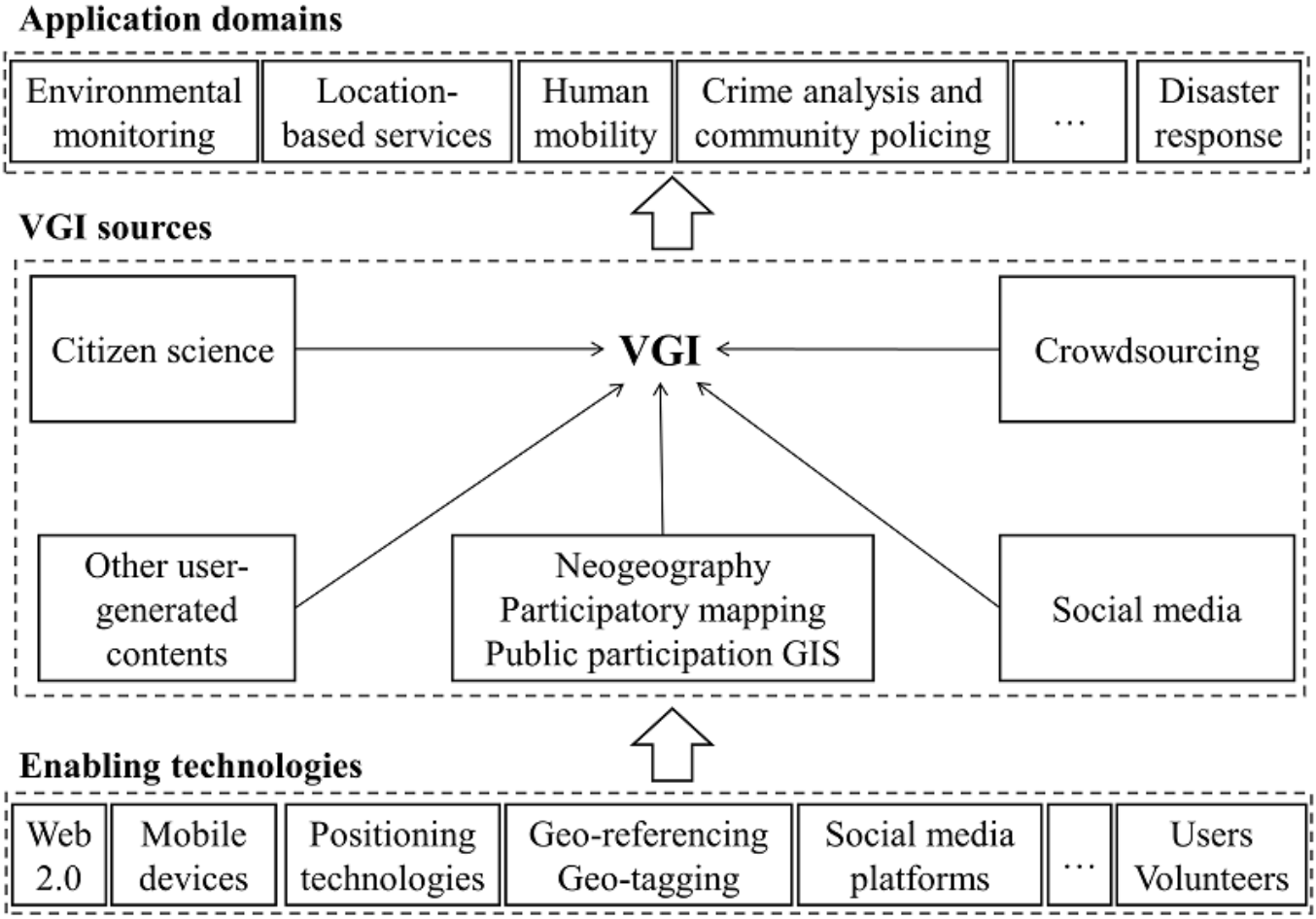


Figure 1. VGI enabling technologies, sources and application domains (see Section 1 for definitions of the terms). Source: author.

3. Types of VGI

According to Sui and Cinnamon (2016), VGI can be loosely grouped into three types: geospatial framework data, gazetteer data, and thematic data. Among the themes of geographic framework data, VGI greatly contributes to producing transportation and road networks data. OSM (www.openstreetmap.org) is an exemplary VGI platform on which volunteer contributors compile detailed streets, roads (and other features) for much of the world by uploading GPS tracks or tracing and digitizing geographic features from high resolution satellite imagery (Haklay and Weber, 2008).

Gazetteer, concerned with associating place names with particular places, is expensive to construct and maintain using traditional methods but well-suited for a VGI approach. Wikimapia (wikimapia.org) is a VGI project that gathers information about places around the world for constructing gazetteers; Volunteers draw polygons representing places in their local areas on an imagery base map and contribute associated place names and descriptions (Ballatore and Jokar Arsanjani, 2019).

Other VGI provides versatile thematic information of geographic phenomena, for example, geo-tagged tweets capturing scenes of a wildfire, and geo-referenced entries reporting sightings of birds. This type of VGI is producing rich geographic information revealing spatiotemporal dynamics of the underlying phenomena and thus is of much interest to a variety of application domains. For instance, geo-tagged social media are used as a new approach for “social sensing” for understanding the socioeconomic environments (Liu et al., 2015); Records contributed by birdwatchers to eBird (ebird.org) on a daily basis are used to study bird distribution and migration (Sullivan et al., 2014).

4. Advantages of VGI

VGI has several advantages as an innovative mechanism of acquiring and compiling geographic data that could reveal spatiotemporal dynamics of social and natural phenomena. First, VGI has the potential of providing geographic data over large areas, as human footprints have reached much of the world. OpenStreetMap, Wikimapia, and eBird are all global-scale VGI projects that compile datasets across the whole world. Moreover, VGI contains rich local information that may span a wide temporal spectrum because citizens as local experts have accumulated knowledge of their environments over long time periods. As such, sightings of wildlife in historical periods are used to study habitat changes over time (Zhang et al., 2018). VGI can also provide timely updated geographic information that is difficult to obtain through traditional geographic data collection protocols (e.g., planned sampling, survey) but can easily be collected by citizen volunteers on the ground (e.g., damage reports after a major disaster). Lastly, VGI features much lower costs compared to traditional data collection protocols. This has made it feasible to produce large-scale geographic datasets through VGI initiatives (Sullivan et al., 2009).

5. Data Quality of VGI

Data quality of VGI is under constant scrutiny. As the general public engaged in creating VGI is not composed of well-trained professionals and their voluntary data collection actions are mostly constrained by internal commitment, data collected by volunteers may or may not be accurate (Goodchild, 2007). Assessment of VGI data quality provides the basic

information for users to evaluate the fitness for use of VGI in applications.

VGI quality is often assessed by examining VGI source credibility (Flanagin & Metzger, 2008) and spatial data quality indicators such as positional accuracy, attribute accuracy, temporal accuracy, semantic accuracy, logical consistency, completeness, and lineage (Goodchild & Li, 2012). Many studies have found that VGI data quality is satisfactory with respect to these dimensions. For instance, Olteanu-Raimond et al. (2016) found that much VGI data was acquired with a positional accuracy that, while less than that typically acquired by professional mapping agencies, exceeded the requirements of the nominal data capture scale used by most agencies. General approaches to ensure the quality of VGI are briefly summarized here based on an abundance of research (Goodchild & Li, 2012; Haklay, 2016; Senaratne, Mobasheri, Ali, Capineri, & Haklay, 2017): (1) "crowdsourcing"–using a group to validate and correct errors made by an individual contributor, (2) "social"–trusted individuals acting as gatekeepers to maintain and control the quality of contributions, (3) "geographic"–use of geographic knowledge to assess data quality, (4) "domain"–use of domain-specific knowledge to assess data quality, (5) data mining–discovering patterns by learning purely from data to assess data quality , (6) "instrumental observation"–removing some aspects of human subjectivity in data collection by relying on accurate equipment to improve data quality, and (7) "process-oriented"–participants going through training before data collection to ensure data quality. Readers interested in the details of each approach are referred to the original references.

Representativeness is yet another important aspect of VGI data quality that is especially relevant to the use of VGI containing thematic information for modeling and inference. The representativeness of VGI refers to the degree to which a "sample" consisting of VGI observations can represent the underlying "population". Observations in a VGI dataset is a sample drawn from the universe of all instances of the underlying geographic phenomenon (i.e., the population) (Jensen & Shumway, 2010). Analyses that involve inferring properties of the underlying population from a sample require the sample to be "representative." For example, the opinion of a larger group of people can be inferred from tweets only if the sampled Twitter users form a "representative" sample of that group. Species distribution modeling requires representative species records as input so that the modeled distribution is indicative of the species real distribution. Assessing the representativeness of VGI provides vital information on deciding whether VGI is suitable for such analyses.

Demographic biases in contributors is the major cause that impedes the representativeness of VGI to represent a larger group of people (Liu, Yuan, & Zhang, 2020; Malik, Lamba, Nakos, & Pfeffer, 2015). The fundamental issue behind such biases is that not all citizens have an equal opportunity to contribute to VGI due to reasons including (but not limited to) digital divide (e.g., urban/rural divide, unequal access to technology) (Hecht & Stephens, 2014; Sui, Goodchild, & Elwood, 2013). As an example, most contributions to eBird are from developed regions of the world, and the most intensively sampled areas are in proximity to large cities (Figure 2) (Zhang, 2020). Spatial bias is another common issue of VGI (Zhang & Zhu, 2018). Individual volunteers decide where to conduct observations and their observation efforts are often 'ad-hoc' and opportunistic in nature, which is radically different from traditional geographic sampling in which observation sites are carefully chosen to ensure the set of observations is representative. As a result, VGI records are often more concentrated in some geographic areas (e.g., populous or more accessible areas) (Zhang, 2020). Due to such spatial bias, VGI may not be representative of the spatial variation of the underlying geographic phenomena (Figure 2). Biases in VGI are widely

recognized and acknowledged, but only a few methods have been developed to tackle such biases for improving the representativeness of VGI observations for more reliable spatial modeling and predictions (Zhang & Zhu, 2019).

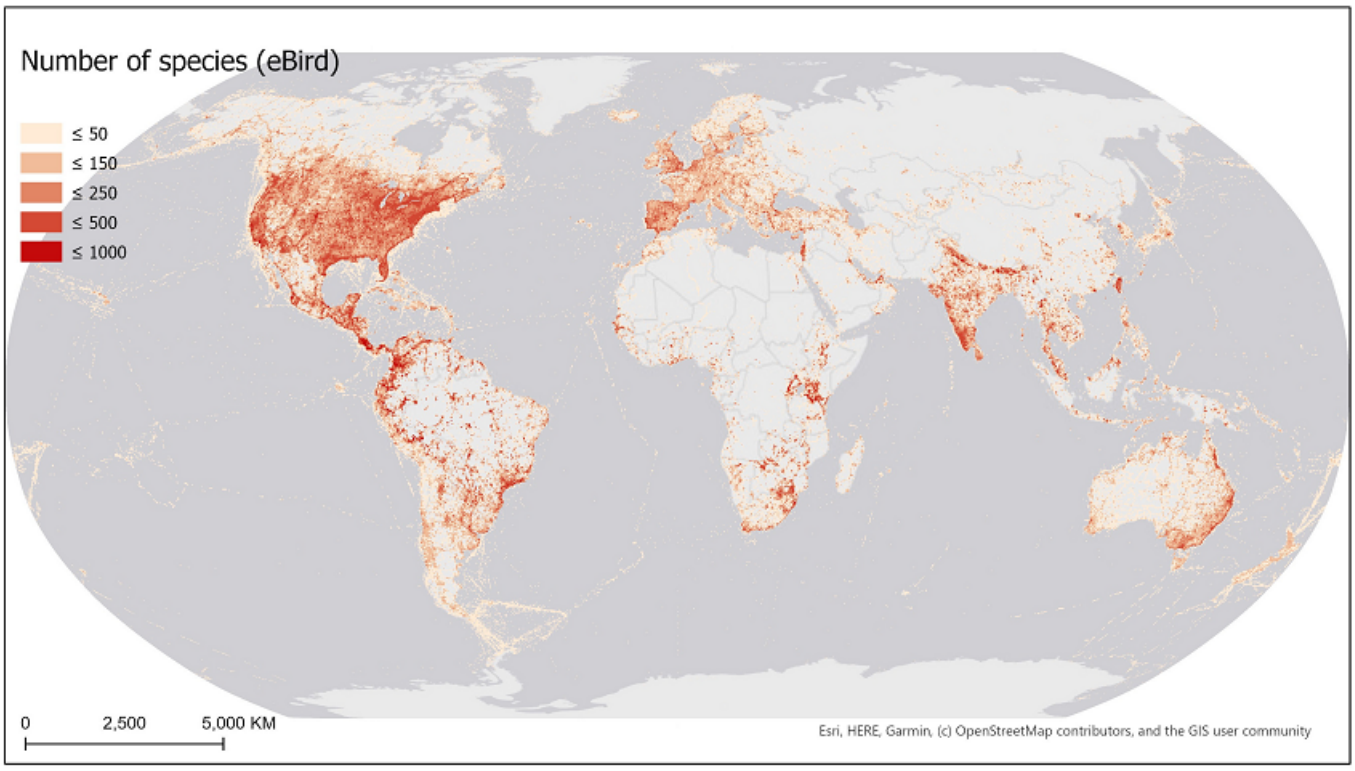


Figure 2. Number of species reported to eBird (as of December 31, 2019) mapped over a grid of 0.25° latitude x 0.25° longitude cells. Reported species are biased towards populous and accessible geographic regions, which do not necessarily represent the real spatial variation of bird diversity. The geographic regions with more species are mostly developed regions of the world and areas in proximity to large cities, which reflects digital divide (e.g., unequal access to technology and infrastructure). Source: author.

6. Privacy and Security

Privacy and security are serious concerns associated with VGI. Volunteers contributing geographic data to a VGI platform or database often expose their locations, actively or passively, willingly or unwillingly. Accumulated VGI contributions make it possible to track locations of individual users, which may pose serious privacy and security concerns to them (Elwood et al., 2012; Sui & Cinnamon, 2016). On one hand, VGI platforms should make VGI contributors fully aware of the intended use of the data they contribute, which is particularly important in cases where VGI is collected passively and users do not fully understand the process and its consequences. On the other hand, VGI contributors need to be vigilant sharing their locations or disclosing any sensitive information (e.g., identification) to reduce the risks of privacy invasion and security infringement. For example, when using mobile apps to contribute VGI, location and time information is often automatically collected at high accuracy. Such information makes it possible to reconstruct

contributor spatiotemporal trajectories, which poses risks to contributors (e.g., stalking). Some VGI mobile apps offer the option to obscure geographic coordinates when submitting observations (e.g., iNaturalist). Contributors may use this option to obscure observations made at sensitive locations (e.g., near their homes).

Moreover, there should be regulations in place for VGI data privacy protection, just as for any other user data. For instance, the usage of OSM data with metadata fields that could potentially reveal contributor identity (e.g., username, user ID) is governed by data protection regulations in the European Union because some OSM contributors live in the European Union. Use of the OSM data may be limited to OSM internal purposes, e.g. quality assurance. Any derived databases and works should be only accessible to OSM contributors.

7. Data Licenses and Copyright

The use of VGI data is often constrained by terms and conditions specified in respective data licenses and related documentations. Most VGI data are open and free for non-commercial uses. However, users may or may not distribute the data depending upon particular data licenses. For instance, users are free to copy, distribute, transmit and adapt OSM data under the same license, as long as credit is attributed to OSM and its contributors. Users can download the publicly available eBird data directly from the site but are prohibited from passing the data to others.

VGI contributors often hold copyright of the creative materials they contribute such as photos, audios and videos although the hosting VGI platform may by default assume the right to use the materials or sublicense the materials to a third party for non-commercial uses.

8. Challenges and Outlook

Data quality of VGI is at the core of VGI applications in various domains. Among the dimensions of VGI data quality discussed above, assessing the fundamental aspects of spatial data quality (e.g., positional accuracy, attribute accuracy) may provide sufficient information for evaluating the fitness for use of the first two types of VGI (geographic framework data, gazetteer data). Nonetheless, these aspects alone provide little insights on the representativeness of VGI observations, which in many cases is crucial for using the third type of VGI (thematic data) in modeling (e.g., inferring the opinion of a larger group of people from tweets; modeling and predicting species distribution from species occurrence data reported by volunteers).

Although various forms of biases in VGI have been identified and widely acknowledged, there are only limited methodological developments for tackling such biases. Research on data quality of VGI currently focuses more on issues at the data collection stage rather than on the impacts of the issues on VGI applications (e.g., modeling). Analytical methods that can accommodate and mitigate the biases are much needed for better using VGI in inferential analyses where the underlying phenomena of interest are inferred from a sample of VGI observations. The use of VGI for inference and modeling adds much value to VGI.

Therefore, addressing the issue of representativeness and biases of VGI is necessary to fulfill the full potential of VGI.

VGI is still an active research field that keeps evolving. While VGI are used in areas that previously relied on traditional data sources and there will always be needs for technical research around data quality, representativeness, etc., VGI is also being used to solve new problems as VGI is producing new data at spatial and temporal scales that were never possible to collect in the past (Goodchild, Aubrecht, & Bhaduri, 2017). For instance, eBird data are used for modeling avian full annual cycle distribution and population trends in the Americas (Fink et al., 2020), and the opportunities offered by crowdsourcing are exploited to generate traffic network databases to aid autonomous driving (Szántó & Vajta, 2019). The past decade or so has witnessed applications of VGI in a wide array of domains. Nonetheless, more future research is desirable to address the fundamental intellectual and scholarly needs that persist in the field, for example, understanding more fully how VGI operates, its implications, assumptions, limitations, affordances, etc., all are intrinsically important issues across VGI applications.

References

Ballatore, A., & Arsanjani, J. J. (2019). Placing Wikimapia: an exploratory analysis. International Journal of Geographical Information Science, 33(8), 1633–1650.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. Annals of the Association of American Geographers, 102(3), 571-590.

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M., & Kelling, S. (2020). Modeling avian full annual cycle distribution and population trends with citizen science data. Ecological Applications, 30(3), 1–16.

Flanagin, A., & Metzger, M. (2008). The credibility of volunteered geographic information. GeoJournal, 72(3), 137–148.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221.

Goodchild, M. F. & Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial Statistics, 1, 110-120.

Goodchild, M. F., Aubrecht, C., & Bhaduri, B. (2017). New Questions and a Changing Focus in Advanced VGI Research. Transactions in GIS, 21(2), 189–190.

Goranson, C., Thihalolipavan, S., & di Tada, N. (2013). VGI and Public Health: Possibilities and Pitfalls. In G. M. Sui D., Elwood S. (Ed.), Crowdsourcing Geographic Knowledge (pp. 329–340). Dordrecht, Netherlands: Springer.

Haklay, M. (2016). Volunteered Geographic Information: Quality Assurance. In International Encyclopedia of Geography: People, the Earth, Environment and Technology (eds D. Richardson, N. Castree, M.F. Goodchild, A. Kobayashi, W. Liu and R.A. Marston).

Haklay, M. & Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing, vol. 7, no. 4, pp. 12-18.

Hecht, B. & Stephens, M. (2014). A Tale of Cities: Urban Biases in Volunteered Geographic Information. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 197-205.

Jelokhani-Niaraki, M., Bastami Mofrad, R., Yazdanpanah Dero, Q., Hajiloo, F., & Sadeghi-Niaraki, A. (2019). A volunteered geographic information system for monitoring and managing urban crimes: a case study of Tehran, Iran. Police Practice and Research, 1–15.

Jensen, R. R., & Shumway, J. M. (2010). Sampling Our World. In B. Gomez & J. P. Jones III (Eds.), Research Methods in Geography: A Critical Introduction (pp. 77–90). John Wiley & Sons.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding Human Mobility from Twitter. PLOS ONE, 10(7), e0131469.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... Shi, L. (2015). Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. Annals of the Association of American Geographers, 105(3), 512–530.

Liu, Y., Yuan, Y., & Zhang, F. (2020). Mining urban perceptions from social media data. Journal of Spatial Information Science, 20(20), 51–55.

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population Bias in Geotagged Tweets. Ninth International AAAI Conference on Web and Social Media, May 26–29, 18–27. Oxford, United Kingdom.

Miller, H. J. and Goodchild, M. F. (2016) Data-Driven Geography. GeoJournal, 80: 449-461.

Olteanu-Raimond, A.-M., Hart, G., Foody, G. M., Touya, G., Kellenberger, T., & Demetriou, D. (2017). The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies. Transactions in GIS, 21(1), 74–90.

Rahimi, S., Mottahedi, S., & Liu, X. (2018). The geography of taste: Using Yelp to study urban culture. ISPRS International Journal of Geo-Information, 7(9), 1–25.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science 31(1), 139-167.

Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. Annals of the Association of American Geographers 96(3), 491-507.

Sui, D. and Cinnamon, J. (2017). Volunteered Geographic Information. In International Encyclopedia of Geography: People, the Earth, Environment and Technology (eds D. Richardson, N. Castree, M.F. Goodchild, A. Kobayashi, W. Liu and R.A. Marston).

Sui, D., Goodchild, M., & Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice (pp. 1–12). Dordrecht: Springer Netherlands.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., … Kelling, S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation, 169, 31–40.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation, 142(10), 2282–2292.

Szántó, M., & Vajta, L. (2019). Introducing CrowdMapping: A Novel System for Generating Autonomous Driving Aiding Traffic Network Databases. 2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), 7–12.

White, J. J. D., & Roth, R. E. (2010). Twitterhitter: Geovisual Analytics for Harvesting Insight from Volunteered Geographic Information. Proceedings of GIScience (2010) in Zurich, Switzerland.

Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science.

Zhang, G. (2020). Spatial and Temporal Patterns in Volunteer Data Contribution Activities: A Case Study of eBird. ISPRS International Journal of Geo-Information, 9(10), 597.

Zhang, G., & Zhu, A.-X. (2018). The representativeness and spatial bias of volunteered geographic information: a review. Annals of GIS, 24(3), 151–162.

Zhang, G., & Zhu, A.-X. (2019). A representativeness-directed approach to mitigate spatial bias in VGI for the predictive mapping of geographic phenomena. International Journal of Geographical Information Science, 33(9), 1873–1893.

Zhang, G., Zhu, A.-X., Huang, Z.-P., Ren, G., Qin, C.-Z., & Xiao, W. (2018). Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena. Transactions in GIS, 22(1), 149–164.