CellPress
OPEN ACCESS

One Earth

Review

Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use

Andrea Ghermandi,[1,2,*] Johannes Langemeyer,[3,4,*] Derek Van Berkel,[5,*] Fulvia Calcagni,[1,4] Yaella Depietri,[2,6] Lukas Egarter Vigl,[7] Nathan Fox,[5,8] Ilan Havinga,[9] Hieronymus Jäger,[10,11] Nina Kaiser,[12,26] Oleksandr Karasov,[13,14,15,16] Timon McPhearson,[17,18,19] Simone Podschun,[20] Ana Ruiz-Frau,[21] Michael Sinclair,[2,22,23] Markus Venohr,[20] and Spencer A. Wood[24,25]

[1]Department of Natural Resources and Environmental Management, University of Haifa, Haifa, Israel
[2]Natural Resources and Environmental Research Center, University of Haifa, Haifa, Israel
[3]Institute of Environmental Science and Technology, Universitat Autònoma de Barcelona, Barcelona, Spain
[4]Department of Geography, Humboldt Universität zu Berlin, Berlin, Germany
[5]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA
[6]Faculty of Architecture and Town Planning, Technion - Israel Institute of Technology, Haifa, Israel
[7]Institute for Alpine Environment, Eurac Research, Bolzano, Italy
[8]School of Geography and the Environment, University of Southampton, Southampton, UK
[9]Environmental Systems Analysis Group, Wageningen University, Wageningen, the Netherlands
[10]ifuplan – Institute for Environmental Planning and Spatial Development, Munich, Germany
[11]Department of Ecology, University of Innsbruck, Innsbruck, Austria
[12]Department of Aquatic Ecology, University of Duisburg-Essen, Essen, Germany
[13]Digital Geography Lab, Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland
[14]Helsinki Institute of Urban and Regional Studies (Urbaria), University of Helsinki, Helsinki, Finland
[15]Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland
[16]Department of Geography, University of Tartu, Tartu, Estonia
[17]Urban Systems Lab, The New School, New York, NY, USA
[18]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
[19]Cary Institute of Ecosystem Studies, Millbrook, New York, NY, USA
[20]Department of Ecohydrology and Biogeochemistry, Leibniz Institute of Freshwater Ecology and Inland Fisheries - IGB, Berlin, Germany
[21]Instituto Mediterraneo de Estudios Avanzados (CSIC), Esporles, Spain
[22]Urban Big Data Centre, University of Glasgow, Glasgow, UK
[23]School of Social & Political Sciences, University of Glasgow, Glasgow, UK
[24]eScience Institute, University of Washington, Seattle, WA, USA
[25]EarthLab, University of Washington, Seattle, WA, USA
[26]Environmental Campus Birkenfeld, University of Applied Sciences Trier, Trier, Germany
*Correspondence: aghermand@univ.haifa.ac.il (A.G.), johannes.langemeyer@uab.cat (J.L.), dbvanber@umich.edu (D.V.B.)
https://doi.org/10.1016/j.oneear.2023.02.008

SUMMARY

Social media data are transforming sustainability science. However, challenges from restrictions in data accessibility and ethical concerns regarding potential data misuse have threatened this nascent field. Here, we review the literature on the use of social media data in environmental and sustainability research. We find that they can play a novel and irreplaceable role in achieving the UN Sustainable Development Goals by allowing a nuanced understanding of human-nature interactions at scale, observing the dynamics of social-ecological change, and investigating the co-construction of nature values. We reveal threats to data access and highlight scientific responsibility to address trade-offs between research transparency and privacy protection, while promoting inclusivity. This contributes to a wider societal debate of social media data for sustainability science and for the common good.

INTRODUCTION

With more than half of the world's population active on social media (SM) networks,[1] unprecedented amounts of user-generated data are opening new frontiers in the investigation of human interactions with the natural environment. Researchers are increasingly turning to these data to investigate social-ecological systems and ecosystem services,[2] analyze climate change discourses,[3] explore urban sustainability,[4] and provide novel insights for ecology and conservation science.[5,6] The interpretation of the digital traces of people's values of nature, as manifested in SM data streams, promises to add insight into individual beliefs and societal processes that might be key to motivating and honing conservation messages.[7] This is particularly important in a context where technological advances generate an "extinction of experiences" and a troubling

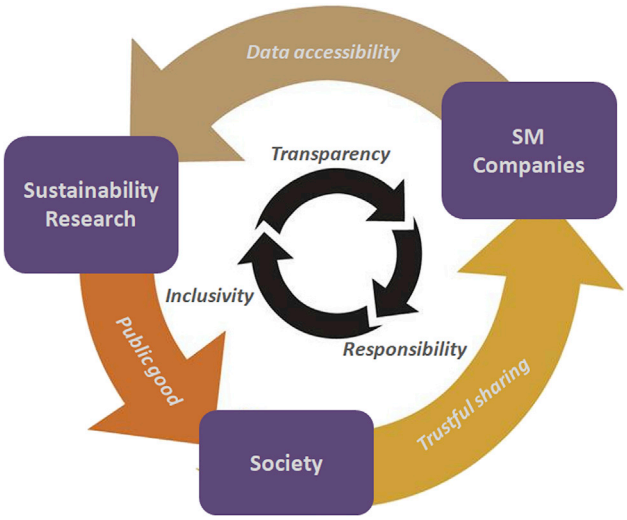

Figure 1. A virtuous cycle for social media (SM) data and sustainability through transparency, inclusivity, and responsible data use

disconnect from nature.[8] Such insights may be particularly valuable at a time when humanity is facing a formidable set of global environmental challenges, as the UN Sustainable Development Goals (SDGs) clearly articulate and address.[9] As a new kind of "environmental information system,"[10] the promise of SM data for tackling such challenges is tantalizing.[11]

The extent to which SM data are accessible to environmental and sustainability researchers in the future is likely to determine whether such potential will be fulfilled. Although largely generated by individual users and organizations, access to SM content is overseen by the private entities that provide the necessary Web-based services. Unless legislated differently, such companies can unilaterally change terms of service at any point in time, restrict accessibility, or apply content filters and censorship, thus hampering the use of these data in research and practice.[12] Current business models, and resulting data collection and sharing practices, have generated a vicious cycle in which user data are treated as a private asset that can be purchased or sold for profit. This has raised public concern and mistrust in SM companies, in turn leading to societal pressure to regulate them, culminating in regulation against the misuse of personal data, such as the 2018 European Union (EU) General Data Protection Regulation. Not only has this threatened the perceived legitimacy of SM companies, but it may also limit the potential public benefits from researching this unique data source. The lost opportunities to use SM data due to access restrictions, whether arising from corporate practices or governmental regulation of data collection and sharing,[13] may be of a comparable size with the harm done by data misuse.[14,15]

The establishment of virtuous cycles for enabling the wide potential of SM research for sustainability will require collaboration between SM companies, environmental researchers, and society at large (Figure 1). More open and meaningful data sharing[16] by SM companies, including granting independent access and analysis by researchers, would reinforce their perceived legitimacy, resulting in a renewed social license to operate. This could then translate into more trustful data sharing by the users who, in turn, would benefit from knowledge generated by sustainability researchers using SM data. Improved collaboration models may require shifts in the current practices both on the side of SM companies and on that of sustainability researchers. While recognition by the former of the potential public good of the data shared by online users is essential,[17] a coordinated and widely shared commitment on the part of individual researchers toward key principles for a responsible, ethical use of the data might be critical in establishing the trust required for the creation of a "data commons" space. This is especially meaningful considering that guidance from ethical review boards is still limited when it comes to SM research.[18] Such commitment would demonstrate the public benefits of such research. Although there are several examples of data-sharing models between academia and the SM industry, including data collaboratives[19] and data philanthropy initiatives,[10] as well as SM companies granting occasional data access to researchers, such disjointed efforts remain reserved for a small group and are insufficient to ensure that the full potential of SM data is brought to fruition.[20]

This critical review aims at contributing to a societal debate around the value of promoting more open access to SM data for research purposes and the ethical challenges associated with it, particularly from the perspective of environmental and sustainability research. Based on the review of 415 studies, we first articulate the potential benefits of SM data in light of the SDG targets and the characteristics that make these data unique and potentially irreplaceable in applications at the nature-society interface. Subsequently, we highlight current restrictions and threats to future data accessibility for sustainability research. Finally, we define sustainability-specific principles for a shared ethical commitment by researchers toward responsible data use and evaluate how the scientific literature has fared against them thus far.

SOCIETAL BENEFITS FROM SM-BASED SUSTAINABILITY RESEARCH

Over the past decade, there has been a rapid increase in environmental and sustainability studies using SM data, with authors frequently acknowledging the novel opportunities that such digital technologies offer in detecting and examining a broad range of human-nature interactions. To showcase the range of societal benefits that may be achieved through SM-based sustainability research, Figure 2 presents a carefully selected list of applications, which are drawn from the 415 studies in the database. For each relevant SDG target, one or multiple fields of application are given, along with a real-world example of how this potential has been realized in one of the investigated studies. The studies referenced in Figure 2 are not chosen based only on their perceived quality and innovation value but also to provide an overview of the broad range of the spatial scales, socio-economic contexts, and environmental issues covered in the dataset.

Figure 2 identifies 12 SDGs (out of 17) and 29 SDG targets (out of 169) that are, usually implicitly, addressed by the reviewed studies. It should be noted that SM-based research in general is most likely relevant for additional SDGs and SDG targets

SDG	SDG target	Example application
2	Sustainable agriculture (2.4)	Assess cultural significance of agricultural landscapes: Chianti region, Italy[21]
		Map and monitor urban farming: four world metropolises[22]
4	Sustainable development education (4.7)	Explore meaning-making about environment and sustainability: young adults in Sweden[23]
6	Safe drinking water (6.1)	Gauge public attitude toward management: water charges in Ireland[24]
		Inform operation of supply systems: catchment hydrology in Italian Alps[25]
	Water quality (6.3)	Promote sustainable sanitation: benefits of nature-based solutions[26]
		Value benefits of quality improvement: recreation in Minnesota lakes[27]
	Water-related ecosystems (6.6)	Inform ecosystem restoration: tourism in a Ramsar wetland in India[28]
		Assess provision of cultural ecosystem services: rivers in Idaho[29]
7	Access to energy (7.1)	Explore public perception of energy supply: hydraulic fracturing[30]
	Share of renewable energy (7.2)	Understand public opinion on renewable energy: local opposition to wind power project in Germany[31]
8	Sustainable tourism (8.9)	Characterize spatial-temporal patterns of tourist visits: Areas Of Interest in six world metropolises[32]
		Analyze tourist movements and choices: tourist routes in NYC[33]
9	Sustainable and clean industries (9.4)	Analyze sustainability marketing communication: Fortune 500 enterprises[34]
10	Social inclusion (10.2)	Address inequality in access to natural areas: green gentrification in Barcelona[35]
11	Public transport (11.2)	Analyze cycling infrastructure and their use: path networks in Belgium[36]
		Plan and improve public transport systems: human mobility in Chicago[37]
	Inclusive and sustainable urbanization (11.3)	Characterize visual quality of urban landscape: public open spaces in Munich, Germany[38]
		Map urban functions and urban land use: use of streets in London and associated semantics[39]
	Cultural and natural heritage (11.4)	Examine use and management of heritage sites: UNESCO World Heritage sites in conflict areas[40]
	Resilience to disasters (11.5)	Detect and characterize flood extent and severity: flooding thresholds in US East Coast[41]
		Social sensing natural hazards for footprint and damage assessment: Hurricane Sandy[42]
		Enhance preparedness, response, recovery: wildfires in Sumatra[43]
	Urban green and public spaces (11.7)	Assess ecosystem services of urban parks and green infrastructure: green spaces in Helsinki[44]
		Evaluate well-being benefits from exposure to nature: Nanjing residents during COVID-19 pandemic[45]
		Understand public opinion, perceptions, satisfaction: green spaces in Dublin[46]
12	Management of chemicals and wastes (12.4)	Monitor solid waste management: odors from landfills in China[47]
		Infer urban air pollution levels: air quality index for Chinese cities[48]
	Corporate sustainable practices (12.6)	Inform corporate sustainability practices: spillover effects of environmental regulation in China[49]
	Sustainable development awareness (12.8)	Uncover public perspectives on sustainability topics: global debate about land grabbing[50]
		Understand perceptions of nature: testing the biophilia hypothesis[51]
		Explore spread of sustainability information: Deepwater Horizon oil spill[52]

(figure continued on next page)

13	Resilience and adaptive capacity (13.1)	Identify mismatches in socio-ecological systems: phenology and visitation in Mount Rainier National Park[53]
	Climate change policies (13.2)	Explore perception of impacts and policies: remarkability of temperature anomalies[54]
	Climate change awareness (13.3)	Analyze online discussions on climate change: communication on weather extremes in China[55]
14	Marine and coastal ecosystems (14.2)	Assess coastal and marine ecosystem services: global coral reef tourism[56]
		Map human interactions with marine species and disturbance to ecosystems: Hawaiian monk seal[57]
	Conservation of coastal areas (14.5)	Assess benefits of marine protected areas: cultural ecosystem services of 14 areas worldwide[58]
15	Terrestrial and freshwater ecosystems (15.1)	Map land use/land cover or geomorphometry: landscape variation and folksonomies in Switzerland[59]
		Analyze public perception and benefits of terrestrial protected areas: recreation on public lands in New Mexico and Washington states[60]
		Investigate human-nature conflicts: unwanted visitors' behavior in South African national park[61]
		Assess cultural ecosystem services: growth in Arctic eco-tourism[62]
		Quantify landscape aesthetic values: the European continent[63]
		Complement traditional monitoring: climate in the UK[64]
	Sustainable forest management (15.2)	Assess cultural ecosystem services of forests and urban vegetation: mangroves in Singapore[65]
	Conservation of mountain ecosystems (15.4)	Assess mountain cultural ecosystem services: changes in landscape value over 150 years in Austria[66]
	Loss of biodiversity (15.5)	Collect information on species ecology and behavior: spatial variation in species traits in Japan[67]
		Map species distribution: UK flowering plants[68]
		Characterize human-wildlife interactions: encounters with giant pandas in China[69]
		Analyze perceptions of biodiversity and endangered species: sentiment towards iconic species[70]
	Protected species trafficking (15.7)	Monitor online wildlife trade: Indonesian songbirds[71]
	Invasive alien species (15.8)	Monitor spread of non-native species: oak processionary in Europe[72]

Notes: = Understanding direct human-nature interactions at scale
= Observing temporal dynamics of social-ecological change
= Investigating the co-construction of meaning and values

Figure 2. Selected examples of applications of social media data to sustainability research and the related Sustainable Development Goal (SDG) target

This chart lists 52 example applications.[21–72] Notes: the icons refer to the most relevant among the themes discussed in section "societal benefits from SM-based sustainability research."

that are not captured in Figure 2 due to this study's primary focus on environmental issues.

The thematic and geographical scope of the studies in Figure 2 is wide, ranging from eco-tourism in the Arctic[62] and India[28] to preparedness and response to environmental disasters in the USA[41] and Indonesia[43]; from perceptions of climate change impacts[55] and related policies[54] to the well-being benefits of nature exposure during the COVID-19 pandemic[45]; from the spatial distribution of animal species and their traits[67] to the monitoring of the wildlife trade.[71] Specific example studies from Figure 2 are discussed in further detail in sections "understanding direct human-nature interactions at scale" and "investigating the co-construction of meaning and values."

The majority of the studies explore issues that are related to either SDG 15 "Life on land" (35%) or SDG 11 "Sustainable cities and communities" (29%) (see also Figure 3). In particular, 24% of

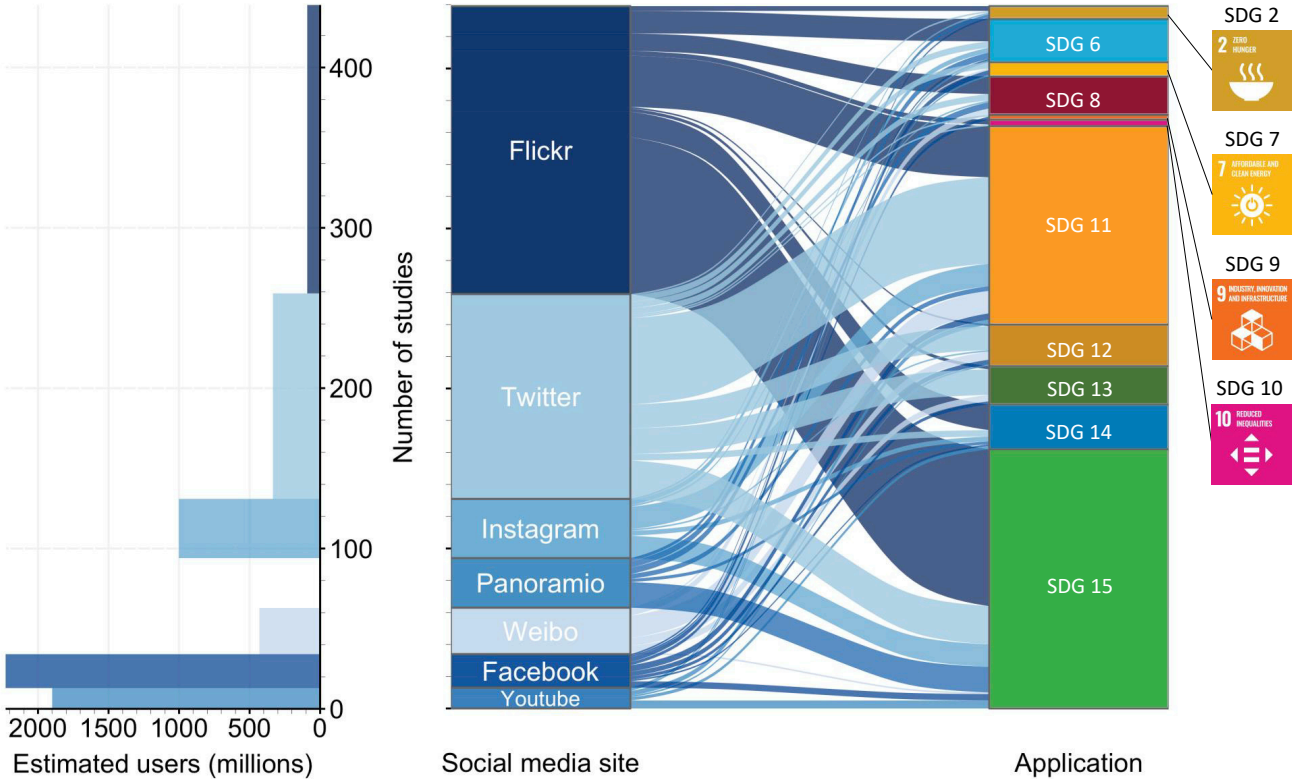

Figure 3. Distribution of studies with data from the seven most frequently investigated SM platforms against the approximate number of platform users and the related UN Sustainable Development Goal (SDG)

Notes: studies relying on data from multiple platforms are assigned to each of the platforms; 47 studies using data from other sources such as blogs (N = 5), TripAdvisor (N = 5), Dianping (N = 4), and Strava (N = 4) are not plotted.

studies find application in the context of SDG target 15.1 "Terrestrial and freshwater ecosystems," examining, among others, questions related to public perceptions and use of terrestrial ecosystems, including protected areas, or the monitoring of environmental quality (see Table S1 for a breakdown of all 415 studies according to SDG target and field of application). Among studies addressing the sustainability of cities and communities (SDG 11), most investigate the use of urban green spaces, including well-being benefits from exposure to nature (11% of studies) or aim at improving the monitoring, response, and resilience to natural disasters such as floods, hurricanes, and wildfires (10% of studies). An additional 6% of studies analyzes tourism patterns, primarily in cities and tourism hotspots (SDG target 8.9 "Sustainable tourism"). The focus on urban areas and terrestrial ecosystems may reflect that SM data are primarily generated in populated or easily accessible areas. Only 6% of studies address topics that are of relevance for marine ecosystems (SDG 14 "Life below water"). Most of the remaining studies do not relate to the spatial dimension or specific geographical location of the data but rather analyze SM content to assess climate change awareness and perceptions (6% of studies; SDG 13 "Climate action"), sustainable development awareness and the management of pollutants (6% of studies; SDG 12 "Responsible consumption and production"), or safe drinking water and sustainable sanitation (6% of studies; SDG 6 "Clean water and sanitation").

Figure 3 shows the distribution of the investigated studies according to the SM platforms from which the data were retrieved. The estimated number of users per platform is also included (see related discussion in section "threats and limitations to the use of SM data in sustainability research"). In the following subsections, we synthesize the information retrieved by identifying three research themes in which SM data can and, to the extent revealed by the investigated studies, already play a role as a source of unique insights for environmental sustainability research.

Understanding direct human-nature interactions at scale

Achieving the SDGs will require large-scale, multi-country efforts as well as granular data for tailoring sustainability efforts. SM data have enabled an unprecedented view of how people interact with natural populations, ecosystems, and biomes, over large spatial scales.[55,63,73] Of the 415 studies examined, about 11% were conducted at the global scale, 5% at continental, and 36% at regional to national scales, a result that is hard to match using traditional survey methods. There are also promising findings that suggest that SM data from different platforms are geographically consistent over large extents; for instance, high correlation between users' posts was found across three different SM platforms (i.e., Flickr, Panoramio, and Instagram) for the entire European continent.[63] Although additional testing

is necessary to confirm this spatial consistency, this potential for replicability may offer unique generalizability about location-specific interactions.[74] Figure 2 offers several examples of studies taking advantage of SM data for research at wide spatial scales, including the analysis of human-nature interactions in coastal ecosystems and marine protected areas (SDG 14),[56,58] the assessment of public awareness surrounding land grabs in Africa and worldwide (SDG 12),[50] and the evaluation of threats to cultural and natural heritage sites located in conflict areas (SDG 11).[40]

The content of texts and images that are shared as SM can offer valuable insights into the motivations, purposes, and perceptions of individual users.[65,75] This is not possible with other emerging "big data" approaches, such as those based on tracking of mobile phone locations, and it can be achieved in a way that is potentially less intrusive than surveys and less prone to biases introduced by the researcher (e.g., questionnaire design and potential interviewer effects) or the respondent (e.g., recall biases). The extraction of semantic information from SM can assist, for instance, in characterizing individual attitudes and engagement with environmental topics such as climate change (SDG 13),[76] understanding behavioral dynamics of visitors in natural and semi-natural areas (SDG 15),[77] and identifying typologies of users based on their interests and cultural background (SDG 6).[78] The current rapid advancements in scalable machine learning tools promises to further enhance our ability to manage and respond to different social and environmental risks.[78,79]

Observing temporal dynamics of social-ecological change

Research is increasingly revealing that SM data can play a unique role in the analysis of the temporal dynamics of social-ecological interactions and environmental change, including disaster risk reduction (SDG 11),[42,54] environmental quality monitoring (SDG 12),[48] and improved transportation and mobility (SDG 11).[37] The quasi-instantaneous nature of SM communication and the speed of data retrieval support high temporal resolution, continuous and (near) real-time analysis of evolving environmental and socio-economic processes,[80] including the use of urban greenspace during the COVID-19 outbreak (SDG 11)[81] and monitoring of invasive species (SDG 15).[82] For events requiring rapid and dynamic responses, such as natural disasters, SM monitoring can play a vital role in capturing and organizing sources from eyewitness accounts, a type of data source that is hardly available with conventional methods.[83] Promising advances in this context include the development of effective, real-time architectures for the analysis of multimodal (visual and textual) SM content[84] and the integration with other data sources, such as remote sensing images.[85]

SM-data-based assessments can also offer a rich historical record of human-nature interactions going back to the mid-2000s, when some of today's leading platforms were launched. Such a relatively long time span compares favorably with that of alternative sources of ecological data such as active crowdsourcing by volunteer observers through platforms such as iNaturalist.[68] The continuous availability of SM data over such time spans may allow updating previous studies with recent data, which is often a limitation with resource-intensive methods such as surveys.

Investigating the co-construction of meaning and values

Unlike in traditional media, or so-called representational media, where "few gatekeepers" control the mass production of popular culture and broadcast it to a passive audience, virtual communities present unique forums where people directly (e.g., through discussions) or indirectly (e.g., by sharing photographs or video contents) communicate and debate their feelings about climate change, perceptions of the environment, and interactions with nature.[3] By sharing nature experiences with a virtual community of peers, people collaboratively generate meaning and ascribe value to nature.[86] SM research is increasingly exploring such co-construction of values and meaning regarding nature and sustainability. Studies have explored, for instance, the expression of emotional responses to climate change-related extreme weather events (SDG 13),[54,55] the circulation and interpretation of environmental information (SDG 12),[52,87] and the shaping of arguments for a stronger appreciation and protection of nature (SDG 12).[51,88] Whether pro-environmental collective meanings (e.g., attachments, commitments, responsibilities, and positive relationships with and within nature) and "digital relational values" are fostered by SM, and whether digital co-creation of societal values regarding nature and sustainability might motivate and sustain public support for ecosystem protection and environmental stewardship,[89] remains an open research frontier.[90]

THREATS AND LIMITATIONS TO THE USE OF SM DATA

Although an in-depth discussion of the limitations of SM data is beyond the scope of this paper, a series of concerns emerge from the reviewed studies, which future SM-based sustainability research will need to grapple with. Biases in SM data may arise at three main levels: users, content, and analysis. The lack of verified personal and socio-demographic information of individual SM *users* has curtailed research to simple classifications of behavioral and socio-demographic types.[78,91] This lack of information may obscure problematic biases (e.g., in geographic representativeness, age, gender, socio-economic condition, education), underpinned by differences in Internet and technology use[92–94] or the appeal that specific platforms have for specific audiences and users.[95,96] Such effects may be exacerbated by relying on a single source of data (see related discussion in section "promoting a virtuous cycle for SM data and sustainability: the role of research"). The *content* of SM posts may also be biased toward subjects or topics that are more likely to be shared because they are perceived to be unusual,[97] valuable,[98] or have a higher social desirability.[86] Verbal communication may amplify tendencies toward homophily and segregation,[99] leading to superficial[100] or polarized discussions where either positive or negative views prevail.[101,102] The design of algorithms purposefully designed to direct users to extreme and emotionally charged content likely contributes to such polarization.[103] Finally, biases may be introduced during the *analysis* of the data. Bias toward very active or very influential users is a common challenge in the field, and methods are needed to control

for and identify how this skews findings.[104,105] Additionally, there is still a limited understanding of how the SM literature may be influenced by the technical complexity of analyzing specific data formats (e.g., YouTube videos), difficulty in coping with SM content in multiple languages, and reliance on specific data science techniques or tools.[106,107] In spatial studies, analysts are often still challenged to achieve sufficient observations for the fine spatial scale and temporal resolutions necessary for decision making.[108] After controlling for active users and the large amount of posts in highly visited locations (e.g., popular attractions, cities), the distribution of SM posts is often highly sparse.[109] Studies often obviate this problem through aggregated analysis at cruder spatial units (e.g., federal parks, Natura 2000), but this may obscure important behavioral-environmental interactions (for example, hiking traffic within protected areas). Combining multiple SM streams[110,111] and complementing with independent monitoring approaches, such as, for example, participatory GIS,[94] sensors,[60] and traditional survey methods,[95] is promising to ameliorate these spatiotemporal biases.

A major threat for the future of SM-based sustainability research, one that is central to the argument developed in this paper, relates to data accessibility. Consistent access to a multiplicity of sources is essential for the successful long-term uptake of this new data type, especially considering the changes in the popularity of individual platforms over time and their appeal to different socio-demographic groups.[1] Accessibility likely already plays a key role in the uptake of data from specific platforms. A majority of the reviewed studies use only two platforms, Flickr and Twitter (see Figure 3), which have long held relatively open data access policies. In spite of their much larger total user base,[5] platforms such as Facebook and Instagram have been relatively underutilized, likely due to restrictions in access. The biggest challenges to data accessibility include (1) shutdown of SM platforms, (2) restrictions and changes in the platforms' terms of services, and (3) censorship and data manipulation.

Shutdown of platforms

SM access can be hindered when online services are discontinued. For example, Panoramio, a photo-sharing platform popular in early research,[11] was shut down in 2016. While much of the data were subsequently integrated into Google Maps, they were no longer available for research usage. Numerous other popular photo-sharing (e.g., Webshots, Ovi Share, Kodiak Gallery) and SM platforms (e.g., Friendster, Myspace) have likewise been discontinued, resulting in mothballing and/or irreversible loss of datasets compiled by users over years of activity. Flickr, a favored source of photographic data for environmental studies, was at risk of shutting down, according to a statement by its CEO in December 2019.[112] In this context, the creation of open repositories, such as the Yahoo Flickr Creative Commons 100M database,[113] represents an important safeguard for maintaining historical SM data records and granting continued future accessibility of data to researchers.

Restrictions in terms of services and changes thereof

Access and terms of use of user-generated data compiled through Web applications are largely determined by the companies that operate individual SM platforms.[114] While restrictions may be dictated by regulation or concerns for data privacy protection, preventing open access is often central to the platforms' business models. The data are sometimes sold outright or indirectly used to generate revenues by granting access to application developers for targeted advertising.[115]

The manner of sharing and actual data availability for researchers are highly variable across platforms. In addition to Flickr, platforms such as Twitter, VKontakte, and Mapillary grant fairly broad access to their data. Terms of service for these platforms generally acknowledge that access to data should not violate laws ruling at the national or supranational level (e.g., EU General Data Protection Regulation) and be limited to non-commercial uses. Twitter and Flickr also stipulate that individuals' privacy and wishes be protected using appropriate techniques, including ensuring that stored data reflect the current online status of content and protecting geographic anonymity in visualizations of data. Other platforms, such as Facebook, have traditionally placed strong limitations on the accessibility and automated retrieval of their content, thereby greatly limiting the use of this potentially extensive data source (Figure 3). Recent initiatives, such as the partnership with Social Science One[19] and the Facebook Data for Good initiative,[116] may point toward an increased, though selective, engagement in data sharing for research purposes. In their current form, however, they also risk eroding research independence insofar as they may skew research toward topics that share common interests with technology companies, remain heavily dependent on the companies' willingness to share and deliver the data, and allow technology companies a veto right on who can receive the data.[117] Among the outdoor recreation and sports apps, only a few offer an application programming interface (API) to facilitate data retrievals, and generally with a very limited set of features (e.g., Strava, Under Armour API for the MapMy apps). Many apps (e.g., Wikiloc, AllTrails) substantially restrict the amount of user-generated content that can be manually retrieved from their Web sites. Such limitations may be set, for instance, because trails and all the digital creations around them, such as photographs and descriptions, are the property and copyright of the authors (Wikiloc, unpublished data on 31st October 2018).

Particularly insidious are changes in SM platforms' terms of services and/or their APIs, because they undermine the replicability of studies and the possibility to update previously collected datasets. For instance, multiple changes in Instagram's API over time have increasingly limited the possibility of retrieving detailed geotagged information, thus substantially limiting its use, despite its great potential.[63] For Flickr, multiple changes in ownership have led to changes in the way users can interact with the API (e.g., accessibility of the Yahoo Where On Earth identifier) and the introduction of a limit of 1,000 photos for free storage in 2019 led to the deletion of large amounts of excess photographs in March 2019. After revelations that its geolocated data could be used to locate secret military bases in January 2018, Strava modified its publicly available heatmap. It no longer displays routes with little activity and refreshes itself monthly to clear any data that might have been made private.[118] Several APIs (e.g., Instagram, Strava) have introduced more restrictive rules that limit access to data that are owned by individual users, thus *de facto* precluding the use of the platforms for any large-scale analysis.

Twitter's API offers an unusual case in which data access for researchers has, so far, broadened over time. For years, Twitter's free, standard API had allowed access to only a fraction of its

real-time data stream and limited historical data, with higher levels of access being achievable only by upgrading to paid API versions. Although free access to premium accounts had occasionally been granted for socio-ecological research,[119] this had raised issues about equitable data access due to the cost of such services, which may not have been affordable to all researchers. In 2019, Twitter also removed the option of precisely locating tweets due to a lack of user engagement with this feature, with potential consequences for research use.[120] In January 2021, however, Twitter set a new standard for broader access to researchers by introducing a new academic research product track, which for the first time allows free full-archive searches for approved researchers.[121] Such an approach could serve as a model for wider open access across SM platforms.

Data manipulation and censorship

State censorship and self-censorship of SM content create biases in the data that are difficult to identify and control for. Self-censorship is understood here as the voluntary removal of content by the SM platforms. It is grounded in the ethical guidelines of each platform and usually affects content prompting violence, nudity, pornography, hate speech, and also certain politically controversial content.[122,123] Excessive self-censorship by SM companies may negatively affect the use of SM content in environmental sustainability research (e.g., by removing pro- or anti-environmental content or representatives).[124,125] Ethical guidelines followed by most of the large SM platforms are primarily established along the ethical and legal standards of Western, liberal democracies, above all the US but increasingly also conforming to the standards of the EU's General Data Protection Regulation. This has raised some criticism for generating a new type of "media imperialism" when these guidelines are applied globally.[126]

Commentators have increasingly argued for reliance on established state laws for guidance on censorship rather than platform guidelines.[127] Beyond ethical and legal considerations, a transparent set of laws would decrease ambiguity regarding content suitability and the criteria for censorship.[128] Where applied, however, state censorship has also lacked transparency, with the potential to hamper, among others, the study of topics such as environmental justice and social-environmental movements.[129] The People's Republic of China, for instance, has strictly regulated platform usage to Sina Weibo and WeChat and prohibited or severely restricted the use of US-based platforms such as Facebook, Instagram, Twitter, and YouTube, which dominate in most other countries.[130] Chinese state censorship has been explained to focus on the potential of posts to encourage collective action[131] but is applied differently throughout the country and by platform.[132,133] Similarly, state control over social media content has reportedly been asserted on the platform VKontakte, a leading SM network in Russian-speaking countries.[130,134] In this context, state censorship over domestic platforms is qualitatively different from that over multinational firms, since it can focus on content removal rather than the less effective content blocking.[130]

PROMOTING A VIRTUOUS CYCLE: THE ROLE OF RESEARCH

SM data misuse and acknowledgment of the influence that digital spheres have on public life and societies have raised concerns in many SM users about how their personal data are handled, including by researchers within technology companies, universities, and other organizations. Such public scrutiny is warranted considering that the involvement of academic researchers may have in the past lent credibility to requests for SM data access that were later revealed to result in questionable data-sharing practices in the Cambridge Analytica scandal.[135]

Gaining public trust in academic use of SM data requires commitment by researchers to fair and ethical use practices and establishing responsible research agendas. Research practices in environmental and sustainability science so far, however, show a mixed picture. A lack of clear regulatory guidelines has resulted in diverse approaches for protecting users' privacy, a central aspect when using users' data in research. While recent scholarship has outlined best practices,[136] our review reveals that consideration of ethical aspects in SM data use in sustainability science is, to a large extent, still lacking. Despite the vast majority of research adhering to ethical standards in protecting the anonymity of individual users (Figure 4), there is a lack of transparency in documenting the methods used for mining these data (fewer than 15% openly sharing study data and/or methods) and little stated concern about safe data storage practices (addressed in only 1.5% of studies). Lack of an explicit awareness of potential ethical issues in data use is apparent in the fact that 85% of the studies do not mention ethical or privacy concerns in their data handling and only 2.4% make direct reference to established guidelines. This is concerning, considering that a sizable fraction of the studies involves handling and analysis of potentially sensitive data, such as personal user identifiers, manual or automated analysis of user-generated textual or photographic content, and socio-demographic information that is usually extracted from the users' public profiles (Figure 5).

SM will undoubtedly play an important role in shaping future discourse and behavior on sustainability issues. This will increasingly entangle researchers using SM data in ethical dilemmas on data use and the communication of results. True to a commitment to moving research-generated knowledge into societal action,[137] sustainability scientists face, we argue, an ethical imperative to contribute to public discussion on research and private sector use, and inquiry into the complex influences and bias within this unique data source to elevate standards of SM data use broadly. Based on the insights derived from the reviewed studies, we explore three principles for a fruitful discussion on the fair and ethical use of SM data in sustainability research: (1) ensuring inclusivity, (2) balancing the needs for research transparency and privacy protection, and (3) safeguarding the ethical responsibility of researchers.

Inclusivity

A major challenge of SM research is assessing whether data are demographically inclusive and representative of opinions, behavior, and perspectives of the focal population. It is often assumed that data are skewed toward younger generations, but surprisingly little is known of these apparent biases. While early overrepresentation of young, tech-savvy early adopters appears to pose less of a challenge due to wide penetration across socio-demographic groups,[138,139] new composition uncertainties have emerged due to shifting popularity and self-selection of SM platforms.[140] Facebook, once the preferred SM for

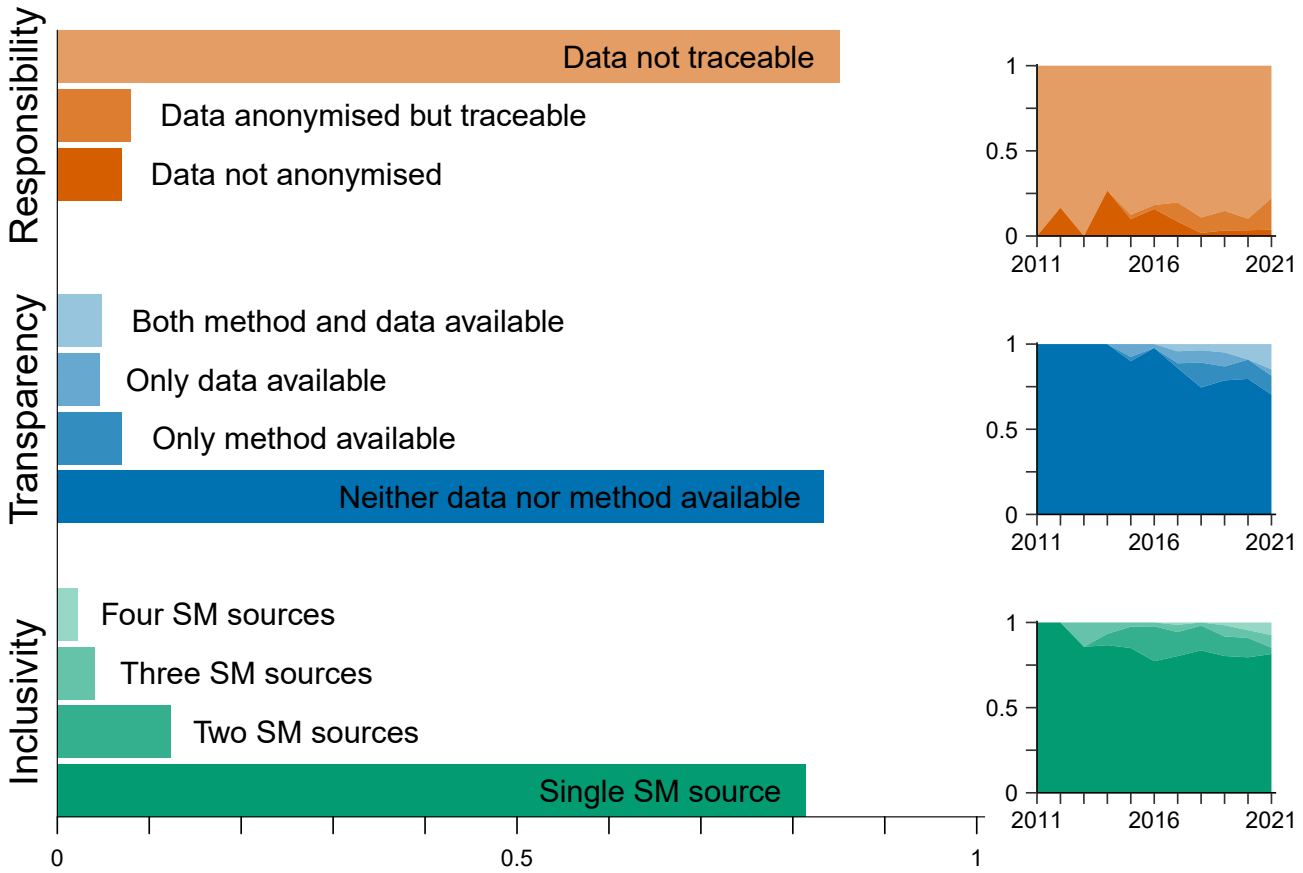

Figure 4. Proportions of reviewed studies (N = 415) engaging with different dimensions of ethical and privacy-related concerns

younger cohorts, is now widely used by adult individuals.[1] The latter are less likely to use platforms such as Instagram, Snapchat, and TikTok.[1] Geographically, there are often insufficient observations for fine-scale analysis,[108] especially for countries where access to the data of popular SM platforms is lacking.[140] The existence of a digital divide among countries with high and low active social network penetration is also evident.[1] Self-censorship, amplification related to "influencers," herd mentality, and platform-specific algorithms are likely to generate biases that are scarcely accounted for in SM research and need investigation with regard to how they affect inclusiveness.

Major efforts should be undertaken to identify, understand, and control representativeness. While much research has confirmed its validity to approximate broad behavioral trends using correlation of independent samples[60,91] and wide geographic analyses across SM platforms,[63] new approaches must assess apparent demographic biases for increased awareness and refinement of the validity of SM data. Techniques that infer users' demographic background using their approximate home location,[91,141] image content,[142] and user-provided profiles[138,139] are promising for filling in these representativeness gaps, and they might be scaled-up for better contextualization of users. Care should be taken that this demographic information is protected due to the possibility of revealing personal information. Mixed-method approaches are likely valuable in this regard, offering the ability to validate qualitative and quantitative assumptions about representativeness and behaviors inferred from SM through first-hand accounts,[94] and gaining informed consent in matters of deeper personal inquiry. Broadening the mix of SM platforms, an aspect that is widely lacking in current research practices (Figure 5), will be imperative for ensuring the geographic and socio-demographic representativeness of data.[60,77] Convenience sampling from single platforms in light of ease of data access should be, as much as possible, avoided. There are inherent challenges in maintaining research relevance given the shifting popularity of platforms that will likely only be addressed given broad buy-in and partnerships between developers and researchers.

Transparency

The open disclosure of data sources and methods has long been a key tenet of sustainability research. Such open science models are increasingly influencing the community of SM researchers.[143] A persistent challenge will be achieving such transparency while protecting the privacy of SM users, given the reliance of many studies on locational attributes and personal profiles from which a range of users' information might be inferred.[138,139] Current best practices for ensuring user privacy, including visualizing location-specific information at aggregated scales, stripping user identifiers, and jittering, are viewed as adequate for protecting users' privacy.[136] Different types of data (e.g., texts, images, videos) will require different strategies, with privacy

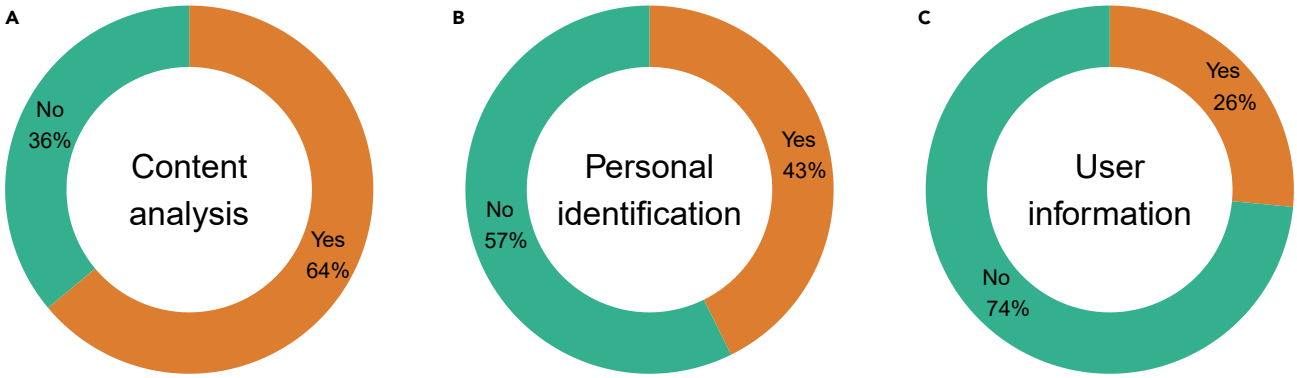

Figure 5. Engagement of reviewed papers with different dimensions of potentially sensitive information associated with social media (SM) data analyses (N = 415)

protection in visual data being, generally speaking, more challenging than textual data.[144] Increasing popularity of SM approaches and accessibility to user-friendly technologies[143] might also touch off additional research that does not adhere to such established procedures. Moreover, growing numbers of studies are likely to increase the scrutiny of these methods, possibly resulting in different interpretations of best practices from research regulating bodies (e.g., institutional review boards) and society.

Anticipating such challenges, we offer three approaches that SM researchers might consider in furthering broadening acceptance of SM-based research:

(1) Publishing detailed data-mining criteria including scripts for API access, timescale, geographic boundaries, and search keywords. This will increase credibility by enhancing reproducibility and ensure the legality of research practices. While documentation might never be fully replicable due to the shifting and ephemeral nature of SM data, it will serve to establish and codify common standards of ethical practice and reduce ethical ambiguities in data retrieval; for instance, in relation to the use of Web scraping. In the long run, this should lead to standardized approaches for documentation of methods, as already adopted in other research fields.[145]

(2) Engaging formal institutional review boards in discussion of the unique character of SM data, best practices, and legal requirements. Review boards will be instrumental in legitimizing SM research, and open dialogue will help in establishing appropriate procedures, outside traditional norms that require informed consent, considering their practical impossibility in the case of SM data. A risk review assessment may better suit research involving SM data than an ethical review procedure.[146]

(3) Undertaking further research into what constitutes reasonable expectations of privacy for data that are publicly available on SM platforms, and the extent to which specific risks to subjects may compare with the importance of the knowledge that may be expected to result from the research. While there remains a debate regarding the extent to which data collected from publicly accessible SM platforms constitute (identifiable) private information (see the 2018 Federal Policy for the Protection of Human Subjects, or "Common Rule"), researchers should minimize risks arising from potential misalignments of SM users' expectations and the use that is made of their data.[136] Only after establishing a strong empirical footing on such matters will scientists be able to broker informed societal discussions on what boundaries should be drawn, and potentially advocate for waivers or relaxation of restrictions due to the public interest of the research.

Responsibility

Among SM researchers, sustainability scientists have a unique responsibility to conduct research that reduces the potential for undesired environmental impacts and, indirectly, the resulting effects on health and well-being of people depending on such environments. While much SM-based research advocates for wider conservation and improved environmental management, the high spatial and temporal resolution of the data also has the potential for misuse and undesired outcomes. For example, while SM data might be used to track illegal trafficking of endangered species,[147] they could also be used to target and exploit these species. Similarly, monitoring of unwanted visitors' behavior in protected areas may perversely end up promoting such behaviors.[61] Models of aesthetic appreciation[63] might be used by developers to find locations for housing and tourism development projects, accelerating amenity migration and rural gentrification,[148,149] and potentially compromising the unique cultural landscapes and natural beauty in receiving areas. Finally, using SM to identify scenic or otherwise special areas can result in rapid increases in visitation, with associated environmental and experiential degradation at sites with limited capacities to accommodate such high use.[150] Researchers should take all necessary precautions to minimize such risks; for example, by not disclosing the precise locations of rare observations such as locations of endangered species, unique natural features, and sensitive ecosystems.

CONCLUSIONS

Within this review, we underscore the potential of SM for sustainability research that serves common interests by creating new

knowledge of human-nature interactions at broad geographical scales, observing temporal dynamics in different social-ecological systems, and understanding how environmental meanings and values are shaped in the digital realm. We show how such insights can be instrumental in addressing the targets set by the UN SDGs. Without downplaying the challenges involved in SM data-based research, which have been extensively investigated in previous studies,[151,152] this form of passive sensing has the potential to usher in a revolution in the current practices of sustainability research, especially in the social sciences,[12] which we consider on par with recent advances in Earth observation for the environmental sciences.[153]

To allow this emerging research field to flourish, broad access to a multiplicity of SM sources, in a way that is consistent over time, is essential. Shutdowns of platforms, increasing data usage restrictions (including payment requirements), and censorship are threatening the successful expansion of this novel research frontier. Different incentives for academics and industry hamper scalable collaborations, and scientists are currently largely dependent on the goodwill of SM platforms to allow free access to these data. There is also concern that societal distrust about how SM data are being used may result in further restrictions on data access. By showing how continued and broad access to SM data can help address questions of great societal importance, this review aims at contributing to breaking this vicious cycle. It offers a novel environment- and sustainability-oriented perspective to the ongoing societal debate on acceptable standards and rules under which data can be ethically utilized.

Sustainability researchers using SM data have a unique role to play as well as unique duties in fostering greater trust and cooperation. A broad endorsement of universal principles and standards in the use of SM data that guarantee inclusivity, as well as carefully balancing the needs for research transparency and privacy protection, will promote high ethical standards and increase the legitimacy and positive impact of resulting research. The shared values and goals of working for a sustainable future may provide common ground for cooperation, and motivation for establishing wide-ranging collaboration between SM companies, academia, and society in a virtuous cycle. While our critical review suggests that sustainability researchers have, on the whole, adhered to ethical standards, an improved standardization of research practices is necessary to rise to the challenge of realizing the potential of these data in an ethical way.

SM data provide an unprecedented platform for the observation of how behavior, narratives, and visions related to the environment and sustainability evolve across cultures and over time. By allowing this novel field of research to realize its full potential, some of the very tools that are often pointed at as being partly responsible for the loss of human-nature interactions might paradoxically turn out to play an important role in counteracting such extinction of experience and promoting a more sustainable society.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Andrea Ghermandi (aghermand@univ.haifa.ac.il).

Material availability

This study did not generate new unique materials.

Data and code availability

References to reviewed articles and our associated evaluated data are deposited in the FAIR aligned ZENODO repository (https://doi.org/10.5281/zenodo.7517193) and are publicly available as of the date of publication. This paper does not report original code.

Review protocol

This critical review relies on an extensive database of studies applying SM data in environmental sustainability research, which were collected and reviewed in full by the authors. Rather than providing a comprehensive summary of all relevant literature as in a systematic review, our objective was to take stock and evaluate the previous body of work in the field in order to promote conceptual innovation from its critical examination.[154] Building on a set of 169 studies collected in a previous systematic review of SM data applications in environmental research,[11] the database includes additional relevant studies that were identified by snowballing previous references and adding further gray and scientific academic articles known to the authors. For studies to be included in our analysis, they had to involve the use of data from one or more SM platforms and investigate human interactions with and/or impacts on the environment. We relied on a broad definition of SM including any Web site or application that enables users to create and share content or to participate in social networking (e.g., blogging sites, recommendation sites, and online forums). We further strengthened the analysis by including insights from additional literature on SM that do not have a direct application to environmental sustainability (e.g., studies on biases in SM data).

The final database consists of 415 studies, which were published between 2011 and 2021. The full references of all studies are provided in the supplemental information. As a group of natural and social scientists using SM data primarily for understanding environmental spatial phenomena, we acknowledge that the sample of studies might inadvertently be skewed toward spatial analysis.

All studies were read in full by the authors and analyzed according to the specific themes of interest for this review. First, each study was classified based on the investigated SM platform(s), the spatial scale of analysis (i.e., local, city/county, regional/national, supranational/continental, global), and the presence of analysis of temporal changes or trends. Subsequently, we characterized how each SM data analysis advanced environmental sustainability research by the way it addressed or contributed to SDG targets. Each study was assigned to the most closely related SDG target only, although it might be relevant to multiple ones, even pertaining to different SDGs (e.g., a study on the cultural significance of small-scale mountain farms could pertain to both SDG targets 2.4 "Sustainable agriculture" and 15.4 "Conservation of mountain ecosystems"). We focused in particular on identifying unique contributions and insights that can be generated with these data, as well as potential threats and limitations associated with their current or future use. In addition, we investigated the ethical standards adopted in the studies during the phases of SM data retrieval, handling, and reporting, with an emphasis on the role of researchers in promoting good practices in applying SM data in environmental sustainability research. For instance, we evaluated whether the authors explicitly acknowledged ethical or privacy concerns in their data handling and whether they relied on existing guidelines and regulations. We also examined whether the data were reported and stored in a way that ensures the anonymity of individual users, and the degree to which the studies appeared to have collected and investigated potentially sensitive information, such as personal identifiers and other user information. Finally, we characterized the degree to which studies made the data and methodology available in an open and transparent way.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.oneear.2023.02.008.

ACKNOWLEDGMENTS

A.G. and M.S. are supported by the Israel Science Foundation through grant no. 2751/16. J.L. is supported by the EU-H2020 through grants #818002, #869324, and the BiodivRestore ERA-Net COFUND (2020–2021) through NICHES. D.V.B. is supported by the US National Science Foundation (NSF) through grant #73133571. A.G., F.C., and J.L. are supported by the German-Israeli Foundation for Scientific Research and Development #I-1533-500.15/2021. S.P. and M.V. are supported by the German Federal Ministry for Education and Research (BMBF) through grant no. 033W046A. T.M. is

supported by the NSF through grants #1444755, #1927167, and #193493. S.A.W. was supported by a Data Science Environments project award from the Gordon and Betty Moore Foundation (Award 2013-10-29) and the Alfred P. Sloan Foundation (Award 3835) to the University of Washington eScience Institute. A.R.-F. was supported by H2020 Marie-Sklodowska-Curie Individual Fellowship (#655475) and Juan de la Cierva-Incorporación Postdoctoral Fellowship (IJC2019-040836-I/AEI/10.13039/501100011033) from the Spanish Government. O.K. was supported by the Academy of Finland and Kone Foundation. The authors thank the Department of Innovation, Research, University and Museums of the Autonomous Province of Bozen/Bolzano (Italy) for covering the open access publication costs.

AUTHOR CONTRIBUTIONS

A.G., J.L., and D.V.B. jointly conceived the study and led the writing of the article with equal contributions. All other authors have contributed to data collection and analysis, interpretation of results, and writing of the article, and are listed in alphabetical order.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. DataReportal (2021). Digital 2021 October global statshot report. https://datareportal.com/reports/digital-2021-october-global-statshot.
2. Zhang, H., Huang, R., Zhang, Y., and Buhalis, D. (2020). Cultural ecosystem services evaluation using geolocated social media data: a review. Tourism Geogr. *24*, 646–668. https://doi.org/10.1080/14616688.2020.1801828.
3. Pearce, W., Niederer, S., Özkula, S.M., and Sánchez Querubín, N. (2019). The social media life of climate change: platforms, publics, and future imaginaries. WIREs Clim. Change *10*, e569. https://doi.org/10.1002/wcc.569.
4. Ilieva, R.T., and McPhearson, T. (2018). Social-media data for urban sustainability. Nat. Sustain. *1*, 553–565. https://doi.org/10.1038/s41893-018-0153-6.
5. Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., and Di Minin, E. (2019). Social media data for conservation science: a methodological overview. Biol. Conserv. *233*, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023.
6. Jarić, I., Correia, R.A., Brook, B.W., Buettel, J.C., Courchamp, F., Di Minin, E., Firth, J.A., Gaston, K.J., Jepson, P., Kalinkat, G., et al. (2020). iEcology: harnessing large online resources to generate ecological insights. Trends Ecol. Evol. *35*, 630–639. https://doi.org/10.1016/j.tree.2020.03.003.
7. Spotswood, E.N., Benjamin, M., Stoneburner, L., Wheeler, M.M., Beller, E.E., Balk, D., McPhearson, T., Kuo, M., and McDonald, R.I. (2021). Nature inequity and higher COVID-19 case rates in less-green neighbourhoods in the United States. Nat. Sustain. *4*, 1092–1098. https://doi.org/10.1038/s41893-021-00781-9.
8. Gaston, K.J., and Soga, M. (2020). Extinction of experience: the need to be more specific. People Nat. (Hoboken). *2*, 575–581. https://doi.org/10.1002/pan3.10118.
9. UN General Assembly (2015). Transforming Our World: The 2020 Agenda for Sustainable Development.
10. Espinoza, M.I., and Aronczyk, M. (2021). Big data for climate action or climate action for big data? Big Data Soc. *8*. 205395172098203. https://doi.org/10.1177/2053951720982032.
11. Ghermandi, A., and Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: a systematic map. Global Environ. Change *55*, 36–47. https://doi.org/10.1016/j.gloenvcha.2019.02.003.
12. Lazer, D.M.J., Pentland, A., Watts, D.J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., et al. (2020). Computational social science: obstacles and opportunities. Science *369*, 1060–1062. https://doi.org/10.1126/science.aaz8170.
13. World Economic Forum (2011). Personal data: the emergence of a new asset class. www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.
14. Kirkpatrick, R. (2021). Unpacking the Issue of Missed Use and Misuse of Data (UN Glob. Pulse Blog). https://www.unglobalpulse.org/2019/03/unpacking-the-issue-of-missed-use-and-misuse-of-data/.
15. Yakowitz, J. (2011). Tragedy of the data commons. Harv JL Tech *25*, 1.
16. World Economic Forum (2014). Rethinking personal data: a new lens for strengthening trust. ww3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.
17. Poom, A., Järv, O., Zook, M., and Toivonen, T. (2020). COVID-19 is spatial: ensuring that mobile Big Data is used for social good. Big Data Soc. *7*. 2053951720952088. https://doi.org/10.1177/2053951720952088.
18. Monkman, G.G., Kaiser, M., and Hyder, K. (2018). The ethics of using social media in fisheries research. Rev. Fish. Sci. Aquacult. *26*, 235–242. https://doi.org/10.1080/23308249.2017.1389854.
19. King, G., and Persily, N. (2020). A new model for industry–academic partnerships. PS Polit. Sci. Polit. *53*, 703–709. https://doi.org/10.1017/S1049096519001021.
20. Levi, M., and Rajala, B. (2020). Alternatives to social science one. PS Polit. Sci. Polit. *53*, 710–711. https://doi.org/10.1017/S1049096520000438.
21. Sottini, V.A., Barbierato, E., Bernetti, I., Capecchi, I., Fabbrizzi, S., and Menghini, S. (2019). Winescape perception and big data analysis: an assessment through social media photographs in the Chianti Classico region. Wine Econ. Pol. *8*, 127–140. https://doi.org/10.1016/j.wep.2019.07.001.
22. Brown, M.E., and McCarty, J.L. (2017). Is remote sensing useful for finding and monitoring urban farms? Appl. Geogr. *80*, 23–33. https://doi.org/10.1016/j.apgeog.2017.01.008.
23. Andersson, E., and Öhman, J. (2017). Young people's conversations about environmental and sustainability issues in social media. Environ. Educ. Res. *23*, 465–485. https://doi.org/10.1080/13504622.2016.1149551.
24. Quinn, M., Lynn, T., Jollands, S., and Nair, B. (2016). Domestic water charges in Ireland - issues and challenges conveyed through social media. Water Resour. Manag. *30*, 3577–3591. https://doi.org/10.1007/s11269-016-1374-y.
25. Giuliani, M., Castelletti, A., Fedorov, R., and Fraternali, P. (2016). Using crowdsourced web content for informing water systems operations in snow-dominated catchments. Hydrol. Earth Syst. Sci. *20*, 5049–5062. https://doi.org/10.5194/hess-20-5049-2016.
26. Ghermandi, A. (2018). Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands. Ecosyst. Serv. *31*, 351–357. https://doi.org/10.1016/j.ecoser.2017.12.012.
27. Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T., and Downing, J.A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. Front. Ecol. Environ. *13*, 76–81. https://doi.org/10.1890/140124.
28. Sinclair, M., Ghermandi, A., and Sheela, A.M. (2018). A crowdsourced valuation of recreational ecosystem services using social media data: an application to a tropical wetland in India. Sci. Total Environ. *642*, 356–365. https://doi.org/10.1016/j.scitotenv.2018.06.056.
29. Hale, R.L., Cook, E.M., and Beltrán, B.J. (2019). Cultural ecosystem services provided by rivers across diverse social-ecological landscapes: a social media analysis. Ecol. Indicat. *107*, 105580. https://doi.org/10.1016/j.ecolind.2019.105580.
30. Jaspal, R., Turner, A., and Nerlich, B. (2014). Fracking on YouTube: exploring risks, benefits and human values. Environ. Values *23*, 501–527. https://doi.org/10.3197/096327114X13947900181473.
31. Reusswig, F., Braun, F., Heger, I., Ludewig, T., Eichenauer, E., and Lass, W. (2016). Against the wind: local opposition to the German Energiewende. Util. Pol. *41*, 214–227. https://doi.org/10.1016/j.jup.2016.02.006.
32. Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Comput. Environ. Urban Syst. *54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001.
33. Hu, F., Li, Z., Yang, C., and Jiang, Y. (2019). A graph-based approach to detecting tourist movement patterns using social media data. Cartogr. Geogr. Inf. Sci. *46*, 368–382. https://doi.org/10.1080/15230406.2018.1496036.
34. Lee, Y. (2017). Corporate sustainable development and marketing communications on social media: fortune 500 enterprises. Bus. Strat. Environ. *26*, 569–583. https://doi.org/10.1002/bse.1936.
35. Amorim Maia, A.T., Calcagni, F., Connolly, J.J.T., Anguelovski, I., and Langemeyer, J. (2020). Hidden drivers of social injustice: uncovering unequal cultural ecosystem services behind green gentrification. Environ. Sci. Pol. *112*, 254–263. https://doi.org/10.1016/j.envsci.2020.05.021.
36. Baker, K., Ooms, K., Verstockt, S., Brackman, P., De Maeyer, P., and Van de Walle, R. (2017). Crowdsourcing a cyclist perspective on suggested

recreational paths in real-world networks. Cartogr. Geogr. Inf. Sci. *44*, 422–435. https://doi.org/10.1080/15230406.2016.1192486.

37. Luo, F., Cao, G., Mulligan, K., and Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: a case study of Chicago. Appl. Geogr. *70*, 11–25. https://doi.org/10.1016/j.apgeog.2016.03.001.
38. Montaño, F. (2018). The use of geo-located photos as a source to assess the landscape perception of locals and tourists – case studies: two public open spaces in Munich, Germany. J. Digit. Landsc. Archit. *3*, 346–355. https://doi.org/10.14627/537642037.
39. Bahrehdar, A.R., Adams, B., and Purves, R.S. (2020). Streets of London: using Flickr and OpenStreetMap to build an interactive image of the city. Comput. Environ. Urban Syst. *84*, 101524. https://doi.org/10.1016/j.compenvurbsys.2020.101524.
40. Levin, N., Ali, S., Crandall, D., and Kark, S. (2019). World Heritage in danger: big data and remote sensing can help protect sites in conflict zones. Global Environ. Change *55*, 97–104. https://doi.org/10.1016/j.gloenvcha.2019.02.001.
41. Moore, F.C., and Obradovich, N. (2020). Using remarkability to define coastal flooding thresholds. Nat. Commun. *11*, 530. https://doi.org/10.1038/s41467-019-13935-3.
42. Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. Sci. Adv. *2*, e1500779. https://doi.org/10.1126/sciadv.1500779.
43. Kibanov, M., Stumme, G., Amin, I., and Lee, J.G. (2017). Mining social media to inform peatland fire and haze disaster management. Soc. Netw. Anal. Min. *7*, 1–19. https://doi.org/10.1007/s13278-017-0446-1.
44. Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T., and Toivonen, T. (2020). Understanding the use of urban green spaces from user-generated geographic information. Landsc. Urban Plann. *201*, 103845. https://doi.org/10.1016/j.landurbplan.2020.103845.
45. Cheng, Y., Zhang, J., Wei, W., and Zhao, B. (2021). Effects of urban parks on residents' expressed happiness before and during the COVID-19 pandemic. Landsc. Urban Plann. *212*, 104118. https://doi.org/10.1016/j.landurbplan.2021.104118.
46. Ghahramani, M., Galle, N.J., Duarte, F., Ratti, C., and Pilla, F. (2021). Leveraging artificial intelligence to analyze citizens' opinions on urban green space. City Environ. Interact. *10*, 100058. https://doi.org/10.1016/j.cacint.2021.100058.
47. Cai, B., Wang, J., Long, Y., Li, W., Liu, J., Ni, Z., Bo, X., Li, D., Wang, J., Chen, X., et al. (2015). Evaluating the impact of odors from the 1955 landfills in China using a bottom-up approach. J. Environ. Manag. *164*, 206–214. https://doi.org/10.1016/j.jenvman.2015.09.009.
48. Wang, Y.D., Fu, X.K., Jiang, W., Wang, T., Tsou, M.-H., and Ye, X.Y. (2017). Inferring urban air quality based on social media. Comput. Environ. Urban Syst. *66*, 110–116. https://doi.org/10.1016/j.compenvurbsys.2017.07.002.
49. Chong, Z., Qin, C., and Ye, X. (2017). Environmental regulation and industrial structure change in China: integrating spatial and social network analysis. Sustainability *9*, 1465. https://doi.org/10.3390/su9081465.
50. Zoomers, A., Gekker, A., and Schäfer, M.T. (2016). Between two hypes: will "big data" help unravel blind spots in understanding the "global land rush?". Geoforum *69*, 147–159. https://doi.org/10.1016/j.geoforum.2015.11.017.
51. Chang, C.C., Cheng, G.J.Y., Nghiem, T.P.L., Song, X.P., Oh, R.R.Y., Richards, D.R., and Carrasco, L.R. (2020). Social media, nature, and life satisfaction: global evidence of the biophilia hypothesis. Sci. Rep. *10*, 4125. https://doi.org/10.1038/s41598-020-60902-w.
52. Starbird, K., Dailey, D., Walker, A.H., Leschine, T.M., Pavia, R., and Bostrom, A. (2015). Social media, public participation, and the 2010 BP Deepwater Horizon oil spill. Hum. Ecol. Risk Assess. *21*, 605–630. https://doi.org/10.1080/10807039.2014.947866.
53. Breckheimer, I.K., Theobald, E.J., Cristea, N.C., Wilson, A.K., Lundquist, J.D., Rochefort, R.M., and HilleRisLambers, J. (2020). Crowd-sourced data reveal social–ecological mismatches in phenology driven by climate. Front. Ecol. Environ. *18*, 76–82. https://doi.org/10.1002/fee.2142.
54. Moore, F.C., Obradovich, N., Lehner, F., and Baylis, P. (2019). Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. Proc. Natl. Acad. Sci. USA *116*, 4905–4910. https://doi.org/10.1073/pnas.1816541116.
55. Wang, J., Obradovich, N., and Zheng, S. (2020). A 43-million-person investigation into weather and expressed sentiment in a changing climate. One Earth *2*, 568–577. https://doi.org/10.1016/j.oneear.2020.05.016.
56. Spalding, M., Burke, L., Wood, S.A., Ashpole, J., Hutchison, J., and Zu Ermgassen, P. (2017). Mapping the global value and distribution of coral reef tourism. Mar. Pol. *82*, 104–113. https://doi.org/10.1016/j.marpol.2017.05.014.
57. Sullivan, M., Robinson, S., and Littnan, C. (2019). Social media as a data resource for #monkseal conservation. PLoS One *14*, e0222627. https://doi.org/10.1371/journal.pone.0222627.
58. de Juan, S., Ospina-Álvarez, A., Villasante, S., and Ruiz-Frau, A. (2021). A Graph Theory approach to assess nature's contribution to people at a global scale. Sci. Rep. *11*, 9118. https://doi.org/10.1038/s41598-021-88745-z.
59. Derungs, C., and Purves, R.S. (2016). Characterising landscape variation through spatial folksonomies. Appl. Geogr. *75*, 60–70. https://doi.org/10.1016/j.apgeog.2016.08.005.
60. Wood, S.A., Winder, S.G., Lia, E.H., White, E.M., Crowley, C.S.L., and Milnor, A.A. (2020). Next-generation visitation models using social media to estimate recreation on public lands. Sci. Rep. *10*, 15419. https://doi.org/10.1038/s41598-020-70829-x.
61. Liang, Y., Kirilenko, A.P., Stepchenkova, S.O., and Ma, S.D. (2020). Using social media to discover unwanted behaviours displayed by visitors to nature parks: comparisons of nationally and privately owned parks in the Greater Kruger National Park, South Africa. Tour. Recreat. Res. *45*, 271–276. https://doi.org/10.1080/02508281.2019.1681720.
62. Runge, C.A., Daigle, R.M., and Hausner, V.H. (2020). Quantifying tourism booms and the increasing footprint in the Arctic with social media data. PLoS One *15*, e0227189. https://doi.org/10.1371/journal.pone.0227189.
63. Van Zanten, B.T., Van Berkel, D.B., Meentemeyer, R.K., Smith, J.W., Tieskens, K.F., and Verburg, P.H. (2016). Continental-scale quantification of landscape values using social media data. Proc. Natl. Acad. Sci. USA *113*, 12974–12979. https://doi.org/10.1073/pnas.1614158113.
64. Jeawak, S.S., Jones, C.B., and Schockaert, S. (2020). Predicting environmental features by learning spatiotemporal embeddings from social media. Ecol. Inform. *55*, 101031. https://doi.org/10.1016/j.ecoinf.2019.101031.
65. Richards, D.R., and Friess, D.A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs. Ecol. Indicat. *53*, 187–195. https://doi.org/10.1016/j.ecolind.2015.01.034.
66. Schirpke, U., Altzinger, A., Leitinger, G., and Tasser, E. (2019). Change from agricultural to touristic use: effects on the aesthetic value of landscapes over the last 150 years. Landsc. Urban Plann. *187*, 23–35. https://doi.org/10.1016/j.landurbplan.2019.03.004.
67. Atsumi, K., and Koizumi, I. (2017). Web image search revealed large-scale variations in breeding season and nuptial coloration in a mutually ornamented fish, Tribolodon hakonensis. Ecol. Res. *32*, 567–578. https://doi.org/10.1007/s11284-017-1466-z.
68. August, T.A., Pescott, O.L., Joly, A., and Bonnet, P. (2020). AI naturalists might hold the key to unlocking biodiversity data in social media imagery. Patterns *1*, 100116. https://doi.org/10.1016/j.patter.2020.100116.
69. Cong, L., Wu, B., Morrison, A.M., Shu, H., and Wang, M. (2014). Analysis of wildlife tourism experiences with endangered species: an exploratory study of encounters with giant pandas in Chengdu, China. Tour. Manag. *40*, 300–310. https://doi.org/10.1016/j.tourman.2013.07.005.
70. Fink, C., Hausmann, A., and Di Minin, E. (2020). Online sentiment towards iconic species. Biol. Conserv. *241*, 108289. https://doi.org/10.1016/j.biocon.2019.108289.
71. Fink, C., Toivonen, T., Correia, R.A., and Di Minin, E. (2021). Mapping the online songbird trade in Indonesia. Appl. Geogr. *134*, 102505. https://doi.org/10.1016/j.apgeog.2021.102505.
72. Daume, S., Albert, M., and von Gadow, K. (2014). Forest monitoring and social media – complementary data sources for ecosystem surveillance? For. Ecol. Manag. *316*, 9–20. https://doi.org/10.1016/j.foreco.2013.09.004.
73. Arkema, K.K., Verutes, G.M., Wood, S.A., Clarke-Samuels, C., Rosado, S., Canto, M., Rosenthal, A., Ruckelshaus, M., Guannel, G., Toft, J., et al. (2015). Embedding ecosystem services in coastal planning leads to better outcomes for people and nature. Proc. Natl. Acad. Sci. USA *112*, 7390–7395. https://doi.org/10.1073/pnas.1406483112.
74. Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat. Hum. Behav. *2*, 637–644. https://doi.org/10.1038/s41562-018-0399-z.
75. Langemeyer, J., Calcagni, F., and Baró, F. (2018). Mapping the intangible: using geolocated social media data to examine landscape

aesthetics. Land Use Pol. *77*, 542–552. https://doi.org/10.1016/j.landusepol.2018.05.049.

76. Maynard, D., Roberts, I., Greenwood, M.A., Rout, D., and Bontcheva, K. (2017). A framework for real-time semantic social media analysis. J. Web Semant. *44*, 75–88. https://doi.org/10.1016/j.websem.2017.05.002.
77. Ghermandi, A., Sinclair, M., Fichtman, E., and Gish, M. (2020). Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing. Global Environ. Change *65*, 102189. https://doi.org/10.1016/j.gloenvcha.2020.102189.
78. Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B., and Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst. Serv. *38*, 100958. https://doi.org/10.1016/j.ecoser.2019.100958.
79. Winder, S.G., Lee, H., Seo, B., Lia, E.H., and Wood, S.A. (2022). An open-source image classifier for characterizing recreational activities across landscapes. People Nat. (Hoboken). *4*, 1249–1262. https://doi.org/10.1002/pan3.10382.
80. Bubalo, M., van Zanten, B.T., and Verburg, P.H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: a review. Landsc. Urban Plann. *184*, 101–111. https://doi.org/10.1016/j.landurbplan.2019.01.001.
81. Lu, Y., Zhao, J., Wu, X., and Lo, S.M. (2021). Escaping to nature during a pandemic: a natural experiment in Asian cities during the COVID-19 pandemic with big social media data. Sci. Total Environ. *777*, 146092. https://doi.org/10.1016/j.scitotenv.2021.146092.
82. Daume, S. (2016). Mining Twitter to monitor invasive alien species — an analytical framework and sample information topologies. Ecol. Inform. *31*, 70–82. https://doi.org/10.1016/j.ecoinf.2015.11.014.
83. Truelove, M., Vasardani, M., and Winter, S. (2015). Towards credibility of micro-blogs: characterising witness accounts. Geojournal *80*, 339–359. https://doi.org/10.1007/s10708-014-9556-8.
84. Moumtzidou, A., Andreadis, S., Gialampoukidis, I., Karakostas, A., Vrochidis, S., and Kompatsiaris, I. (2018). Flood relevance estimation from visual and textual content in social media streams. In Companion Proceedings of the Web Conference 2018 WWW '18 (International World Wide Web Conferences Steering Committee), pp. 1621–1627. https://doi.org/10.1145/3184558.3191620.
85. Rosser, J.F., Leibovici, D.G., and Jackson, M.J. (2017). Rapid flood inundation mapping using social media, remote sensing and topographic data. Nat. Hazards *87*, 103–120. https://doi.org/10.1007/s11069-017-2755-0.
86. Calcagni, F., Amorim Maia, A.T., Connolly, J.J.T., and Langemeyer, J. (2019). Digital co-construction of relational values: understanding the role of social media for sustainability. Sustain. Sci. *14*, 1309–1321. https://doi.org/10.1007/s11625-019-00672-1.
87. Haider, J. (2016). The shaping of environmental information in social media: affordances and technologies of self-control. Environ. Commun. *10*, 473–491. https://doi.org/10.1080/17524032.2014.993416.
88. Wu, Y., Xie, L., Yuan, Z., Jiang, S., Liu, W., and Sheng, H. (2020). Investigating public biodiversity conservation awareness based on the propagation of wildlife-related incidents on the Sina Weibo social media platform. Environ. Res. Lett. *15*, 094082. https://doi.org/10.1088/1748-9326/ab9ed1.
89. Andersson, E., Tengö, M., McPhearson, T., and Kremer, P. (2015). Cultural ecosystem services as a gateway for improving urban sustainability. Ecosyst. Serv. *12*, 165–168. https://doi.org/10.1016/j.ecoser.2014.08.002.
90. Langemeyer, J., and Calcagni, F. (2022). Virtual spill-over effects: what social media has to do with relational values and global environmental stewardship. Ecosyst. Serv. *53*, 101400. https://doi.org/10.1016/j.ecoser.2021.101400.
91. Sinclair, M., Mayer, M., Woltering, M., and Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. J. Environ. Manag. *263*, 110418. https://doi.org/10.1016/j.jenvman.2020.110418.
92. Arts, K., van der Wal, R., and Adams, W.M. (2015). Digital technology and the conservation of nature. Ambio *44*, 661–673. https://doi.org/10.1007/s13280-015-0705-1.
93. Dylewski, Ł., Mikula, P., Tryjanowski, P., Morelli, F., and Yosef, R. (2017). Social media and scientific research are complementary—YouTube and shrikes as a case study. Naturwissenschaften *104*, 48. https://doi.org/10.1007/s00114-017-1470-8.
94. Depietri, Y., Ghermandi, A., Campisi-Pinto, S., and Orenstein, D.E. (2021). Public participation GIS versus geolocated social media data to assess urban cultural ecosystem services: instances of complementarity. Ecosyst. Serv. *50*, 101277. https://doi.org/10.1016/j.ecoser.2021.101277.
95. Heikinheimo, V., Minin, E.D., Tenkanen, H., Hausmann, A., Erkkonen, J., and Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: a comparison of social media data and visitor survey. ISPRS Int. J. Geo-Inf. *6*, 85. https://doi.org/10.3390/ijgi6030085.
96. Agryzkov, T., Martí, P., Tortosa, L., and Vicent, J.F. (2017). Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain). Int. J. Geogr. Inf. Sci. *31*, 100–121. https://doi.org/10.1080/13658816.2016.1188931.
97. Becken, S., Stantic, B., Chen, J., Alaei, A.R., and Connolly, R.M. (2017). Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. J. Environ. Manag. *203*, 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007.
98. Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. Landsc. Urban Plann. *142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022.
99. Williams, H.T., McMurray, J.R., Kurz, T., and Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environ. Change *32*, 126–138. https://doi.org/10.1016/j.gloenvcha.2015.03.006.
100. Cooper, G., Green, N., Burningham, K., Evans, D., and Jackson, T. (2012). Unravelling the threads: discourses of sustainability and consumption in an online forum. Environ. Commun. *6*, 101–118. https://doi.org/10.1080/17524032.2011.642080.
101. Jang, S.M., and Hart, P.S. (2015). Polarized frames on "climate change" and "global warming" across countries and states: evidence from Twitter big data. Global Environ. Change *32*, 11–17. https://doi.org/10.1016/j.gloenvcha.2015.02.010.
102. Papworth, S.k., Nghiem, T.p. l., Chimalakonda, D., Posa, M.r. c., Wijedasa, L.s., Bickford, D., and Carrasco, L.r. (2015). Quantifying the role of online news in linking conservation research to Facebook and Twitter. Conserv. Biol. *29*, 825–833. https://doi.org/10.1111/cobi.12455.
103. Vaidhyanathan, S. (2018). Antisocial Media: How Facebook Disconnects Us and Undermines Democracy (Oxford University Press).
104. Dalrymple, K.E., Shaw, B.R., and Brossard, D. (2013). Following the leader: using opinion leaders in environmental strategic communication. Soc. Nat. Resour. *26*, 1438–1453. https://doi.org/10.1080/08941920.2013.820812.
105. Wood, S.A., Guerry, A.D., Silver, J.M., and Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. Sci. Rep. *3*, 2976. https://doi.org/10.1038/srep02976.
106. Ghermandi, A., Depietri, Y., and Sinclair, M. (2022). In the AI of the beholder: a comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces. Landsc. Urban Plann. *217*, 104261. https://doi.org/10.1016/j.landurbplan.2021.104261.
107. Palomino, M., Taylor, T., Göker, A., Isaacs, J., and Warber, S. (2016). The online dissemination of nature–health concepts: lessons from sentiment analysis of social media relating to "nature-deficit disorder". Int. J. Environ. Res. Publ. Health *13*, 142. https://doi.org/10.3390/ijerph13010142.
108. Wilkins, E.J., Wood, S.A., and Smith, J.W. (2021). Uses and limitations of social media to inform visitor use management in parks and protected areas: a systematic review. Environ. Manag. *67*, 120–132. https://doi.org/10.1007/s00267-020-01373-7.
109. Zapata-Caldas, E., Calcagni, F., Baró, F., and Langemeyer, J. (2022). Using crowdsourced imagery to assess cultural ecosystem services in data-scarce urban contexts: the case of the metropolitan area of Cali, Colombia. Ecosyst. Serv. *56*, 101445. https://doi.org/10.1016/j.ecoser.2022.101445.
110. Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., and Toivonen, T. (2017). Instagram, Flickr, or Twitter: assessing the usability of social media data for visitor monitoring in protected areas. Sci. Rep. *7*, 17615. https://doi.org/10.1038/s41598-017-18007-4.
111. Ghermandi, A. (2022). Geolocated social media data counts as a proxy for recreational visits in natural areas: a meta-analysis. J. Environ. Manag. *317*, 115325. https://doi.org/10.1016/j.jenvman.2022.115325.
112. MacAskill, D. (2019). The World's Most-Beloved, Money-Losing Business Needs Your Help (Flickr Blog). https://blog.flickr.net/en/2019/12/19/the-worlds-most-beloved-money-losing-business-needs-your-help/.
113. Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., and Li, L.-J. (2016). YFCC100M: the new data in multimedia research. Commun. ACM *59*, 64–73. https://doi.org/10.1145/2812802.

114. Freelon, D. (2018). Computational research in the post-API age. Polit. Commun. *35*, 665–668. https://doi.org/10.1080/10584609.2018.1477506.

115. Kozlowska, I. (2018). Facebook and Data Privacy in the Age of Cambridge Analytica (Henry M Jackson Sch. Int. Stud). https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/.

116. Meta data for good home. https://dataforgood.facebook.com/.

117. Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. Inf. Commun. Soc. *22*, 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447.

118. Lumb, D. (2018). After exposing secret military bases, Strava restricts data visibility. Engadget. https://www.engadget.com/2018-03-13-after-exposing-secret-military-bases-strava-restricts-data-visi.html.

119. Ruiz-Frau, A., Ospina-Alvarez, A., Villasante, S., Pita, P., Maya-Jariego, I., and de Juan, S. (2020). Using graph theory and social media data to assess cultural ecosystem services in coastal areas: method development and application. Ecosyst. Serv. *45*, 101176. https://doi.org/10.1016/j.ecoser.2020.101176.

120. Hu, Y., and Wang, R.-Q. (2020). Understanding the removal of precise geotagging in tweets. Nat. Hum. Behav. *4*, 1219–1221. https://doi.org/10.1038/s41562-020-00949-x.

121. Twitter (2021). Twitter API for academic research. https://developer.twitter.com/en/products/twitter-api/academic-research.

122. Hendriks, C.M., Duus, S., and Ercan, S.A. (2016). Performing politics on social media: the dramaturgy of an environmental controversy on Facebook. Environ. Polit. *25*, 1102–1125. https://doi.org/10.1080/09644016.2016.1196967.

123. Tai, Y., and Fu, K.W. (2020). Specificity, conflict, and focal point: a systematic investigation into social media censorship in China. J. Commun. *70*, 842–867. https://doi.org/10.1093/joc/jqaa032.

124. Marwick, A.E., and Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc. *13*, 114–133. https://doi.org/10.1177/1461444810365313.

125. Poberezhskaya, M. (2014). Reflections on climate change and new media in Russia: challenges and opportunities and new media. Digit. Icons Stud. Russ. Eurasian Cent. Eur. New Media, 37–50.

126. Boyd-Barrett, O., and Mirrlees, T. (2019). Media Imperialism: Continuity and Change (Rowman & Littlefield).

127. Hintz, A. (2015). Social media censorship, privatized regulation and new restrictions to protest and dissent. In Critical perspectives on social media and protest: Between control and emancipation (Rowman & Littlefield), pp. 109–125.

128. Heins, M. (2013). The brave new world of social media censorship. Harv. Law Rev. Forum *127*, 325–330.

129. Poupin, P. (2021). Social media and state repression: the case of VKontakte and the anti-garbage protest in Shies. In Far Northern Russia (First Monday). https://doi.org/10.5210/fm.v26i5.11711.

130. Pan, J. (2017). How market dynamics of domestic and foreign social media firms shape strategies of Internet censorship. Probl. Post-Communism *64*, 167–188. https://doi.org/10.1080/10758216.2016.1181525.

131. King, G., Pan, J., and Roberts, M.E. (2014). Reverse-engineering censorship in China: randomized experimentation and participant observation. Science *345*, 1251722. https://doi.org/10.1126/science.1251722.

132. Bamman, D., O'Connor, B., and Smith, N. (2012). Censorship and Deletion Practices in Chinese Social Media (First Monday). https://doi.org/10.5210/fm.v17i3.3943.

133. Zhang, S.I. (2020). China's social media platforms: Weibo. In Media and Conflict in the Social Media Era in China (Springer Singapore), pp. 21–40. https://doi.org/10.1007/978-981-15-7635-5.

134. Golovchenko, Y. (2022). Fighting propaganda with censorship: a study of the Ukrainian ban on Russian social media. J. Polit. *84*, 639–654. https://doi.org/10.1086/716949.

135. Richterich, A. (2018). How data-driven research fuelled the Cambridge Analytica controversy. Partecip. E Conflitto *11*, 528–543. https://doi.org/10.1285/i20356609v11i2p528.

136. Di Minin, E., Fink, C., Hausmann, A., Kremer, J., and Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. Conserv. Biol. *35*, 437–446. https://doi.org/10.1111/cobi.13708.

137. Kates, R.W. (2011). What kind of a science is sustainability science? Proc. Natl. Acad. Sci. USA *108*, 19449–19450. https://doi.org/10.1073/pnas.1116097108.

138. Ribeiro, F.N., Benevenuto, F., and Zagheni, E. (2020). How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. In 12th ACM Conference on Web Science WebSci '20 (Association for Computing Machinery), pp. 325–334. https://doi.org/10.1145/3394231.3397923.

139. Wood-Doughty, Z., Xu, P., Liu, X., and Dredze, M. (2021). Using noisy self-reports to predict Twitter user demographics. In Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media, pp. 123–137. https://doi.org/10.18653/v1/2021.socialnlp-1.11.

140. Yuan, S., and Lou, C. (2020). How social media influencers foster relationships with followers: the roles of source credibility and fairness in parasocial relationship and product interest. J. Interact. Advert. *20*, 133–147. https://doi.org/10.1080/15252019.2020.1769514.

141. Lenormand, M., Luque, S., Langemeyer, J., Tenerelli, P., Zulian, G., Aalders, I., Chivulescu, S., Clemente, P., Dick, J., van Dijk, J., et al. (2018). Multiscale socio-ecological networks in the age of information. PLoS One *13*, e0206672. https://doi.org/10.1371/journal.pone.0206672.

142. Mashhadi, A., Winder, S.G., Lia, E.H., and Wood, S.A. (2021). No walk in the park: the viability and fairness of social media analysis for parks and recreational policy making. In Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM 2021), pp. 409–420.

143. Fox, N., August, T., Mancini, F., Parks, K.E., Eigenbrod, F., Bullock, J.M., Sutter, L., and Graham, L.J. (2020). "photosearcher" package in R: an accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX *12*, 100624. https://doi.org/10.1016/j.softx.2020.100624.

144. Chen, Y., Sherren, K., Smit, M., and Lee, K.Y. (2021). Using social media images as data in social science research. New Media Soc. 14614448211038760. https://doi.org/10.1177/14614448211038761.

145. Grimm, V. (2020). The ODD protocol: an update with guidance to support wider and more consistent use. Ecol. Model. *428*, 109105. https://doi.org/10.1016/j.ecolmodel.2020.109105.

146. Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: an application of contextual integrity. Soc. Media Soc. *4*. 205630511876830. https://doi.org/10.1177/2056305118768300.

147. Di Minin, E., Fink, C., Hiippala, T., and Tenkanen, H. (2019). A framework for investigating illegal wildlife trade on social media with machine learning. Conserv. Biol. *33*, 210–213. https://doi.org/10.1111/cobi.13104.

148. Van Berkel, D.B., Tabrizian, P., Dorning, M.A., Smart, L., Newcomb, D., Mehaffey, M., Neale, A., and Meentemeyer, R.K. (2018). Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosyst. Serv. *31*, 326–335. https://doi.org/10.1016/j.ecoser.2018.03.022.

149. Nelson, P.B., Oberg, A., and Nelson, L. (2010). Rural gentrification and linked migration in the United States. J. Rural Stud. *26*, 343–352. https://doi.org/10.1016/j.jrurstud.2010.06.003.

150. Kibby, M. (2020). Instafamous: social media influencers and Australian beaches. In Writing the Australian Beach: Local Site, Global Idea, E. Ellison and D.L. Brien, eds. (Springer International Publishing), pp. 57–70. https://doi.org/10.1007/978-3-030-35264-6_4.

151. Hargittai, E. (2020). Potential biases in big data: omitted voices on social media. Soc. Sci. Comput. Rev. *38*, 10–24. https://doi.org/10.1177/0894439318788322.

152. Tufekci, Z. (2014). Big questions for social media big data: representativeness, validity and other methodological pitfalls. In Eighth International AAAI Conference on Weblogs and Social Media.

153. Nagaraj, A., Shears, E., and de Vaan, M. (2020). Improving data access democratizes and diversifies science. Proc. Natl. Acad. Sci. USA *117*, 23490–23498. https://doi.org/10.1073/pnas.2001682117.

154. Grant, M.J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf. Libr. J. *26*, 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.