societies

MDPI

Review

User Spatial Content in Social Research: Approaches, Opportunities, and Challenges

Ciro Clemente De Falco

Department of Social Sciences, University of Naples "Federico II", 80138 Napoli, Italy; ciroclemente.defalco@unina.it

Abstract: The availability of user-generated spatial data (user spatial content, USC) has transformed social science research, enabling the real-time, large-scale exploration of socio-spatial dynamics. This article traces the evolution from volunteered geographic information (VGI) to USC, highlighting their multidimensional nature and epistemological significance. Brief examples underscore USC's potential for capturing the interplay between territorial factors, digital activity, and social phenomena, ranging from mapping urban vitality to tracking large-scale crises. However, the recent tightening of data access in the post-API era demands a rethinking of research approaches. Alternatives such as data donation, dedicated applications, and geoparsing can maintain the viability of USC-driven analyses. Overall, this article underlines the need for diversified, ethical, and methodologically sound strategies to harness USC's value in understanding the digitally intertwined realities of contemporary society.

Keywords: user spatial content; post-API era; spatial data

1. Introduction

The availability of spatial data produced by users has revolutionized contemporary social science investigations, enhancing how researchers comprehend and examine complex relationships among society, space, and networks. This content, such as social media updates, geotagged photographs, or contributions to collaborative mapping initiatives, provides more than just new sources of information. This presents a unique opportunity to examine social and spatial dynamics in real time and on a large scale. The sheer volume and diverse geographic origins of this content have paved the way for ambitious research in this area. These studies can now capture intricate patterns and explore all countries in unprecedented detail. User-generated content has enabled researchers to examine the interplay between virtual interactions and physical spaces, thereby fostering new avenues for cross-disciplinary research. However, new obstacles arise as these research opportunities expand, particularly platform restrictions and growing limitations in accessing data through free application programming interfaces (APIs). This trend threatens research that relies heavily on the widespread availability of geographically tagged information. Therefore, there is a need to develop innovative approaches to continue using these data. Thus, this paper aims to (a) briefly reconstruct the debate on the concepts regarding geographic data from users, thereby proposing a concept that is considered more semantically correct; (b) highlight their value by viewing the different approaches and related epistemological characteristics generated; (c) demonstrate the potential of USC-based analyses through three illustrative research examples that showcase diverse contexts and methodological perspectives; and (d) identify alternative strategies for addressing emerging challenges, ensuring the sustainability of research in an increasingly limited data access context.

check for updates

Academic Editors: Felice Addeo and Gabriella Punziano

Received: 21 February 2025
Revised: 31 March 2025
Accepted: 3 April 2025
Published: 8 April 2025

Citation: De Falco, C.C. User Spatial Content in Social Research: Approaches, Opportunities, and Challenges. *Societies* **2025**, *15*, 96. https://doi.org/10.3390/soc15040096

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. From Volunteered Geographic Information (VGI) to User Spatial Content (USC)

2.1. The Origin and Characteristics of VGI

In a well-known article, Goodchild [1] gave notoriety to the volunteer geographic information (VGI) concept by framing it as a type of user-generated content (UGC). The proposal was born within the emerging debate on the type of content web users produce concerning Web 2.0 and the phenomenon of prosumerism [2]. However, what is VGI? VGI is content created by web users that concerns the territory and contains geographic information. The possibility of its creation was enabled by the emergence and/or spread of specific technologies, such as GPS and georeferencing (which allow direct measurement and recording, even with mobile phones and cameras, of the position on the Earth's surface with remarkable precision), geotagging (which involves adding geographic information to content, such as photos), and high-resolution graphics and broadband connections (which allow for quality maps and faster uploading of geographic content). Thus, geographical information, which used to be the prerogative of states and national institutes, can now be produced by web users without special technical training. All it takes is a mobile phone (and its technology) and an Internet connection to produce geographic information. From this perspective, the paradigm shift is radical, and it is comprehensible that the person who coined the term "data revolution" is a geographer. Since the early 2000s, thousands of users have started to produce multiple geographic information sources [1]: (a) they build maps, contributing to projects such as OpenStreetMap, by adding roads, buildings, and local details, and even filling gaps in official maps; (b) they describe places with tools such as Wikimapia, by selecting geographical areas, adding descriptions, and linking additional information to enrich the understanding of places; (c) they geotag content both photographic (as on Flickr (San Francisco, USA) and textual (as on X (formerly Twitter,(San Francisco, USA); and (d) they create geographical mash-ups, with tools such as Google Maps, overlaying information from different sources.

The varied production of geographic information by web users has become the object of analysis aimed at identifying and distinguishing differences.

The first question concerns the motivations that drive users to produce such data. Coleman and colleagues [3] identified four main motivations. The first is related to general interest, which includes the attraction to technological novelty, the desire for personal learning and the pleasure derived from it, the convenience of sharing information, and personal satisfaction. The second most common motivation is altruism. It regards the improvement in fundamental datasets for the common good, contribution to the community, and alignment with altruistic and reciprocal visions. The third motivation, related to egotism, focuses on building social capital, self-promotion, and reputation. Finally, the fourth motivation relates to power, that is, the desire to influence political processes or achieve specific political goals.

The second line of thought focuses on the potential and limitations of VGI data. Gouveia and Fonseca [4] highlighted some critical issues, including the heterogeneity in data quality and lack of metadata. Furthermore, if VGI considers community initiatives, there is a need for training and supervision to ensure continued commitment from volunteers. These technical challenges are compounded by social issues, such as the risk of vandalism and the not in my backyard (NIMBY) effect, which can undermine the community purpose of these initiatives. However, VGI also offers significant advantages: it makes it possible to monitor neglected areas by institutions at low cost, helps to raise public awareness of environmental issues, and provides access to rich and diverse data through voluntary participation, thus expanding the available information base and providing new perspectives for spatial analysis.

2.2. Towards USC

The final consideration concerns the various forms of VGI available on the Internet, which enable users to incorporate geographical data into their content in diverse ways and for different purposes. This reasoning was incorporated in the proposal by Craglia and colleagues [5], who identified four types of VGI based on two intersecting criteria. The first criterion focuses on whether the content produced by the user includes territorial elements directly or indirectly. The second criterion considers whether the geographical information linked to the user's content is explicitly or implicitly expressed. Among the four types identified by the authors, only the first fully matched the traditional concept of VGI, in which geographical information is the primary focus. The other three types broaden the definition by including content in which geographical information can either be derived indirectly or play a secondary role. This typology thus extends the meaning of voluntary, shifting the emphasis from content explicitly centered on a territory to content that at least mentions it. From a geographical perspective, the analysis transitioned from focusing solely on content with direct georeferencing to including content with indirect georeferencing. This classification is helpful because it encapsulates both the classic VGI as a description of a geographical place uploaded on a collaborative mapping platform (direct–direct) and a tweet where the user talks about an activity in a city location (indirect–indirect). The focus on VGI prompted an exploration of the possibilities offered by the users' geographical data, which led to identifying and classifying other data sources. These sources do not fit the previous classification because they are not voluntary. In some cases, users do not willingly provide geographic information, but these data are collected either passively or indirectly. Examples are geographic data from mobile phone calls or even data on individuals' movements that apps such as Google record automatically and continuously after a generic concession by the user. Because it is impossible to label these data as VGI, a broader concept is needed. In the literature, there are several terms that can include the two classes of data just described, such as big spatial data or big geospatial data. These labels have the quality of placing the user's voluntariness or otherwise in producing geographic information in the background. However, they are overly inclusive, as they encompass spatial data that users do not produce.

Moreover, not all VGI is necessarily "big". Among the terms proposed, it is also possible to find crowdsourced geographic information (CGI), which emphasizes the process of data collection by eliminating the problem of voluntariness; however, the term points to the concept of "crowd", implying widespread and collective participation, which, however, is not always the basis of the data.

The term "user spatial content" describes geographical data produced voluntarily or involuntarily by a variable number of users. The word "user" clearly indicates to whom the data pertain, without necessarily implying voluntariness or a specific number of contributors. The concept of "spatial", instead of "geographic", more effectively encompasses unstructured information about places. Finally, "content", instead of "information", does not assume any form of processing applied to spatial data. For conceptual reasons, instead of VGI or CGI, from now on, we talk about USC. However, as is shown in the following paragraphs, the central aspects of USC concern the type of use and the way it is collected.

3. Epistemological Orientation of USC

Multiple reviews have highlighted the roles of different types of USC as a data source for analyzing a wide range of social science phenomena, from quality-of-life studies to risk assessment, monitoring, management, activism, and social movements. Moreover, studies have been conducted on linguistics, nutrition, and travel habits. Many phenomena can be investigated using USC but listing them was beyond the scope of this study. One

way to systematize the literature on the subject is through an epistemological lens. This criterion has gained prominence with the data revolution, as the advent of big data has sparked epistemological debate between data-driven and theory-driven approaches. The data-driven approach asserts that "data can speak for themselves", suggesting a decreasing need for theory. This perspective promotes the idea that empirical observations can be made independent of theoretical frameworks [6]. However, this claim has sparked debate, as critics argue that data interpretation inherently requires theoretical guidance to contextualize findings and derive meaningful insights. Furthermore, this approach recognizes that data are not neutral but are inherently shaped by the theoretical assumptions embedded in their construction. Empirical evidence is always the result of an interplay between theory, evidence, and a broader knowledge framework that includes decisions and assumptions underlying the creation of specific datasets [7]. Two key points must be emphasized in this debate. The first arises upstream of the cognitive process and the other downstream. The first relates to the fact that the data are neither neutral nor objective but the result of a process that incorporates theory, methods, and techniques with different gradations. Data are not collected but re constructed. The second point concerns the data interpretation phase: empirical evidence can be sterile or easily misinterpreted without a theoretical orientation. This issue is particularly relevant when working with USC in the social sciences, where the concept of territory plays a central role. A territory that is far from being a neutral container actively shapes and influences social phenomena. Multiple theoretical frameworks address how territory interacts with and impacts these phenomena, underscoring the need for theories guiding the analysis of USC. Durkheim [8] demonstrated how the social environment shapes individual behavior, providing a foundational framework for understanding space as a sociological construct rather than merely a physical dimension. In this sense, the ecological approach allows researchers to explore how space characteristics influence phenomena, thus distinguishing between data-driven perspectives (which ignore these elements) and theory-driven perspectives (which integrate them). Integrating theory-driven perspectives requires combining two data types to explore possible relationships: online data (particularly USC data) and offline data (such as secondary data collected in the field or socioeconomic data on users).

Approaches to Using USC

The two main families of approaches are data-driven and theory-driven. In the data-driven approach, the collected data are treated as self-reliant for interpretation. Although this approach, as noted above, may lack the ability to achieve analytical depth, it has been proven to be effective in specific contexts. For instance, geolocated data from social media have been utilized to support the management of critical situations, such as emergencies. Event detection algorithms [9] integrated into social media monitoring systems harness geolocated data to identify areas affected by catastrophic events, enabling timely intervention. Furthermore, these algorithms detect emerging urban dynamics requiring immediate attention, such as unusual activity patterns or localized disorders [10].

In theory-driven approaches, the data are framed within a pre-existing theoretical framework in which the context and sociological characteristics of space play a relevant role. In these approaches, space is conceived as a geographical and sociological construct that influences and shapes the phenomena studied. Thus, USC data analysis is closely linked to territorial analysis in these approaches. Within this family, it is possible to distinguish at least three approaches that differ in the relationship structuring between online and offline phenomena: (a) explanatory, (b) same plane, and (c) reconstruction.

In the explanatory approach, the socioeconomic characteristics of the ecological unit, such as neighborhoods or cities, are considered determinants that can influence the phe-

nomena detected by the USC and vice versa. Some qualitative studies have focused on comprehending how the online world can influence the offline world, seeking to understand how digital representations on social media can alter the perception and meaning of physical environments through visualization and designation. This approach explores how representation spaces can transform spatial practices [11]. For example, Sutko and de Souza Silva [12] investigated the connections between social and spatial dimensions through geosocial applications and services, highlighting the impact of these technologies on the social production of space and spatial production of society, focusing on the transformation of relational dynamics, such as the sociability analyzed by Simmel [13]. Other quantitative studies have focused on explaining the variability in phenomena investigated through statistical models in which socioeconomic (offline) variables are considered independent. This growing body of research, articulated at multiple levels of geographic detail, has investigated how the spatial dimension influences online events, particularly on social platforms. This type of study follows the tradition of authors such as Durkheim [14], who adopted an ecological approach to link and explain social phenomena through their spatialization and territorialization [15].

In the so-called "same plane" approach, online and offline spaces are treated as independent variables influencing each other, thus overcoming the traditional separation of physical and digital reality. This approach emphasizes how both contexts are part of the same sociological reality and can mutually impact behaviors and interactions. The study conducted by Ristea and colleagues [16] examined the relationship between socioeconomic variables and the content of tweets regarding crime. The researchers used the socioeconomic characteristics of the ecological unit and the content of tweets as independent variables to explain phenomena such as the incidence of criminal acts. The study reflected the "same plane" approach because the space's communicative and socioeconomic aspects were considered in the same way in the explanatory model, without a clear distinction between online and offline. This integration of communicative and social elements echoes American sociological tradition, with a perspective that valorizes both communicative and socioeconomic contexts.

In the reconstruction approach, spatial characteristics are used descriptively or discursively to interpret phenomena rather than the main variables in analytical models. This approach offers a qualitative narrative of social space, often without applying statistical techniques or using operational definitions of spatial characteristics. Beiró [17] explored the movements of social classes within a city and interpreted how different social classes moved according to the characteristics of neighborhoods and streets. In this case, the characterization of space was not based on rigorous quantitative analysis but rather on a discursive reconstruction of the relationship between space and social structure. The socioeconomic elements of space were employed as a narrative context to describe phenomena rather than variables to be measured statistically. This approach focuses on a qualitative understanding of sociological space, using interpretation to portray how spatial features influence social behavior.

As discussed, these three approaches highlight varying degrees of integration between online and offline phenomena. However, the explanatory approach stands out for its analytical depth and capacity to link social phenomena to spatial dimensions using statistical models and quantitative analysis. This approach explains the variability in online events through offline socioeconomic variables and offers a robust framework for understanding how territorial structures influence digital dynamics. Thus, the explanatory approach is particularly effective in unveiling the interactions between digital narratives and spatial contexts, thus providing a richer understanding of the spatialization and territorialization of social phenomena. Therefore, it is crucial to explore how this approach is applied in

different studies to reveal the potential of geo-media in social research. In the next section, we examine some empirical examples that illustrate the explanatory approach in action, demonstrating its effectiveness in territorializing digital narratives and linking online dynamics to offline structures.

4. Explanatory Approaches to Socio-Spatial Dynamics: Three Case Studies Using Geolocated Twitter Data

The examples we examined adopted an explanatory approach and relied on a single data source, X (formerly Twitter). Geolocated tweets were used to demonstrate how offline spatial characteristics influence digital social dynamics. There are three primary examples: in the first, we illustrate how USC can be used for cognitive purposes when analyzing communicative phenomena; in the second, we focus on the study of urban phenomena; and, in the third, we emphasize the methodological potential of USC analysis.

Before illustrating the individual examples, it should be clarified that we do not aim to provide an exhaustive overview of the state of the art on USC but rather to highlight the potential of this type of data from an explanatory perspective. For this reason, research dealing with heterogeneous phenomena was selected to demonstrate the versatility of the approach in correlating digital dynamics with offline characteristics in different contexts.

In particular, the studies presented are from previous research, all designed from the perspective of the same "explanatory frame". Moreover, the analyses were based on data from a single platform, X (). This "invariability" was due both to practical reasons and methodological considerations: on the one hand, Twitter has long provided relatively easy access to geolocated data and related metadata, facilitating large-scale collection and management; on the other hand, the nature of the information provided has been proven to be particularly suitable for an explanatory approach.

The first example derives from a study conducted with Punziano and Trezza [18], who analyzed approximately 12,000 geolocated tweets related to the pandemic and posted during that period. The study aimed to show how offline determinants, such as the spread of contagion and region-level regulatory measures shape digital narratives. To achieve this, the researchers grouped tweets by region and applied automated content analysis techniques, revealing a close interplay between digital narratives and the spatial dimensions of the pandemic. Specifically, the geography of contagion (COVID-19 spread) showed that narratives of fear and worry dominated in regions with higher contagion rates, whereas more neutral or resilient discourses emerged in less-affected areas. This distribution reflected the local experience of risk, indicating that the spatialization of contagion guided collective emotional processing. Similarly, the geography of containment measures (COVID-19 measures) demonstrated that normative differences across regions (red, orange, and yellow zones) influenced digital narratives. More critical or negative discourses emerged in the red zones, while moderate or informative narratives predominated in the yellow zones, highlighting how mobility restrictions shape social interpretations and emotional responses. Finally, the geography of emergent narratives (COVID-19 issues) showed that recurring topics evolved over time in response to territorial shifts in contagion and regulatory measures. The study confirmed how digital narratives adapt to local changes by employing topic modeling and clustering, thus territorializing collective discourse according to lived experiences.

The second study, conducted by Lenzi and Mari [19] on urban vitality in Rome's neighborhoods, collected about three million geolocated tweets in Rome between 1 January 2020 and 31 March 2023 to investigate how digital dynamics mirrored the distribution of urban vitality. These high-vitality areas clustered around monuments, bars, clubs, and shopping centers, confirming urban attractions' importance in generating social and digital

activities. Such locations reflected the concentration of Twitter users, correlating with infrastructural features that fostered sociability and public interaction. By examining tweeting patterns during the pandemic, the study revealed a clear reduction in social activity in central areas and a relative increase in residential zones, showing the effects of lockdown measures. The work demonstrated how the spatialization of digital narratives provided new perspectives on the relationship between physical space and digital discourse, thus confirming the potential of USC.

In the last study, titled "Geo-Social Media and Socio-Territorial Distribution: A Study on the Italian Case", De Falco and Ferracci [20] highlighted the representativeness constraints and potential biases in social media data by analyzing about 90,000 geolocated tweets from 2018 to 2020, obtained from the Twita database. The focus was on two spatial dimensions: first, city size, where large municipalities display a proportional increase in tweet volume compared to smaller ones and rural areas; and, second, the center–periphery relationship, where the data point to a concentration of users in the central districts of large cities, contrasting with peripheral zones. This polarization reflected socioeconomic inequalities and distinct social practices related to urban space usage. Notably, neighborhoods with a higher social advantage index showed a denser presence of Twitter users, indicating that local socioeconomic conditions affected the access to and use of digital technologies.

These examples demonstrate how the explanatory approach effectively correlates digital social dynamics with offline territorial characteristics using USC data. These instances vividly demonstrate the potential and adaptability of utilizing USC to examine spatial–social phenomena. One of the main advantages of USC analysis is its ability to collect data with high detail and immediacy, allowing researchers to swiftly capture and examine spatialized social dynamics in real time and on a large scale. As shown in the first example, USC analysis offers innovative insights into complex and rapidly changing phenomena such as pandemics, enabling researchers to link digital narratives with offline factors such as the spread of contagion and regulatory actions. Similarly, the second study highlights how USC methods can detect subtle spatial patterns and hotspots of social activity, providing detailed interpretations of urban vitality and shifts in public space usage during extraordinary events, such as lockdowns. The third example further illustrates the analytical power of USC to investigate socioeconomic disparities and spatial inequalities in digital behaviors, offering valuable explanatory frameworks that effectively connect digital interactions with physical spaces.

However, it is important to recognize critical methodological challenges concerning bias and representativeness. Geolocated data from platforms such as X (formerly Twitter) are inherently selective, often reflecting younger, digitally active, urban populations with socioeconomic advantages. As a result, spatial analyses risk systematically under-representing marginalized groups, rural populations, and socioeconomically disadvantaged communities, as explicitly noted in the third study. Dependence on data from a single platform exacerbates these limitations, as each platform attracts specific demographics and encourages particular digital behaviors, potentially diminishing the generalizability and robustness of the findings.

In conclusion, despite these important limitations, USC approaches can still represent a valuable source of insight for social research, provided they are used cautiously and critically. However, in the post-API era [21], restricted access to geolocated data jeopardizes the USC's potential. To preserve this approach's explanatory power, it is crucial to adapt methodological frameworks and consider alternative data sources, ensuring continued insight into the link between digital narratives and physical space.

5. What Is the Future of USC Use in the Social Sciences

5.1. From Data Explosion to the Post-API Era

USC can be derived from multiple sources, including collaborative mapping projects, points of interest, smartphones, and social platforms. For a certain time, social platforms were widely considered rich data sources. In the social sciences, discussions around data collection from platforms recognized two key periods. The first, emerging at the beginning of the data revolution, is characterized by a "data explosion", where access to platform data seemed boundless. The second period, by contrast, has been described in various ways—"post-API", "API-closure", or, more ironically, "API-calypse"—reflecting the significant restrictions that followed.

The first period was characterized by relatively easy access to data. APIs allowed users with technical expertise to quickly and freely retrieve vast amounts of information, including geolocated data. Many companies and organizations provided free API, encouraging a surge in research that leveraged this abundance of data for various applications.

Among the platforms most frequently used by Western researchers, as we have just seen in the examples, Twitter allowed users to geotag their posts by providing latitude and longitude information. Although only around 1 percent of users enabled geotagging, the sheer volume of geotagged data generated was substantial, making them valuable resources for research purposes. This widespread interest in Twitter data led to the creation of specialized databases that included geotagged tweets. For instance, the Italian Twitter experience aimed to compile a database containing all tweets shared in Italy, whereas the Leibniz Institute developed TweetsKB, a public corpus of anonymized and annotated tweets that provided researchers with access to a structured and rich dataset for their analyses.

Therefore, this period was characterized by easily obtainable data, including USC data. Owing to multiple factors, including the increasing focus on user privacy, platforms' policies for data sharing have changed. In some cases, increasingly stringent restrictions have been imposed to the point of complete shutdown.

Facebook (Menlo Park, Usa) initially allowed generous access to user data through the Graph API, allowing developers and researchers to easily retrieve detailed information such as users' social networks and personal interests [22]. However, the Cambridge Analytica scandal in 2018, in which data from 87 million users were misused, was a radical turning point [23]. Since then, Facebook has imposed increasingly strict restrictions: previously available endpoints have been removed, accessible information has drastically reduced, and obtaining data have become complex and highly regulated [24]. Currently, to access data via API, developers must pass a rigorous approval process, clearly demonstrating the intended use of the data and often obtaining direct validation from the meta team [25]. Facebook is among the most emblematic examples of API restriction; however, it is equally important to discuss Twitter within the USC discourse. Twitter has been a favored resource for searching for years owing to its broad access to data via API, including the ability to collect geolocated tweets [26]. Although not directly involved in the Cambridge Analytica scandal, the platform gradually introduced restrictive measures as early as 2019 to avoid similar abuses and protect user privacy [27]. With the transition to API v2 in 2021, many previously free features were eliminated or scaled back [28]. The most significant change, however, occurred in February 2023, after Elon Musk acquired the platform, when Twitter permanently discontinued free access to its API, forcing users and researchers to adopt a pay-only model [29].

5.2. New Sources for USC

With the closing of API, 0-cost research on USC has not necessarily halted. There are other ways of collecting geolocated data without paying for them. The most recent are

(a) data donation, (b) app development, and (c) geoparsing and geocoding. Data donation is an ethical and innovative method for collecting digital traces left by users during online interactions. In social research, data donations allow researchers to obtain granular and accurate data that can be integrated into other data sources or used individually. In Europe, data donation is based on the possibility offered by GDPR for users to request a copy of their personal data (data download package (DDP)) from digital platforms and then decide whether and which of these data are shared with researchers. This method appears to be ethical because it allows data to be collected with the active consent of users, overcoming the data access limitations encountered with "platform-centric" collection methods such as scraping or API. As indicated by Carrière et al. [30], this procedure has limitations. First, voluntary participation may introduce selection bias because users who choose to share their data may not be representative of the general population. In addition, access to DDPs may vary among platforms and not all allow easy data export, reducing the possibility of obtaining complete and consistent datasets that limit the generalizability of the results.

Additionally, the technical complexity of extracting and sharing data download packages (DDPs) requires a certain level of technical expertise. Skills are also required to manage and analyze large amounts of unstructured data. Finally, access to DDPs can vary among platforms and not all allow easy data export, reducing the possibility of obtaining complete and consistent datasets. Finally, protecting user privacy remains a crucial challenge because even with active consent, sharing granular data can expose sensitive or identifying information. These challenges underscore that while data donation offers significant benefits, it also presents critical issues that need to be addressed. Creating an app that, once downloaded by users, allows them to collect their movements and statuses from major social media platforms is a strategy that could prove complex but fruitful. The goal of such an app would be to collect geolocated data and information related to the content shared by users, such as posts, status updates, images, or reactions, and then use it for various types of analysis and research, allowing for extremely detailed and up-to-date data in real time. In addition, the app could engage users by offering personalized features, such as reports on their behaviors or suggestions based on the data collected. This approach not only increases user interest but also offers added value. In addition to the aforementioned issues, data donation presents additional critical challenges that must be addressed. Social platforms may not allow access to content; therefore, strategies must be developed to address these limitations. In addition, the design should be carefully calibrated to meet the limits imposed by the GDPR, which requires transparency, privacy protection, and the ability of users to withdraw their consent at any time. However, developing and maintaining apps requires significant investment, from the technical infrastructure to ensuring data protection and legal compliance. While challenges related to privacy, trust, data quality, and costs must be addressed, the potential value of real-time user-generated data can justify this effort. Geoparsing and geocoding are complementary processes that transform geographical references into structured geospatial data. Specifically, geoparsing identifies and interprets places mentioned in unstructured texts, distinguishes entities such as cities, regions, or monuments, and resolves contextual ambiguities [31]. Geocoding, on the other hand, associates these references with precise geographic coordinates (latitude and longitude) using databases such as GeoNames or OpenStreetMap. Together, these algorithms make it possible to map geographic information and thus identify it from the content shared on the Web. This procedure implements what is called indirect georeferencing, as opposed to direct georeferencing, where latitude and longitude are immediately available. Both geoparsing and geocoding have critical aspects that make them simple and automatic procedures. Geoparsing, or the recognition of toponyms (place names) in a text, can be considered a specific category of named entity recognition and classification (NERC). How-

ever, distinguishing places from other ambiguous terms requires correctly interpreting the linguistic context. In contrast, geocoding (or toponym resolution) focuses on associating a toponym with its spatial footprint, such as a geographic coordinate pair. Again, the main challenge lies in resolving ambiguities: many place names are homonyms, and the textual context is not always sufficient to determine their exact location. Both processes require advanced systems and the strong integration of NLP tools and geospatial datasets to ensure accuracy and efficiency. Moreover, in addition to inheriting all the limitations of geolocated data seen earlier, these techniques assume the use of a source as a source of textual data from social media, which are increasingly difficult to find.

In light of these considerations, it is clear that each of the three strategies offers unique opportunities and challenges for detecting CGI. The first two focus on users' active participation in the data collection process and their explicit authorization. In contrast, geoparsing and geocoding rely on platform-centered logic to obtain data. Beyond this difference in approach, strategies vary in terms of their strengths and limitations. The first two methods provide richer datasets but make it more challenging to achieve consistent samples. Conversely, geoparsing and geocoding techniques ensure greater sample consistency but at the cost of reduced data richness. From a logistical and technical standpoint, all strategies present challenges, although the first two require significantly more effort. Ultimately, there is no universally preferred approach; the choice depends on the specific objectives of the research as well as the time and resources available.

Beyond these methodological reflections, ethical issues must be given due attention. Methods that involve user engagement, such as data donations or custom applications, can collect more detailed data with explicit consent. However, they also pose a risk of sampling bias and require stringent privacy measures, particularly when dealing with potentially identifiable information. On the other hand, platform-based techniques such as geoparsing and geocoding can collect data without direct user consent, which raises concerns about surveillance, data sovereignty, and the potential to perpetuate existing biases. In all approaches, researchers should aim for transparency, data minimization, and strong anonymization practices to protect individual rights and build public trust. Tackling these ethical issues is not merely a regulatory requirement but an essential part of ensuring that research based on USC serves the wider social good without compromising the privacy or autonomy of the individuals involved.

6. Conclusions

User spatial content is a valuable source of data for multiple areas of research, offering the opportunity to integrate spatial dimensions within network dynamics and social phenomena. This integration makes it possible to combine the two worlds previously considered separate: online and offline. As highlighted, there are several approaches to analyzing these two dimensions and their interactions, each with unique methodological perspectives and potential applications in the social science context. However, in recent years, access to this data, particularly from social media, has become increasingly difficult. The increasing proprietary closure of platforms, accompanied by the divestment of free APIs, has made it nearly impossible to continue using USC derived from social media, such as Twitter. This closure is particularly significant because, despite methodological issues, such data have played a crucial role in numerous academic studies. Their large numbers and heterogeneous geographic origins have enabled ambitious research aimed at studying entire national territories, providing insights into complex and territorially distributed phenomena. In addition to the loss of data access, there is a gradual obsolescence of procedures and methodological approaches specifically developed to leverage USC from certain platforms. This issue underscores the inherent risk of relying on private infrastructure for

scientific research, as these platforms can change policy access or discontinue operations without prior notification. Therefore, it is important to identify new ways of sourcing USC and continue to exploit its potential. As discussed, three alternative strategies have shown promise: data donations offer an ethical and transparent approach but present challenges related to the representativeness of the sample and the technical complexity of data processing; dedicated app development can provide detailed and real-time data but requires significant resources for design, implementation, and compliance with privacy regulations; and geoparsing and geocoding represent viable tools for extracting geographic information from textual content, although they are constrained by the availability of sources and the quality of unstructured data. Each solution has advantages and limitations that must be carefully considered according to the research objectives and available resources. Adopting diversified strategies and investing in building alternative infrastructure could be crucial steps in ensuring the sustainability of USC-based research, preserving its enormous potential for studying the interactions between spaces, networks, and social phenomena.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: I thank the anonymous reviewers and the editorial staff for their constructive feedback and support.

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

References

1. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [CrossRef]
2. Ritzer, G.; Jurgenson, N. Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'. *J. Consum. Cult.* **2010**, *10*, 13–36. [CrossRef]
3. Coleman, D.J.; Sabone, B.; Nkhwanana, N.J. Volunteering geographic information to authoritative databases: Linking contributor motivations to program characteristics. *Geomatica* **2010**, *64*, 27–39.
4. Gouveia, C.; Fonseca, A. New approaches to environmental monitoring: The use of ICT to explore volunteered geographic information. *GeoJournal* **2008**, *72*, 185–197. [CrossRef]
5. Craglia, M.; Ostermann, F.; Spinsanti, L. Digital Earth from vision to practice: Making sense of citizen-generated content. *Int. J. Digit. Earth* **2012**, *5*, 398–416. [CrossRef]
6. Anderson, C. The end of theory: The data deluge makes the scientific method obsolete. *Wired Mag.* **2008**, *16*, 16-07.
7. Amaturo, E.; Aragona, B. Per un'epistemologia del digitale: Note sull'uso di big data e computazione nella ricerca sociale. *Quad. Di Sociol.* **2019**, *81*, 71–90. [CrossRef]
8. Durkheim, E. *La Divisione del Lavoro Sociale*; Edizioni di Comunità: Milano, Italy, 1962.
9. Nurwidyantoro, A.; Winarko, E. Event detection in social media: A survey. In Proceedings of the International Conference on ICT for Smart Society, Jakarta, Indonesia, 13–14 June 2013; pp. 1–5.
10. Wei, L.; Magee, D.R.; Cohn, A.G. An anomalous event detection and tracking method for a tunnel look-ahead ground prediction system. *Autom. Constr.* **2018**, *91*, 216–225.
11. Rzeszewski, M. Geosocial capta in geographical research—A critical analysis. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 18–30. [CrossRef]
12. Sutko, D.M.; de Souza e Silva, A. Location-aware mobile media and urban sociability. *New Media Soc.* **2011**, *13*, 807–823. [CrossRef]
13. Simmel, G. *The Sociology of Georg Simmel*; Wolff, K., Translator; Free Press: New York, NY, USA, 1950.
14. Durkheim, E. *Suicide: A Study in Sociology*; Free Press: New York, NY, USA, 1951.
15. Zajczyk, F. *La Conoscenza Sociale del Territorio: Fonti e Qualità dei Dati (Vol. 186)*; FrancoAngeli: Milano, Italy, 1991.
16. Ristea, A.; Andresen, M.A.; Leitner, M. Using tweets to understand changes in the spatial crime distribution for hockey events in Vancouver. *Can. Geogr.* **2018**, *62*, 338–351. [CrossRef] [PubMed]
17. Beiró, M.G.; Bravo, L.; Caro, D.; Cattuto, C.; Ferres, L.; Graells-Garrido, E. Shopping mall attraction and social mixing at a city scale. *EPJ Data Sci.* **2018**, *7*, 28. [CrossRef]

18. De Falco, C.C.; Punziano, G.; Trezza, D. Follow the Geographic Information: The Challenges of Spatial Analysis in Digital Methods. *Athens J. Soc. Sci.* **2023**, *10*, 45–58.
19. De Falco, C.C.; Mari, F.; Lenzi, F.R. Urban Vibrancy in Roma's Neighbourhoods: A Case Study. In *Rethinking Social Theories and Methods in a Digital Society*; McGraw-Hill Education: Milano, Italy, 2024.
20. De Falco, A.; De Falco, C.C.; Ferracci, M. Geo-Social Media and Socio-Territorial Distribution: A Study on the Italian Case. *Ital. Sociol. Rev.* **2022**, *12*, 685.
21. Trezza, D. To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. *Front. Sociol.* **2023**, *8*, 1145038. [CrossRef]
22. Venturini, T.; Rogers, R. "API-based research" or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. *Digit. J.* **2019**, *7*, 532–540. [CrossRef]
23. Isaak, J.; Hanna, M.J. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer* **2018**, *51*, 56–59. [CrossRef]
24. Bruns, A. After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. In *Disinformation and Data Lockdown on Social Platforms*; Routledge: London, UK, 2021; pp. 14–36.
25. Meta, 2023: "Meta for Developers: Accessing Data Responsibly". Available online: https://developers.facebook.com/docs/development/maintaining-data-access/ (accessed on 1 March 2025).
26. Bruns, A.; Weller, K. Twitter as a first draft of the present: And the challenges of preserving it for the future. In Proceedings of the 8th ACM Conference on Web Science, Hannover, Germany, 22–25 May 2016; pp. 183–189.
27. Puschmann, C. An end to the wild west of social media research: A response to Axel Bruns. *Inf. Commun. Soc.* **2019**, *22*, 1582–1589. [CrossRef]
28. Barrie, C.; Ho, J.C.T. AcademictwitteR: An R package to access the Twitter Academic Research Product Track v2 API endpoint. *J. Open Source Softw.* **2021**, *6*, 3272.
29. Announcing New Access Tiers for the Twitter API. 2023. Available online: https://devcommunity.x.com/t/announcing-new-access-tiers-for-the-twitter-api/188728 (accessed on 1 March 2025).
30. Carrière, T.C.; Boeschoten, L.; Struminskaya, B.; Janssen, H.L.; de Schipper, N.C.; Araujo, T. Best practices for studies using digital data donation. *Qual. Quant.* **2024**, *59*, 389–412. [CrossRef]
31. Middleton, S.E.; Kordopatis-Zilos, G.; Papadopoulos, S.; Kompatsiaris, Y. Location extraction from social media: Geoparsing, location disambiguation, and geotagging. *ACM Trans. Inf. Syst. (TOIS)* **2018**, *36*, 40.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.