Received: 1 October 2023 | Revised: 19 December 2023 | Accepted: 23 December 2023

DOI: 10.1111/cobi.14257

Conservation Biology

METHOD

# A protocol for harvesting biodiversity data from Facebook

**Shawan Chowdhury**[1,2,3,4] | **Sultan Ahmed**[5] | **Shofiul Alam**[5] | **Corey T. Callaghan**[6] | **Priyanka Das**[5] | **Moreno Di Marco**[7] | **Enrico Di Minin**[8,9,10] | **Ivan Jarić**[11,12] | **Mahzabin Muzahid Labi**[5] | **Md. Rokonuzzaman**[5] | **Uri Roll**[13] | **Valerio Sbragaglia**[14] | **Asma Siddika**[5] | **Aletta Bonn**[1,2,3]

[1]Institute of Biodiversity, Friedrich Schiller University Jena, Jena, Germany

[2]Department of Biodiversity and People, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

[4]School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

[5]Department of Zoology, University of Dhaka, Dhaka, Bangladesh

[6]Department of Wildlife Ecology and Conservation, Fort Lauderdale Research and Education Center, University of Florida, Davie, Florida, USA

[7]Department of Biology and Biotechnologies Charles Darwin, Sapienza University of Rome, Rome, Italy

[8]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

[9]Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, Finland

[10]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

[11]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Gif-sur-Yvette, France

[12]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic

[13]Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[14]Department of Marine Renewable Resources, Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

**Correspondence**
Shawan Chowdhury, Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Straße 159, 07743 Jena, Germany.
Email: dr.shawanchowdhury@gmail.com

**Funding information**
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: DFG-FZT 118, 202548816

## Abstract

The expanding use of community science platforms has led to an exponential increase in biodiversity data in global repositories. Yet, understanding of species distributions remains patchy. Biodiversity data from social media can potentially reduce the global biodiversity knowledge gap. However, practical guidelines and standardized methods for harvesting such data are nonexistent. Following data privacy and protection safeguards, we devised a standardized method for extracting species distribution records from Facebook groups that allow access to their data. It involves 3 steps: group selection, data extraction, and georeferencing the record location. We present how to structure keywords, search for species photographs, and georeference localities for such records. We further highlight some challenges users might face when extracting species distribution data from Facebook and suggest solutions. Following our proposed framework, we present a case study on Bangladesh's biodiversity—a tropical megadiverse South Asian country. We scraped nearly 45,000 unique georeferenced records across 967 species and found a median of 27 records per species. About 12% of the distribution data were for threatened species, representing 27% of all species. We also obtained data for 56 DataDeficient species for Bangladesh. If carefully harvested, social media data can significantly reduce global biodiversity knowledge gaps. Consequently, developing an automated tool to extract and interpret social media biodiversity data is a research priority.

**KEYWORDS**
Bangladesh, citizen science, crowdsourcing, Facebook, iEcology, megadiverse countries, social media, tropics, Wallacean shortfall

## INTRODUCTION

Amid the sixth mass extinction, many species worldwide are dramatically declining—28% of all assessed species on the International Union for Conservation of Nature (IUCN) Red List are threatened with extinction (Dirzo et al., 2014; Murali et al., 2023; Pimm et al., 2014). About 48% of species are declining, 49% are in stable condition, and only 3% are increasing (Finn et al., 2023). The Living Planet Index Report reveals an average 69% decrease in monitored wildlife populations since 1970 (WWF, 2022). However, global assessments are highly biased toward certain taxa and regions (Miqueleiz et al., 2020). For example, 85% of described reptile species have been assessed by IUCN (Meiri et al., 2023), whereas ∼1% of arthropod species have been assessed (Cardoso et al., 2011). This severe discrepancy reflects a long-known bias in research interests (Cardoso et al., 2011; Di Marco et al., 2017) and can be partially attributed to missing or inadequate distribution data that are fundamental to species threat assessments (Beck et al., 2014; Hughes et al., 2021). For example, a taxonomic revision and distribution sample of beetle species shows that 53% of 186 species are known only from a single locality, and 13% are known only from a single specimen (Stork, 1997). These biases are known as Linnean shortfalls (taxonomic knowledge gaps) and Wallacean shortfalls (distribution knowledge gaps) (Diniz-Filho et al., 2023; Hortal et al., 2015). Poor representation of species records is also prominent in the most extensive biodiversity repository—Global Biodiversity Information Facility (GBIF)—which contains locality data for only 10% of described insect species (Chowdhury, Zalucki, et al., 2023). In addition, the taxonomic bias in GBIF is increasing with time, mostly due to the accumulation of distribution data for birds (Troudet et al., 2017; Tang et al., 2021).

To bridge this data gap, many initiatives are using knowledge from the general public (Hochkirch et al., 2021). These data collection initiatives are commonly known as citizen science or community science, in which people share their species observation records through different, often online, applications (e.g., iNaturalist, eBird; Bela et al., 2016; Callaghan et al., 2021; Chandler et al., 2017; Conrad & Hilchey, 2011; Pocock et al., 2019). These observations, in many cases, are eventually deposited in large online repositories (e.g., GBIF; Callaghan et al., 2022; Roy et al., 2018). There are hundreds of such applications globally that have greatly improved understanding of biodiversity patterns in recent years. Since 2007, there has been a 12-fold increase in biodiversity data in GBIF through community science applications (e.g., iNaturalist, eBird) (Heberling et al., 2021) (GBIF contains >2.3 billion species occurrence records as of 9 June 2023). Despite increases in the volume of biodiversity data available, geographic and taxonomic knowledge gaps on species distributions remain (Hughes et al., 2021), even for well-represented taxa (e.g., birds) in well-sampled regions (e.g., Germany) (Bowler et al., 2022). Most biodiversity observation records are from Europe and North America, resulting in major sampling and observation biases (Ramírez et al., 2022). Although most species occur in tropical forests, knowledge of the biodiversity of these regions is extremely limited (Chowdhury, Aich, et al., 2023; Collen et al., 2008; Hortal et al., 2015; Hughes et al., 2021; Meyer et al., 2015). Although it comprises <2% of Earth's area, nearly 36% of the biodiversity data in GBIF come from the United States. Conversely, Brazil, which has a similar land area to the United States and is the most biodiverse country on Earth, is represented by only 0.8% of the records in GBIF (accessed on 22 May 2023). Therefore, new approaches and methods are needed to overcome the Linnean and Wallacean shortfalls, and changing societal preferences through community science activities (e.g., by targeting less-represented species) can play a crucial role (Troudet et al., 2017).

With the increasing popularity of social media and the growing availability of digital phones and fast internet, many people post biodiversity observations on different social media platforms (e.g., Facebook, X) (Andrachuk et al., 2019; Di Minin et al., 2015; Toivonen et al., 2019) that do not necessarily make it to GBIF. Among these platforms, Facebook has become the most popular social media network (Ortiz-Ospina, 2019). There are thousands of biodiversity observation groups on Facebook globally with a wealth of species distribution information, often with more in-depth data than are available in global biodiversity repositories (Chowdhury, Aich, et al., 2023). For example, by scraping a single Facebook group for the butterflies of Bangladesh (Butterfly Bangladesh, https://www.facebook.com/groups/488719627817749), Chowdhury, Alam, et al. (2021) obtained about 35 times more distribution records from Facebook than were deposited in GBIF at the time. Facebook also contains data regarding many unique species that are absent from GBIF altogether (Chowdhury, Aich, et al., 2023). Such data are also key for improved spatial conservation prioritization (Chowdhury et al., 2024) and can help track range-shifting species (Chowdhury, Braby, et al., 2021). The utility of Facebook as a biodiversity repository is possible due to the volunteer contributions of moderators and administrators of Facebook groups who help users identify species (Chowdhury, Aich, et al., 2023; Marcenò et al., 2021). Moreover, many scientists are unaware of the great potential biodiversity data held on social media platforms or are unfamiliar with methods to extract such data (Chowdhury et al., 2024).

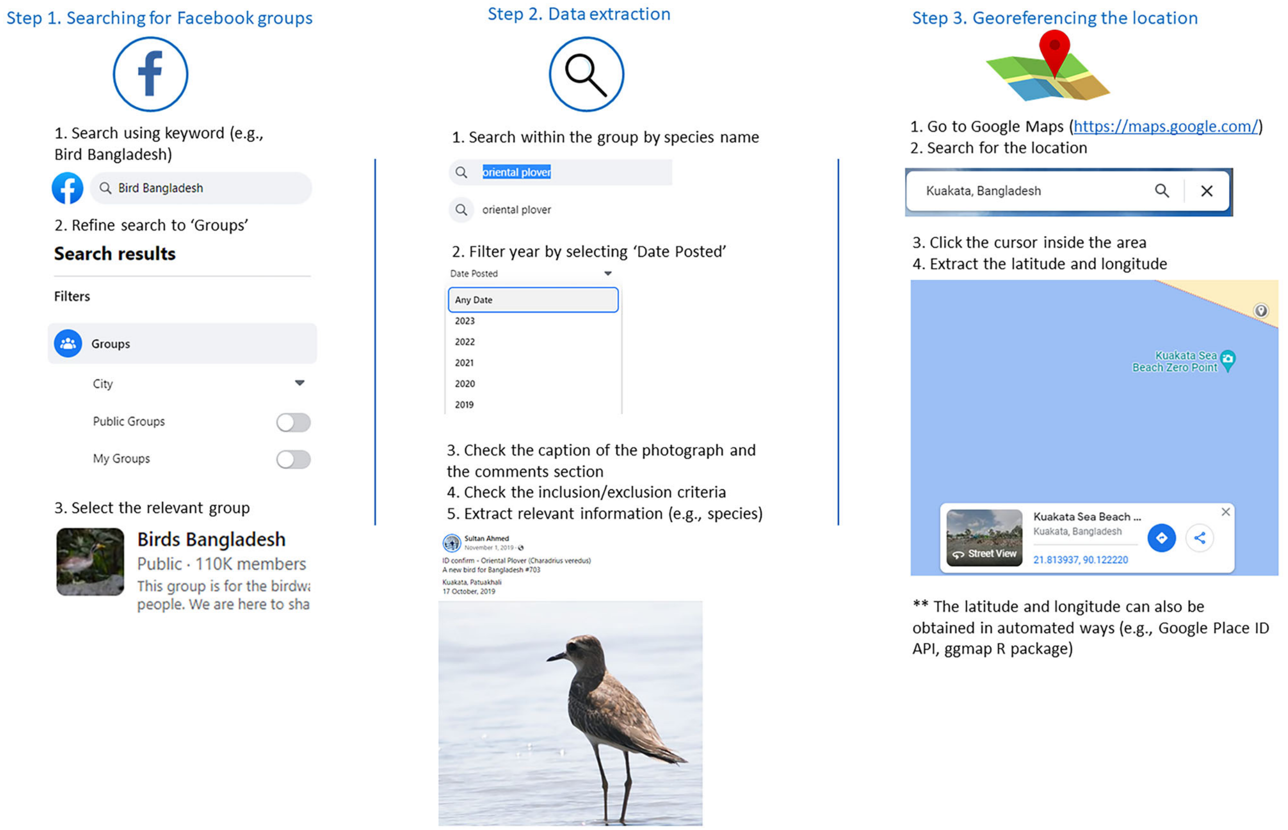


**FIGURE 1** Steps to extract species distribution records from Facebook and the inclusion and exclusion criteria used to filter species photographs for different animal groups of Bangladesh.

Several studies have explored the importance of using Facebook data to fill the global biodiversity data shortfall (e.g., Chamberlain, 2018; Chowdhury, Aich, et al., 2023; Marcenò et al., 2021; O'Neill et al., 2023). Although having a standard data collection protocol is essential to improve conservation assessments, there is currently no such protocol that researchers can follow to collate species distribution data from Facebook. We devised a thorough method for harvesting high-quality biodiversity data from Facebook groups accessed with the permission of the groups' administrators. We applied this method to Bangladesh as a case study.

The method includes a keyword formulation process and protocols for searching Facebook groups, extracting and filtering species photographs and georeferencing location information. We considered key points that future researchers should follow to avoid some commonly encountered problems when using these data sources to help expedite the process and to follow the data privacy rules when using species locality data in conservation science. Finally, we considered the potential of species distribution data obtained from Facebook.

## METHODS

Extracting species locality information from Facebook involves 3 steps: group selection, data extraction, and georeferencing the record location (Figure 1a).

### Group selection (Step 1)

Before starting the data extraction process, it is important to filter the relevant Facebook groups using a systematic search. Search keywords could contain a combination of taxon and country names. For example, if there is an interest in extracting data for birds from Bangladesh, the keyword could be "*bird Bangladesh*." Singular and plural forms and capitalization of words do not make any substantial difference, but word order does. We, therefore, recommend conducting multiple searches with the keywords in different combinations (e.g., *Bangladesh bird*, *bird Bangladesh*). Because there may be multiple groups for each country and taxa combination, this approach will allow researchers to get more extensive results. Given English is not widely spoken in many areas of the world, we recommend using a multilingual search (Chowdhury, Gonzalez, et al., 2022) when listing relevant Facebook groups. After formulating keywords, the search can be conducted using the Facebook search function to identify relevant groups (Figure 1).

### Data extraction (Step 2)

When extracting data, the authors should carefully consider the privacy of the group (public, private, or secret) and maintain the data usage policy (Di Minin et al., 2021). Authors should also prepare a complete list of species that need a manual search.

When a relevant group is selected, the next step is to extract species information. This can be done by searching for species locality records inside individual groups. Searching within Facebook groups is similar to searching in engines such as Google Scholar: species names can be searched with the search function of the selected Facebook group (magnifying glass icon inside the Facebook group), which includes options to restrict the search (e.g., year) (Figure 1).

During data extraction (from posts, but not photographs), problems that may be encountered include species misidentification in the caption, inconsistencies in species' names (some users use the scientific name, whereas others use the English or local names), Facebook becomes too slow when the keyword searches produce too many results, and search results include erroneous species due to partial keyword overlap (e.g., butterfly named common Pierrot could appear when searching for "*common jay*"). To handle misidentification issues, we recommend double-checking species identifications before extracting data. In active Facebook groups, moderators, administrators, or other users often verify individual photographs and provide suggestions when required. This is particularly relevant when they help users identify photographs and confirm locality information. Such information can often be found in the comment section of each photograph. To control naming inconsistencies, we recommend searching by scientific name, English names, and local names for each species in each group. Another issue is that the same photographs might appear in different searches due to partial keyword overlap. To control this issue, researchers could use a unique identifier for the users and then remove duplicate information. To expedite the search process during an individual search, we recommend restricting the search by year with the date posted tab on the left side of the screen. Here, individual years (e.g., 2021) can be filtered for each search. The process should be repeated until there are no further search results. Finally, duplicate records should be removed.

## Georeferencing (Step 3)

Facebook lacks an automated georeferencing system. Using the location information from species photographs (Figure 1), any mapping software (e.g., Google Map, Google Earth, ArcGIS) can be used to match the latitude and longitude information. Conveniently, the Google Place ID application programing interface (https://developers.google.com/maps/documentation/places/web-service/place-id) enables automatic georeferencing for many locations. It can also be done with the geocode function of the ggmap R package (Kahle & Wickham, 2013).

## Case study of Bangladesh

We extracted biodiversity data from Bangladesh's dedicated Facebook groups. We compiled a checklist of the animals of Bangladesh with a known conservation status (1,619 species) from the most recent Red List of Bangladesh (IUCN Bangladesh, 2015). From this red-list database, we extracted taxonomic group name (e.g., Birds, Butterflies), Order, Family, scientific name, English common name, local common name, and the IUCN Red List assessment status in Bangladesh. While cross-checking the data, we noticed that the Family information for butterflies was old and did not match the family names in GBIF. We updated Family names following the GBIF taxonomy to address this problem. Specifically, we updated the information for 4 families (Acraeidae, Amathusiidae, Danaidae, and Satyridae) and moved them under the family Nymphalidae. Finally, we removed regionally extinct species (IUCN Bangladesh, 2015) from the species list.

## Group selection (Step 1)

We searched for each taxonomic group (e.g., bird, butterfly) and added the country name (Bangladesh) at the end (e.g., bird Bangladesh, butterfly Bangladesh). From our previous experiences, we were aware that some Facebook groups contain the term *biodiversity*. Consequently, we also searched Facebook for the keyword "*biodiversity Bangladesh*." Altogether, our search keywords included the following 7 combinations: "*amphibian Bangladesh*," "*bird Bangladesh*," "*butterfly Bangladesh*," "*crustacean Bangladesh*," "*fish Bangladesh*," "*mammal Bangladesh*," and "*reptile Bangladesh*." Based on the search results, we filtered the most popular Facebook groups, for each taxon, based on moderation activity (e.g., whether group moderators help users with species identification), group rules (if the group has strict rules about the location and date of the photographs), and group activities (if members post every day).

## Data extraction (Step 2)

We followed a range of approaches when extracting species distribution data from Facebook groups. Before starting the data extraction process, we conducted a test search with some common and rare species. To control for naming inconsistencies, we searched each species with its scientific name, English name, and local name in Bengali, which led to 3 rounds of searching for each species in each group. To expedite the search process, we restricted searches by year with the date posted tab. We filtered individual years (e.g., 2021) for each search, checked until there were no remaining records for that year, and extracted all relevant results for that year. Finally, we removed duplicate records from the compiled datasheet (see above).

From each post, we extracted the following information: species name (search keyword), life stage (e.g., adult, egg), date (day, month, year), location information (name of the place), and name of the photographer. For quality control, we carefully checked whether the species identification was correct before extracting its information. Given that the group users, including the administrators and moderators, verify species identification and comment on misidentified species, and that some users might forget to update post captions, we also checked the comments of individual photographs. We excluded photographs if

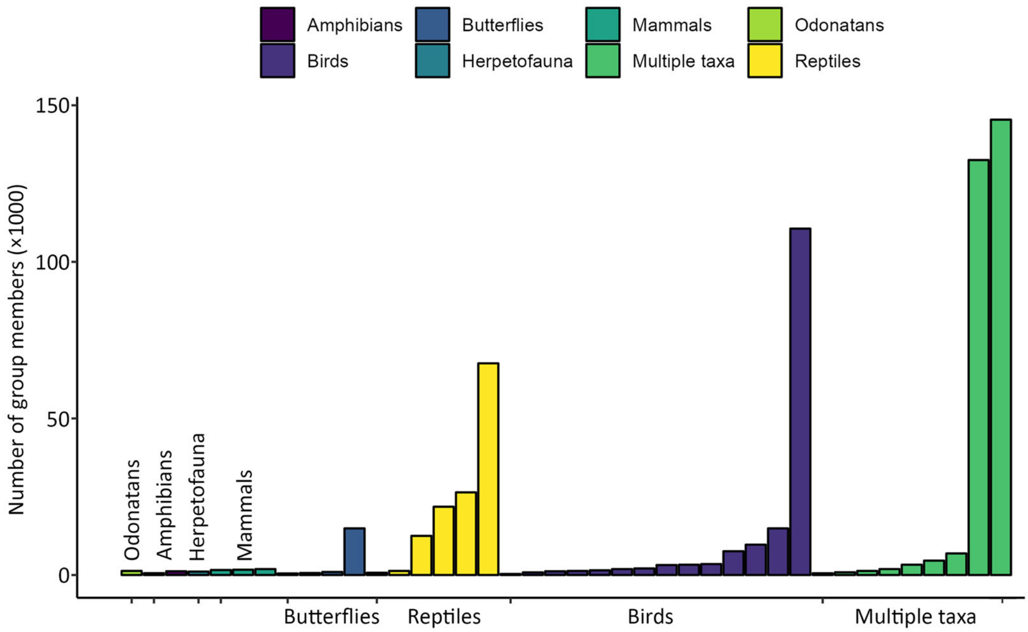


**FIGURE 2** The number and membership details of the available Facebook groups for Bangladeshi animals (herpetofauna, amphibians, and reptiles; multiple taxa, group focused on >2 taxa).

the photograph was not from Bangladesh and was not clear enough to identify the subject up to the species level (even if someone [poster or other users from the group] identified the species); if subject was not identified to the species level; if the locality information was missing; and if the area of the location was over 100 km$^2$.

## Georeferencing (Step 3)

Using the location information from each post, we searched it in Google Maps (https://www.google.com/maps), selected a random point in that area (clicked on the location), and extracted the latitude and longitude (in decimals). Considering that we discarded photographs if the specified location was an area over 100 km$^2$, the precision uncertainty of these extracted distribution records was within 10 km.

# RESULTS

## Data harvested from Bangladesh

Overall, we identified 41 Facebook groups for different animals in Bangladesh (Appendix S1). Although 14 of these groups focused on birds, only one focused on dragonflies and damselflies (Figure 2). There were only 4 groups with >50,000 members. Several of these groups focused on multiple taxa, one on birds, and one on reptiles (Figure 2; Appendix S1).

Of the 41 Facebook groups, we chose the 7 most popular (based on our experience): Birds Bangladesh (https://www.facebook.com/groups/2403154788), Deep Ecology and Snake Rescue Foundation (https://www.facebook.com/groups/959896627527624), Biodiversity of Bangladesh, (https://www.facebook.com/groups/249240636186853), Butterfly Bangladesh, (https://www.facebook.com/groups/488719627817749), Mammals of Bangladesh, (https://www.facebook.com/groups/647662968655338), Amphibians and Reptiles of Bangladesh, (https://www.facebook.com/groups/560709511527645), and Biodiversity of Greater Kushtia (https://www.facebook.com/groups/244807066739477).

From these 7 groups, we collated 44,726 occurrence records for 967 species, ranging from 1 to 719 records per species. These data included 45 amphibian species, 494 bird species, 265 butterfly species, 72 mammal species, and 91 reptile species (Figures 3a & 4). We could not locate any species occurrence records for fishes or crustaceans. Our extracted Facebook data are publicly available (Chowdhury, Aich, et al., 2022).

Although the median number of occurrence records per species was 27, it varied substantially among taxa. The median number of occurrence records per species was 54 for birds, 33 for butterflies, 5 for mammals, and 3 for reptiles. There were 196 species with 80 or more records, of which only 7 were butterflies (no threatened species), and the rest were birds (one threatened species, *Threskiornis melanocephalus* [Vulnerable]). The following were the most popular species from each group: *Alcedo atthis* (birds, 719 records), *Danaus chrysippus* (butterflies, 107 records), *Prionailurus viverrinus* (mammals, 66 records), *Xenochrophis piscator* (reptiles, 39 records), and *Duttaphrynus melanostictus* and *Polypedates leucomystax* (amphibians, 14 records).

The number of occurrence records grew substantially with time, with some random fluctuations (Figure 3b). Although Facebook started in the early 2000s, we obtained many records

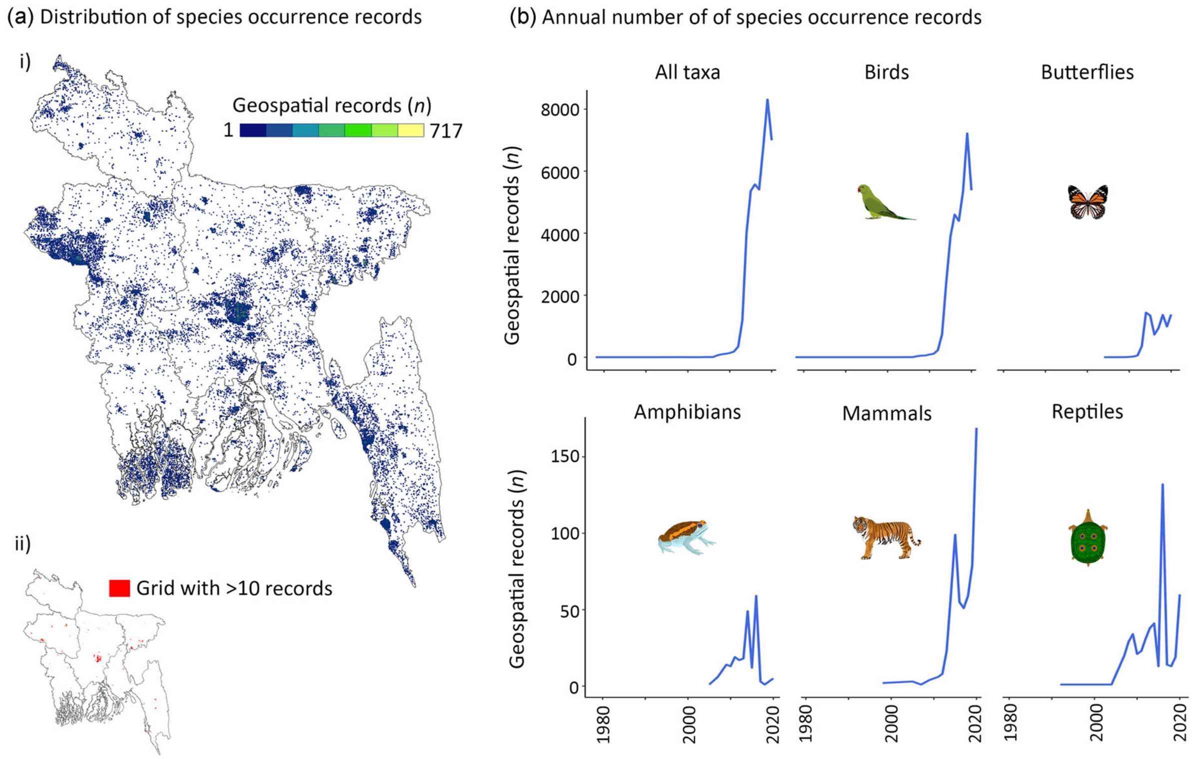


**FIGURE 3** The (a) density (at 1 km$^2$) (i, includes all species occurrence records; ii, includes grid cells with >10 species occurrence records) and (b) yearly growth of species occurrence records obtained from Facebook for different animal groups in Bangladesh.

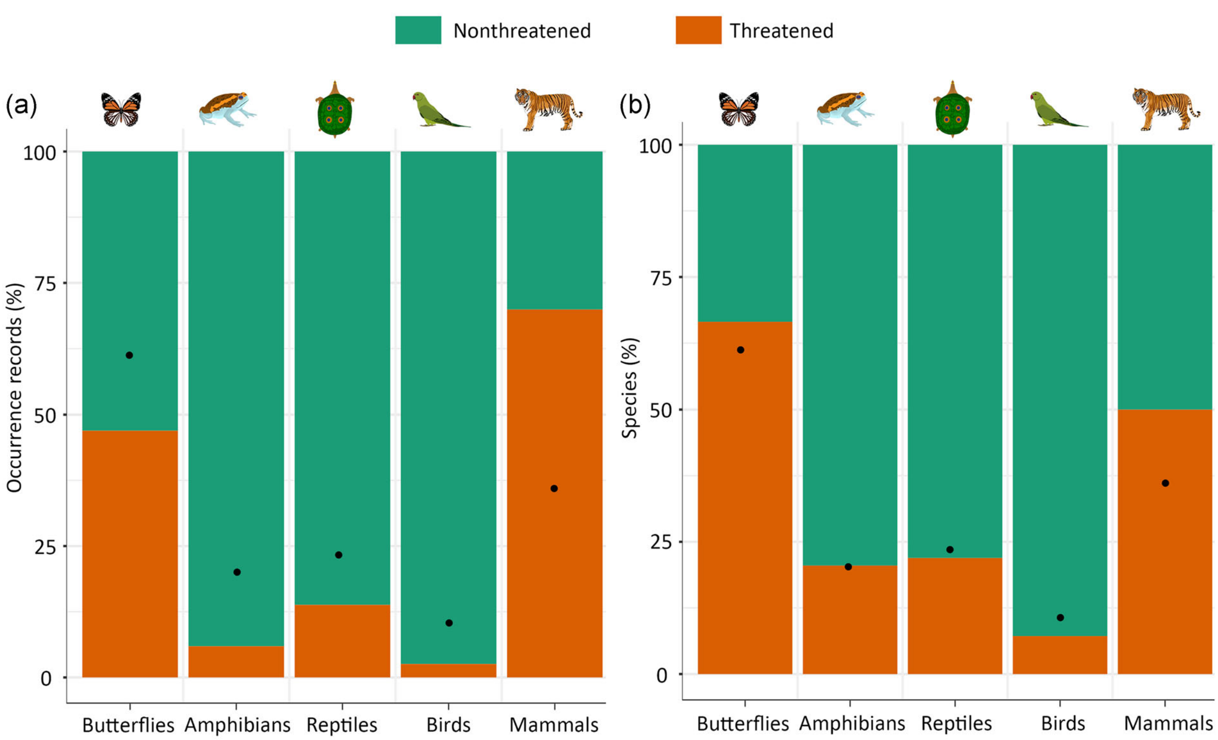


**FIGURE 4** Percentage of threatened and nonthreatened species in (a) occurrence records and (b) species records obtained from Facebook for different animal groups in Bangladesh (black dots, percentages of species that are nationally threatened [IUCN Bangladesh, 2015]).

before that date. For birds, there were records available from 1978 (users shared their past observations). For other taxa, the records started in 1992 for reptiles, 1998 for mammals, 2004 for butterflies, and 2005 for amphibians. For all taxa, the number of species occurrence records increased markedly from the start date to early 2020; however, the numbers rapidly declined afterwards, possibly due to the COVID-19 pandemic (Figure 3b).

We obtained distribution data for 260 threatened species (27% of species), 651 nonthreatened species (67% of species), and 56 data-deficient species (6% of species) (Appendix S2). Although 12% of the species distribution data were for threatened species, it varied substantially across taxa. For mammals, 70% of the occurrence records were for threatened species, 47% for butterflies, 14% for reptiles, 6% for amphibians, and 3% for birds (Figure 4a). The pattern was different when considering the percentage of species recorded for each taxa, both for threatened and nonthreatened species. For butterflies, we obtained records for twice as many species classified as threatened (167) than nonthreatened (84). For other groups, threatened species contributed 46% of all mammal species, 20% for reptiles, 18% for amphibians, and 7% for birds (Figure 4b).

## DISCUSSION

The increasing number of community science applications is contributing to a sharp increase in the available data that can be used to obtain species distribution records (Heberling et al., 2021). Despite this increase, a substantial bias remains in understanding global biodiversity distribution—the distribution of tropical species remains overlooked (Di Marco et al., 2017; Hortal et al., 2015; Hughes et al., 2021; Kühl et al., 2020; Chowdhury, Fuller et al., 2023). Although iNaturalist contains 16,444 observations (until 2022) from Bangladesh (for all taxa) (accessed on 19 December 2023), we obtained nearly 45,000 records for 967 species from Bangladesh (for some selected taxa), of which 27% were nationally threatened, and many were Data Deficient. iNaturalist is an increasingly important source of biodiversity data for many parts of the world (Callaghan et al., 2022; Di Cecco et al., 2021; Soroye et al., 2022), yet we obtained nearly 3 times the number of iNaturalist records from Facebook; even Facebook data did not include all groups and taxa. This highlights the importance of extracting biodiversity data from Facebook (Chowdhury et al., 2024; Chowdhury, Aich, et al., 2023, Chowdhury, Fuller et al., 2023). Over time, data increased sharply for all taxa; however, there was a substantial decline in the amount of data in 2021 during the COVID-19 pandemic. Our step-by-step guidelines for extracting species locality data could help future researchers obtain local and global biodiversity data that could aid in conservation assessments.

Facebook data have the potential to improve knowledge of biodiversity. For example, additional biodiversity data could improve the information on Data Deficient species (Cazalis et al., 2023). However, there are some problems users should consider (Chowdhury, Aich, et al., 2023; Di Minin et al., 2015; Toivonen et al., 2019). First, extracting data from Facebook is a time-consuming and multistep process. On average, it took us nearly 33 min to complete the data extraction process for a single species (i.e., ∼532 h in total for all species). And if a species has different names, the number of manual searches increases. Second, Facebook photographs do not contain precise or automatic geolocation functions, which results in coordinate uncertainties. Third, users' interests in certain taxa could lead to limited data for other taxa (e.g., no Bangladeshi Facebook groups on crustaceans or fishes). Fourth, maintaining the quality of posts in Facebook groups requires active moderation activities, which include a high level of taxonomic expertise and spending time reviewing every photograph. Although this is typical for Bangladesh, where administrators, moderators, and users help maintain every photograph's quality, this might not be the case for many other countries. Although automatic extraction is not yet possible, developing such tools to expedite the process would be useful. Likewise, developing tools that automatically deposit species distribution records from Facebook into global biodiversity repositories could be instrumental in harvesting this source for biodiversity records (Correia et al., 2021; Jarić et al., 2020). If more resources are allocated for this endeavor, people can be appointed to check Facebook posts regularly, help administrators and moderators maintain group quality, extract data regularly, and create a local database that could eventually be deposited into global biodiversity repositories, such as GBIF.

When using social media data for research purposes, inherent risks must be addressed to protect individuals from potential physical or psychological harm, whether intentional or unintentional (Di Minin et al., 2021). To mitigate these risks and ensure user safety, it is important to adopt practices such as data minimization, anonymization, and strict data management protocols. Employing risk-based approaches, such as conducting data privacy impact assessments, can aid in identifying and minimizing privacy risks for social media users. Besides, when sharing the observation records, authors should carefully consider whether they should share threatened species' locations (Sbragaglia et al., 2021). This not only showcases accountability but also ensures compliance with data protection laws that safeguard the privacy of individuals involved in the research process (Di Minin et al., 2021). Although open data can revolutionize research efforts, it can create major problems in conserving threatened species (e.g., poaching, disturbance; Bergman et al., 2022; Di Minin et al., 2015, 2022; Lindenmayer & Scheele, 2017). The regional researchers and legal authorities should carefully deal with this issue.

Scientists worldwide are formulating innovative approaches to improve understanding of species distributions, yet there remain substantial taxonomic and regional gaps and biases. Several studies show Facebook's importance in extracting biodiversity records (e.g., Chowdhury et al., 2024; Chowdhury, Aich, et al., 2023; Chowdhury, Alam, et al., 2021; Chowdhury, Braby, et al., 2021); however, previously, there has been no standard method that researchers can follow. We provided detailed guidelines on preparing a list of relevant Facebook groups and recommend approaches to search efficiently. Although the entire process requires time, it was worth the effort, given that we obtained thousands of species occurrence records for a tropical, biodiverse country, Bangladesh. Such biodiversity data could help support conservation efforts and assess population status more accurately with relatively little effort for species that otherwise would have little data or protection (Chowdhury, Fuller et al., 2023). We recommend a coordinated effort to create a global database containing information from biodiversity groups on Facebook.

## ACKNOWLEDGMENTS

S.C. and A.B. gratefully acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) and the sMon project funded by the German Research Foundation (DFG-FZT 118, 202548816). VS is supported by a Ramón y Cajal research fellowship (RYC2021-033065-I) granted by the Spanish Ministry of Science and Innovation and he also acknowledges the Spanish government through the 'Severo Ochoa Centre of Excellence' accredition (CEX2019-000928-S).

## OPEN RESEARCH BADGES

This article has earned Open Data and Open Materials badges. Data and materials are available at https://doi.pangaea.de/10.1594/PANGAEA.948104.

## ORCID

*Shawan Chowdhury* https://orcid.org/0000-0003-2936-5786
*Shofiul Alam* https://orcid.org/0000-0002-0354-4658
*Corey T. Callaghan* https://orcid.org/0000-0003-0415-2709
*Priyanka Das* https://orcid.org/0000-0002-4000-874X
*Moreno Di Marco* https://orcid.org/0000-0002-8902-4193
*Enrico Di Minin* https://orcid.org/0000-0002-5562-318X
*Ivan Jarić* https://orcid.org/0000-0002-2185-297X
*Mahzabin Muzahid Labi* https://orcid.org/0000-0002-5338-6141
*Md. Rokonuzzaman* https://orcid.org/0000-0003-4477-8546
*Uri Roll* https://orcid.org/0000-0002-5418-1164
*Valerio Sbragaglia* https://orcid.org/0000-0002-4775-7049
*Aletta Bonn* https://orcid.org/0000-0002-8345-4600

## REFERENCES

Andrachuk, M., Marschke, M., Hings, C., & Armitage, D. (2019). Smartphone technologies supporting community-based environmental monitoring and implementation: A systematic scoping review. *Biological Conservation*, *237*, 430–442.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15.

Bela, G., Peltola, T., Young, J. C., Balázs, B., Arpin, I., Pataki, G., Hauck, J., Kelemen, E., Kopperoinen, L., Van Herzele, A., Keune, H., Hecker, S., Suškevičs, M., Roy, H. E., Itkonen, P., Külvik, M., László, M., Basnou, C., Pino, J., & Bonn, A. (2016). Learning and the transformative potential of citizen science. *Conservation Biology*, *30*(5), 990–999.

Bergman, J. N., Buxton, R. T., Lin, H. Y., Lenda, M., Attinello, K., Hajdasz, A. C., Rivest, S. A., Tran Nguyen, T., & Bennett, J. R. (2022). Evaluating the benefits and risks of social media for wildlife conservation. *Facets*, *7*(1), 360–397.

Bowler, D. E., Callaghan, C. T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Temporal trends in the spatial bias of species occurrence records. *Ecography*, *2022*(8), Article e06219.

Callaghan, C. T., Mesaglio, T., Ascher, J. S., Brooks, T. M., Cabras, A. A., Chandler, M., Cornwell, W. K., Ríos-Málaver, I. C., Dankowicz, E., Dhiya'ulhaq, N. U., Fuller, R. A., Galindo-Leal, C., Grattarola, F., Hewitt, S., Higgins, L., Hitchcock, C., Hung, K.-L. J., Iwane, T., & Kahumbu, P. (2022). The benefits of contributing to the citizen science platform iNaturalist as an identifier. *PLoS Biology*, *20*(11), Article e3001843.

Callaghan, C. T., Poore, A. G., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J., VergÉs, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, *71*(1), 55–63.

Cardoso, P., Erwin, T. L., Borges, P. A., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, *144*, 2647–2655.

Cazalis, V., Santini, L., Lucas, P. M., González-Suárez, M., Hoffmann, M., Benítez-López, A., Pacifici, M., Schipper, A. M., Böhm, M., Zizka, A., Clausnitzer, V., Meyer, C., Jung, M., Butchart, S. H. M., Cardoso, P., Mancini, G., Akçakaya, H. R., Young, B. E., Patoine, G., & Di Marco, M. (2023). Prioritizing the reassessment of data-deficient species on the IUCN Red List. *Conservation Biology*, *37*(6), Article e14139.

Chamberlain, J. (2018). Using social media for biomonitoring: How Facebook, Twitter, Flickr and other social networking platforms can provide large-scale biodiversity data. In D. Bohan, A. Dumbrell, G. Woodward, & M. Jackson (Eds.), *Advances in ecological research* (Vol. 59, pp. 133–168). Academic Press.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294.

Chowdhury, S., Fuller, R. A., Rokonuzzaman, Md., Alam, S., Das, P., Siddika, A., Ahmed, S., Labi, M. M., Chowdhury, S. U., Mukul, S. A., Böhm, M., & Hanson, J. O. (2023). Insights from citizen science reveal priority areas for conserving biodiversity in Bangladesh. *One Earth*, *6*(10), 1315–1325. https://doi.org/10.1016/j.oneear.2023.08.025

Chowdhury, S., Zalucki, M. P., Hanson, J. O., Tiatragul, S., Green, D., Watson, J. E., & Fuller, R. A. (2023). Three-quarters of insect species are insufficiently represented by protected areas. *One Earth*, *6*(2), 139–146.

Chowdhury, S., Aich, U., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A., Ahmed, S., Labi, M. M., Marco, M. D., Fuller, R. A., & Callaghan, C. T. (2023). Increasing biodiversity knowledge through social media: A case study from tropical Bangladesh. *BioScience*, *73*(6), 453–459.

Chowdhury, S., Fuller, R., Ahmed, S., Alam, S., Callaghan, C., Das, P., Correia, R. A., Di Marco, M. D., Di Minin, E. D., Jarić, I., Labi, M. M., Ladle, R. J., Rokonuzzaman, M., Roll, U., Sbragaglia, V., Siddika, A., & Bonn, A. (2024). Using social media records to inform conservation planning. *Conservation Biology*, *38*(1), Article e14161.

Chowdhury, S., Gonzalez, K., Aytekin, M. Ç. K., Baek, S. Y., Bełcik, M., Bertolino, S., Duijns, S., Han, Y., Jantke, K., Katayose, R., Lin, M.-M., Nourani, E., Ramos, D. L., Rouyer, M.-M., Sidemo-Holm, W., Vozykova, S., Zamora-Gutierrez, V., & Amano, T. (2022). Growth of non-English-language literature on biodiversity conservation. *Conservation Biology*, *36*(4), Article e13883.

Chowdhury, S., Aich, U., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A., Ahmed, S., Labi, M. M., Marco, M. D., Callaghan, C. T., & Fuller, R. A. (2022). *Spatial occurrence data for the animals of Bangladesh derived from Facebook*. Pangaea. https://doi.org/10.1594/PANGAEA.948104

Chowdhury, S., Alam, S., Chowdhury, S. U., Rokonuzzaman, M., Shahriar, S. A., Shome, A. R., & Fuller, R. A. (2021). Butterflies are weakly protected in a mega-populated country, Bangladesh. *Global Ecology and Conservation*, *26*, Article e01484.

Chowdhury, S., Braby, M. F., Fuller, R. A., & Zalucki, M. P. (2021). Coasting along to a wider range: Niche conservatism in the recent range expansion of the Tawny Coster, Acraea terpsicore (Lepidoptera: Nymphalidae). *Diversity and Distributions*, *27*(3), 402–415.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science*, *1*(2), 75–88.

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, *176*, 273–291.

Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C., Mittermeier, J. C., Roll, U., Soriano-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, *35*(2), 398–411.

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, *71*, 1179–1188.

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., Maina, J. M., Possingham, H. P., von Bieberstein, K. R., Venter, O., & Watson, J. E. M. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, *10*, 32–42.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437–446.

Di Minin, E., Selier, J., Louis, M., & Bradshaw, C. J. (2022). Dismantling the poachernomics of the illegal wildlife trade. *Biological Conservation*, *265*, Article 109418.

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, Article 63.

Diniz Filho, J. A. F., Jardim, L., Guedes, J. J., Meyer, L., Stropp, J., Frateles, L. E. F., Pinto, R. B., Lohmann, L. G., Tessarolo, G., de Carvalho, C. J., Ladle, R. J., & Hortal, J. (2023). Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*, *15*(2), Article e59566. https://doi.org/10.21425/F5FBG59566

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*(6195), 401–406.

Finn, C., Grattarola, F., & Pincheira-Donoso, D. (2023). More losers than winners: Investigating Anthropocene defaunation through the diversity of population trends. *Biological Reviews*, *98*(5), 1732–1748. https://doi.org/10.1111/brv.12974

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6), Article e2018093118.

Hochkirch, A., Samways, M. J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P., Cumberlidge, N., Stephenson, P. J., Seddon, M. B., Clausnitzer, V., Borges, P. A. V., Mueller, G. M., Pearce-Kelly, P., Raimondo, D. C., Danielczak, A., & Dijkstra, K. D. B. (2021). A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, *35*(2), 502–509.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 523–549.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259–1269.

IUCN Bangladesh. (2015). *Red list of Bangladesh: A brief on assessment result 2015*. International Union for Conservation of Nature, Bangladesh Country Office.

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., Soriano-Redondo, A., Souza, A. T., & Roll, U. (2020). iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*(7), 630–639.

Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

Kühl, H. S., Bowler, D. E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., Fernández, N., Guerra, C. A., Henle, K., Herbinger, I., Isaac, N. J. B., Jansen, F., König-Ries, B., Kühn, I., Nilsen, E. B., Pe'er, G., Richter, A., Schulte, R., … Bonn, A. (2020). Effective biodiversity monitoring needs a culture of integration. *One Earth*, *3*, 462–474.

Lindenmayer, D., & Scheele, B. (2017). Do not publish. *Science*, *356*(6340), 800–801.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., Gristina, A. S., Agrillo, E., Bonari, G., Giusso del Galdo, G., Ilardi, V., Landucci, F., & Guarino, R. (2021). Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology*, *58*(10), 2018–2028.

Meiri, S., Chapple, D. G., Tolley, K. A., Mitchell, N., Laniado, T., Cox, N., Bowles, P., Young, B. E., Caetano, G., Geschke, J., Böhm, M., & Roll, U. (2023). Done but not dusted: Reflections on the first global reptile assessment and priorities for the second. *Biological Conservation*, *278*, Article 109879.

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*, Article 8221.

Miqueleiz, I., Bohm, M., Ariño, A. H., & Miranda, R. (2020). Assessment gaps and biases in knowledge of conservation status of fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(2), 225–236.

Murali, G., Iwamura, T., Meiri, S., & Roll, U. (2023). Future temperature extremes threaten land vertebrates. *Nature*, *615*(7952), 461–467.

O'Neill, D., Häkkinen, H., Neumann, J., Shaffrey, L., Cheffings, C., Norris, K., & Pettorelli, N. (2023). Investigating the potential of social media and citizen science data to track changes in species' distributions. *Ecology and Evolution*, *13*(5), Article e10063.

Ortiz-Ospina, E. (2019). *The rise of social media. Our world in data*. https://ourworldindata.org/rise-of-social-media

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), Article 1246752.

Pocock, M. J., Roy, H. E., August, T., Kuria, A., Barasa, F., Bett, J., Githiru, M., Kairo, J., Kimani, J., & Kinuthia, W. (2019). Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of Applied Ecology*, *56*, 274–281.

Ramirez, F., Sbragaglia, V., Soacha, K., Coll, M., & Piera, J. (2022). Challenges for marine ecological assessments: Completeness of findable, accessible, interoperable, and reusable biodiversity data in European seas. *Frontiers in Marine Science*, *8*, Article 802235.

Roy, H., Groom, Q., Adriaens, T., Agnello, G., Antic, M., Archambeau, A. S., Bacher, S., Bonn, A., Brown, P., Brundu, G., López, B. C., Cleary, M., Cogălniceanu, D., de Groot, M., De Sousa, M., Deidun, A., Essl, F., Pečnikar, Ž. F., … Gazda, A. (2018). Increasing understanding of alien species through citizen science (Alien-CSI). *Research Ideas and Outcomes*, *4*, Article e31412.

Sbragaglia, V., Coco, S., Correia, R. A., Coll, M., & Arlinghaus, R. (2021). Analyzing publicly available videos about recreational fishing reveals key ecological and social insights: A case study about groupers in the Mediterranean Sea. *Science of The Total Environment*, *765*, 142672. https://doi.org/10.1016/j.scitotenv.2020.142672

Soroye, P., Edwards, B. P., Buxton, R. T., Ethier, J. P., Frempong-Manso, A., Keefe, H. E., Berberi, A., Roach-Krajewski, M., Binley, A. D., Vincent, J. G., Lin, H.-Y., Cooke, S. J., & Bennett, J. R. (2022). The risks and rewards of community science for threatened species monitoring. *Conservation Science and Practice*, *4*, Article e12788.

Stork, N. E. (1997). Measuring global biodiversity and its decline. In M. L. Reaka-Kudla, D. E. Wilson, & E. O. Wilson (Eds.), *Biodiversity II: Understanding and protecting our biological resources* (pp. 41–68). Joseph Henry Press.

Tang, B., Clark, J. S., & Gelfand, A. E. (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics*, *28*(3), 609–630. https://doi.org/10.1007/s10651-021-00508-1

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298–315.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), Article 9132.

World Wide Fund For Nature (WWF). (2022). *Living Planet Report 2022 – Building a naturepositive society*. Author.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Chowdhury, S., Ahmed, S., Alam, S., Callaghan, C. T., Das, P., Di Marco, M., Di Minin, E., Jarić, I., Labi, M. M., Rokonuzzaman, Md., Roll, U., Sbragaglia, V., Siddika, A., & Bonn, A. (2024). A protocol for harvesting biodiversity data from Facebook. *Conservation Biology*, e14257. https://doi.org/10.1111/cobi.14257