

## PERSPECTIVE

## Innovation in Practice

# Overcoming biodiversity blindness: Secondary data in primary citizen science observations

Nadja Pernat<sup>1,2</sup>  | Susan Canavan<sup>3,4</sup>  | Marina Golivets<sup>5</sup>  | Jasmijn Hillaert<sup>6</sup>  |  
 Yuval Itescu<sup>7,8,9</sup>  | Ivan Jarić<sup>10,11</sup>  | Hjalte M. R. Mann<sup>12</sup>  | Pavel Pipek<sup>3,13</sup>  |  
 Cristina Preda<sup>14</sup>  | David M. Richardson<sup>3,15</sup>  | Heliana Teixeira<sup>16</sup>  |  
 Ana Sofia Vaz<sup>17,18,19</sup>  | Quentin Groom<sup>20</sup> 

## Correspondence

Nadja Pernat

Email: [nadjapernat@uni-muenster.de](mailto:nadjapernat@uni-muenster.de)

## Funding information

European Cooperation in Science and Technology, Grant/Award Number: COST Action CA17122

Handling Editor: Mark O'Connell

## Abstract

1. In the face of the global biodiversity crisis, collecting comprehensive data and making the best use of existing data are becoming increasingly important to understand patterns and drivers of environmental and biological phenomena at different scales.
2. Here we address the concept of secondary data, which refers to additional information unintentionally captured in species records, especially in multimedia-based citizen science reports. We argue that secondary data can provide a wealth of ecologically relevant information, the utilisation of which can enhance our understanding of traits and interactions among individual organisms, populations and biodiversity dynamics in general.
3. We explore the possibilities offered by secondary data and describe their main types and sources. An overview of research in this field provides a synthesis of the results already achieved using secondary data and different approaches to information extraction.
4. Finally, we discuss challenges to the widespread use of secondary data, such as biases, licensing issues, use of metadata and lack of awareness of this trove of data due to a missing common terminology, as well as possible solutions to overcome these barriers.
5. Although the exploration and use of secondary data is only emerging, the many opportunities identified show how these data can enrich biodiversity research and monitoring.

## KEYWORDS

artificial intelligence, biodiversity, citizen science, community science, ecological interactions, invasion science, monitoring

For Affiliation refer page on 10

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Ecological Solutions and Evidence* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

# 1 | THE UNTAPPED INFORMATION IN EXISTING BIODIVERSITY DATA

Citizen science contributes enormously to biodiversity monitoring (Chandler et al., 2017) by providing data that are potentially as useful as those collected by professional scientists, especially for research over large spatial and temporal extents (Callaghan et al., 2020). However, there are still major gaps in the taxonomic, temporal and spatial coverage of biodiversity data to track changes in species' abundance and distribution (Amano et al., 2016; Feldman et al., 2021; Wetzel et al., 2018). Furthermore, biodiversity encompasses not only the diversity of organisms but also the diversity of interactions between them and with their environment, which has been given scant attention by citizen science (Chandler et al., 2017; Groom et al., 2021). Ecological interactions are the foundation of ecology and the architecture of ecosystems, but observing these relationships is challenging (Jordano, 2016). The sheer amount and complexity of interactions as well as the detection probability in the field limits studies with respect to number of species, locations and time periods that can be studied.

However, both occurrence and interaction data are needed to understand, address and mitigate the consequences of the five major drivers of biodiversity loss (Díaz et al., 2019). For instance, land-use change may affect host-vector dynamics (Spence Beaulieu et al., 2019), pollution may lead to adaptations in species traits (Rech et al., 2022), climate change can affect trophic cascades and distribution shifts (van Gils et al., 2016), overexploitation may change former mutualisms (Speziale et al., 2018) and biological invasions can result in novel plant-pollinator networks (Parra-Tabla & Arce-Gómez, 2021). The variety of direct and indirect effects of biotic and abiotic interactions are difficult to study, which poses a challenge that can only be addressed through global collective effort (Díaz et al., 2019).

To achieve more accurate and reliable monitoring and interaction data, we need to improve and integrate current methods (Besson et al., 2022; Kühl et al., 2020; van Klink et al., 2022), to better analyse existing data (Probert et al., 2022), and to extract more information from already collected data (Johnston et al., 2022). In the latter case, one untapped source of abundant data is the corpus of digital images and other media, generated and shared on citizen science platforms. As of May 2023, four such websites alone (Artportalen, iNaturalist, Observation.org, and Pl@ntNet) had collectively published over 78 million images through the Global Biodiversity Information Facility (GBIF). Almost all these photos were taken of organisms or signs of their presence in situ (i.e. nests, faeces, tracks, etc.) and thus may capture ecologically relevant information as a by-product. This type of additional information has been termed 'secondary data' (Callaghan et al., 2021, see Box 1 for details). Having recognised the potential, a growing number of studies have examined ecological questions using incidentally captured information. Yet, so far, the scientific community has been relatively blind to the opportunities to extend

## BOX 1 What are secondary data?

Secondary data in the context of citizen science in biodiversity research, as we define it, refers to a subset of information that is unintentionally captured alongside primary data. Primary citizen science data collected for a specific research focus, such as monitoring the distribution of a species, provide information on the location and date of record of the species in addition to evidence of its occurrence. This primary data of 'what?', 'when?' and 'where?' are the intended focus of many ad hoc observing portals. In contrast, secondary data are ancillary details that are also present in the materials collected but were *not the intended subject* of the study. Indeed, the observer may have been unaware of the secondary information they evidenced.

Secondary data can offer valuable opportunities for additional research and analysis, enriching our understanding of ecosystems' functioning, population dynamics, natural behaviours and environmental conditions. They represent any retrievable pieces of information that can be seen on an image or in a video, heard on an acoustic recording or be included in a descriptive text. The information they contain may relate to features of individuals or populations, biotic interactions (including human-nature interactions), landscape and environmental conditions, or any other biotic or abiotic features or their combination (Figure 1). We recognise, of course, that the subject of investigation may be something other than the mere detection of species. In complex citizen science programmes, the primary data can only be separated from secondary data with reference to the objectives of the project. For example, in the COASST project (Parrish et al., 2017), citizen scientists record bird carcasses on beaches, providing not only an image but also record a wealth of information about the morphology of the carcass and the state of the environment. In addition to the primary information collected, the images of dead birds may contain even more information than the research project anticipated, such as the presence of necrophagous species. Thus, secondary data are data that the methodology was not intended to capture, though there are no sharp demarcation lines.

our knowledge of biodiversity from secondary data. Therefore, we assume a 'biodiversity blindness', drawing on the concept to describe the public's lack of attention to the presence and diversity of plant and animal life (Moscoe & Hanes, 2019).

Following our definition of secondary data (Box 1), we do not explicitly refer to the metadata (e.g. timestamp or geolocation) associated with the occurrence record, though they also represent potential secondary data sources and play a major role in selecting



**FIGURE 1** Examples of secondary data captured in images vouchering for species occurrences. (a) background habitat of a fire salamander (*Salamandra salamandra*), and the unintended record of another species (the fungus) (b) the interaction between human and red fox (*Vulpes vulpes*), (c) interaction between a leaf miner and the host plant, (d) predation by the Mexican grass-carrying wasp *Isodontia mexicana*. Image copyrights: © <https://www.inaturalist.org/photos/251871116>, <http://creativecommons.org/publicdomain/zero/1.0/> (a) © juniper\_likethetree <https://www.inaturalist.org/observations/155228701>, <https://creativecommons.org/licenses/by/4.0/> (b) ©ahabo <https://www.inaturalist.org/photos/257468598>, <https://creativecommons.org/licenses/by/4.0/> (c) © Mattia Menchetti <https://www.inaturalist.org/observations/129293402>, <https://creativecommons.org/licenses/by/4.0/> (d).

datasets and their analyses. An important element in utilising secondary data is the explorative character of this research method, as the nature of this unintentionally recorded information may be unclear. The additional information may document biotic interactions and co-occurrences, which could provide not only important ecological information of the observed species but also records of the by-catch for respective monitoring programmes. Secondary data may also include details on morphology, behaviour, habitat and various other aspects of a species' traits and ecology.

This paper explores the opportunities and pitfalls of extracting secondary data from multimedia records of biodiversity. We also examine how advances in artificial intelligence can and might accelerate data extraction. Our goal is to illuminate the hidden treasure of biodiversity data contained in citizen science multimedia records and available openly to the scientific community. While the efforts to explore and exploit the realm of secondary data are still in their infancy, they already demonstrate numerous opportunities to enrich and inform biodiversity research.

## 2 | RESEARCH OPPORTUNITIES AND TYPES OF SECONDARY DATA

Extracting secondary data from existing citizen science sources helps to address universal challenges of biodiversity research, such

as taxonomic bias, detectability of species and their interactions, and recognition of spatio-temporal dynamics.

Taxonomic bias towards charismatic or well-known species in citizen science data poses a challenge for researchers and limits the possibilities of launching projects that deal with less popular, cryptic or under-researched species. In addition, although simple, unstructured programmes generate high numbers of citizen science observations, information that is more complex to record, such as an individual's health condition, is not purposefully collected and therefore not formally documented. Similarly, studies that are less engaging, such as those that are time-consuming, physically challenging or in less attractive localities are discriminated against. Extracting secondary data from existing observations could be fruitful to fill such data needs. For example, diurnal or seasonal activity patterns or vocal characteristics of rare species could be retrieved from soundscapes or the background of audio recordings of focal species. From images, occurrences of less charismatic arthropods or pathogenic fungi living on photographed plants could be extracted. In the latter case, in 2010, citizen scientists monitored and scored leaf damage on horse-chestnut trees (*Aesculus hippocastanum*) caused by the leaf-miner *Cameraria ohridella* in Great Britain (Pocock & Evans, 2014). Today, the infestation could also be detected as secondary data in images on which horse-chestnut trees are the primary observation, thereby improving the data situation at low additional costs and resource expenditure.

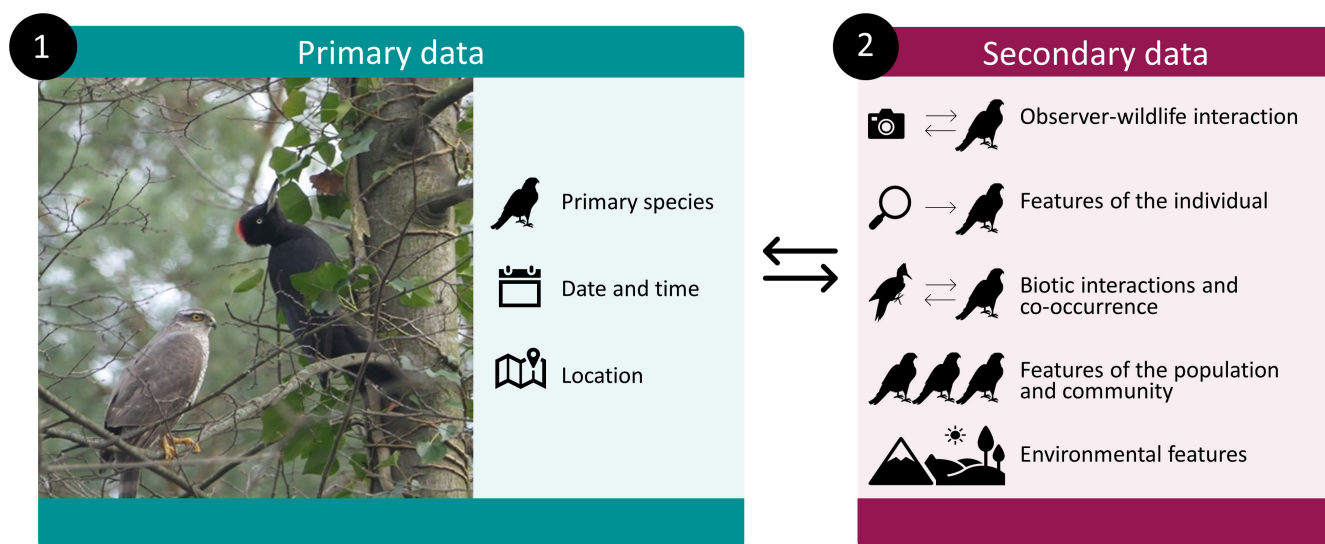
Using secondary data to take advantage of the taxonomic bias towards well-documented species also brings additional research opportunities. When observations for a given species are widely available, one can extract data on multiple aspects of interest, such as morphological traits or biotic interactions, without spending time and resources on launching and running a new raw data collection campaign. For example, Putman et al. (2021) used images of the secretive but thoroughly documented lizard *Elgaria multicarinata* that were primarily collected for determining the species' distribution in Southern California. From the images, the authors assessed predation pressure and health condition by measuring the lizards' tails and by looking for ectoparasites in the animal ear regions. Likewise, citizen science photos have been used to identify subtle morphological differences between two very similar species of grasshopper and to establish their distributions (Pélissier et al., 2023).

Coincidental evidence can mitigate low detection probabilities. By identifying a rare species in primary observations of other species, whether that is through biotic interaction or an incidental co-occurrence, the pool of observations can be enlarged. A citizen science project in Australia has shown that co-occurrence of a common and rare possum species can lead to more detections of the latter (Steven et al., 2021). Aside from potentially increasing sample sizes of monitoring data for the benefit of statistical analyses, we can improve our understanding of ecological impacts on other species, including people. We envisage application in pollination dynamics, invasion impact or climate change research. For example, we can potentially study the preferred flower species and colour in a network of native and exotic bumblebees and host plants (Catron et al., 2023; Fontúrbel et al., 2023). Another example is hair loss in moose (*Alces alces*) and wapiti (*Cervus canadensis*) caused by the expanding distribution of winter tick

(*Dermacentor albipictus*) due to climate warming in Yukon, Canada (Chenery, 2023). Serendipity is a factor as well to reveal ecological interactions; Rosa et al. (2022) not only found new and supposedly extinct species as primary observations, but also novel predatory interactions that were accidentally captured in the iNaturalist images of marine snails. Such chance discoveries based on the background information could be especially useful in invasion science, where secondary data may reveal new or hidden invasions or previously undocumented ecological processes that facilitate or hinder invasions.

Extracting secondary data from a series of observations across space and time can also support efforts to move from a mere single-species snapshot (an occurrence record) to spatio-temporal biodiversity dynamics. Using timestamps and geolocation metadata of citizen science observations to investigate spatial and temporal dynamics has been successfully applied before (e.g. Feldman et al., 2021; Newson et al., 2016). Given a sufficient temporal span and frequency of observations, we suggest linking secondary data to such a stamp to obtain a variety of observable dynamics. For example, a series of landscape images would not only contribute to monitoring data (e.g. the abundance and distribution of species on the images), but can also be useful for studying phenological dynamics at the community level (Hofmeester et al., 2020).

Figure 2 illustrates how secondary data can add contextual dimensions to primary species observations, thereby mitigating 'biodiversity blindness' by expanding on the information in citizen science multimedia records beyond geographical locations and time. This context applies to different scales and scopes of consideration, specifically on the level of individuals, populations, communities, the surrounding environment and the human dimension. For each of these levels, Table 1 gives extensive lists of types of information contained in secondary data.



**FIGURE 2** Illustrative relationship between primary and secondary data as sources for different types of information. Image copyright: © zilpzalp17, <https://www.inaturalist.org/observations/148103702>, <https://creativecommons.org/licenses/by-nc/4.0/>.



**TABLE 1** Types of information extractable from secondary data from citizen science projects. As the literature on secondary data is scarce and obscure due to a missing common terminology, we also listed examples that used secondary information in combination with other sources and approaches (e.g. iEcology or literature). The table groups the publications by level (human–nature interactions, features of the individual, etc.) as described. For each study it gives the feature of interest (extracted data) and the data elements used to extract it as well as a short description (study example) of the content and the methods and sources used (source and extraction method).

Data element	Extracted data	Study example	Source and extraction method <sup>a</sup>	Reference
<i>Observer/human–nature interactions</i>				
Interpretation of photo	Animal behaviour	Human–coyote encounter in urban environments, assessing timing, level of urbanisation and aggressiveness	iNaturalist; manual extraction by visual inspection and categorisation of coyote behaviour	Drake et al. (2021)
Interpretation of photo	Association with human infrastructure	Investigations of wrens nesting in stable or unstable structures by examining citizen science images	iNaturalist, eBird, WikiAves and field data; manual extraction by visual inspection of nesting sites	Alexandrino et al. (2022)
Interpretation of photo	Cause of death	Taxa and seasonality of roadkills in the USA	Tagged iNaturalist images ('roadkill', 'dead on road'); manual extraction by visual inspection, for example, species identification	Unger (2022)
<i>Features of the individual</i>				
Time and interpretation of photo	Diurnal patterns	Diurnal pattern of a lizard	iNaturalist, HerpMapper; manual extraction by visual inspection of daytime and light on the images	Blais and Shaw (2018)
Date and interpretation of photo	Phenology	Fine-scale delimitation of the flowering phenology of <i>Yucca</i> species	iNaturalist; manual extraction by applying a scoring rubric and consensus method	Barve et al. (2020)
		Progression of phenological stages of <i>Alliaria petiolata</i>	iNaturalist; deep learning algorithm to classify phenological stages	Reeb et al. (2022)
Identification and interpretation of photo	Hybrids	Assessing hybridisation by colour variation among grass snakes across Europe	iNaturalist; manual extraction by visual inspection following an identification key for several coloured features	Fritz and Ihlow (2022)
Measurement and interpretation of photo	Morphological traits	Pigmentation of wings in male damselflies across broad spatial scales	iNaturalist; half-automated extraction of size and place of pigmentation using visual inspection and ImageJ	Drury et al. (2019)
		Melanisation of grey squirrels in relation to urbanisation and temperature	iNaturalist and SquirrelMapper; manual extraction of squirrel colour by visual inspection in the form of a citizen science project	Cosentino and Gibbs (2022)
		Length ratio measurements to identify two closely related grasshopper species	iNaturalist; length ratios from photos using ImageJ	Pélissié et al. (2023)
Interpretation of photo	Behavioural traits	Warming-up behaviour of differently coloured rattlesnakes and their corresponding survival rates	iNaturalist; manual extraction and classification of colour, living or dead and background	Rhodes et al. (2022)
Date and interpretation of photo	Life history	Breeding biology of swallow-tailed hummingbirds	WikiAves, iNaturalist, eBird; manual extraction of classes of breeding stages (nesting, eggs, fledglings, etc.) and using dates for calculation of breeding timings	Turella et al. (2022)
Interpretation of photo	Health	Body condition of seals	Applying images of the New Zealand Leopard Seal Photograph Library (NZLSPL) to a self-developed scoring system	Warren (2021)

(Continues)

TABLE 1 (Continued)

Data element	Extracted data	Study example	Source and extraction method <sup>a</sup>	Reference
Measurement and interpretation of photo	Injuries/damage	Frequency of wing damage in migrating butterflies	Comparing images of butterfly wings with a reference image and estimating frequency/coverage of damage using Inkscape and MATLAB	Korkmaz et al. (2022)
<i>Biotic interactions and co-occurrence</i>				
Identification	Flower visitation/pollination	Identification of host flowers of the introduced wasp <i>Isodontia mexicana</i> with an iNaturalist-Pl@ntnet-workflow	iNaturalist; automatic data exchange tool approach validated by visual inspection	Pernat et al. (2022)
Identification and interpretation of photo	Species conflicts	Competition for nest cavities between bee species and wild honey bee colonies in its native and introduced ranges	iNaturalist and literature review; from images manual extraction of country, combs and entrance characteristics of nests	Saunders et al. (2021)
Identification and interpretation of photo	Diet	Analysing prey items in correlation with the age and sex of sparrowhawks by images from different sources	Google Images, Macaulay Library, iNaturalist, BirdGuides, Facebook, and Twitter by automatic or manual retrieval; extraction of sparrowhawk age, sex, and prey species by visual inspection	Panter and Amar (2021)
Identification and interpretation of photo	Host preferences	Host preferences and phenological peak of hairworms	iNaturalist and literature review; manual extraction by visual inspection of infected host species	Doherty et al. (2021)
Interpretation of photo	Disease	The proportion of <i>Caryophyllaceae</i> plants infected with anther-smut disease in eastern USA, comparing herbarium and citizen science data	iNaturalist and herbarium data; manual extraction and classification—infected or not—by visual inspection	Kido and Hood (2020)
Identification and interpretation of photo	Predation	Using images to investigate the spectrum of prey of assassin bugs (Hemiptera: Reduviidae)	Flickr, iSpot Nature, BugGuide, NatureWatch, Google Images; extraction of predation events and preyed species by visual inspection	Hernandez et al. (2019)
Measurement and interpretation of photo	Parasitism	Assessing parasitism and predation of lizards in an urban environment by measuring tail/body ratio and number of ticks	iNaturalist, HerpMapper; manual extraction of tick number by visual inspection, half-automated extraction of relative body size with measurement via ImageJ	Putman et al. (2021)
Identification and interpretation of photo	Basibionts	Occurrence and environmental conditions of epibiotic algae on gastropods detected from images of a corresponding citizen science project in Japan	iNaturalist, manual extraction of the proportion of infected fish species by visual inspection	Happel (2019)
Identification and interpretation of photo	Cause of death	Quantity and causes of shark stranding on a global scale	Dedicated citizen science project collecting images via Google Form, e-mail and Twitter; manual extraction of substrate and condition by visual inspection, coverage of epibiotic algae measured with ImageJ	Kagawa et al. (2020)
<i>Features of the population and community</i>				
Measurement and interpretation of photo	Mating system	Evolution of mating-related wing ornamentation of dragonflies influenced by climate warming	Literature, iNaturalist, YouTube, Twitter, Facebook, Instagram; manual extraction of species, sex and potential cause of stranding	Wosnick et al., 2022
Measurement and interpretation of photo	Mating system	Evolution of mating-related wing ornamentation of dragonflies influenced by climate warming	iNaturalist; half-automated extraction of wing pigmentation with ImageJ (following Drury et al., 2019)	Moore et al. (2021)

TABLE 1 (Continued)

Data element	Extracted data	Study example	Source and extraction method <sup>a</sup>	Reference
Quantification and identification	Population size	Decline of introduced llama population in Italy estimated by recognition of individuals on images	iNaturalist, Facebook, Twitter; manual extraction of coat pattern and age (by body size) by visual inspection	Gargioni et al. (2021)
Quantification and identification	Sex ratio, age groups	Sex ratio and proportion of life stages of manta rays	images of MantaMatcher (citizen science) combined with YouTube, Facebook, Instagram, Flickr and Vimeo and private collections; extraction of sex and life stages by visual inspection	Knochel et al. (2022)
<b>Environmental features</b>				
Interpretation of photo	Habitat type	Classification of marine biotic habitats from image background and evaluating results with professional reef life survey	iNaturalist; manual extraction and classification of biotic habitats by visual inspection	Bolt et al. (2022)
Interpretation of photo	Cause of death	Correlating weather and catastrophic events with images of dead birds	iNaturalist project dedicated to bird mortality	Yang et al. (2021)
Interpretation of photo and location	Substrate	Substrate choices of native oyster species <i>Ostrea lurida</i>	iNaturalist, fieldwork and literature; manual extraction of the substrate information by visual inspection of images in combination with observer notes and Google Earth	Kornbluth et al. (2022)

<sup>a</sup>Only the sources and methods used to obtain secondary data whose data type was the subject of the study (feature, interaction, etc.) are listed. In most cases, the analyses also used geo-locations, dates and metadata that were part of the primary observation.

### 3 | SECONDARY DATA ARE SLOWLY DIFFUSING INTO THE SCIENTIFIC LITERATURE

Studies using secondary data (Table 1) have mostly focused on the extraction of morphological information, such as the pigmentation on wings of Calopterygidae damselflies (Drury et al., 2019), coloration patterns of grass snakes (Fritz & Ihlow, 2022), and intra- and interspecific variabilities in coloration of birds and plants (Laitly et al., 2021). Some studies also used secondary data to assess human–nature interactions such as bat handling during the COVID crisis (Van der Jeucht et al., 2021), to classify marine habitats using image backgrounds (Bolt et al., 2022), and to identify plants visited by hummingbirds (Marín-Gómez et al., 2022). Secondary data from citizen science are often combined with iEcology or culturomics data sources (Jarić et al., 2020) or museum collections (Box 2). Examples include a dietary study of African snakes (Maritz & Maritz, 2020), arthropod parasitism by hairworms (Doherty et al., 2021) and the distribution of anther-smut disease in the *Caryophyllaceae* plant family (Kido & Hood, 2020).

Citizen science has already proven useful in mapping and tracking biological invasions (Encarnação et al., 2021). The additional information that comes with secondary data could reveal even more aspects of the invasion process, thereby supporting invasive species management. For example, first approaches explored the host plants of introduced pollinators (Bila Dubaïc et al., 2022; Guariento et al., 2019; Pernat et al., 2022) and cavity occupancy by wild honey bees (*Apis mellifera*) in Australia (Saunders et al., 2021).

Reanalysis of images has also been used for trait-based studies to characterise, for example particulate matter in the global oceans (Trudnowska et al., 2021) and the feeding habits of marine copepods

#### BOX 2 Secondary data, conservation culturomics and iEcology—What is the difference?

The use of secondary data in research is similar to the emerging areas of conservation culturomics and iEcology (Jarić et al., 2020). Culturomics seeks to understand human culture through the quantitative analysis of changes in word frequencies in large bodies of digital texts (Michel et al., 2011). In the context of biodiversity, the emergent area of 'conservation culturomics' focuses on the relationship between people and nature (Ladle et al., 2016), informed by contents of various types of online data. iEcology, on the other hand, is an umbrella term for analysing various types of digital data generated or collected for purposes other than ecological research to obtain insights into ecological questions. In contrast, in citizen science projects, people consciously contribute to the goal of a particular activity, such as biodiversity monitoring or invasive species detection (Marchante et al., 2023).

(Vilgrain et al., 2021), although these studies did not use citizen science data. The potential for extracting functional traits from images, either directly measured or inferred by combining visible features with context metrics from the metadata, has been thoroughly considered for plankton (Orenstein et al., 2022).

#### 4 | SECONDARY DATA EXTRACTION COULD BE ACHIEVED ALONG A GRADIENT OF HUMAN AND ARTIFICIAL INTELLIGENCE

As approaches to obtain secondary data are just emerging, such data are still mainly extracted manually. This can be challenging when thousands of images need to be interpreted and evaluated. For example, the aforementioned study of anther-smut infection within the *Caryophyllaceae* examined 79,801 iNaturalist images (Kido & Hood, 2020). There is much to be gained from automation that could scale up the process to millions of images, particularly for pre-selecting images and recognising relevant image features. For example, computer vision could be used to extract and analyse information on colour in images, for example, greenness of plants (Yuke, 2019), and deep learning models to detect, count and classify specific features of interest (Bjerger et al., 2023; Mann et al., 2022). Likewise, algorithms and pretrained dictionaries in Natural Language Processing could leverage the use of textual content, such as image captions, commentaries and tags in secondary data. Automated systems would also facilitate real-time analysis of biodiversity dynamics, making them particularly useful for informing decision-makers regarding effects of conservation efforts or as early warning tools (van Klink et al., 2022).

Despite the obvious appeal of machine learning for automatic data extraction from citizen science sources, several obstacles lie before its full potential can be realised. Developing robust models that effectively handle diverse and noisy datasets is challenging and resource-intensive. Nevertheless, for some tasks, existing tools may be customised or applied directly. Multiple trained deep learning models to screen multimedia for human or natural objects are freely available. For instance, object detection models, which are often pretrained and benchmarked on the COCO dataset (Lin et al., 2014) containing 80 different object categories (including birds and other animals), may already provide relevant secondary data output. Moreover, models exist for specific groups of organisms and data types: [Merlin Bird ID](#) and [BirdNET](#) (Kahl et al., 2021) for bird detection based on sound (the former can identify species also from images), [PI@ntNet API](#) for plants, Bjerger et al. (2023) created a test dataset for insects, [FishID](#) for fish species in images; [MegaDetector](#) or [TrapTagger](#) for animals in camera trap photos; and [BatDetect2](#) and [BatNet](#) for bats (Aodha et al., 2022; Krivek et al., 2023) in sound recordings. Additionally, customised models can be trained on open datasets, for example, [FathomNet](#) for marine organisms (Katija et al., 2022), [PI@ntNet](#) for plants and [iNaturalist](#) for a range of different species. Importantly, even with readily available models, manual resources and expertise are

required to ensure the anticipated model behaviour and performance on new data.

In other cases, models and analysis pipelines may need to be developed from scratch. Where models or training data are not available, the cost-benefit ratio of developing new artificial intelligence models should be weighed against the use of human-mediated approaches. For example, Mann et al. (2022) developed an approach to automatically detect flowering plants in images, which were then examined by citizen scientists for the rare presence of insects.

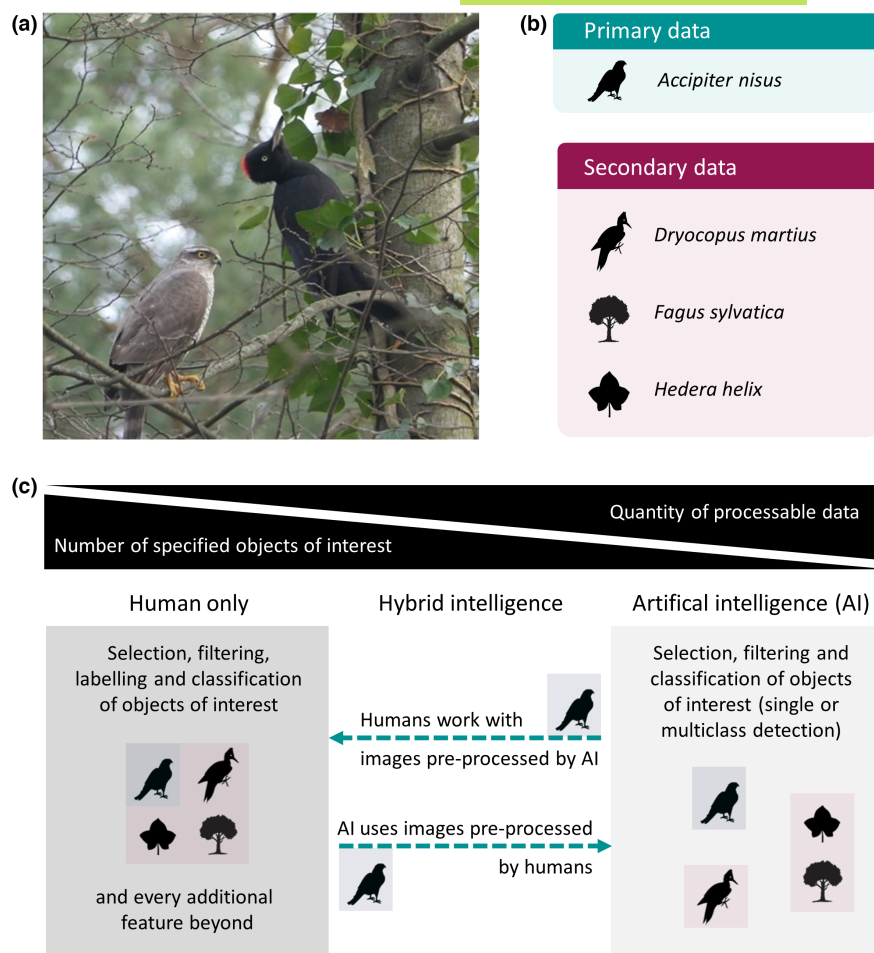
Efficient processing is relevant when dealing with large amounts of data, but it is critical to consider the resources needed. Developing custom automated methods and their broader usefulness and applicability versus setting up and maintaining manual processing pipelines (e.g. citizen science projects or recruiting and managing volunteers) may differ in terms of time, costs and personnel demands as well as the output quality. In any case, to address the uncertainty in exploratory analyses of secondary data variables, that is, to get an idea of what kind of additional information primary datasets contain, a subset of data would most often be analysed manually. This pre-processing can inform researchers about which methodologies to apply for larger scale extraction of information.

We expect that future secondary data extraction will be performed on a continuum between fully human and fully automated approaches with the respective advantages and disadvantages along this spectrum. Hybrid intelligence, that is, the combination of deep learning and human diligence (Mann, 2022; Rafner et al., 2021), can be effectively used to extract and analyse secondary data. Primary data can be filtered manually and, if necessary, annotated or immediately tested for relevant secondary data in the case of an existing algorithm. Conversely, one or more features can be selected from images (or other types of media) by an algorithm to be processed afterwards by a human (e.g. for annotation, validation or analysis; Figure 3).

Another challenge to apply artificial intelligence in secondary data studies is not knowing which data variables to look for and how to select or develop a potential identification algorithm or, simply put, how to search for the unknown unknowns. A human eye is able to identify the unexpected while the algorithm only recognises what is expected of it, that is, what it was trained for. In order to leverage the power of artificial intelligence for effective data extraction, data collection generally needs to be guided by precise research questions and must be based on a priori identification of variables of interest. As detection models trained to recognise an increasing number of objects, or segmentation models able to distinguish different areas in images (Kirillov et al., 2023) emerge, these may become increasingly relevant for exploration of secondary data without predetermined knowledge on what to look for. As artificial intelligence tools develop to analyse, visualise and synthesise multimodal data (audio, video, text, etc.), it is likely that they will be able to recognise multiple features of interest. Also, identifying patterns across media types, including insights overlooked by a researcher, and even suggesting mechanisms and hypotheses that could explain such patterns are within reach. Application of these models should



**FIGURE 3** Interaction and possible applications of methods to extract/retrieve secondary data. Our sample image of the primary observation of a raptor (a) also happened to include a black woodpecker and two plant species as secondary data (b). These exemplary secondary data could be extracted by humans or artificial intelligence only or by both in hybrid intelligence approaches (c). (image: © zilpzalp17, <https://www.inaturalist.org/observations/148103702>, <https://creativecommons.org/licenses/by-nc/4.0/>).



always be done with careful consideration of potential model biases that can skew results.

## 5 | WHY ARE WE NOT THERE YET?

Citizen science multimedia records are clearly more than meets the eye. To protect biodiversity, it is not only essential to inventory and monitor species, but also to understand the ecological networks they are part of. By giving many examples of current and possible future areas of application we demonstrated how secondary data offer the opportunity to extend and complement systematically collected interaction and monitoring data. Although we are convinced of the great potential in the untapped information, we still see some challenges to overcome and specific pitfalls to address.

Similar to the early days of the citizen science movement, the issue of bias can cast doubt on this new resource. Indeed, we suspect a similar bias in secondary data as in primary data (Isaac et al., 2014). Secondary data, however, would be less influenced by known recording behaviour (e.g. aesthetic preferences or charisma of observed target species) and more affected by previously less considered human actions. Staging of observed species in a particular location and environments, and cultural differences in what can be appropriately photographed are imaginable examples. In these

cases, scientists using secondary data can benefit from accelerated development and discussions in analysing opportunistic data to correct for bias (Johnston et al., 2022). Transparent handling of potential biases should be a given in both metadata and corresponding publications.

Of greater concern are biases from data generated by citizen science projects with unknown scientific goals. For example, when for a project citizens document a particular plant solely in forest habitat, higher-level analysis of that plant's habitats based on image backgrounds would document an unrepresentative proportion of this plant in forests. Therefore, the source of data should be known, that is, in unclear cases, the project organisers would also need to be contacted or the images excluded from the analysis. A thorough critique of data provenance may be particularly necessary if counterintuitive trends or patterns emerge during initial spatio-temporal visualisation or classification of data.

Major obstacles preventing the breakthrough of secondary data research concern methods for filtering and processing the vast amounts of data. Filtering primary data for the desired information is mainly a matter of metadata. More precise information provided by metadata for each image, text file, audio recording and other data types can speed up data acquisition and minimise errors. Citizen science platforms such as iNaturalist allow users to add more details about their records in corresponding observation fields or select

specific features from the list of annotations. However, free-text fields are arbitrary, and feature lists are confining, and may not include, for example, possible interactions between species.

Since improving metadata according to the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles is a worldwide effort, this problem will hopefully be solved with time. Similarly, the development of new and better machine-aided object recognition will allow large amounts of secondary data to be processed automatically in the future. In addition, approaches have been developed to not only recognise objects in multimedia data sources, but also to differentiate by anomalies or other species-specific features such as plant or animal colours (e.g. Hantak et al., 2022; Perez-Udell et al., 2023). Ultimately, the ever-improving models used to generate ecological networks also help to turn information into knowledge.

A more pressing issue is the legality and ethical defensibility of using millions of secondary data sources for purposes other than those intended when the primary observation was recorded and posted. Considerations of ethical and privacy issues are not exclusive to secondary data. They are also pertinent regarding the primary data used in iEcology and culturomics (Jarić et al., 2020), from which secondary data can be derived. While much of primary data (e.g. online texts, images, videos or audio recordings) are publicly available, and in many cases, people have given consent to their availability (e.g. by registering in citizen science or social media platforms), researchers are required to pay careful consideration to how they collect, use and share these data (Di Minin et al., 2021; Thompson et al., 2021; Zimmer, 2010).

Ethics are particularly relevant when dealing with online data from social media, where work is often used or distributed without the owners' consent. In fact, most social media platforms allow posting nearly any content, as they are not able to automatically identify copyrighted material. This issue is less prominent on citizen science platforms, as the users have stronger control of posted data and media licensing. But the way such information is shared and scraped still opens various possibilities of copyright infringement in the digital space. As such, there is a considerable uncertainty regarding situations in which acquiring permission and crediting authorship becomes mandatory. If not Creative Commons, checking licences can become a time-consuming process. Especially, when dealing with big data that are derived from multiple sources, it may be highly unfeasible to directly contact the media owners to get permission for use (Leighton et al., 2016).

Ethical issues have to be carefully considered and are especially delicate when secondary data allow recognition of people, or allow the identification of contentious human interactions, such as illegal fisheries (Sbragaglia et al., 2021), poaching or trade in wild organisms (Di Minin et al., 2019; Zimmer, 2010). Di Minin et al. (2021) have suggested a set of guidelines that can help address ethical concerns in research when using such data. Likewise, while publicly sharing species location information is useful for research, disclosing the location and identification of rare or threatened species can become a threat to their conservation (Lindenmayer & Scheele, 2017). Although citizen science platforms such as iNaturalist already

consider 'taxon geoprivacy' as a way to safeguard the locations of species 'at risk', a sensitive species as secondary data would still come with full coordinates if not recognised as such.

Finally, it is most important that awareness of the existence and potential of secondary data grows among scientists. With our contribution, we aim to open the eyes of the scientific community to overcome 'biodiversity blindness' and acknowledge the wealth of information far beyond the location and date of a species observation in the millions of freely available multimedia files. Besides being blind to this treasure of data, studies and projects dedicated to the topic may also not be seen as such due to a lack of common terminology. Therefore, we would like to establish the term secondary data as proposed by Callaghan et al. (2021) or at least stimulate a discussion about terminology, so that a corresponding field of research can grow.

It should be clear to the community that this approach applies not only to data from citizen science, social media or webpages, but also to data collected by scientists in the field or in the laboratory. The multiple benefits demonstrated here should convince people to make the (raw) data available to the public according to the FAIR principle, be it via GBIF, GitHub or other openly accessible repositories. As with all new and innovative methods, a transition period will be necessary before this approach is fully integrated into the research toolkit. Again, we draw comparisons with iEcology, culturomics and citizen science in that secondary data are utilised in a complementary and supportive way to other data sources and verified with ground truthing. Efforts to explore and use secondary data, although still in their early stages, are already demonstrating many ways to enrich and inform biodiversity research.

## AUTHOR CONTRIBUTIONS

Nadja Pernat conceived the ideas and planned and facilitated a 3-day workshop in November 2022 that was attended by all authors; Nadja Pernat, Yuval Itescu, Jasmijn Hillaert, Cristina Preda and Marina Golivets researched the reviewed articles; Nadja Pernat, Jasmijn Hillaert and Susan Canavan created the visualisations; Nadja Pernat, Quentin Groom and David M. Richardson led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## AFFILIATIONS

<sup>1</sup>Institute of Landscape Ecology, University of Münster, Münster, Germany;

<sup>2</sup>Centre for Integrative Biodiversity Research and Applied Ecology (CIBRA), University of Münster, Münster, Germany; <sup>3</sup>Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic; <sup>4</sup>School of Natural Sciences, University of Galway, Galway, UK; <sup>5</sup>Helmholtz Centre for Environmental Research – UFZ, Halle, Germany; <sup>6</sup>Research Institute of Nature and Forest, Brussels, Belgium; <sup>7</sup>Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany; <sup>8</sup>Freie Universität Berlin, Berlin, Germany; <sup>9</sup>Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel; <sup>10</sup>CNRS, AgroParisTech, Ecologie Systématique Evolution, Université Paris-Saclay, Gif-sur-Yvette, France; <sup>11</sup>Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic; <sup>12</sup>Department of Ecoscience, Aarhus University, Aarhus, Denmark; <sup>13</sup>Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic; <sup>14</sup>Faculty

of Natural and Agricultural Sciences, Ovidius University of Constanta, Constanta, Romania; <sup>15</sup>Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa; <sup>16</sup>Centre for Environmental and Marine Studies and Department of Biology, University of Aveiro, Aveiro, Portugal; <sup>17</sup>CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Porto, Portugal; <sup>18</sup>BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal; <sup>19</sup>NBI, Natural Business Intelligence, Tec Labs 1.2.1, Campus da Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal and <sup>20</sup>Meise Botanic Garden, Brussels, Belgium

## ACKNOWLEDGEMENTS

This work originated from a workshop supported by the COST Action *Increasing understanding of alien species through citizen science* (COST Action CA17122). DMR acknowledges support from the Centre for Invasion Biology, Stellenbosch University, Mobility 2020 project no. CZ.02.2.69/0.0/0.0/18\_053/0017850 (Ministry of Education, Youth and Sports of the Czech Republic) and long-term research development project RVO 67985939 (Czech Academy of Sciences). ASV acknowledges support from the FCT–Portuguese Foundation for Science and Technology through the program Stimulus for Scientific Employment–Individual Support (contract reference 2020.01175.CEECIND/CP1601/CT0009), and project ClimateMedia–Understanding climate change phenomena and impacts from digital technology and social media (contract reference 2022.06965.PTDC). IJ and PP acknowledge support from the Czech Science Foundation (project no. 23-07278S). HT acknowledges FCT/MCTES support to CESAM (UIDP/50017/2020+UIDB/50017/2020+LA/P/0094/2020). NP acknowledges the Animal Ecology Lab (Sascha Buchholz, Hilke Hollens-Kuhr) and the University of Münster for their support in running the workshop and Emily Brunner for co-producing the illustrations. We thank Tom August for helpful comments on the manuscript as well as Wouter Koch and an anonymous reviewer for their valuable advice on improving the article. Open Access funding enabled and organized by Projekt DEAL.







## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

This work does not contain any original data.

## ORCID

Nadja Pernat  <https://orcid.org/0000-0003-2244-1002>  
 Susan Canavan  <https://orcid.org/0000-0002-7972-7928>  
 Marina Golivets  <https://orcid.org/0000-0003-1278-2663>  
 Jasmijn Hillaert  <https://orcid.org/0000-0002-3761-8846>  
 Yuval Itescu  <https://orcid.org/0000-0002-6301-7112>  
 Ivan Jarić  <https://orcid.org/0000-0002-2185-297X>  
 Hjalte M. R. Mann  <https://orcid.org/0000-0002-4768-4767>  
 Pavel Pipek  <https://orcid.org/0000-0003-1116-1013>  
 Cristina Preda  <https://orcid.org/0000-0002-3513-7344>  
 David M. Richardson  <https://orcid.org/0000-0001-9574-8297>  
 Heliana Teixeira  <https://orcid.org/0000-0001-8525-9967>

Ana Sofia Vaz  <https://orcid.org/0000-0001-6588-4650>

Quentin Groom  <https://orcid.org/0000-0002-0596-5376>

## REFERENCES

- Alexandrino, E. R., Silva, G. A., Corbo, M. C., Demuner, B. A., & Szabo, J. K. (2022). Urban southern house wren (*Troglodytes musculus*) nesting in apparently unsuitable human-made structures: Is it worth it? *Ornitologia Neotropical*, 33(1), 44–52. <https://doi.org/10.58843/ornneo.v33i1.879>
- Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial gaps in global biodiversity information and the role of citizen science. *BioScience*, 66, 393–400. <https://doi.org/10.1093/biosci/biw022>
- Aodha, O. M., Martínez Balvanera, S., Damstra, E., Cooke, M., Eichinski, P., Browning, E., Barataud, M., Boughey, K., Coles, R., Giacomini, G., Swiney, G., Mac, M. C., Obrist, M. K., Parsons, S., Sattler, T., & Jones, K. E. (2022). Towards a general approach for bat echolocation detection and classification. *BioRxiv*, 2022.12.14.520490 <https://doi.org/10.1101/2022.12.14.520490>
- Barve, V. V., Brenskelle, L., Li, D., Stucky, B. J., Barve, N. V., Hantak, M. M., McLean, B. S., Paluh, D. J., Oswald, J. A., Belitz, M. W., Folk, R. A., & Guralnick, R. P. (2020). Methods for broad-scale plant phenology assessments using citizen scientists' photographs. *Applications in Plant Sciences*, 8(1), e11315. <https://doi.org/10.1002/aps3.11315>
- Besson, M., Alison, J., Bjerger, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, 25(12), 2753–2775. <https://doi.org/10.1111/ele.14123>
- Bila Dubaić, J., Lanner, J., Rohrbach, C., Meimberg, H., Wyatt, F., Čačija, M., Galešić, M., Ješovnik, A., Samurović, K., Plečaš, M., Raičević, J., & Četković, A. (2022). Towards a real-time tracking of an expanding alien bee species in Southeast Europe through citizen science and floral host monitoring. *Environmental Research Communications*, 4, 085001. <https://doi.org/10.1088/2515-7620/ac8398>
- Bjerger, K., Alison, J., Dyrmann, M., Frigaard, C. E., Mann, H. M. R., & Høye, T. T. (2023). Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, 2(3), e0000051. <https://doi.org/10.1371/journal.pstr.0000051>
- Blais, B. R., & Shaw, C. J. (2018). In the heat of the night: Assessing nocturnal activity of the desert iguana (*Dipsosaurus dorsalis*). *Sonoran Herpetologist*, 31(4), 65–70.
- Bolt, M., Callaghan, C., Poore, A., Vergés, A., & Roberts, C. (2022). Using the background of fish photographs to quantify habitat composition in marine ecosystems. *Marine Ecology Progress Series*, 688, 167–172. <https://doi.org/10.3354/meps14027>
- Callaghan, C. T., Poore, A. G. B., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J. L., Vergés, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, 71, 55–63. <https://doi.org/10.1093/biosci/biaa131>
- Callaghan, C. T., Roberts, J. D., Poore, A. G. B., Alford, R. A., Cogger, H., & Rowley, J. J. L. (2020). Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodiversity and Conservation*, 29, 1323–1337. <https://doi.org/10.1007/s10531-020-01937-3>
- Catron, K. A., Hennen, D. A., Wagner, J. F., Brown, B. L., Weber, D. C., & Kuhar, T. P. (2023). Analysis of host plant availability and use by 2 species of soldier beetles (Coleoptera: Cantharidae) using community-gathered digital specimens. *Annals of the Entomological Society of America*, 116, 305–313. <https://doi.org/10.1093/aesa/saad016>
- Chandler, M., See, L., Copas, K., Bonde, A. M. Z., 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. <https://doi.org/10.1016/j.biocon.2016.09.004>
- Cheney, E. S. (2023). Detecting and monitoring wildlife parasites: Determining the current extent and future impact of the winter tick (*Dermacentor albipictus*) in the Yukon and beyond (order no. 30247039). Available from ProQuest Dissertations & Theses Global (2832671224). Retrieved from <https://www.proquest.com/dissertations-theses/detecting-monitoring-wildlife-parasites/docview/2832671224/se-2>
- Cosentino, B. J., & Gibbs, J. P. (2022). Parallel evolution of urban–rural clines in melanism in a widespread mammal. *Scientific Reports*, 12, 1752. <https://doi.org/10.1038/s41598-022-05746-2>
- 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. <https://doi.org/10.1111/cobi.13708>
- Di Minin, E., Fink, C., Hiiipala, T., & Tenkanen, H. (2019). A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation Biology*, 33, 210–213. <https://doi.org/10.1111/cobi.13104>
- Díaz, S. M., Settele, J., Brondizio, E., Ngo, H., Guèze, M., Agard, J., Arneeth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., ... Zayas, C. N. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. [https://ipbes.net/sites/default/files/inline/files/ipbes\\_global\\_assessment\\_report\\_summary\\_for\\_policymakers.pdf](https://ipbes.net/sites/default/files/inline/files/ipbes_global_assessment_report_summary_for_policymakers.pdf)
- Doherty, J.-F., Fillion, A., Bennett, J., Raj Bhattarai, U., Chai, X., de Angeli Dutra, D., Donlon, E., Jorge, F., Milotic, M., Park, E., Sabadel, A. J. M., Thomas, L. J., & Poulin, R. (2021). The people vs science: Can passively crowdsourced internet data shed light on host–parasite interactions? *Parasitology*, 148, 1313–1319. <https://doi.org/10.1017/S0031182021000962>
- Drake, D., Dubay, S., & Allen, M. L. (2021). Evaluating human–coyote encounters in an urban landscape using citizen science. *Journal of Urban Ecology*, 7, juaa032. <https://doi.org/10.1093/jue/juaa032>
- Drury, J. P., Barnes, M., Finneran, A. E., Harris, M., & Grether, G. F. (2019). Continent-scale phenotype mapping using citizen scientists' photographs. *Ecography*, 42, 1436–1445. <https://doi.org/10.1111/ecog.04469>
- Encarnação, J., Teodósio, M. A., & Morais, P. (2021). Citizen science and biological invasions: A review. *Frontiers in Environmental Science*, 8, 602980. <https://doi.org/10.3389/fenvs.2020.602980>
- Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS One*, 16, e0234587. <https://doi.org/10.1371/journal.pone.0234587>
- Fontúrbel, F. E., Sepúlveda, I. B., Muschett, G., Carvallo, G. O., Vieli, L., & Murúa, M. M. (2023). Do exotic plants and flower colour facilitate bumblebee invasion? Insights from citizen science data. *Flora*, 298, 152200. <https://doi.org/10.1016/j.flora.2022.152200>
- Fritz, U., & Ihlow, F. (2022). Citizen science, taxonomy and grass snakes: iNaturalist helps to clarify variation of coloration and pattern in *Natrix natrix* subspecies. *Vertebrate Zoology*, 72, 533–549. <https://doi.org/10.3897/vz.72.e87426>
- Gargioni, C., Monaco, A., Francesco Ficetola, G., Lazzeri, L., & Mori, E. (2021). From the Andes to the Apennines: Rise and fall of a free-ranging population of feral llamas. *Animals*, 11(3), 857. <https://doi.org/10.3390/ani11030857>
- Groom, Q., Pernat, N., Adriaens, T., de Groot, M., Jelaska, S. D., Marčiulynienė, D., Martinou, A. F., Skuhrovec, J., Tricarico, E., Wit, E. C., & Roy, H. E. (2021). Species interactions: Next-level citizen science. *Ecography*, 44, 1781–1789. <https://doi.org/10.1111/ecog.05790>
- Guariento, E., Lanner, J., Staggl, M. A., & Kranebitter, P. (2019). *Megachile sculpturalis* (Smith, 1853) (Hymenoptera: Megachilidae), the giant resin bee new to South Tyrol with a newly described plant species interaction. *Gredleriana*, 19, 209–215. <https://doi.org/10.5281/ZENODO.3565365>
- Hantak, M. M., Guralnick, R. P., Zare, A., & Stucky, B. J. (2022). Computer vision for assessing species color pattern variation from web-based community science images. *iScience*, 25(8), 104784. <https://doi.org/10.1016/j.isci.2022.104784>
- Happel, A. (2019). A volunteer-populated online database provides evidence for a geographic pattern in symptoms of black spot infections. *International Journal for Parasitology: Parasites and Wildlife*, 10, 156–163. <https://doi.org/10.1016/j.ijppaw.2019.08.003>
- Hernandez, M., Masonick, P., & Weirauch, C. (2019). Crowdsourced online images provide insights into predator–prey interactions of putative natural enemies. *Food Webs*, 21, e00126. <https://doi.org/10.1016/j.fooweb.2019.e00126>
- Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., André, H., Linnell, J. D. C., & Croomsigt, J. P. G. M. (2020). Using by-catch data from wildlife surveys to quantify climatic parameters and timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, 6, 129–140. <https://doi.org/10.1002/rse2.136>
- Isaac, N. J. B., Strien, A. J., August, T. A., Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5, 1052–1060. <https://doi.org/10.1111/2041-210X.12254>
- 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., Ladle, R., Soriano-Redondo, A., Souza, A. T., & Roll, U. (2020). iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, 35, 630–639. <https://doi.org/10.1016/j.tree.2020.03.003>
- Johnston, A., Matechou, E., & Dennis, E. B. (2022). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, 14, 103–116. <https://doi.org/10.1111/2041-210X.13834>
- Jordano, P. (2016). Chasing Ecological Interactions. *PLoS Biology*, 14, e1002559. <https://doi.org/10.1371/journal.pbio.1002559>
- Kagawa, O., Uchida, S., Yamazaki, D., Osawa, Y., & Ito, S. (2020). Citizen science via social media revealed conditions of symbiosis between a marine gastropod and an epibiotic alga. *Scientific Reports*, 10(1), 19647. <https://doi.org/10.1038/s41598-020-74946-5>
- Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236. <https://doi.org/10.1016/j.ecoinf.2021.101236>
- Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B., & Bell, K. L. C. (2022). FathomNet: A global image database for enabling artificial intelligence in the ocean. *Scientific Reports*, 12(1), 15914. <https://doi.org/10.1038/s41598-022-19939-2>
- Kido, A., & Hood, M. E. (2020). Mining new sources of natural history observations for disease interactions. *American Journal of Botany*, 107, 3–11. <https://doi.org/10.1002/ajb2.1409>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. *arXiv*, 2304.02643 <https://doi.org/10.48550/ARXIV.2304.02643>
- Knochel, A. M., Cochran, J. E. M., Kattan, A., Stevens, G. M. W., Bojanowski, E., & Berumen, M. L. (2022). Crowdsourced data reveal multinational connectivity, population demographics, and possible nursery ground of endangered oceanic manta rays in the Red Sea. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 32(11), 1774–1786. <https://doi.org/10.1002/aqc.3883>



- Korkmaz, R., Rajabi, H., Eshghi, S., Gorb, S. N., & Büscher, T. H. (2022). The frequency of wing damage in a migrating butterfly. *Insect Science*, 30, 1507–1517. <https://doi.org/10.1111/1744-7917.13153>
- Kornbluth, A., Perog, B. D., Crippen, S., Zacherl, D., Quintana, B., Grosholz, E. D., & Wasson, K. (2022). Mapping oysters on the Pacific coast of North America: A coast-wide collaboration to inform enhanced conservation. *PLoS One*, 17(3), e0263998. <https://doi.org/10.1371/journal.pone.0263998>
- Krivek, G., Gillert, A., Harder, M., Fritze, M., Frankowski, K., Timm, L., Meyer-Olbersleben, L., von Lukas, U. F., Kerth, G., & van Schaik, J. (2023). BatNet: A deep learning-based tool for automated bat species identification from camera trap images. *Remote Sensing in Ecology and Conservation*. <https://doi.org/10.1002/rse2.339>
- Kühl, H. S., Bowler, D. E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., Fernández, N., Guerra, C. A., Henle, K., Herbing, 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. <https://doi.org/10.1016/j.oneear.2020.09.010>
- Ladle, R. J., Correia, R. A., Do, Y., Joo, G.-J., Malhado, A. C. M., Proulx, R., Roberge, J. M., & Jepson, P. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment*, 14, 269–275. <https://doi.org/10.1002/fee.1260>
- Laitly, A., Callaghan, C. T., Delhey, K., & Cornwell, W. K. (2021). Is color data from citizen science photographs reliable for biodiversity research? *Ecology and Evolution*, 11, 4071–4083. <https://doi.org/10.1002/ece3.7307>
- Leighton, G. R. M., Hugo, P. S., Roulin, A., & Amar, A. (2016). Just Google it: Assessing the use of Google images to describe geographical variation in visible traits of organisms. *Methods in Ecology and Evolution*, 7, 1060–1070. <https://doi.org/10.1111/2041-210X.12562>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common Objects in Context. *arXiv*, 1405.0312 <https://doi.org/10.48550/ARXIV.1405.0312>
- Lindenmayer, D., & Scheele, B. (2017). Do not publish. *Science*, 356(6340), 800–801. <https://doi.org/10.1126/science.aan1362>
- Mann, H. (2022). *Automatic monitoring of plant-pollinator interactions with computer vision and deep learning* (PhD thesis). Aarhus University. [https://pure.au.dk/portal/files/272488583/HjalteMRMann\\_PhD\\_Thesis\\_20220620.pdf](https://pure.au.dk/portal/files/272488583/HjalteMRMann_PhD_Thesis_20220620.pdf)
- Mann, H. M. R., Iosifidis, A., Jepsen, J. U., Welker, J. M., Loonen, M. J. J. E., & Høye, T. T. (2022). Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning. *Remote Sensing in Ecology and Conservation*, 8, 765–777. <https://doi.org/10.1002/rse2.275>
- Marchante, H., Jubase-Tshali, N., Pauchard, A., Richardson, D. M., & Marchante, E. (2023). *Public participation in the management of invasive Australian Acacia species: Contributions from citizen science, volunteer groups, and public awareness raising*. Richardson, D. M., Marchante, E. & Le Roux, J. J. (eds). *Wattles, Australian Acacias Around the World*, CABI (in press).
- Marín-Gómez, O. H., Rodríguez Flores, C., & Arizmendi, M. d. C. (2022). Assessing ecological interactions in urban areas using citizen science data: Insights from hummingbird-plant meta-networks in a tropical megacity. *Urban Forestry & Urban Greening*, 74, 127658. <https://doi.org/10.1016/j.ufug.2022.127658>
- Maritz, R. A., & Maritz, B. (2020). Sharing for science: High-resolution trophic interactions revealed rapidly by social media. *PeerJ*, 8, e9485. <https://doi.org/10.7717/peerj.9485>
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. B. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>
- Moore, M. P., Hersch, K., Sricharoen, C., Lee, S., Reice, C., Rice, P., Kronick, S., Medley, K. A., & Fowler-Finn, K. D. (2021). Sex-specific ornament evolution is a consistent feature of climatic adaptation across space and time in dragonflies. *Proceedings of the National Academy of Sciences of the United States of America*, 118(28), e2101458118. <https://doi.org/10.1073/pnas.2101458118>
- Moscoe, L. J., & Hanes, M. M. (2019). Taste of life: Science outreach made delicious. *Plants, People, Planet*, 1, 183–187. <https://doi.org/10.1002/ppp3.42>
- Newson, S. E., Moran, N. J., Musgrove, A. J., Pearce-Higgins, J. W., Gillings, S., Atkinson, P. W., Miller, R., Grantham, M. J., & Baillie, S. R. (2016). Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes. *Ibis*, 158, 481–495. <https://doi.org/10.1111/ibi.12367>
- Orenstein, E. C., Ayata, S.-D., Maps, F., Becker, É. C., Benedetti, F., Biard, T., de Garidel-Thoron, T., Ellen, J. S., Ferrario, F., Giering, S. L. C., Guy-Haim, T., Hoebeke, L., Iversen, M. H., Kjørboe, T., Lalonde, J.-F., Lana, A., Laviale, M., Lombard, F., Lorimer, T., ... Irissou, J.-O. (2022). Machine learning techniques to characterize functional traits of plankton from image data. *Limnology and Oceanography*, 67, 1647–1669. <https://doi.org/10.1002/lno.12101>
- Panther, C. T., & Amar, A. (2021). Sex and age differences in the diet of the Eurasian Sparrowhawk (*Accipiter nisus*) using web-sourced photographs: Exploring the feasibility of a new citizen science approach. *Ibis*, 163(3), 928–947. <https://doi.org/10.1111/ibi.12918>
- Parra-Tabla, V., & Arceo-Gómez, G. (2021). Impacts of plant invasions in native plant-pollinator networks. *New Phytologist*, 230, 2117–2128. <https://doi.org/10.1111/nph.17339>
- Parrish, J. K., Litle, K., Dolliver, J., Hass, T., Burgess, H. K., Frost, E., Wright, C. W., & Jones, T. (2017). Defining the baseline and tracking change in seabird populations: The Coastal Observation and Seabird Survey Team (COASST). In J. A. Cigliano & H. L. Ballard (Eds.), *Citizen science for coastal and marine conservation* (pp. 19–38). Routledge.
- Pélissier, M., Nabholz, B., Labadessa, R., & Ancillotto, L. (2023). Hidden in plain sight: Unveiling the distributions of green-winged grasshoppers (*Aiolopus* spp.) with citizen-science data. *Journal of Zoology*, 320, 301–307. <https://doi.org/10.1111/jzo.13086>
- Perez-Udell, R. A., Udell, A. T., & Chang, S. M. (2023). An automated pipeline for supervised classification of petal color from citizen science photographs. *Applied Plant Science*, 11(1), e11505. <https://doi.org/10.1002/aps3.11505>
- Pernat, N., August, T., Groom, Q. J., Memedemin, D., & Reyserhove, L. (2022). An iNaturalist-PI@ntNet-workflow to identify plant-pollinator interactions—A case study of *Isodontia mexicana*. *BioHackrXiv* <https://doi.org/10.37044/osf.io/em3rk>
- Pocock, M. J. O., & Evans, D. M. (2014). The success of the horse-chestnut leaf-miner, *Cameraria ohridella*, in the UK revealed with hypothesis-led citizen science. *PLoS One*, 9, e86226. <https://doi.org/10.1371/journal.pone.0086226>
- Probert, A. F., Wegmann, D., Volery, L., Adriaens, T., Bakui, R., Bertolino, S., Essl, F., Gervasini, E., Groom, Q., Latombe, G., Marisavljevic, D., Mumford, J., Pergl, J., Preda, C., Roy, H. E., Scalera, R., Teixeira, H., Tricarico, E., Vanderhoeven, S., & Bacher, S. (2022). Identifying, reducing, and communicating uncertainty in community science: A focus on alien species. *Biological Invasions*, 24, 3395–3421. <https://doi.org/10.1007/s10530-022-02858-8>
- Putman, B. J., Williams, R., Li, E., & Pauly, G. B. (2021). The power of community science to quantify ecological interactions in cities. *Scientific Reports*, 11, 3069. <https://doi.org/10.1038/s41598-021-82491-y>
- Rafner, J., Gajdacz, M., Kragh, G., Hjorth, A., Gander, A., Palfi, B., Berditchevskaia, A., Grey, F., Gal, K., Segal, A., Walmsley, M., Miller, J. A., Dellerman, D., Haklay, M., Michelucci, P., & Sherson, J. (2021). Revisiting citizen science through the lens of hybrid intelligence. *arXiv*, 2104.14961 <https://doi.org/10.48550/arXiv.2104.14961>

- Rech, F., Narimanov, N., Bauer, T., & Schirmel, J. (2022). Urbanization increases fluctuating asymmetry and affects behavioral traits of a common grasshopper. *Ecology and Evolution*, 12, e9658. <https://doi.org/10.1002/ece3.9658>
- Reeb, R. A., Aziz, N., Lapp, S. M., Kitzes, J., Heberling, J. M., & Kuebbing, S. E. (2022). Using convolutional neural networks to efficiently extract immense phenological data from community science images. *Frontiers in Plant Science*, 12, 787407. <https://doi.org/10.3389/fpls.2021.787407>
- Rhodes, C., Haunfelder, W., & Carlson, B. E. (2022). Citizen science reporting indicates geographic and phenotypic drivers of road use and mortality in a threatened rattlesnake. *Current Zoology*, 69, 264–276. <https://doi.org/10.1093/cz/zoac050>
- Rosa, R. M., Cavallari, D. C., & Salvador, R. B. (2022). iNaturalist as a tool in the study of tropical molluscs. *PLoS One*, 17, e0268048. <https://doi.org/10.1371/journal.pone.0268048>
- Saunders, M. E., Goodwin, E. K., Santos, K. C., Sonter, C. A., & Rader, R. (2021). Cavity occupancy by wild honey bees: Need for evidence of ecological impacts. *Frontiers in Ecology and the Environment*, 19, 349–354. <https://doi.org/10.1002/fee.2347>
- Sbragaglia, V., Correia, R. A., & Di Minin, E. (2021). Responsible use of social media data is needed: A reply to Maya-Jariego et al. "Plenty of black money: Netnography of illegal recreational underwater fishing in southern Spain". *Marine Policy*, 134, 104780. <https://doi.org/10.1016/j.marpol.2021.104780>
- Spence Beaulieu, M. R., Hopperstad, K., Dunn, R. R., & Reiskind, M. H. (2019). Simplification of vector communities during suburban succession. *PLoS One*, 14, e0215485. <https://doi.org/10.1371/journal.pone.0215485>
- Speziale, K. L., Lambertucci, S. A., Gleiser, G., Tella, J. L., Hiraldo, F., & Aizen, M. A. (2018). An overlooked plant–parakeet mutualism counteracts human overharvesting on an endangered tree. *Royal Society Open Science*, 5, 171456. <https://doi.org/10.1098/rsos.171456>
- Steven, R., Van Helden, B. E., Tulloch, A. I., Barnes, M., Close, P. G., & Fuller, R. A. (2021). Exploring the ability of urban householders to correctly identify nocturnal mammals. *Urban Ecosystems*, 24, 1359–1369. <https://doi.org/10.1007/s11252-021-01118-2>
- Thompson, R. M., Hall, J., Morrison, C., Palmer, N. R., & Roberts, D. L. (2021). Ethics and governance for internet-based conservation science research. *Conservation Biology*, 35(6), 1747–1754. <https://doi.org/10.1111/cobi.13778>
- Trudnowska, E., Lacour, L., Ardyna, M., Rogge, A., Irisson, J. O., Waite, A. M., Babin, M., & Stemmann, L. (2021). Marine snow morphology illuminates the evolution of phytoplankton blooms and determines their subsequent vertical export. *Nature Communications*, 12, 2816. <https://doi.org/10.1038/s41467-021-22994-4>
- Turella, I. Z., Da Silva, T. L., Rumpel, L., & Marini, M. Â. (2022). Breeding biology of swallow-tailed hummingbird (*Eupetomena macroura*) based on citizen science data. *Ornithology Research*, 30(3), 181–189. <https://doi.org/10.1007/s43388-022-00098-x>
- Unger, S. (2022). On the road again: Touring iNaturalist for roadkill observations as a new tool for ecologists. *Journal of Wildlife and Biodiversity*, 6, 72–86.
- Van der Jeucht, L., Groom, Q., Agosti, D., Phelps, K., Reeder, D., & Simmons, N. B. (2021). Using iNaturalist to monitor adherence to best practices in bat handling. *Biodiversity Data Journal*, 9, e68052. <https://doi.org/10.3897/BDJ.9.e68052>
- van Gils, J. A., Lisovski, S., Lok, T., Meissner, W., Ożarowska, A., De Fouw, J., Rakhimberdiev, E., Soloviev, M. Y., Piersma, T., & Klaassen, M. (2016). Body shrinkage due to Arctic warming reduces red knot fitness in tropical wintering range. *Science*, 352(6287), 819–821. <https://doi.org/10.1126/science.aad6351>
- van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F., Høye, T. T., Jongejans, E., Menz, M. H. M., Miraldo, A., Roslin, T., Roy, H. E., Ruczyński, I., Schigel, D., Schäffler, L., Sheard, J. K., Svenningsen, C., Tschan, G. F., Wäldchen, J., ... Bowler, D. E. (2022). Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology & Evolution*, 37, 872–885. <https://doi.org/10.1016/j.tree.2022.06.001>
- Vilgrain, L., Maps, F., Picheral, M., Babin, M., Aubry, C., Irisson, J., ... Ayata, S.-D. (2021). Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. *Limnology and Oceanography*, 66, 1155–1167. <https://doi.org/10.1002/lno.11672>
- Warren, J. (2021). *A comparison of statistical techniques for evaluating body condition in New Zealand leopard seals (Hydrurga leptonyx) using citizen science data*. (MSc Thesis, 139 pp.). University of York. <https://etheses.whiterose.ac.uk/28944/>
- Wetzel, F. T., Bingham, H. C., Groom, Q., Haase, P., Köljal, U., Kuhlmann, M., Martin, C. S., Penev, L., Robertson, T., Saarenmaa, H., Schmeller, D. S., Stoll, S., Tonkin, J. D., & Häuser, C. L. (2018). Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. *Biological Conservation*, 221, 78–85. <https://doi.org/10.1016/j.biocon.2017.12.024>
- Wosnick, N., Leite, R. D., Giareta, E. P., Morick, D., & Musyl, M. (2022). Global assessment of shark strandings. *Fish and Fisheries*, 23(4), 786–799. <https://doi.org/10.1111/faf.12648>
- Yang, D., Yang, A., Yang, J., Xu, R., & Qiu, H. (2021). Unprecedented migratory bird die-off: A citizen-based analysis on the spatiotemporal patterns of mass mortality events in the western United States. *GeoHealth*, 5(4), e2021GH000395. <https://doi.org/10.1029/2021GH000395>
- Yuke, Z. (2019). Greenness index from phenocams performs well in linking climatic factors and monitoring grass phenology in a temperate prairie ecosystem. *Journal of Resources and Ecology*, 10(5), 481–493. <https://doi.org/10.5814/j.issn.1674-764x.2019.05.003>
- Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12, 313–325. <https://doi.org/10.1007/s10676-010-9227-5>

**How to cite this article:** Pernat, N., Canavan, S., Golivets, M., Hillaert, J., Itescu, Y., Jarić, I., Mann, H. M. R., Pipek, P., Preda, C., Richardson, D. M., Teixeira, H., Vaz, A. S., & Groom, Q. (2024). Overcoming biodiversity blindness: Secondary data in primary citizen science observations. *Ecological Solutions and Evidence*, 5, e12295. <https://doi.org/10.1002/2688-8319.12295>