



# Comparing two crowdsourcing platforms: assessing their potential for mapping Antarctica

Ayşe Giz Gulnerman<sup>1</sup> · Muge Senel<sup>2</sup> · Ozan Deniz Gokduman<sup>2</sup>

Received: 29 February 2024 / Accepted: 20 June 2024  
© The Author(s) 2024

## Abstracts

Geographical data acquisition in Antarctic regions is challenging due to the lack of human habitation, harsh environmental conditions, and limited accessibility. This research explores and evaluates the capability of two crowdsourcing platforms in mapping facilities across Antarctic regions. The study presents crowdsourcing projects related to polar regions in the literature. The methodology section outlines the data acquisition techniques employed by Flickr and Happywhale, and the spatial evaluation methods applied to the collected data. In the implementation and results section, the spatiotemporal potential of the data obtained from the two identified crowdsourced platforms is assessed, and the results based on spatial statistical methods are compared. In the discussion and conclusion section, the contribution of the two identified crowdsourced platforms to mapping activities is evaluated in terms of spatial, temporal, and content differences. This study reveals that Happywhale offers data with higher spatial consistency, considering seasonal representation and spatial autocorrelation. Additionally, content restrictions and reliance on GPS enhance spatial accuracy in Happywhale. At the same time, the liberation of data production leads to lower quality but increased quantity, diversity, and spatial coverage, as observed in Flickr. By comparing two crowdsourced platforms, this study enhances data acquisition and evaluation potential in Antarctica.

**Keywords** Crowdsourcing · Volunteered geographic information · Geo-social media · Polar science · Antarctica

## Introduction

Antarctica is the fifth-largest of the seven continents in the world and is predominantly covered by a polar desert (Shapley 2013). The continent was first discovered in the early nineteenth century, leading several countries to embark on expeditions to Antarctica. These expeditions resulted in new maps and charts about the continent, yet they also raised increasing questions about its geography. Rack (2018) notes that the current view of Antarctica is largely inherited from these expedition activities, including mapping and photography taken by Antarctic photographers. The initial efforts to map Antarctica were driven

by the necessity for precise cartography to serve mercantile interests, notably those of mammal hunting companies, as well as the ambition to conquer and colonize new territories. Eventually, the Antarctic land gained scientific importance, and international collaboration was initiated to explore the continent, focusing on major scientific problems such as terrestrial magnetism, meteorology, and geology. Although the land is vast and available technology was limited initially, it was later enhanced with aerial photogrammetry and satellite imagery (Li et al. 2023). Recently, geographic information of Antarctica is served as an open-source dataset by means of Quantarctica which is an integrated mapping environment for Antarctica (Matsuoka et al. 2021). Quantarctica data collection is composed of data from nine scientific fields which is majorly retrieved or extracted from satellite imageries. Despite the significant effort put into generating geographic information for Antarctica through remote sensing technologies, on-site sensors are necessary to observe animal behaviors, monitor environmental changes, and track climatic conditions, among other tasks. Yet the vast area cannot be covered by these sensors alone and still there is high data

---

Communicated by: Hassan Babaie

✉ Ayşe Giz Gulnerman  
ayse.gulnerman@hbv.edu.tr

<sup>1</sup> Land Registry and Cadastre Department, Ankara Hacı Bayram Veli University, Ankara, Türkiye

<sup>2</sup> Geomatics Engineering Department, Istanbul Technical University, Istanbul, Türkiye

demand for exploring the continent (Dong et al. 2022). However, crowdsourced images, which are referred to as human-based sensors, can provide an additional source of Antarctic geospatial information based on human visits to Antarctica. This study investigates and compares the data usability of two crowdsourcing platforms over Antarctica, considering data collection system design, data retrieval techniques, and their spatial footprint.

Geo-crowdsourcing embodies the data sourced by humans through platforms, thanks to the development in the web technologies and the advent of global navigation satellite systems (GNSS) (Goodchild 2007a, b; Hall et al. 2010; Kullenberg & Kasperowski 2016; Zhang et al. 2011). The widespread use of smart devices with GNSS embedding has enriched geo-crowdsourced data through various applications of crowdsourcing (Haklay et al. 2008; Turner 2006). These technological developments led to the new mapping tools with various themes and concepts, termed with several different names such as; “citizen science”, “neogeography”, and “collective intelligence”. Later, a unified name is proposed and widely accepted for these new sources of data which is “Volunteered Geographic Information” (VGI) (See et al. 2016; Steiger et al. 2016). In the last decades, VGI is diversely adopted for various themes and purposes. The majority of these uses are for urban areas due to data abundance, where populations are dense and highly active over space (Kanhare 2013). Yet, the quality of VGI is an ongoing debate and research area that should be considered prominently before its use. The usability of VGI in less human-populated areas is a greater issue due to low visitation rate (Daymond et al. 2023). The limited presence of people in a space may reduce the completeness of spatial data, but it can also lead to data bias in the regions where data is provided. Antarctica is one of the most special cases in this perspective due to the lack of native human habitation.

Antarctica attracts a growing number of researchers and tourists exploring different regions across the continent. The exploration of the continent began in the nineteenth century, with modern tourist visits starting in the 1960s through sea cruises. Whether in tourism or science, all human activities are regulated by the Antarctic Treaty to ensure responsible and sustainable activities in Antarctica (Frame et al. 2022). Recently, the majority of these visits are operated by members of the International Association of Antarctica Tour Operators (IAATO) in cooperation with Antarctic Treaty to preserve the unique and fragile environment. IAATO reports that Antarctic visits reached over 100 k tourists in 2023, indicating a growing number of human visitors (sensors) to the continent. However, the coverage of these tourist visits is limited to some regions of Antarctica. According to records and statistics from IAATO (2022), over 90% of tourist visits are confined to the Antarctic Peninsula, located between

60 and 70 southern latitudes and closest to Chile and Argentina. On the other hand, research facilities are distributed across the continent. Figure 1 depicts Antarctic continent with place names and COMNAP (The Council of Managers of National Antarctic Programs) Antarctic facilities (Matsuoka et al. 2021; Norwegian Polar Institute 2018; COMNAP 2017). Although COMNAP facilities are located in different regions of the continent, they are concentrated on the Antarctic Peninsula. In addition to these, not all facilities operate year-round, leaving some regions, especially during the winter season, apart from the Antarctic Peninsula in complete darkness.

The number of human visits to the Antarctic regions raises the question of whether this potential could be utilized for retrieving additional geographic information across Antarctica by means of VGI platforms. To investigate, first, in the literature review, we delve into VGI platforms and types, along with related VGI studies in polar regions. Second, in the “Materials and Methods” section, we provide details regarding our data retrieval methods from two determined VGI platforms and the techniques used for assessment. Third, in the “Implementation and Results” section, we present the specifics of the retrieved data and the assessments conducted. Lastly, we conclude our study with the “Discussion and Conclusion” section, where we explain how the two VGI platforms differ based on their VGI type, context, and design, and how VGI can be effectively leveraged for comprehensive analysis.

This paper aims to shed light on the potential of VGI platforms in understanding the Antarctic regions, acknowledging the unique challenges posed by this remote and data-scarce environment. The study particularly discusses the potential coverage, positional accuracy, and spatial repeatability of data when using photographs for mapping the Antarctic continent. This discussion is conducted comparatively by focusing on two VGI platforms. While researching VGI platforms, we encountered challenges related to data accessibility. Despite the presence of numerous VGI platforms, we decided to compare only two: Flickr and Happywhale. Flickr is a social media platform focused on photo sharing, while Happywhale is a global data center monitoring whales and possessing comprehensive data for polar regions. Therefore, comparing data from a customized web page and a social media application for a specific Antarctic region is crucial for understanding the degree of overlap between social media data and citizen science platform data. There are no limitations on the Flickr API, and all photographs produced for Antarctica has been accessed using the techniques mentioned in “Data retrieval methods” of this paper. Happywhale functions as a citizen science data center with no limitations on data sharing as well. All data produced for Antarctica has also been accessed from the Happywhale platform. Since these platforms share both

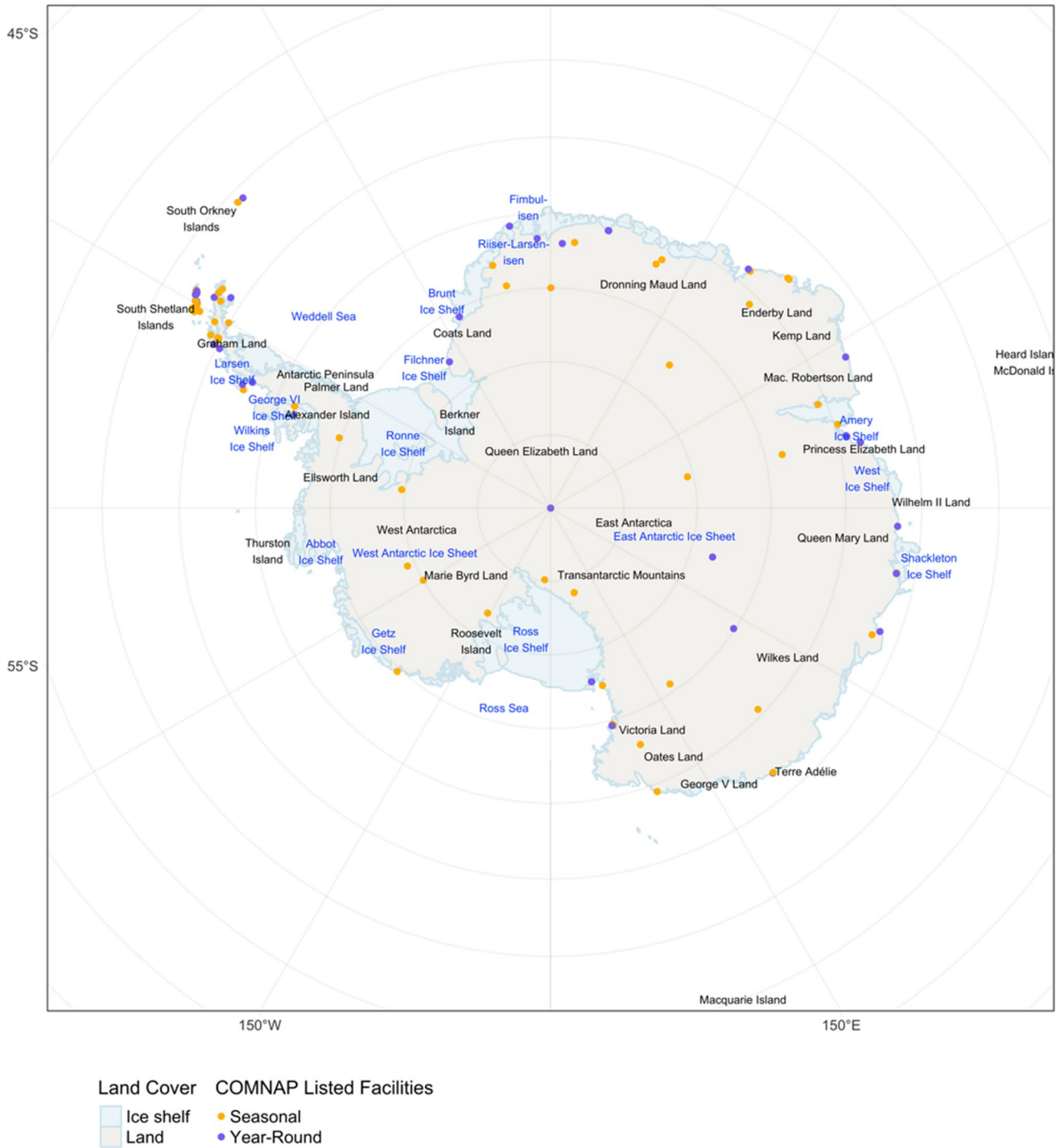


Fig. 1 Map of the Antarctic continent with placenames and listed facilities

commonalities and differences with social media, comparing data from both for a specific region is a unique and logical approach. Other social media platforms such as Facebook and Instagram were not used in this study due to data access restrictions.

### Literature review

The term VGI encompasses several perspectives and has a diverse background. The assets of VGI, such as crowd profiles, the technology used, and its design, identify features

of platforms including data accountability, continuity, scope, diversity, and detail (Ball 2002; Hecht and Stephens 2014; Gulnerman et al. 2021). The product of crowdsourcing platforms with the mapping technologies is called volunteered geographic information (VGI). In this paper we accept that VGI has three types (Elwood 2008; Hecht & Stephens 2014; Gulnerman et al. 2021); 1-Citizen Science VGI, 2-Peer Production VGI, and 3-social media VGI (Table 1).

The main distinctions among these types are the motivation of volunteers, platform design, and outcomes, as described in Table 1. Different features of VGI types enable varying contribution levels to different projects. Therefore, these assets become significant characteristics in addressing geographical data requirements across a range of research studies. The flexibility of providing data on various topics, especially in polar regions where data scarcity is a prevalent challenge, is invaluable. However, it is necessary to anticipate that these platforms may not provide comprehensive data across all regions and time intervals. In this context, we present a review of projects in polar regions based on VGI types in the following three subsections.

### Citizen science projects on polar regions

The early form of citizen science emerged with urban planning meetings with the public for decisions making and mostly called public participation (PP). With the advent of Geographic Information Systems (GIS) and its integration with the public participation, the terminology has changed as public participation geographic information systems (PPGIS) (Sieber 2006). Lately, the methodological use of PPGIS crossed the urban borders, used for remote areas

such as Antarctica, Arctic and even for other planets other than earth. Goodchild (2007a, b) defines “*citizens as sensors*” capturing their environment and sharing information and ideas via internet platforms, widely referred to as citizen science. Currently, citizen science has its own pervasive platforms, such as Zoouniverse, CitizenScience, and SciStarter, where project holders are enrolled and carry out their projects through these platforms. In Table 2, we present citizen science projects on polar regions carried out on these popular Citizen Science platforms.

Two projects conducted on the Zoouniverse platform cover partial spatial regions in the Arctic. While one of these projects aims to monitor the lifestyles and habitats of polar bears, the other aims to digitally label plant species and monitor plant diversity. Camera data placed in various regions is used in polar bear observation, while data collected by volunteers visiting the region is used in the study of plant diversity. Another platform Citizenscience acts as a collection website for projects. MAPPPD (Mapping Application for Penguin Populations and Projected Dynamics) project is listed on that platform, aiming to detect and count penguin colonies using remote sensing data and field research data. Another project named Penguin Watch on the SciStarter platform focuses on explaining the lives of penguins through the use of annotated images taken from nesting sites. Kickstarter one of the popular citizen science platforms has two projects on polar regions. The first one is campaigning for a book fund to document the ice change and understand climate change using Arctic field research data. The second one aims to fund side cost of an Antarctic Expedition for a photographic study of the station and the surrounding landscape. In addition, there are two projects

**Table 1** Types of VGI (Ball 2002; Elwood 2008; Hecht & Stephens 2014; Gulnerman et al. 2021)

Features of VGI Types			
	Motivation of Volunteers	Platform Design	Outcomes
Citizen Science	<ul style="list-style-type: none"> <li>conscious contribution</li> <li>possibility of self-interest</li> </ul>	<ul style="list-style-type: none"> <li>platforms oriented to a specific area and topic for decision making</li> </ul>	<ul style="list-style-type: none"> <li>views of observations related to a geographic location or area for supporting decision-making processes for environmental and urban planning</li> </ul>
Peer Production	<ul style="list-style-type: none"> <li>conscious contribution</li> <li>self-entertainment</li> <li>producing base maps</li> </ul>	<ul style="list-style-type: none"> <li>platforms provide sets of tools to add geographic features to the worldwide map</li> </ul>	<ul style="list-style-type: none"> <li>base map layers;               <ul style="list-style-type: none"> <li>streets</li> <li>buildings</li> <li>point of interests</li> <li>etc</li> </ul> </li> </ul>
Social Media	<ul style="list-style-type: none"> <li>unconscious contribution</li> <li>entertainment</li> <li>networking</li> <li>marketing</li> <li>information sharing</li> </ul>	<ul style="list-style-type: none"> <li>platforms provide generic tools for sharing information in a multimodal (text, photo, video) way with additional geographic location features, and without a topic limit</li> </ul>	<ul style="list-style-type: none"> <li>uncertain (amount, content, reliability (of)) geo-referenced data is used in various projects;               <ul style="list-style-type: none"> <li>monitoring user behavior in space for urban planning</li> <li>real-time information for crisis response</li> <li>community engagements between for collaborating residents and local authorities</li> <li>etc</li> </ul> </li> </ul>

**Table 2** Citizen science platforms and projects related to polar regions

Project Name	Platform	Data Type	Spatial Scope	Ref
The Arctic Bears Project	Zooniverse	Text, Image, Video	Hudson Bay, Arctic	URL 1 <a href="https://www.zooniverse.org/projects/douglas-clark/the-arctic-bears-project">https://www.zooniverse.org/projects/douglas-clark/the-arctic-bears-project</a>
Notes from Nature	Zooniverse	Text, Location	North Canada, Arctic	URL 2 <a href="https://www.zooniverse.org/projects/cmnbotany/notes-from-nature-expedition-arctic-botany">https://www.zooniverse.org/projects/cmnbotany/notes-from-nature-expedition-arctic-botany</a>
MAPPPD	Citizenscience	Text, Location	Antarctica Peninsula	URL 3 <a href="https://www.citizenscience.gov/catalog/44/#">https://www.citizenscience.gov/catalog/44/#</a>
Penguin Watch	SciStarter	Text, Image	Antarctica	URL 4 <a href="https://scistarter.org/penguinwatch">https://scistarter.org/penguinwatch</a>
Plan D: A Race Against Time	Kickstarter	Text, Image	Arctic	URL 5 <a href="https://www.kickstarter.com/projects/raceagainsttime/plan-d-a-race-against-time?ref=discovery&amp;term=polar">https://www.kickstarter.com/projects/raceagainsttime/plan-d-a-race-against-time?ref=discovery&amp;term=polar</a>
Portraits of Place in Antarctica II	Kickstarter	Text, Image	US Research Stations, Antarctica	URL 6 <a href="https://www.kickstarter.com/projects/1322320805/portraits-of-place-in-antarctica-ii?ref=discovery&amp;term=antarctic">https://www.kickstarter.com/projects/1322320805/portraits-of-place-in-antarctica-ii?ref=discovery&amp;term=antarctic</a>
Crowdsourcing Arctic Navigation Using Multispectral Ice Classification and GNSS	Stanford University	Text, Location	Arctic	Reid et al. (2014)
Happywhale	Happywhale	Text, Location	World seas, Antarctica	URL 7 <a href="http://happywhale.com/browse">happywhale.com/browse</a>

that have their own platforms for citizen science. The first, a research project from Stanford University, aims to find the safest route using geo-referenced data with GNSS, enabling reliable ship-to-ship aiding. The high-integrity sharing of ice data offers a framework for performing path planning in a reliable and automated-systematic way (Reid et al. 2014). The second project is Happywhale, the only website dedicated to understanding marine environments globally. The platform collects and indexes whale photos, especially those with unique IDs, to understand their global movements. In reviewing these projects, citizen science projects in polar regions primarily encompass two key areas: monitoring animal populations (including bears, penguins, and whales) and studying climate change, specifically ice dynamics and ice classification. These projects adopt diverse data types, and their coverage of project areas is also diverse.

### Peer production projects on polar regions

OpenStreetMap (OSM) is a widely recognized VGI platform offering a range of digital tools to volunteers for generating geographical data (Mooney et al. 2010). Through the platform, volunteers can contribute geographical data using various methods, including uploading GNSS tracks, manually editing maps based on their local knowledge, or digitizing from satellite imagery (Grinberger et al. 2021).

This platform is used for geographical data production in many regions around the world. However, especially in areas where geographical data is not available due to various reasons, data production, particularly in times of urgent need after disasters, is carried out through digitization using satellite images (Ahmouda et al. 2018; Poiani et al. 2016). The OSM platform hosts and provides data, including polar regions such as the Arctic, and even a small amount of data from Antarctica (Schott et al. 2022). In these areas where polar deserts are widespread, generally data production by volunteers is carried out to be done through digitization using satellite images. Therefore, there are limited studies carried out with OSM in polar regions. Xu et al. (2022) adopt OSM, Sentinel-1, Sentinel-2, and ArcticDEM for mapping man-made impervious areas in the circumpolar Arctic to provide insights regarding environmental sustainability. Similarly, Liu et al. (2023) used OSM as auxiliary data to remove errors in the detection process of impervious and non-impervious surface areas from satellite imagery in the Arctic Circle.

### Social media projects on polar regions

Social media (SM) is used as a crowd-sourcing platform for various themes, i.e., disaster management (Wang and Ye 2018; Xiao et al. 2015; Gulnerman et al. 2021), urban

human mobility research (Kang et al. 2020; Huang and Wong 2016; Gulnerman 2021), and urban design impacts on human mental health (Garimella et al. 2016; Reichert et al. 2020). Facebook, Instagram, Twitter, Pinterest, Flickr, Reddit, and YouTube are the actual source of various data, such as images, videos, texts, locations, and videos. Extracting information from these platforms becomes more evident when there is an abundance of data. On the other hand, the places visited less, unpopular or low accessible are not considered in studies, yet the available limited data still have potential to draw some valuable inferences. There is limited research on polar regions related with SM data. Two different groups carried out studies on polar tourists with the SM data. The first study by Runge et al. (2020) explore the tourism activity change (with examining TripAdvisor platform data) over Arctic and reveal the human footprint increase in the region. The second study by He and Liu (2023) discusses the influence of tourist over Antarctica by adopting Chinese SM platforms (Zhihu and Mafengwo). Another study which is not directly adopting and analysing SM data but also related with polar regions is carried out by LaRue et al. (2020). According to this study, polar projects is proposed to disseminate on SM. In these three studies, SM is seen as an axillary source for promoting and integrating their projects on polar studies. Therefore, social media data has not been adopted and empirically evaluated for its potential contribution to geographical data in polar regions.

## Materials and methods

There are diverse data retrieval methods from VGI platforms. As noted, this study investigates and compares the data usability of crowdsourcing platforms over Antarctica, considering data collection system design, data retrieval techniques, and their spatial footprint. Data retrieval is the first step for data evaluation. In sub-"Data retrieval programs and programming libraries", we reviewed the accessibility of social media (SM) data with available toolsets

and alternative programming libraries. Following that, we present our data retrieval methods over two freely accessible VGI platforms: Flickr (SM based) and Happywhale (citizen science-based) in sub-"Data retrieval methods". Finally, we introduce VGI data evaluation techniques to investigate the retrieved data in sub-"Data exploration and analysis".

## Data retrieval programs and programming libraries

Accessing SM data is possible through different programs. These data retrieval programs present data in various forms depending on the scope of the study related to SM data. Additionally, while some of these programs retrieve data from a single SM source, others can pull data from multiple SM platforms. Table 3 provides information about the programs that retrieve data from SM along with the platforms from which they collect data. Various social media data retrieval programs have different features and limitations depending on their intended purposes. NodeXL (URL 8 <https://download.geofabrik.de/>) is designed for assessing friendship networks and presents data in a graph format. Crowdtangle (URL 9 <https://www.smrfoundation.org/>) focuses on investigating critical topics, such as elections and racial justice through influential accounts. Facepager (URL 10 <https://www.crowdtangle.com/>) is a program capable of crawling data from multiple SM platforms within the limitations of each platform. Twitter Advanced Search (URL 11 <https://github.com/strohne/Facepager>) is a tool provided by Twitter that allows users to query data based on content, language, accounts, reaction types, numbers (retweets, likes, replies), and time interval. One Million Tweet Map (URL 12 <https://twitter.com/search-advanced>) offers an analytical interface for Twitter analytics, providing spatial clusters and sentiment views based on queries related to hashtags, time, and spatial extent. Additionally, there are web apps for different SM platforms (URL 13 <https://onemilliontweetmap.com/>, 14 <https://fdown.net/>, 15 <https://www.storysaver.net/>)

**Table 3** Social media data retrieval applications

No	Application	Data Sources	Reference
1	NodeXL	Flickr, Twitter, Youtube, Wikipedia	URL 8 <a href="https://download.geofabrik.de/">https://download.geofabrik.de/</a>
2	Crowdtangle	Facebook, Instagram, Reddit	URL 9 <a href="https://www.smrfoundation.org/">https://www.smrfoundation.org/</a>
3	Facepager	Twitter, YouTube, Websites	URL 10 <a href="https://www.crowdtangle.com/">https://www.crowdtangle.com/</a>
4	Twitter Advanced Search	Twitter	URL 11 <a href="https://github.com/strohne/Facepager">https://github.com/strohne/Facepager</a>
5	One Million Tweet Map	Twitter	URL 12 <a href="https://twitter.com/search-advanced">https://twitter.com/search-advanced</a>
6	Facebook Video Downloader	Facebook	URL 13 <a href="https://onemilliontweetmap.com/">https://onemilliontweetmap.com/</a>
7	Story Saver	Instagram	URL 14 <a href="https://fdown.net/">https://fdown.net/</a>
8	Twitter Video Downloader	Twitter	URL 15 <a href="https://www.storysaver.net/">https://www.storysaver.net/</a>
9	Flickr downloader	Flickr	URL 16 <a href="https://twittervideodownloader.com/">https://twittervideodownloader.com/</a>

that allow users to download videos or stories one by one using the appropriate links and usernames.

There are various popular methods for programmatically downloading SM data. Table 4 lists the libraries used in R and Python programming languages. These libraries have different approaches to data querying and managing large amount of data requests. The Twitter (Gentry et al. (2016)), Tweepy (URL 17 <https://flickrdownloadr.com/>), rtweet (Kearney 2019) libraries for Twitter data allow users to download data with attributes such as username, content, creation date, latitude, longitude, and more. However, these libraries impose restrictions on users to prevent misuse. Instagram data is collected using instaR (URL 18 <https://www.tweepy.org/>) and Instaloader (URL 19 <https://github.com/pablobarbera/instaR>), and although these libraries do not require additional API (Application Programming Interface) credentials, users need to have an Instagram profile. Instagram's strict bot detector may consider profiles with low or no activity as bots, leading to automatic IP blocking after a certain number of queries. Repeated IP blocks may eventually result in account removal. While Rfacebook, was useful for downloading Facebook data, it has not been maintained since 2020 (URL 20 <https://instaloader.github.io/>, URL 21 <https://developers.facebook.com/docs/graph-api/>). Additionally, instead of separately accessing Facebook and Instagram platforms, Meta allows access to all this data using their Graph API (URL 20 <https://instaloader.github.io/>) for academic purposes). FlickrAPI (URL 22 <https://github.com/pablobarbera/Rfacebook>) has encountered some maintenance issues causing changes in data quantity based on the specified time granularity, as we have tested and observed during our data retrieval within this study.

## Data retrieval methods

In this study, citizen science projects in polar regions, data production within the scope of peer-production, and finally, social media platforms have been discussed. Two platforms have been selected from these VGI-based data platforms, which have been ongoing for years and provide uninterrupted access to historical image data: Flickr (SM-based

VGI) and Happywhales (Citizen science-based VGI). Image data from both platforms covering the Antarctic continent was obtained using the Python programming language. From the Flickr platform, images were obtained through the freely accessible FlickrAPI. Images from the HappyWhale platform were provided through HTTPS POST requests. According to Flickr's terms of service and API terms, we ensured that our data use of Flickr data aligns with their policies. Our search does not include individual data explicitly; however, implicit privacy issues that may be encountered during the process of further assessments were considered. To manage, organize, store, and analyze the large number of obtained images, a PostgreSQL database was created, and database tables with similar data structures were established for both platforms. The retrieved data from the two platforms have different standardized metadata structures due to their platform-specific designs (Fig. 2). The standardization issue caused by different metadata structures was addressed with platform-specific automation steps and approaches.

Images shared on the Flickr platform are provided through static URLs via the Flickr API, as long as they are not deleted by their owners. Image retrieval through the Flickr API was carried out by searching with keywords or within a predefined bounding box for Antarctica (WGS84 bounds: -180.0 -90.0, 180.0 -60.0) (URL 23 <https://www.flickr.com/services/api/>). The date range for data acquisition was set from the launch date of Flickr in 2004 until the data acquisition year 2023. However, it was observed that for any given date range, the Flickr API provides a maximum of 4000 image results. Therefore, the algorithm used for data retrieval was adjusted to split the date range into shorter intervals, ensuring that the results obtained for each interval do not exceed 4000 images. One of the input parameters, deltaTime, determines the width of these time intervals. However, the start date (startDate) and end date (endDate) of the interval are automatically updated by the algorithm during the search process. This process continues until a time span is created whose start date is on or after the maxDate.

Another aspect considered in the algorithm design was pagination. Flickr can display a maximum of 250 geotagged (or 500 non-geotagged) images on one page. Therefore, the

**Table 4** R and Python libraries designed for the acquisition and analysis of SM data

No	Library	Programming Language	Data Sources	Reference
1	Twitter	R	Twitter	Gentry et al. (2016)
2	Tweepy	Python	Twitter	URL 17 <a href="https://flickrdownloadr.com/">https://flickrdownloadr.com/</a>
3	rtweet	R	Twitter	Kearney (2019)
4	instaR	R	Instagram	URL 18 <a href="https://www.tweepy.org/">https://www.tweepy.org/</a>
5	Instaloader	Python	Instagram	URL 19 <a href="https://github.com/pablobarbera/instaR">https://github.com/pablobarbera/instaR</a>
6	Rfacebook	R	Facebook	URL 21 <a href="https://developers.facebook.com/docs/graph-api/">https://developers.facebook.com/docs/graph-api/</a>
7	flickrAPI	Python	Flickr	URL 22 <a href="https://github.com/pablobarbera/Rfacebook">https://github.com/pablobarbera/Rfacebook</a>



Fig. 2 Examples of metadata table of (a) Flickr, (b) Happywhale

algorithm takes into account the number of pages the results consist of for the specified date range and adds the photos from each page to the content list. Not only the basic information of the photos but also the Exif data available on Flickr is added to the list. Another challenge encountered with the Flickr API was that repeated searches for the same date range could yield different results. To overcome this issue, searches were conducted weekly, daily, and hourly between 2004–2023, and the result table was created after removing duplicate entries. The overall process of retrieving Flickr data is outlined in the Algorithm 1 (Appendix A).

The algorithm designed to obtain images from Happywhale addresses various issues. Although consistent results were obtained in repeated data retrieval searches on the Happywhale platform, this platform also has a query size limit like the Flickr API, and this limit is 15,000 records. The algorithm designed to retrieve data from Happywhale overcomes this limit by using shorter date intervals. Unlike the first algorithm, this algorithm adopts an approach to minimize the number of requests sent to the server by using a dynamic interval instead of a fixed time interval. The reason for this choice is that as the data on the Happywhale platform approaches the present day, it becomes more frequent.

Using a dynamic time interval aims to minimize the number of requests sent to the server. The algorithm takes the start (minDate) and end dates (maxDate) as input from the user and outputs the images in this date range to the user. The overall process of retrieving data from the Happywhale platform is provided in the Algorithm 2 (Appendix B).

### Data exploration and analysis

The analysis of data obtained from Flickr and Happywhale platforms aims to investigate variations in data quantity based on the activities of tourists, researchers, or personnel in Antarctica, a region characterized by a lack of human habitation. Some research bases operate only during the summer season, while others remain open throughout the year. Additionally, tourism activities on the continent mostly occur during the summer season. Consequently, it is expected that data obtained during the winter months will be less. Taking into account Antarctica's two seasons, information on the year and season was extracted from the date column in the data tables and added to new columns. When determining the seasonal range, the summer season covers the period from November to April, while the winter



season covers the period from May to October. The examination of the spatial distribution of data and the investigation of its spatial consistency aim to understand whether there are changes in the spatial distribution of data between summer and winter months. Spatial distribution maps and spatial correlation analyses are employed for this purpose. In regions where positive spatial autocorrelation is identified, the evaluation of the data will depend on the presence of a research base facility or whether it is a popular tourist visitation center. With these fundamental insights, the goal is to make inferences about the consistency of data obtained from VGI platforms.

**Data representation** Various methods and techniques are available for the visual representation and analysis of geographic data. In addition to geographic data, Volunteered Geographic Information (VGI) platforms provide various information such as place names, dates, and IP addresses (Owuor & Hochmair 2020). For example, geotagged data obtained can be extracted, visualized, and analyzed using the R programming language. In this study, line graphs were used to depict the distribution of images over the years. Additionally, spatial distribution maps were employed to visualize the spatial representation of the images. Data obtained from VGI platforms often offer extensive geographic coverage. The spatial representation of data from VGI platforms can provide valuable insights into the accuracy of the data generated (Zanten et al. 2016). Within the study, the spatial representation of data contributes to interpreting the accuracy of the data, taking into account the accessibility conditions of the Antarctica continent.

**Spatial autocorrelation analysis** R programming language provides several libraries performing spatial analysis tools. The “spdep” (Bivand & Wong 2018; Bivand 2022; Pebesma & Bivand 2023) library serve functions that measures spatial autocorrelation based on feature locations. The spatial autocorrelation computation in the package is based on Global Moran’s I and Local Moran’s I methodologies (Anselin 1995). While the former computation gives overall positive or negative autocorrelation test results, the latter return the local autocorrelation result over predetermined fishnet for the specific area. By this way Local Moran’s I computation results can be visualized for further spatial exploration. Spatial autocorrelation is a valuable tool for illustrating the spatial variation present in a map. Positive spatial autocorrelation refers that locations close together having similar values. Negative spatial autocorrelation refers that locations close together having more dissimilar values than those locations further away (Dormann et al. 2007). Local Moran’s I is useful to identify spatial clusters and spatial outliers and show them on the map. The following steps were applied in order to utilize these geoprocessing functions.

- 1- CSV files (image coordinates) are imported into R Studio, and latitude and longitude columns are designated as coordinates to convert the data frame into a spatial point data frame (sf). This data uses the World Geodetic System 1984 (EPSG:4326) as the coordinate reference system (CRS) since both VGI platforms (Flickr, Happywhale) uses GPS systems for geotagging.
- 2- Image spatial point data coordinates is transformed from World Geodetic System 1984 (EPSG:4326) to Antarctica Stereographic Projection (3031) to make the CRS same as the base map of Antarctica.
- 3- Simple base map of Antarctica polygon shapefile is obtained from Quantarctica which is predefined Antarctica Stereographic Projection (EPSG:3031) and imported in R Studio.
- 4- A 50×50 fishnet grid is created based on the base map of Antarctica. Other sizes of fishnet were tested but the results indicate similarity and this size demonstrates better visual readability.
- 5- Antarctica Stereographic Projection (3031) is defined in the generated grid to make the coordinate system same as the image point dataset.
- 6- Grid and image point data are spatially joined. A column named "count" is automatically created in the spatial joined grid layer.
- 7- Then the Global Moran’s I process is performed with moran.test function over the “count” feature. The resulting report contains values for spatial autocorrelation. (Global Moran’s I computation steps below are followed)
  - a. fishnet polygons are defined as neighboring polygons
  - b. fishnet polygons are initialized equally weighted
  - c. weighted polygons are calculated based on number of images intersecting each polygon
  - d. Moran’s I test was applied
- 8- After that, Local Moran’s I process is applied with localmoran function, and spatial clusters and spatial outliers are displayed as LISA (Local Indicators of Spatial Association) cluster maps on the grid adopting tmap library functions.

## Implementation and results

### General data overview

The number of Antarctic images obtained by the text method from the Flickr platform was nearly 88 K, and the number of images obtained by the bbox method was approximately 46 K. 55% of the images obtained by the text method and 13% of the images obtained using the

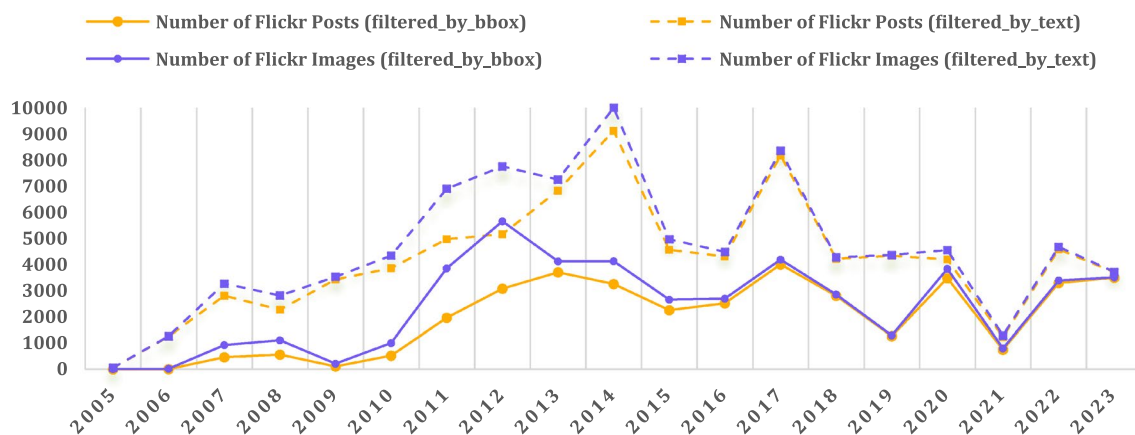
bbox method lacked image coordinates. In contrast, all Happywhale images had embedded coordinates (Table 5).

Figures 3 and 4 illustrated the number of photos from the Flickr and Happywhale platforms, respectively, based on data retrieval methodology. The line graphs in both figures depicted data sharing trends over the years. Figure 5 demonstrated the number of actual visitors to Antarctica (URL 25 <https://www.npolar.no/quantarctica/>). The information covered the summer season visit activities, and although we could not directly compare years, we observed

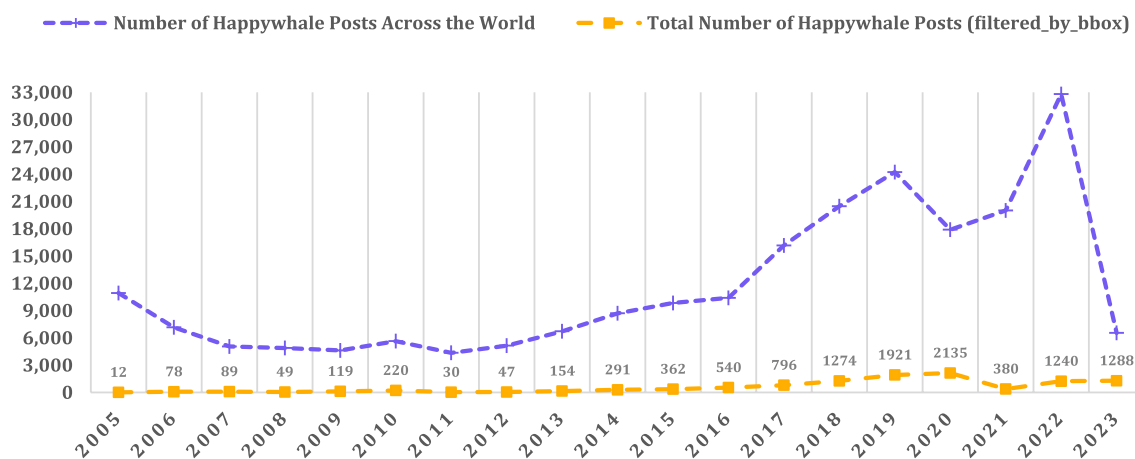
the seasonality. Additionally, this line graph showed that the majority of visitors solely visited the Antarctic Peninsula, but in the Covid year (2020–2021 season), there were no visitors both on the continent and the peninsula. During the Covid-19 period, when travel to Antarctica was restricted, both platforms showed a decreasing trend in post amounts. However, outside of this period, Flickr data exhibited gradual fluctuations from 2005 to 2020, while Happywhale data demonstrated a slight increase. It appeared that the total number of posts on both platforms decreased during

**Table 5** Summary of retrieved data (geotagged)

Id	Platform	Image Retrieving Technique	Filtering Method	Number of Images	Number of Posts
1	Flickr	Flickr API	text	87,894	79,118
2	Flickr	Flickr API	bbox	46,268	37,575
3	Happywhale	Python	complete data	251,279	251,279
4	Happywhale	Python	bbox	11,408	11,408



**Fig. 3** Number of Flickr posts and images by year based on filtering methods (bbox or text)



**Fig. 4** Number of Happywhale posts by year based on bbox filtering method and unfiltered posts

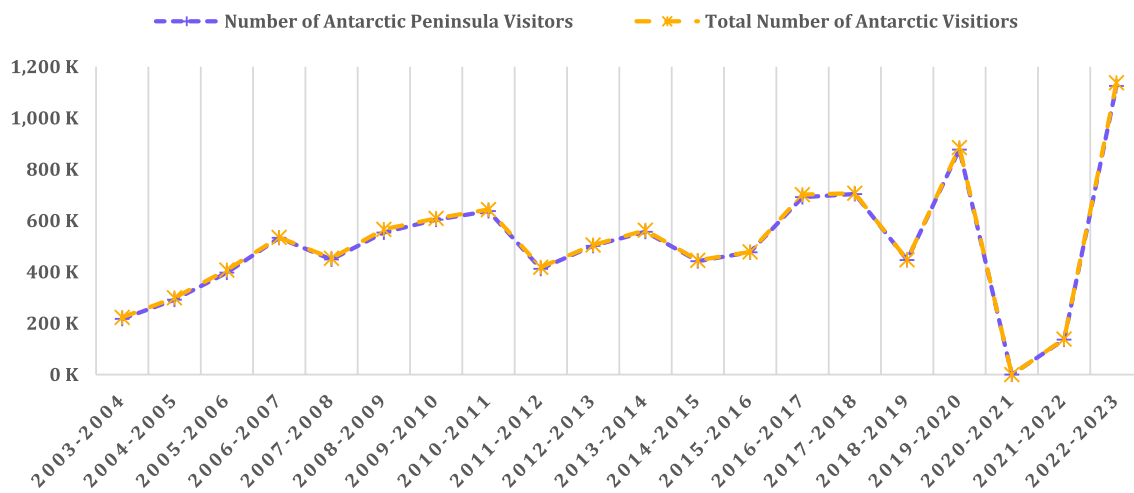


Fig. 5 Number of visitors during summer season based on Antarctic Peninsula and Continental (URL 25 <https://www.npolar.no/quantarctica/>)

the pandemic period. On the other hand, the count of Happywhale images increased, while Flickr images decreased (Figs. 3–4).

### Spatial data manipulation

To compare Flickr and Happywhale data, we considered the Antarctic bounding box (bbox) in different seasons. Table 6 presented the seasonal counts of retrieved data from each platform. The Flickr data retrieved using text filtering was contained within the Flickr bbox data, resulting in a slightly lower number. Therefore, for the spatial comparison, we focused on the Flickr bbox and Happywhale bbox data. All analyses in this study had a common output: the number of images in Antarctica was much higher in the summer season than in the winter season on both platforms. Additionally, we investigated participation bias by considering the users' contribution amounts. Table 6 presents the number of users and their contribution percentage to the retrieved dataset based on each activity quartile group by season. According to the table, the first quartile (Q1) represents the most active data-sharing group, while the fourth represents the least active (Q4). Based on this information, Flickr (text) has the most biased dataset, with over 93% of data produced by the most active 144 users in the summer season. Unsurprisingly,

Flickr (bbox) ranks second in this bias comparison, with over 93% of data produced by 144 users. On the other hand, the Happywhale platform has the least biased dataset, with nearly 65% of data produced by 1182 users in the summer season. When we look at the winter seasons, we observe slightly less contribution bias for Flickr datasets but much less bias for the Happywhale dataset. Although the contribution amounts of users are highly skewed in both platforms based on activity quartiles, none of the users in the activity quartiles can be accounted for as bot users, since a human can post this number of posts.

The footprints of retrieved image data from both platforms were further examined within the Antarctic bbox (Fig. 6). To understand the overall spatial distribution within Antarctic bbox by season, we used the image date taken and followed the methodology to produce spatial plots in Fig. 6(a-1) and (b-1). The first observation from the plots is that Flickr images covered a wider area, including the inner part of the Antarctic mainland, whereas Happywhale images only surrounded the Antarctic mainland. This observation highlighted two key points: the spatial focus of Happywhale data, and the wide coverage of the Flickr platform. Given that the Happywhale platform contains data related exclusively to whales and marine life, the spatial distribution map aligned with the content's intended purpose. In contrast,

Table 6 Number of images within the Antarctic bbox by season

Id	Platform (Filtering Method)	Number of Images		Number of Users		% data amount generated by Users based on Activity Quartiles (Q1/Q2/Q3/Q4)				
		SS	WS	Total	SS	WS	SS	WS		
1	Flickr (text)	38,020	1673	820	616	260	93.9/ 5.0/ 0.6/ 0.4	83.6/ 6.7/ 4.9/ 4.8		
2	Flickr (bbox)	44,560	1708	749	578	224	93.6/ 5.3/ 0.8/ 0.3	84.4/ 7.5/ 4.8/ 3.3		
3	Happywhale (bbox)	10,792	256	4893	4730	221	64.9/ 13.1/ 11.0 / 11.0	35.2/ 21.5/ 21.5/ 21.9		

\*Summer season (SS), \*\*Winter Season (WS)

the comprehensive nature of Flickr data resulted in images being distributed within the inner mainland of Antarctica, raising questions about their accountability due to the difficulties in reaching these remote areas. The second observation reinforced the first observation with respect to the seasons. During the winter season, very few people can stay in the Antarctic regions, and we did not expect to find data in the inner mainland of Antarctica (Fig. 1). Surprisingly, Flickr contained image data located in the inner mainland of Antarctica, which further emphasized the issue of unaccountability. As anticipated, Happywhale has only a small amount of image data located in the Antarctic Peninsula region, which served as the main gateway to enter the Antarctic regions. This region also housed permanent research base stations (Fig. 1) belonging to several countries, such as Chile, Argentina, Russia, China, and others (COMNAP 2017).

To further assess the spatial distribution of retrieved data, Global Moran's I and Local Moran's I computations were applied to seasonal datasets from both platforms. For Global Moran's I computation the steps were followed as mentioned in the 2.3 section for each dataset.

Test results given in Table 7 show that all Moran's I values are positive indicating that there was positive spatial autocorrelation in the data and  $p$ -values (less than 0.001) supported that test results were statistically significant. The standard deviate value is also another strong indicator of spatial autocorrelation. Expectation values were close to zero which meant there was little or no expectation for no spatial autocorrelation. On the other hand, variance values were all positive but varied between datasets, showing that the strength of the spatial pattern between seasons differed. As a result, all values supported that there was statistically significant positive spatial autocorrelation in the data while the strength of this spatial pattern varied across seasonal datasets.

Following the spatial autocorrelation test with Global Moran's I, Local Moran's I analysis was applied to the datasets to measure and visualize local spatial autocorrelation. The computation of Local Moran's I returned values ( $I_i$ ) and  $p$ -values, which assessed whether a specific feature had similar or dissimilar values compared to its neighboring features and indicated statistical significance, respectively. Similarly, LISA cluster visually represents the

spatial patterns of similarity or dissimilarity among neighboring features, along with their statistical significance. The analysis results were displayed as LISA cluster maps with "high-high" (HH) or "low-low" (LL) classes for spatial association, and "high-low" (HL) or "low-high" (LH) for spatial outliers. LISA cluster maps are represented for Flickr in Fig. 6(a-2) and (a-3), and for the Happywhale dataset in Fig. 6(b-2) and (b-3). When assessing LISA cluster maps, it was evident that in the summer dataset, Flickr images were concentrated in the Antarctic Peninsula region with "HH" spatial association. Spatial association was also observed in various regions as "LL" which means low number of images seen the neighboring areas. In the winter dataset, Flickr images exhibited "HH" spatial association in the inner mainland of Antarctica and several isolated areas. The inner mainland association could be caused by manual geotagging through the Flickr places library during data upload or the presence of researchers at the Amundsen-Scott South Pole Station (US) during harsh winter conditions.

Happywhale images in both the summer and winter datasets displayed "HH" clusters in the Antarctic Peninsula, corresponding to the high visitor rate in this area. Notably, the spatial association areas decreased in the Happywhale winter dataset. On the distribution map, it was evident that Happywhale data primarily surrounded the coastal area of the Antarctic Continent; however, only spatial association seen as "LL" clusters was observed from other sides of the continent within the limited area.

Considering the locations of the stations operating in Antarctica, the seasons of operation, the limitations of visitor numbers to Antarctica, and the places visited, it can be argued that the spatial manipulation of the Flickr data was evident in the study results. On the other hand, it can be concluded that the spatial coverage of Happywhale data was less than that of Flickr data. Evaluating the results, the inference was made, particularly about the low spatial accountability of the Flickr data. Therefore, in image processing studies, it should be considered that the content of images may not be directly related to the geotagged location. However, this situation was not encountered in the Happywhale platform because the Happywhale platform provided location data directly with the image itself and largely avoided encouraging manual tagging that could lead to spatial manipulation.

**Table 7** Moran's I computation results over Flickr and Happywhale datasets by season

Platform (season)	Moran's I	$p$ _value*	Standard deviate	Expectation	Variance
Flickr bbox (summer)	0,108	<0.001	17.172	-4.002e-04	4.003e-05
Flickr bbox (winter)	0,099	<0.001	10.513	-4.002e-04	8.872e-05
Happywhale bbox (summer)	0.129	<0.001	19.238	-4.002e-04	4.551e-05
Happywhale bbox (winter)	0,086	<0.001	17.260	-4.002e-04	2.480e-05

\*Based on common practice  $p$ -value is reported "<0.001" when it is smaller

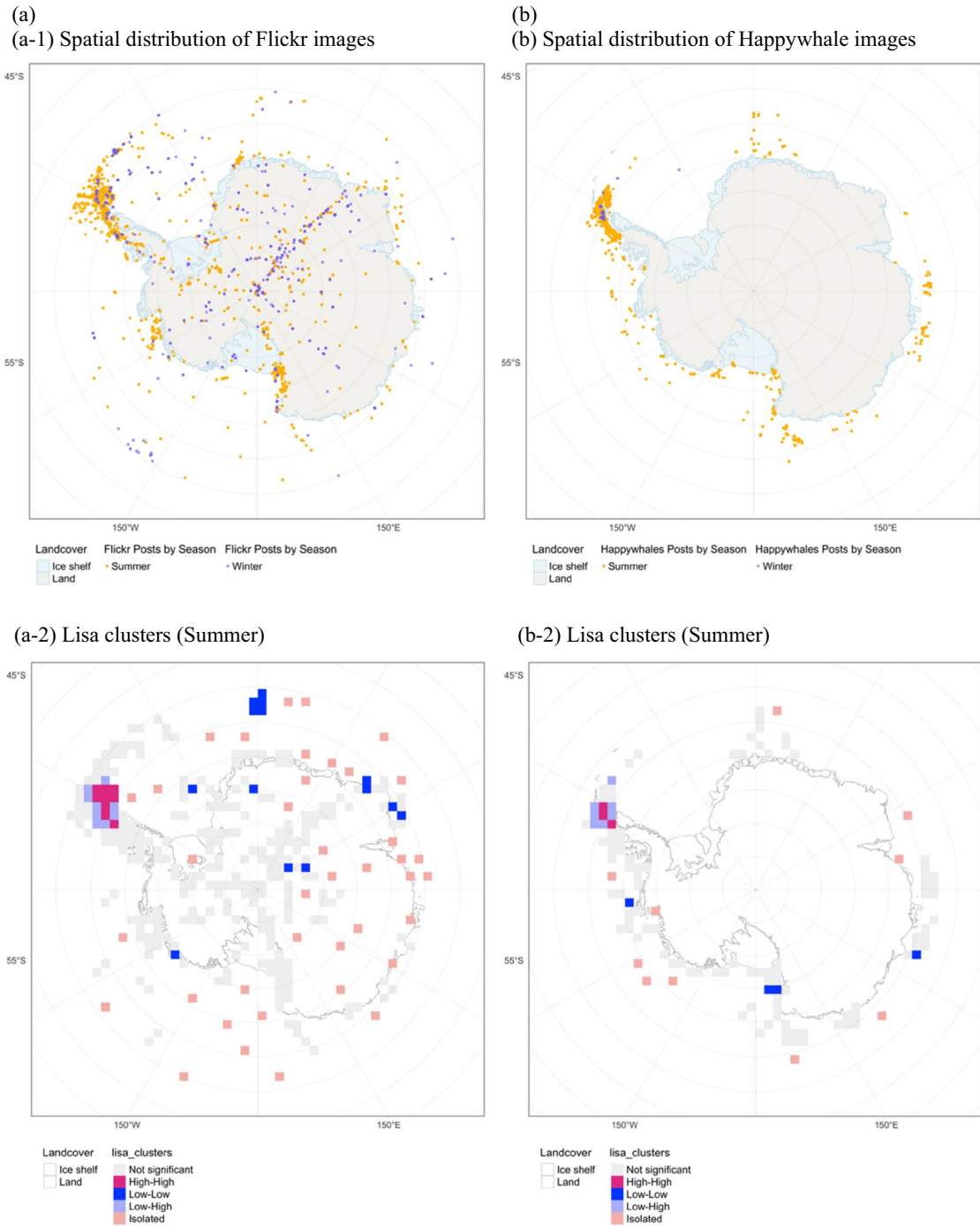


Fig. 6 Spatial distribution of (a) Flickr images filtered by Antarctic bbox, (b) Happywhale images filtered by Antarctic bbox

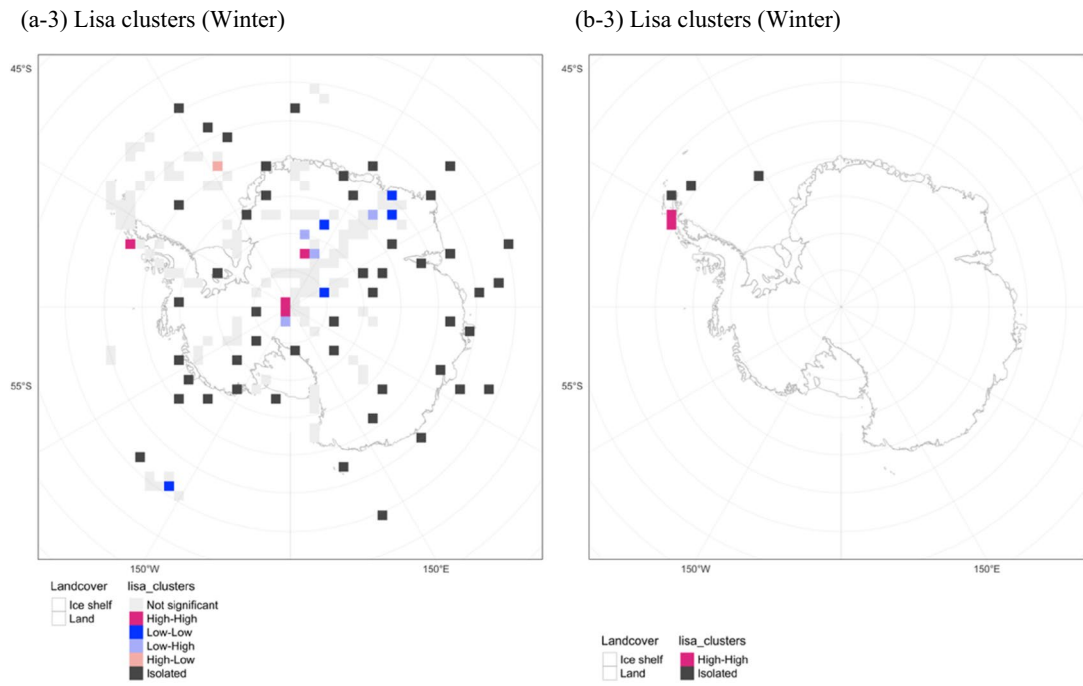


Fig. 6 (continued)

## Discussion and conclusion

VGI, seen as an opportunity to access affordable information from recent sensor and web technologies, is considered to have potential to contribute to various studies as a data source (Tsou 2015). However, there are significant challenges to overcome for this data to be usable. One of the main challenges is the lack of awareness and consciousness among volunteers about data quality, leading to potential issues related to the quality of the data. Volunteers contribute to VGI platforms, whether consciously or unconsciously, use provided platform tools to produce data. These tools, depending on whether platforms allow data manipulation by volunteers who provide data unconsciously, cause uncertainties in data quality. While VGI platforms can offer pre-standardized data production tools covering a specific thematic area or spatial region, providing flexibility. However, in many social media platforms, challenges arise due to unstructured text content, the absence of topic and location restrictions, and the option for users to manually select location information from the places library. This flexibility raises questions about the usability of data obtained from VGI platforms. The situation poses significant challenges in managing large datasets collected by researchers, particularly in areas where data scarcity exists (Hecht & Stephens 2014). Obtaining data from VGI platforms and understanding the usability of the obtained data require expertise.

This study presents VGI-based projects, data platforms, and research as a valuable collection, especially for polar regions where data scarcity may occur. It evaluates various

data retrieval techniques from VGI-based platforms, considering the type of data they offer, thematic areas of study, and data presentation capacities. Despite the claim that most data retrieval techniques from VGI platforms are designed to be free and user-friendly, the study reveals that accessing a vast amount of data without specialized expertise is challenging (Tsou 2015). In this study, several social media platforms such as Facebook, Instagram, and Twitter were considered for use; however, we encountered data accessibility challenges and restrictions. For example, due to Instagram's restrictions on data accessibility, we were only able to retrieve less than 1000 photographs from Instagram using the Instaloader app. We attempted to obtain more data from different times using our project accounts on Instagram, but our account and also our IP address was ultimately banned by Instagram. Also, we cannot access Twitter and Facebook data. Twitter reduced free API access for academic use last year, and collecting such data with a paid API is expensive. We also applied for the Meta API to access Facebook data. However, the Meta API was limited to trending topics and some other information, and did not provide location or photograph data posted by individual accounts. On the other hand, we needed to have all data produced from the platforms to compare, which we obtained from Flickr and Happywhale. In addition, Flickr and Happywhale are directly photograph-oriented platforms, while Twitter and Facebook platforms primarily encourage text-based data production. We aimed to retrieve and assess image datasets to consider the potential coverage, positional accuracy, and spatial repeatability of data when using images in the possible

contribution of mapping activities in Antarctica. Therefore, choosing photograph-oriented Flickr and Happywhale platforms for comparison aligns well with the aim of this study.

VGI platforms are experiencing rapid growth thanks to the prevalence of sensor technology in our daily lives, leading to an abundance of data. It's important to note that the significance lies not only in the quantity of the contributing volunteers or the data they provide but also in the design of the crowdsourcing platforms. The design of these platforms can cause or help preventing data quality issues as it is evidently seen in this study results. In the context of mapping with VGI data, researchers often encounter issues related to locational accuracy. This problem arises from the fact that volunteers are sometimes allowed to manipulate their location data through certain VGI platforms (Middleton et al. 2018). For example, this manipulation can lead to the misrepresentation, as seen in the distribution map of Flickr (Fig. 6(a-1)), which had tags with several places in Antarctica from far away. Since, it is evident that there could be no human in the inner mainland part of Antarctica in the winter season. Consequently, this manipulation can also distort the perception of the extracted information time. In other words, these platforms often result in volunteers providing data with spatiotemporal uncertainty. Unfortunately, uncertainty is not adequately addressed by researchers most of the time, and it largely depends on the data intensity. However, especially in the case of popular locations and direct tags, spatial bias tends to be pervasive (Gulnerman et al. 2020). While there are numerous data quality studies and a few studies that address this issue methodologically (Senaratne et al. 2017), the problem can be initially resolved through platform design (Ogie & Forehead 2017; Hochmair et al. 2018).

In this study, we observe that the design of the Happywhale platform allows volunteers to share data with specific, accurate locations, unlike the data from Flickr. The Happywhale platform does not provide a places library for volunteers to associate their posts with distant locations. Instead, the platform provides "Image Submission Guidelines" that instruct volunteers on the content, metadata requirements, and standards for photos, including guidance on how to handle GPS features. Additionally, platform design might also play important roles in contribution bias in the dataset. While the Happywhale dataset has similar content (such as sea life), the Flickr dataset contains diverse content such as research vessels, base stations, country flags, and researchers' photos. When summarizing the data and data producers, we observe that the Happywhale platform dataset is produced by a larger group of users than the Flickr dataset. Moreover, contribution bias observed in Flickr datasets is much more pronounced than in the Happywhale dataset. This is highly likely caused by the platform design on data upload, depending directly on GPS locations and the content restrictions in Happywhale, unlike the Flickr platform.

This study does not compare the content of the images but rather their potential to cover the Antarctic continent, their

spatial consistency, and spatial associations. While conducting this comparison, we aim to compare a citizen science-based platform and a social media platform in terms of their differentiating potential for area coverage, data amount, user contribution, spatial data consistency, and platform design restrictions on data production and providing. The comparison presented in this study involves the evaluation of data platforms and the data obtained within them, with the results being presented comparatively. It is emphasized that the features of platforms for studies conducted in citizen science are strictly determined, thereby preventing data manipulation. However, it is discussed that the Flickr platform allows data manipulation, leading to distortions in spatial and content terms. The examination conducted within Flickr and Happywhale platforms contributes significantly to demonstrating the differences in social media and citizen science platforms due to their specific features.

The availability of multimodal data contributes to improving data consistency, whereas single-modal data lacks substantial validation for VGI (Hao & Wang 2020). Nonetheless, basic tests can be conducted using single-modal data to infer spatial patterns within the overall dataset. Furthermore, the spatial movement of a volunteer can provide an intrinsic quality measure depending on trajectory velocity (Gengec 2023). Several conclusions can be drawn from the spatial trajectory of volunteers: Volunteers who remain stationary for extended periods and frequently post may be bots or a group account sharing data from the exact same location. Moving volunteers can be classified as either consistent or inconsistent based on the velocity of their movements. Inconsistently moving volunteers, as determined by velocity, may be either bots or individuals associating random locations by using features like the places library. It is evident that there are numerous uncertainties associated with these individual inferences when considering only single-modal data. However, the use of multimodal data can enhance the investigation of data consistency through crosschecking. For instance, the content of a photograph can be compared with the location information of its surroundings, and any mismatches can be easily excluded from the dataset (Can et al. 2019).

Initially, volunteered geographic information and platforms emerged to enrich urban area data content. Recent studies show that their potential goes beyond urban areas and should not be limited to them (Yan et al. 2020). Even regions located far from urban centers but visited by people can benefit from VGI, and this is of great significance for such areas, considering limited alternative data sources. To explore and develop such approaches, this study focuses on the Antarctic continent, one of the world's largest regions with limited monitoring. This study evaluates the current projects and data sources related to VGI types in Polar Regions. Additionally, it tests and implements data retrieval techniques in the Antarctic region. Subsequently, datasets retrieved from two distinct platforms are evaluated and compared. As this study contributes to the assessment of VGI in the Antarctic region,

it will be further enhanced in the subsequent phases of the project through image processing. The geographic features extracted from the image data are expected to contribute to the creation of an Antarctic Geographic Map. While testing and comparing the results of this study, it is important to note the limitations in data retrieval due to the use of unpaid data retrieval methods and restrictions imposed by platforms on data sharing. However, despite these limitations, it is shown that valuable insights can still be derived from multiple spots in Antarctica using the techniques and datasets obtained. Researchers can adopt both social media and citizen-science based datasets to investigate and understand some parts of the Antarctic regions. This potential can also be enhanced by researchers disseminating their work and informing Antarctic visitors (volunteers who can contribute to citizen science or social media platforms) about data collection and sharing. In this way, both data capacity and data quality can be improved by the united effort of researchers and volunteers.

Researchers in the fields of geography, tourism, and environmental studies often encounter challenges with data availability, particularly for remote and uninhabited regions of the world. Recently, researchers have begun exploring social media platforms as potential data sources for natural areas (Mota & Pickering 2018). Oteros-Rozas et al. (2018) found

that Flickr contains more images of nature than other platforms. Pickering et al. (2020) conducted research using social media images to understand temporal patterns in visitor views of the highest mountain (alpine area) in Australia, which receives nearly 100,000 visitors annually. Another study retrieved and compared Flickr and public participation GIS data on social values for remote coastal regions. This study revealed that the Flickr and PPGIS datasets depict different aspects of social values, with Flickr being more biased and restricted by accessibility. Although there have been several studies conducted in remote areas using social media data, the availability of such data remains limited. Despite biases and spatial coverage issues in these studies, the areas under investigation are smaller than Antarctica and receive more visitors. Consequently, more data is available in these other remote regions. However, our goal is to contribute, even regionally, to geographical data in Antarctica. The data obtained in this study has demonstrated the availability of data from Flickr and Happywhale sources for only certain regions of Antarctica. Therefore, we plan to increase awareness to enhance voluntary contributions of social media data in the Antarctic continent. In our future studies, we aim to provide a social media hashtag and an information guide to encourage visitors to create social media content for Antarctica.

## Appendix 1: Data Retrieval Algorithm from Flickr Platform

---

### Algorithm 1: Extract flicker data

---

```

Input: text, bbox, minDate, maxDate, deltaTime
Output: Paired set of image and exif data
startDate ← minDate
endDate ← min(startDate + deltaTime - 1 second, maxDate)
content ← set
page ← 1
while startDate < maxDate do
  photos ← flickr.search(text, bbox, page, startDate, endDate)
  foreach photo in photos do
    exif ← flickr.getExif(photo.id)
    put (photo, exif) in content
  end
  pages ← photos.pages
  if pages > 1 then
    for page ← 2 to pages by 1 do
      photos ← flickr.search(text, bbox, page, startDate, endDate)
      foreach photo in photos do
        exif ← flickr.getExif(photo.id)
        put (photo, exif) in content
      end
    end
  end
  startDate ← endDate
  endDate ← min(startDate + deltaTime - 1 second, maxDate)
  page ← 1
end
return content

```

---



## Appendix 2: Data Retrieval Algorithm from Happywhale Platform

---

### Algorithm 2: Extract happywhale data

---

```

Input: minDate, maxDate
Output: Geotagged images retrieved from HappyWhale
startDate  $\leftarrow$  minDate
endDate  $\leftarrow$  maxDate
content  $\leftarrow$  set
while startDate < endDate do
  response  $\leftarrow$  requests.post(startDate, endDate)
  case QuerySizeLimitExceeded do
    | endDate  $\leftarrow$  min(avg(startDate, endDate), endDate)
  end
  otherwise do
    | put response in content
    | startDate  $\leftarrow$  endDate
    | endDate  $\leftarrow$  min(startDate + 1 year, endDate)
  end
end
return content

```

---

**Acknowledgements** We would like to acknowledge the invaluable assistance provided by ChatGPT, a language model developed by OpenAI, in generating and refining the content presented in this article. ChatGPT played a significant role in enhancing the clarity, elegance, and academic tone of the writing. Its contributions in providing informative and well-structured paragraphs greatly improved the overall quality of the article. We extend our sincere appreciation to OpenAI for developing this powerful language model, which has proved to be an invaluable resource in our research endeavors.

**Author contributions** Conceptualization, AGG; methodology, AGG; software, AGG, MS and ODG; validation, AGG; formal analysis, AGG, MS; investigation, AGG, MS and ODG; writing—original draft preparation, AGG, MS and ODG; visualization, AGG.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This study was carried out under the auspices of the Presidency of the Republic of Türkiye, supported by the Ministry of Industry and Technology (TUBİTAK, project no. 122Y193), and coordinated by TUBİTAK MAM Polar Research Institute.

**Data availability** No datasets were generated or analysed during the current study.

### Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

- Ahmouda A, Hochmair HH, Cvetojevic S (2018) Analyzing the effect of earthquakes on OpenStreetMap contribution patterns and tweeting activities. *Geo-Spatial Inf Sci* 21(3):195–212
- Anselin L (1995) Local indicators of spatial association—LISA. *Geogr Anal* 27(2):93–115
- Ball J (2002) Towards a methodology for mapping 'regions for sustainability' using PPGIS. *Prog Plan* 58(2):81–140
- Bivand R (2022) R packages for analyzing spatial data: A comparative case study with areal data. *Geogr Anal* 54(3):488–518
- Bivand RS, Wong DW (2018) Comparing implementations of global and local indicators of spatial association. *TEST* 27(3):716–748
- Can R, Kocaman S, Gokceoglu C (2019) A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. *ISPRS Int J Geo Inf* 8(7):300
- Council of Managers of National Antarctic Programs (COMNAP). (2017). Antarctic Station Catalogue. [https://static1.squarespace.com/static/61073506e9b0073c7eaaf464/t/611497cc1ece1b43f0eeca8a/1628739608968/COMNAP\\_Antarctic\\_Station\\_Catalogue.pdf](https://static1.squarespace.com/static/61073506e9b0073c7eaaf464/t/611497cc1ece1b43f0eeca8a/1628739608968/COMNAP_Antarctic_Station_Catalogue.pdf)
- Daymond T, Andrew ME, Kobryn HT (2023) Crowdsourcing social values data: Flickr and public participation GIS provide different perspectives of ecosystem services in a remote coastal region. *Ecosyst Serv* 64:101566
- Dong Y, Zhao J, Li C, Liao M (2022) Gapless-REMA100: A gapless 100-m reference elevation model of Antarctica with voids filled by multi-source DEMs. *ISPRS J Photogramm Remote Sens* 186:70–82

- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30(5):609–628
- Elwood S (2008) Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal* 72(3):133–135
- Frame B, Liggett D, Lindström K, Roura RM, van der Watt LM (2022) Tourism and heritage in Antarctica: exploring cultural, natural and subliminal experiences. *Polar Geogr* 45(1):37–57
- Garimella VRK, Alfayad A, Weber I (2016) Social media image analysis for public health. In Proc 2016 CHI Conf Hum Factors Comput Syst (pp. 5543–5547)
- Genec AGG (2023) Exploring crowdsourcing accountability for mapping Antarctica: a case study using 5 years of social media data. *Turkish J Earth Sci* 32(8):1041–1051
- Gentry J, Gentry MJ, RSQLite S, Artistic RL (2016) Package ‘twitteR’. Cran. r-project, <https://cran.irsn.fr/web/packages/twitteR/twitteR.pdf>
- Goodchild MF (2007a) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69:211–221
- Goodchild MF (2007b) Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *Int J Spatial Data Infrastruct Res* 2(2):24–32
- Grinberger AY, Schott M, Raifer M, Zipf A (2021) An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Trans GIS* 25(2):622–641
- Gulnerman AG, Karaman H, Pekaslan D, Bilgi S (2020) Citizens’ spatial footprint on Twitter—anomaly, trend and bias investigation in Istanbul. *ISPRS Int J Geo Inf* 9(4):222
- Gulnerman AG (2021) Changing pattern of human movements in Istanbul during covid-19. In: Gervasi O et al (eds) Computational science and its applications – ICCSA 2021. ICCSA 2021. Lecture Notes in Computer Science, vol 12957. Springer, Cham. [https://doi.org/10.1007/978-3-030-87013-3\\_17](https://doi.org/10.1007/978-3-030-87013-3_17)
- Gulnerman AG, Karaman H, Basiri A (2021) New age of crisis management with social media. In: Mobasheri A (eds) Open source geospatial science for urban studies. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham. [https://doi.org/10.1007/978-3-030-58232-6\\_8](https://doi.org/10.1007/978-3-030-58232-6_8)
- Haklay M, Singleton A, Parker C (2008) Web mapping 2.0: The neogeography of the GeoWeb. *Geogr Compass* 2(6):2011–2039
- Hall GB, Chipeniuk R, Feick RD, Leahy MG, Deparday V (2010) Community-based production of geographic information using open source software and Web 2.0. *Int J Geogr Inf Sci* 24(5):761–781
- Hao H, Wang Y (2020) Leveraging multimodal social media data for rapid disaster damage assessment. *Int J Disaster Risk Reduction* 51:101760
- He Y, Liu C (2023) Environmental Impacts and Sustainable Development of Antarctic Tourism: The Chinese Tourists’ Perspectives. In *Towards Sustainable and Resilient Tourism Futures*. Erich Schmidt Verlag GmbH & Co. KG, Berlin, pp 53–72
- Hecht B, Stephens M (2014) A tale of cities: Urban biases in volunteered geographic information. In Proc Int AAAI Conf Web Soc Med 8(1):197–205
- Hochmair HH, Juhász L, Cvetojevic S (2018) Data quality of points of interest in selected mapping and social media platforms. In: Kiefer P, Huang H, Van de Weghe N, Raubal M (eds) Progress in location based services 2018. LBS 2018. Lecture Notes in Geoinformation and Cartography. Springer, Cham. [https://doi.org/10.1007/978-3-319-71470-7\\_15](https://doi.org/10.1007/978-3-319-71470-7_15)
- Huang Q, Wong DW (2016) Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *Int J Geogr Inf Sci* 30(9):1873–1898
- IAATO (2022) Overview of Antarctic Tourism: A Historical Review of Growth, the 2021–22 Season, and Preliminary Estimates for 2022–23. ATCM 2022, <https://iaato.org/wp-content/uploads/2022/06/ATCM44-IAATO-Overview.pdf>
- Kang Y, Gao S, Liang Y, Li M, Rao J, Kruse J (2020) Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scientific Data* 7(1):390
- Kanhere SS (2013) Participatory sensing: crowdsourcing data from mobile smartphones in urban spaces. In: Hota C, Srimani PK (eds) Distributed computing and internet technology. ICDCIT 2013. Lecture Notes in Computer Science, vol 7753. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-36071-8\\_2](https://doi.org/10.1007/978-3-642-36071-8_2)
- Kearney MW (2019) rtweet: Collecting and analyzing Twitter data. *J Open Source Softw* 4(42):1829
- Kullenberg C, Kasperowski D (2016) What is citizen science?—A scientometric meta-analysis. *PLoS ONE* 11(1):e0147152
- LaRue MA, Ainley DG, Pennycook J, Stamatiou K, Salas L, Nur N, Barrington L (2020) Engaging ‘the crowd’ in remote sensing to learn about habitat affinity of the Weddell seal in Antarctica. *Remote Sens Ecol Conserv* 6(1):70–78
- Li Y, Qiao G, Popov S, Cui X, Florinsky IV, Yuan X, Wang L (2023) Unmanned aerial vehicle remote sensing for antarctic research: a review of progress, current applications, and future use cases. *IEEE Geosci Remote Sens Mag* 10(1):73–93
- Liu Z, Yang J, Huang X (2023) Landsat-derived impervious surface area expansion in the Arctic from 1985 to 2021. *Sci Total Environ* 905:166966
- Matsuoka K, Skoglund A, Roth G, de Pomereu J, Griffiths H, Headland R, Melvær Y (2021) Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands. *Environmen Modell Softw* 140:105015
- Middleton SE, Kordopatis-Zilos G, Papadopoulos S, Kompatsiaris Y (2018) Location extraction from social media: Geoparsing, location disambiguation, and geotagging. *ACM Transact Inf Syst (TOIS)* 36(4):1–27
- Mooney P, Corcoran P, Winstanley A (2010) A study of data representation of natural features in openstreetmap. In Proc GISci 150:150–156
- Mota VT, Pickering C (2018) How can we use social media to know more about visitors to natural areas? The 9th International Conference on Monitoring and Management of Visitors in Recreational and Protected Areas (MMV9), Bordeaux, France. [https://mmv.boku.ac.at/refbase/files/teles\\_da\\_mota\\_vanessa\\_pickering-2018-how-use-social-media.pdf](https://mmv.boku.ac.at/refbase/files/teles_da_mota_vanessa_pickering-2018-how-use-social-media.pdf)
- Norwegian Polar Institute (2018) Get Started with Quantarctica. Retrieved from [https://tryggve.npolar.no/WEB/quantarctica/Quantarctica\\_GetStarted.pdf](https://tryggve.npolar.no/WEB/quantarctica/Quantarctica_GetStarted.pdf)
- Ogie RI, Forehead H (2017) Investigating the accuracy of georeferenced social media data for flood mapping: The PetaJakarta. org case study. In 2017 4th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). IEEE, pp 1–6. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1266&context=smartpapers>
- Oteros-Rozas E, Martín-López B, Fagerholm N, Bieling C, Plieninger T (2018) Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecol Ind* 94:74–86
- Owuor I, Hochmair HH (2020) An Overview of Social Media Apps and their Potential Role in Geospatial Research. *ISPRS Int J Geo Inf* 9(9):526. <https://doi.org/10.3390/ijgi9090526>
- Pebesma E, Bivand R (2023) Spatial data science: with applications in R. Chapman and Hall/CRC
- Pickering C, Walden-Schreiner C, Barros A, Rossi SD (2020) Using social media images and text to examine how tourists view and value the highest mountain in Australia. *J Outdoor Recreat Tour* 29:100252
- Poiani TH, Rocha RDS, Degrossi LC, De Albuquerque JP (2016) Potential of collaborative mapping for disaster relief: a case study

- of OpenStreetMap in the Nepal earthquake 2015. In: 2016 49th Hawaii International Conference on System Sciences (HICSS) IEEE, Koloa, pp 188–197. <https://doi.org/10.1109/HICSS.2016.31>
- Rack U (2018) Exploring and mapping the Antarctic: histories of discovery and knowledge. In: The Routledge Handbook of the Polar Regions, 1st Edt. Routledge, pp 34–44
- Reichert M, Braun U, Lautenbach S, Zipf A, Ebner-Priemer U, Tost H, Meyer-Lindenberg A (2020) Studying the impact of built environments on human mental health in everyday life: methodological developments, state-of-the-art and technological frontiers. *Curr Opin Psychol* 32:158–164
- Reid T, Walter T, Enge P, Fowler A (2014) Crowdsourcing arctic navigation using multispectral ice classification and GNSS. Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014), Tampa, pp 707–721
- Runge CA, Hausner VH, Daigle RM, Monz CA (2020) Pan-Arctic analysis of cultural ecosystem services using social media and automated content analysis. *Environ Res Commun* 2(7):075001
- Schott M, Lautenbach S, Größchen L, Zipf A (2022) Openstreetmap Element Vectorisation-a Tool for High Resolution Data Insights and its Usability in the Land-Use and Land-Cover Domain. *Int Arch Photogramm Remote Sens Spat Inf Sci* 48:395–402
- See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, Rutzinger M (2016) Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS Int J Geo-Inf* 5(5):55
- Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay M (2017) A review of volunteered geographic information quality assessment methods. *Int J Geogr Inf Sci* 31(1):139–167
- Shapley D (2013) *The Seventh Continent: Antarctica in a Resource Age*, 1st ed. RFF Press. <https://doi.org/10.4324/9781315064352>
- Sieber R (2006) Public participation geographic information systems: A literature review and framework. *Ann Assoc Am Geogr* 96(3):491–507
- Steiger E, Westerholt R, Zipf A (2016) Research on social media feeds – A GIScience perspective. In: Capineri C, Haklay M, Huang H, Antoniou V, Kettunen J, Ostermann F, Purves R (eds) *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press, London, pp 237–254. <https://doi.org/10.5334/bax.r>
- Tsou MH (2015) Research challenges and opportunities in mapping social media and Big Data. *Cartogr Geogr Inf Sci* 42(sup1):70–74
- Turner A (2006) Introduction to neogeography. O'Reilly Media, Inc. URL 1 <https://www.zooniverse.org/projects/douglas-clark/the-arctic-bears-project>. Accessed 10/09/2022
- URL 10 (n.d.) <https://www.crowdtangle.com/>. Accessed 10/09/2022
- URL 11 (n.d.) <https://github.com/strohne/Facepager>. Accessed 10/09/2022
- URL 12 (n.d.) <https://twitter.com/search-advanced>. Accessed 10/09/2022
- URL 13 (n.d.) <https://onemilliontweetmap.com/>. Accessed 10/09/2022
- URL 14 <https://fdown.net/>. Accessed 10/09/2022
- URL 15 (n.d.) <https://www.storysaver.net/>. Accessed 10/09/2022
- URL 16 (n.d.) <https://twittervideodownloader.com/>. Accessed 10/09/2022
- URL-17 <https://flickrdownloader.com/>. Accessed 10/09/2022
- URL 18 <https://www.tweepy.org/>. Accessed 10/09/2022
- URL 19 <https://github.com/pablobarbera/instaR>. Accessed 10/09/2022
- URL 2 <https://www.zooniverse.org/projects/cmnbotany/notes-from-nature-expedition-arctic-botany>. Accessed 10/09/2022
- URL 20 <https://instaloader.github.io/>. Accessed 10/09/2022
- URL 21 <https://developers.facebook.com/docs/graph-api>. Accessed 10/09/2022
- URL 22 <https://github.com/pablobarbera/Rfacebook>. Accessed 10/09/2022
- URL 23 <https://www.flickr.com/services/api/>. Accessed 10/09/2022
- URL 25 <https://www.npolar.no/quantarctica/>. Accessed 10/03/2022
- URL 26 <https://iaato.org/information-resources/data-statistics/visitor-or-statistics/visitor-statistics-downloads/>. Accessed 10/03/2022
- URL 3 <https://www.citizenscience.gov/catalog/44/>. Accessed 10/11/2022
- URL 4 <https://scistarter.org/penguinwatch>. Accessed 10/11/2022
- URL 5 <https://www.kickstarter.com/projects/raceagainsttime/plan-d-a-race-against-time?ref=discovery&term=polar>. Accessed 10/11/2022
- URL 6 <https://www.kickstarter.com/projects/1322320805/portraits-of-place-in-antarctica-ii?ref=discovery&term=antarctic>. Accessed 10/11/2022
- URL 7 [happywhale.com/browse](https://happywhale.com/browse). Accessed 10/09/2022
- URL 8 (n.d.) <https://download.geofabrik.de/> (13/07/2023)
- URL 9 (n.d.) <https://www.smrfoundation.org/> (10/09/2022)
- URL 24 <https://epsg.io/3031>. Accessed 10/03/2022
- Van Zanten BT, Van Berkel DB, Meentemeyer RK, Smith JW, Tieskens KF, Verburg PH (2016) Continental-scale quantification of landscape values using social media data. *Proc Natl Acad Sci* 113(46):12974–12979
- Wang Z, Ye X (2018) Social media analytics for natural disaster management. *Int J Geogr Inf Sci* 32(1):49–72
- Xiao Y, Huang Q, Wu K (2015) Understanding social media data for disaster management. *Nat Hazards* 79:1663–1679
- Xu X, Liu C, Liu C, Hui F, Cheng X, Huang H (2022) Fine-resolution mapping of the circumpolar Arctic Man-made impervious areas (CAMI) using sentinels. *OpenStreetMap and ArcticDEM Big Earth Data* 6(2):196–218
- Yan Y, Feng CC, Huang W, Fan H, Wang YC, Zipf A (2020) Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *Int J Geogr Inf Sci* 34(9):1765–1791
- Zhang D, Guo B, Yu Z (2011) The emergence of social and community intelligence. *Computer* 44(7):21–28

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.