



The Verification of Ecological Citizen Science Data: Current Approaches and Future Possibilities

EMILY BAKER

JONATHAN P. DRURY

JOHANNA JUDGE

DAVID B. ROY

GRAHAM C. SMITH

PHILIP A. STEPHENS

**Author affiliations can be found in the back matter of this article*

REVIEW AND
SYNTHESIS PAPER

][ubiquity press

ABSTRACT

Citizen science schemes enable ecological data collection over very large spatial and temporal scales, producing datasets of high value for both pure and applied research. However, the accuracy of citizen science data is often questioned, owing to issues surrounding data quality and verification, the process by which records are checked after submission for correctness. Verification is a critical process for ensuring data quality and for increasing trust in such datasets, but verification approaches vary considerably between schemes. Here, we systematically review approaches to verification across ecological citizen science schemes that feature in published research, aiming to identify the options available for verification, and to examine factors that influence the approaches used. We reviewed 259 schemes and were able to locate verification information for 142 of those. Expert verification was most widely used, especially among longer-running schemes, followed by community consensus and automated approaches. Expert verification has been the default approach for schemes in the past, but as the volume of data collected through citizen science schemes grows and the potential of automated approaches develops, many schemes might be able to implement approaches that verify data more efficiently. We present an idealised system for data verification, identifying schemes where this system could be applied and the requirements for implementation. We propose a hierarchical approach in which the bulk of records are verified by automation or community consensus, and any flagged records can then undergo additional levels of verification by experts.

CORRESPONDING AUTHOR:

Emily Baker

Conservation Ecology Group,
Department of Biosciences,
Durham University, GB

emily.baker@durham.ac.uk

KEYWORDS:

citizen science; ecology;
big data; verification;
crowdsourcing; data quality

TO CITE THIS ARTICLE:

Baker, E, Drury, JP, Judge, J, Roy, DB, Smith, GC and Stephens, PA. 2021. Maintaining User Engagement in an Infectious Disease Surveillance-Related Citizen Science Project. *Citizen Science: Theory and Practice*, 6(1): 12, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.351>

INTRODUCTION

In the current polarised political and media environment (Iyengar and Massey 2019), with public access to a vast choice of information sources (Huber et al. 2019), there is an increasing need for effective public engagement and science communication. There is, therefore, an argument for the democratisation of science, to make information accessible to everyone, to engage the public in scientific issues, and to involve them in scientific research endeavours (Mason and Garbarino 2016; Salomon et al. 2018). Democratizing science in ecology and conservation has the potential to increase understanding of environmental issues and scientific research methods, catalysing bottom-up action, greater environmental stewardship, and ecological conservation. Furthermore, scientists can involve the public in the research process through gaining insight into local knowledge and value systems, and through volunteer contributions to data collection and interpretation (Kimura and Kinchy 2016). Involving the public in research can be a highly effective means of public engagement and science communication, as it involves sustained, longer-term engagements. Also, there is often a two-way dialogue in which both the public and researchers can provide input and feedback, consulting and collaborating on the research (Pace et al. 2010; NCCPE 2016). One way that public engagement is increasingly embedded in ecological research is through data collection by members of the public. For ecology and conservation, specifically, the public can contribute to species monitoring and biological recording, documenting species' occurrences to track species' distribution, abundance, and/or phenology (Sutherland, Roy, and Amano 2015).

Volunteers play a key role in biological recording and have been contributing to ecological datasets for centuries (Silvertown 2009; Dickinson, Zuckerman, and Bonter 2010; Miller-Rushing, Primack, and Bonney 2012; Pocock et al. 2015). This process falls under the overarching term citizen science which broadly encompasses any volunteer involvement in science (Roy et al. 2012). The term was coined in the 1990s as a strategy for improving public trust and understanding in science (Woolley et al. 2016). More recently, the term has been adopted to describe a range of initiatives and research endeavours across disciplines (Woolley et al. 2016), with citizen science now featuring more in published literature (Kullenberg and Kasperowski 2016). Within the field of ecology, in addition to biological recording, citizen science schemes can also include tasks such as identifying species from photographic records or digitising data associated with specimen collections (Roy et al. 2012).

Citizen science recording schemes have collected some of the longest-running time-series datasets of species

populations (Devictor, Whittaker, and Beltrame 2010). Such datasets play a key role in assessments of species' changes in relation to pervasive anthropogenic pressures such as climate change, pollution, invasive species, and urbanisation (Sutherland, Roy, and Amano 2015). Biological recording benefits from contributions by volunteers because those contributions increase the geographical range and temporal span over which species can be recorded, providing long-term species-distribution datasets that can be used to assess and compare ecological trends (Pocock et al. 2015). These recording schemes typically rely on ad hoc, opportunistic records, although there are examples of hypothesis-led citizen science schemes, as well as schemes that have set up standardized monitoring protocols (Sewell, Beebe, and Griffiths 2010; Flower, Jones, and Bernede 2016).

Data quality is a concern with citizen science data, as generally unstructured sampling protocols can introduce bias and noise (Isaac et al. 2014; Pescott et al. 2015; Kamp et al. 2016). This can present challenges when analysing citizen science datasets and can limit the scientific questions that can be addressed (Isaac et al. 2014). The accuracy of citizen science data has also been questioned, owing to issues surrounding validation and verification (Kosmala et al. 2016a). Validation is a process through which records are checked to ensure the data have been submitted correctly. Verification is the process of checking records for correctness; within ecological citizen science schemes, this generally means confirming species identity (Tweddle et al. 2012). Verification is a critical process for ensuring data quality of, and trust in, citizen science datasets (Gilfedder et al. 2019), enabling those datasets to be used in environmental research, management, and policy development (Tweddle et al. 2012).

In this review, we explore the different approaches that published citizen science schemes use to verify their data, the breadth of information they use to verify each record, and the citizen science scheme attributes that may influence choice of verification approach. Our aims are to identify the options available for verification of citizen science data and to examine whether citizen science schemes are using the most suitable verification approach to maximise confidence in, and validity of, the data, whilst also ensuring efficient verification of records.

SYSTEMATIC REVIEW METHOD LITERATURE SEARCH

To survey the verification approaches across existing citizen science schemes, we conducted this review based on the systematic review protocol developed by the Collaboration for Environmental Evidence (2018). The search terms we

used were replicated from a review of the diversity and evolution of citizen science programmes carried out by Pocock et al. (2017). These terms were “citizen science,” “take part AND (nature OR environment),” “volunteer based monitoring,” “public participation in scientific research,” and “participatory science.” We also used the search term “volunteer.” Searches were carried out in October and November 2019 using Web of Science, and were filtered by “ecology,” “zoology,” “entomology,” and “ornithology.” To ensure that our keyword searches in Web of Science were not missing large components of the literature that might be found elsewhere, additional searches for the terms “ecology AND (volunteer OR citizen science)” were carried out using Google Scholar, and the first 100 results were reviewed.

We excluded papers if there were no mentions of a specific citizen science scheme, or if volunteers had been recruited to assist with the research but the contributions did not continue beyond the study and were not linked to a particular scheme. For example, Flaherty and Lawton (2019) requested, using various media outlets, information on grey squirrel, red squirrel, and pine marten sightings by the general public; public sightings were used alongside hair tube and live trapping surveys to assess species distributions. In another example, data were collected from recreational anglers to combine with mark-recapture data to estimate populations of fish species (Lyon et al. 2019). These volunteer contributions were only for the duration of the study and were not linked to any particular scheme. We also excluded review papers, or results that discussed citizen science from a theoretical point of view. Finally, we excluded papers if the citizen science scheme focused on collecting data solely on the abiotic environment. These schemes included those collecting data on water quality (Křeček et al. 2019) or on soil quality (Bone et al. 2012). Where papers had used data from multiple schemes, we recorded all of the schemes included in the research. Citizen science schemes nested within a larger citizen science initiative or repository were considered separately if the paper identified the specific scheme. For example, Snapshot Serengeti (Swanson et al. 2015), Penguin Watch (Jones et al. 2018), and Season Spotter (Kosmala et al. 2016b) were referenced specifically, even though they all fall under the Zooniverse citizen science community, and therefore we recorded them as separate schemes. By contrast, Torney et al. (2019) referenced only the Zooniverse, and therefore the Zooniverse was also recorded. The search yielded 434 papers (see Supplemental File 1 for full reference list), which drew on 259 citizen science schemes (see Supplemental File 2 for full list of schemes).

The search strategy aimed to encompass a broad range of citizen science programmes, including recording

schemes that do not identify as a citizen science scheme but do fit the definition of citizen science. It is, of course, likely that schemes will have been overlooked by the searches—most notably, schemes that have not led to published outputs. The term citizen science has been widely used only in recent decades, although volunteers have been contributing to ecological datasets for centuries (Silvertown 2009; Dickinson, Zuckerberg, and Bonter 2010; Miller-Rushing, Primack, and Bonney 2012; Pocock et al. 2015), and therefore such volunteer contributions may not be linked to a specific volunteer recording scheme and are not referenced in literature. Furthermore, schemes may not provide information on the citizen science scheme attributes or verification approach publicly, and therefore would not be included in the results of this literature review. Although these searches did identify some schemes from non-English-speaking communities and regions, the search strategy is inherently biased towards schemes that operated in English (Pocock et al. 2017). These biases in the search methodology should not systematically impact the conclusions of the review.

IDENTIFYING VERIFICATION APPROACHES AND CITIZEN SCIENCE SCHEME ATTRIBUTES

Verification approaches used by citizen science schemes were not always documented in the paper itself. Therefore, we carried out searches to obtain information on verification approaches and the information used to verify records, as well as citizen science scheme attributes, in both academic and non-academic search engines. We obtained this information from either the published literature in which the scheme featured or the scheme’s public online platform, which may be a website specifically for a scheme, or a web page embedded within a larger website (see Supplemental File 2 for full list of schemes, attributes, and sources).

For each citizen science scheme, we identified the following attributes: number of species recorded through the scheme, number of occurrence records collected through the scheme, data type, number of participants, geographical extent, and duration in years. Data type refers to the amount of information or evidence needed to submit an occurrence record to a scheme. For example, some schemes require photos, recordings, or physical specimens to be submitted before an occurrence can be confirmed. Other schemes allow indirect or direct sightings to be submitted without further evidence. Indirect sightings include observations such as mammal tracks or dung at a given location. Direct sightings refer to a species being observed but the minimum information required for an occurrence to be submitted is species name, location, and date.

DATA ANALYSIS

We performed simple analyses to investigate two questions. First, we asked what attributes of schemes influence whether we were able to find information on their approaches to verification. Second, using those schemes for which we were able to find information on approaches to verification, we asked which attributes of the schemes influenced the approaches that were used.

Some attribute categories included very few schemes. Therefore, we aggregated some categories for our analysis. Specifically, we classified numbers of participants as either $\leq 1,000$ or $> 1,000$; numbers of records as either ≤ 1 million or > 1 million; and data type as either “No evidence” (for reports of direct or indirect sightings without physical evidence) or “Evidence available” (for those data points associated with specimens, photographs, or recordings).

To assess whether scheme attributes influence whether or not we were able to find information on verification approaches, we focused on schemes for which all scheme attributes were available. Inevitably, this biased the data towards schemes with more complete and accessible information. However, this was necessary for a complete investigation of which scheme attributes seemed most predictive of whether verification information could be identified, and still resulted in reasonable sample sizes of schemes with and without verification information. Using this focused dataset, we ran a binary logistic regression including the main effects of geographic scale, participant numbers, record numbers, species numbers, data type (all categorical), and scheme duration (continuous). We used the dredge function from package MuMIn (Barton 2020) to determine the most informative models nested within this global model.

To assess which scheme attributes appear to influence verification approach, we used multinomial regression (function `multinom` from package `nnet`; Venables and Ripley 2002). Specifically, we modelled the probability that expert, automated, community consensus, or other verification approaches would be used as a function of the same scheme attributes included in the saturated binary logistic regression. Once again, we focused on only those schemes for which all attributes were available. Some schemes used more than one approach, in which case those schemes appeared in our data set once for each approach used. The dredge function was used again for model selection, considering main effects only.

RESULTS

SUMMARY OF CITIZEN SCIENCE RECORDING SCHEMES

Of the 259 citizen science schemes, the focal taxa were birds (N = 97), invertebrates (N = 67), mammals (N = 24),

plants and fungi (N = 17), and amphibians and reptiles (N = 8). As well, there were schemes that allowed any taxa to be recorded (N = 27) and schemes that focused on marine taxa (N = 9). There were also schemes that recorded invasive species (N = 6) and schemes that recorded roadkill (N = 4). There was substantial variation in the number of species recorded through the schemes. Where this information was available (N = 203), 68 schemes had recorded 1–10 species, 50 schemes had recorded 11–100 species, 59 had recorded 101–1,000 species, 15 had recorded 1,001–10,000 species, and 11 had recorded more than 10,000 species.

Of the schemes for which record number was available (N = 140), 12 schemes had fewer than 1,000 records, 95 schemes had between 1,000 and 1 million records and 33 had more than 1 million records. The data type submitted with each record varied across schemes: 18 allowed indirect sightings to be submitted, 165 required direct sightings to be submitted, 51 required photo submissions, 10 required recordings, and 15 required specimens to be submitted.

To determine the number of citizen scientists involved in each scheme, we included both those who collected data and registered users who may verify data. Of the schemes for which this information was available (N = 165), 76 had between 1 and 1,000 participants, 86 had between 1,000 and 1 million participants, and 3 had more than 1 million participants.

In terms of geographical extent, 17 schemes collected data at a global, cross-continental scale. Across the remaining schemes, 34 operated across multiple countries within the same continent, 125 schemes collected data at a country level, and 83 schemes operated at a regional level (i.e., the level of a region within a country). There were schemes operating on every continent besides Antarctica, with 106 in Europe, 96 in North America, 17 in Oceania, 10 in Asia, 8 in Africa, and 5 in South America.

The schemes we reviewed spanned a wide range of ages. Of schemes where duration was available (N = 225), 90 schemes had been running for less than 10 years, 64 had been running for between 10 and 20 years, 34 had been running between 20 and 30 years, and 37 schemes had been running for longer than 30 years.

APPROACHES TO DATA VERIFICATION IN CITIZEN SCIENCE SCHEMES

Across the 259 citizen science schemes, no information was found on verification approach for 117 of the schemes. Within the schemes for which verification information was found, 118 schemes relied on expert verification, 24 verified data through community consensus, and 14 used automated approaches, which encompassed algorithmic approaches without human classification. Several of the schemes used multiple verification approaches, and all of

the schemes that used automation to verify data used at least one other method of verification on a subset of the data. Most commonly, automation was used alongside expert verification. Other verification approaches included using existing independent (West 2012) or expert (Hof and Bright 2016) datasets to confirm the likely accuracy of citizen-submitted records, and carrying out follow-up surveys in a subset of locations (Kabat et al. 2012).

The information used to verify citizen science data refers to the record-level information that is used by citizen science schemes when carrying out data verification of species occurrences. This was categorised as species, environmental context, and recorder expertise. Species information is based on ease of identification (Pocock et al. 2015), confusion with other species (Siddharthan et al. 2016), rarity, and co-occurrence with other species. Environmental context takes into account the time, date, and location of the observation and, therefore, whether the species' occurrence was likely given the time of day, season (Dennis et al. 2016), habitat (Pocock et al. 2015), documented range of the species (Terry, Roy, and August 2020), and phenology (Roy and Sparks 2000). Attributes of the recorder that are of interest could include the experience and expertise of the individual submitting the record. This can be considered qualitatively when submitting the record, by asking the recorder to state their confidence in identification (Waetjen and Shilling 2017) or experience with biological recording (Bates et al. 2015). Recorder expertise can also be quantified after record submission, using metrics such as how long the individual has been participating in the scheme, volume of records submitted, and accuracy of previously submitted records (Yu, Wong, and Hutchinson, 2010; August et al. 2020). Schemes can also use novel approaches to account for recorder expertise. One example of this is iSpot, in which recorders develop a taxon-specific reputation via points earned once records they have submitted are verified as correct by other participants (Silvertown et al. 2015).

Schemes were allocated to one or more of these categories based on information provided by the scheme on its verification approach. For many schemes, these details were not publicly available. Furthermore, individual expert verifiers may take into account all, or a combination, of these factors on a record-by-record basis, using their regional and taxonomic expertise as well as their personal knowledge of individual contributors' abilities to identify species correctly. Therefore, it is unlikely that we were able to catalogue for our analysis all of the information considered by schemes and verifiers. Across the schemes for which the required information was available, 105 used information on the species itself, 86 considered the environmental context, and 13 used information on

recorder expertise. The majority of schemes used species information and environmental information together.

CITIZEN SCIENCE SCHEME ATTRIBUTES AND VERIFICATION APPROACH

We restricted our analysis to 103 schemes with complete information on scheme attributes. As expected, this biased schemes towards those with available verification information (all data: schemes with verification information = 142, schemes without = 117; complete attribute data: schemes with verification information = 73, schemes without = 30; Fisher's test, $p = 0.006$). Nevertheless, we were still able to model the propensity for verification information to be found. The best-performing model (based on Akaike information criterion) included data type, number of records, and scheme duration (*Figure 1*). Only more complex versions of the same model had $\Delta AIC_c < 6$, and ΔAIC_c for the null model was > 8 .

Using the 73 schemes for which scheme attributes and verification approach were found, we modelled the factors that best predicted the verification approaches used. Among the schemes we considered, 61 used expert approaches, 7 used automated approaches, 12 used community consensus approaches, and 8 used other approaches. Given the low sample sizes, there was limited evidence of clear predictive effects of scheme attributes. Among the models examined, only those including number of participants, data type, or both, performed better than the null (ΔAIC_c for the null model was 1.9). Recognising that these are weakly supported effects, we nonetheless note that a model including both number of participants and data type suggests that: (i) automated approaches are used only for schemes with more participants and are slightly more common for schemes without physical evidence; (ii) community consensus approaches are more common for schemes with more participants and for which evidence is available; (iii) expert approaches are more common in schemes with fewer participants, but for schemes with more participants, they are more common when no physical evidence is available; and (iv) other approaches are most common for schemes with a smaller number of participants and for which no tangible evidence is available (*Figure 2*).

DISCUSSION

With data quality as a key concern across citizen science datasets, there is a need to ensure validity and increase trust of these data through verification. This review identifies patterns in approaches to data verification among citizen science schemes. By identifying the range of approaches

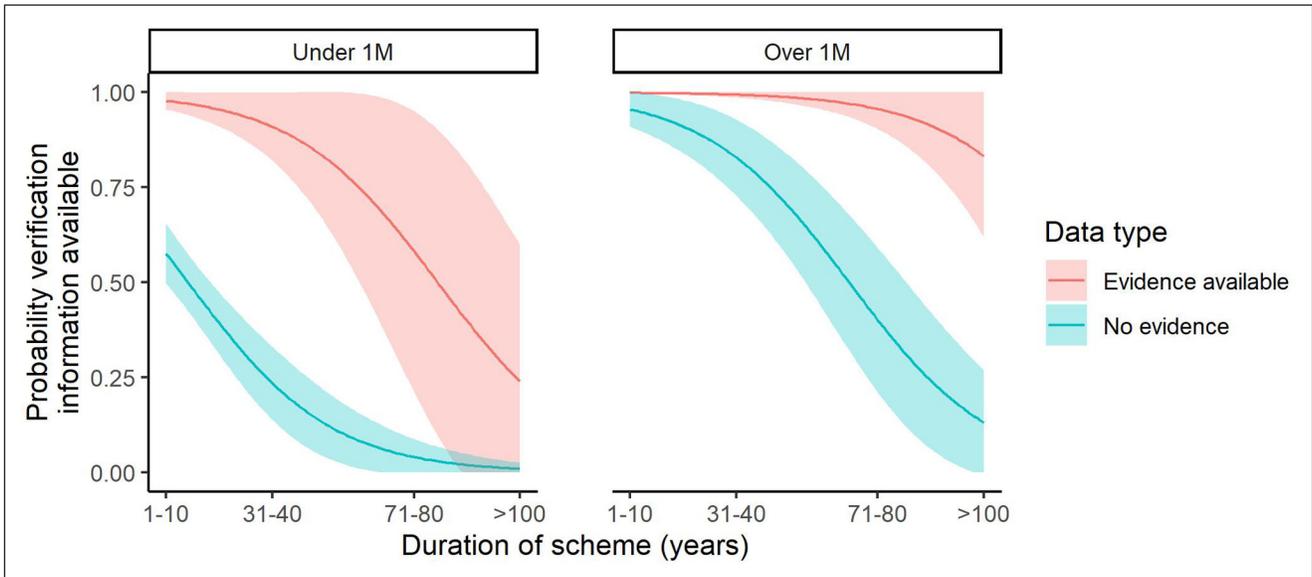


Figure 1 The probability of verification information being found given the numbers of participants (left panel, ≤ 1 million [M]; right panel, > 1 million [M]), duration of schemes, and data type. Fitted probabilities (lines) and standard errors (filled polygons) are estimated using the best-performing binary logistic regression model.

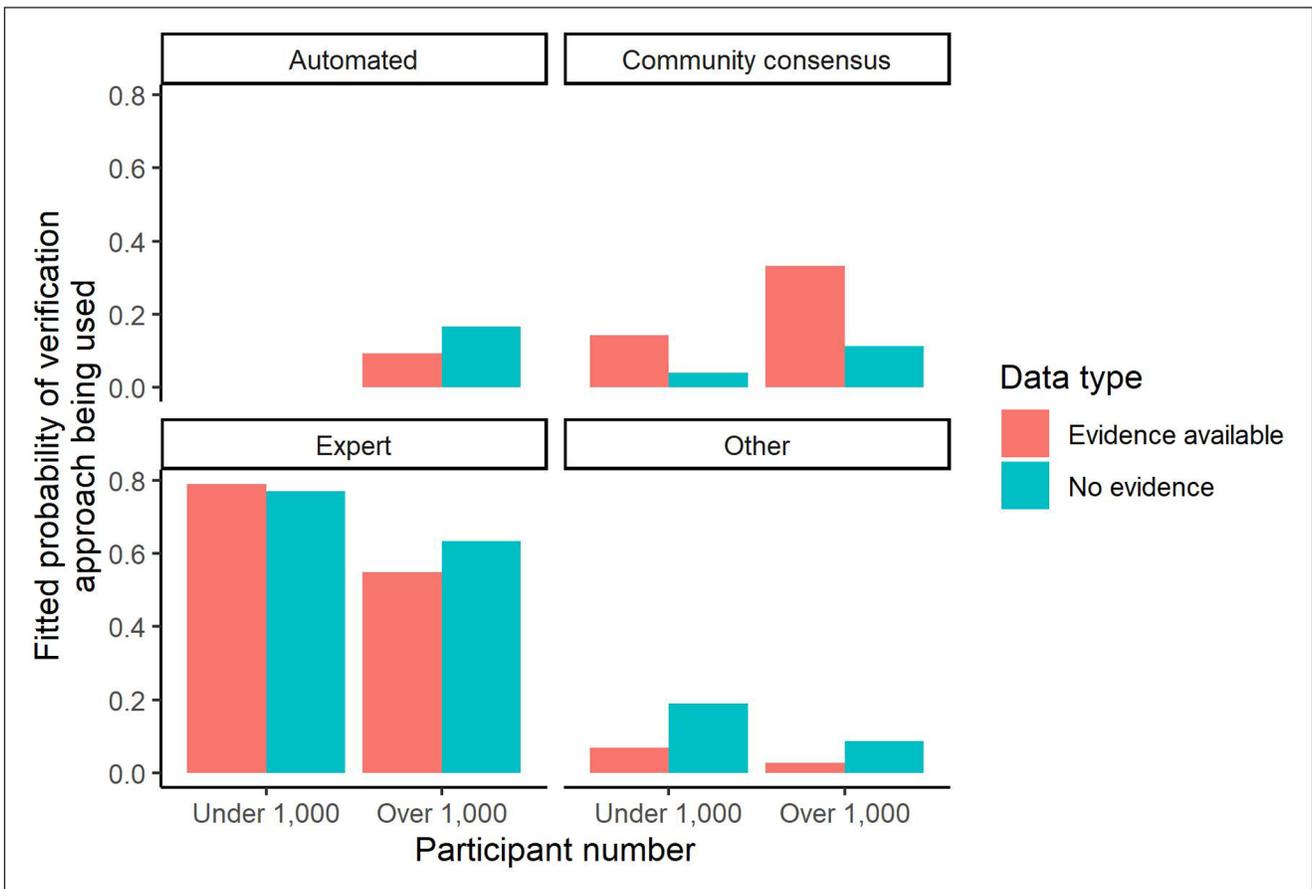


Figure 2 The probability of each verification approach (see panel headings) being used for schemes with different numbers of participants and different data types. Fitted probabilities (filled columns) are estimated using the best-performing parameters in multinomial regressions.

available and by considering scheme attributes that appear to contribute to choices in verification approach, we demonstrate the options available to both new and existing schemes. Here, we also present an idealised system for data verification, identifying where and how such a system could be implemented within citizen science schemes.

EXISTING PATTERNS IN VERIFICATION OF CITIZEN SCIENCE DATA

No information on data verification was found for over 40% of the schemes we reviewed. Our analyses suggest that information on verification was less likely to be found for older schemes, schemes with fewer participants, and schemes that do not require the contribution of physical evidence (specimens, photos, or recordings). Lack of available verification information does not mean that no verification is carried out; for schemes that lack a web presence and do not report verification methods in publications, verification methods are simply not publicly available or therefore are hard to identify. There may, however, be schemes that do not consider verification, trusting the recorders' abilities to report species correctly (Wiggins et al. 2011). This may be justifiable if schemes specifically recruit knowledgeable volunteers (Gardner 2019) or provide training to volunteers before surveying (Smale et al. 2019). Some citizen science schemes focus recording effort on selected days annually (Chase and Levine 2016). In these cases, volunteers may be joined and led by an expert (Chase and Levine 2016) and therefore errors could be identified and corrected, in person, during the data collection. Smaller-scale citizen science schemes may focus on collaborative, community-based approaches with small numbers of participants (Tweddle et al. 2012). In these cases, there may be an established trust amongst members, or verification may happen more informally between participants. Acknowledging this, there is still an imperative to report on verification methods to increase trust in the dataset and to benefit end users of the data. Arguably, this imperative is even more pronounced for those schemes that do not require physical evidence, for which verification information is currently harder to find. If there is transparency in verification approach, then the data quality can be better understood, and potential error or bias can be quantified and accounted for in analyses of the data (Pocock et al. 2014).

Where verification information was available, expert verification was the most common approach. Verification by experts, although not flawless (McBride, Fidler, and Burgman 2012), has a high accuracy (Yu et al. 2012), and therefore may be a more suitable approach to obtain the level of data quality required for published research outputs (Bonter and Cooper 2012). Furthermore, schemes that monitor rare

(Donnelly et al. 2014) or invasive species, for which accuracy of individual records is crucial to guide management actions (Crall et al. 2011; Pusceddu et al. 2019), require expert verification to pinpoint occurrences and ensure high-quality data. Expert verification can be time consuming for large datasets (Kelling et al. 2011; Yu et al. 2012), and schemes that operate at a large geographic scale rely on extensive networks of taxonomic and regional experts (Sutherland, Roy, and Amano 2015). A lack of verifiers in certain regions or with particular specialisms can lead to gaps in verified data (Bonter and Cooper 2012). As a result, there can be a significant time lag between submission and verification of records (Bonter and Cooper 2012).

Community consensus was the second most common verification approach. It was more common among schemes with a larger number of participants and for schemes that required evidence to be submitted with each record. Community consensus may be preferable for schemes with sufficient participants, as crowdsourcing the assessment of physical evidence spreads the task of verification across a greater number of individuals, and can be particularly useful when verifying camera trap datasets, which can rapidly grow to very large sizes (Swanson et al. 2016; Hsing et al. 2018). Community consensus approaches can also be used alongside automated approaches in a hierarchical verification system (Green et al. 2020). Once multiple users have classified a record, consensus algorithms can be applied to analyse classifications and to categorise confidence in a record (Siddharthan et al. 2016; Hsing et al. 2018). Community consensus approaches also have the potential to enhance public engagement and community development. Diversifying the tasks in which citizen scientists can be involved can make the scheme more accessible to those who do not have the access or mobility to go to areas where they can record species (Borden et al. 2013). When using community consensus approaches, expert verification may still be required if datasets contain species that are less straightforward to identify, such as commonly-confused species pairs (Hsing et al. 2018). This approach relies on a large number of citizen scientists investing time in the scheme (Swanson et al. 2016; Hsing et al. 2018), and therefore may not be suitable for schemes with smaller numbers of users. Furthermore, if community consensus approaches are used for schemes that operate on a global scale and record many species, the community may not have the local knowledge required to verify records for species that are less straightforward to identify or are less well known amongst the general public (Swanson et al. 2015). As a result, the verified data in these schemes may be skewed toward widely recognised, charismatic species.

Perhaps unsurprisingly, owing to their recent emergence, automated approaches were not widely used among the

subset of citizen science schemes reviewed. Schemes that used automation, did so in conjunction with other methods including, most frequently, expert verification. Automation is typically the first step in the verification process, with records being checked for a range of attributes. These include whether they are in the expected geographical and temporal range, if the species is particularly rare, or for schemes that ask for the number of individuals of a species recorded, whether that number is unusually high (Kelling et al. 2011; Bonter and Cooper 2012; Yu et al. 2012; Pocock et al. 2015). Any records that do not meet set criteria are flagged and then sent to expert verifiers (Kelling et al. 2011; Bonter and Cooper 2012; Yu et al. 2012). Automation reduces the burden on expert verifiers by decreasing the volume of records that require verification. Automated approaches are widely applicable across citizen science schemes and can be applied to records for a huge diversity of taxa (Pocock et al. 2015). Automation is the most time-efficient way of verifying citizen science data, allowing data to be reviewed in real time as records are submitted, as well as—potentially—providing citizen scientists with immediate feedback on their submissions (Kelling et al. 2011; Bonter and Cooper 2012; Yu et al. 2012; van der Wal et al. 2016). From the perspective of participant involvement, having rapid feedback on submitted records has the potential to strengthen engagement and to increase motivation to continue recording (Rotman et al. 2014). Although automation can reduce the number of records that require expert review, careful consideration of the verification rules is required to reduce the burden on experts without leading to classification errors (Yu et al. 2012).

With the distributions and abundances of many species changing rapidly in response to persistent anthropogenic environmental change, timely and accurate verification is important to ensure the availability of up-to-date biodiversity information (Sutherland, Roy, and Amano 2015). Verification by experts has perhaps been the default approach for citizen science schemes in the past (Silvertown 2009; Kosmala et al. 2016a). With the growing volume of citizen science data that has been and will continue to be collected, there is an argument for schemes to explore and implement other verification approaches that allow large quantities of data to be verified more efficiently. The most appropriate verification approach may vary from scheme to scheme, and research may be required to assess the risks or rewards of alternative approaches. Expert verification is likely always to be required for a subset of the data, but given the emergence of community consensus and automated verification in recent decades (Newman et al. 2012), these approaches should be carefully considered for schemes moving forward. As the position of citizen science in ecological research evolves, with new schemes

continually being established, verification approaches must evolve to suit the needs of schemes whilst also ensuring data quality and accuracy of records.

RECOMMENDATIONS FOR VERIFICATION OF CITIZEN SCIENCE DATA

Our review highlights the range of verification methods used by different citizen science schemes. In some cases, this variation might reflect deliberate and informed choices based on what works best given the attributes of different schemes. In others, it is likely that choices reflect historical contingency, or cost and ease of implementation. Some schemes may be limited to a certain approach due to available resources, time, or personnel. Others may feel bound to a verification approach in order to maintain consistency over time. In those situations, retrospective application of new methods, or calibration by running two systems in tandem, might provide reassurance to enable the implementation of new approaches.

Whilst a range of factors may influence choice of, or lack of, verification approach, transparency of documentation of verification approaches is required to increase confidence in citizen science as a means of collecting reliable data. Therefore, we recommend that citizen science schemes publicly report their verification approach. Schemes that lack a platform on which this information can be made readily available should ensure that published research clearly identifies whether and how the data were verified.

An idealised system for verification

Considering the options available for verification and the attributes that may contribute to the choice of verification approach, we have outlined a hierarchical system for verification (summarised in [Figure 3](#)). This approach considers the data that can be used to verify records, where automated and community consensus approaches can be implemented, and when expert verification may still be required.

When verifying records, schemes should consider the breadth of information available to improve verification, making use of all data that accompanies each record (see [Figure 3](#)). Ideally, recorders should submit the maximum available evidence with each record, such as photos or recordings, assuming the user interface through which volunteers submit records is fit for purpose. Submitting photos or other evidence may not be possible for every scheme, particularly those centred around annual count events, such as the Batumi Raptor Count (Wehrmann et al. 2019) or Christmas Bird Count (Meehan, Michel and Rue 2019), where large numbers of species are recorded during a constrained period. Furthermore, requiring more information to be submitted with every species record may

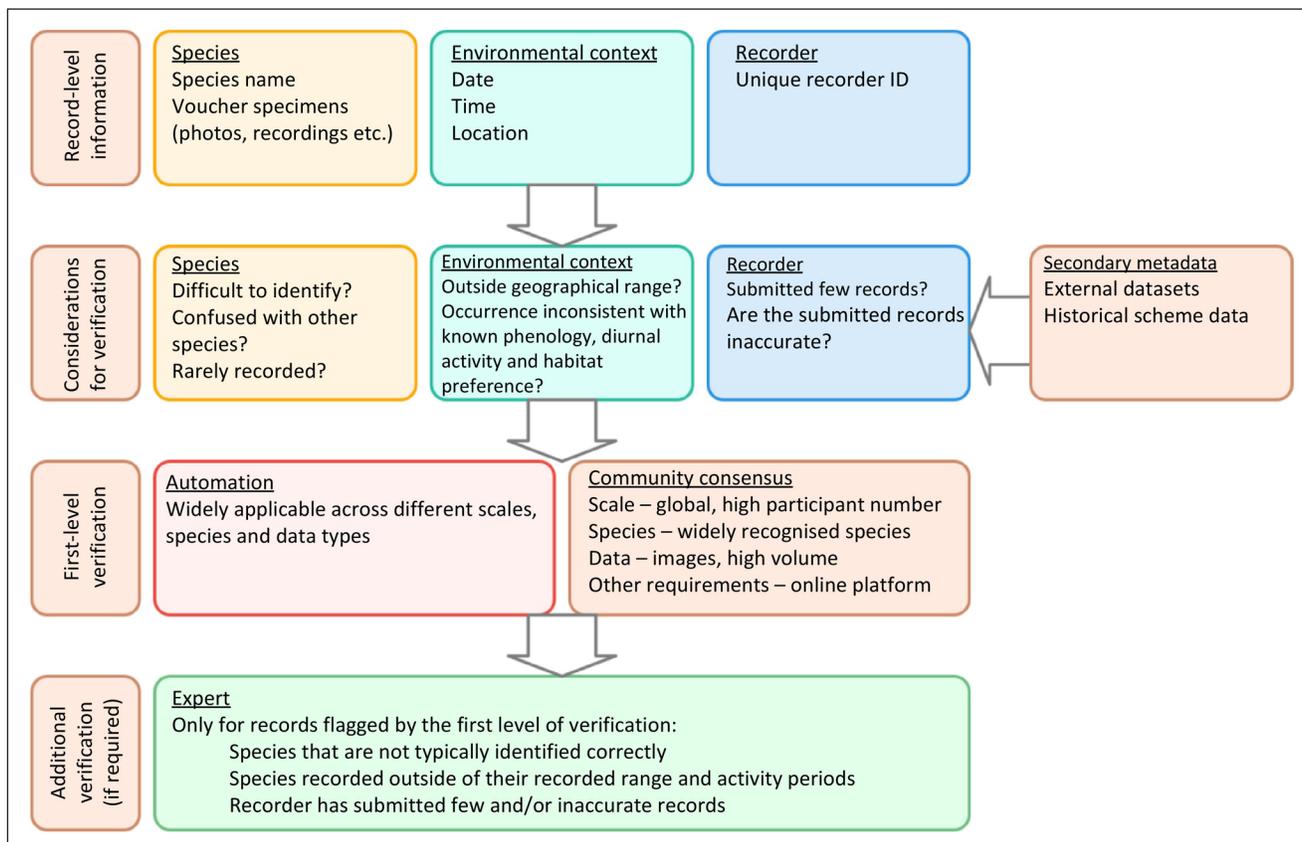


Figure 3 Summary of recommendations for an idealised system for verification of ecological citizen science data. Considerations for verification highlight some of the questions that can be answered using the record-level information and secondary metadata. If the answer to these questions is yes, then we propose further levels of verification may be required. First-level verification indicates the attributes of schemes that could use community consensus and automated approaches. Additional verification highlights the kinds of records that may be flagged and therefore will need to be reviewed by experts.

discourage volunteers from taking part, creating a trade-off between data completeness and data volume. For many schemes, the minimum amount of information required is date, location, and species name. For other schemes, indirect sightings can be submitted, particularly those recording mammal species, which are often less abundant, frequently nocturnal, and less likely to be observed directly. Verification approaches need to be developed and applied in view of the minimum amount of information that typically comes with each record. Even with the limited record-level information that may accompany each record, verification approaches can still take into account information on the species, the environmental context, and the recorder (see [Figure 3](#)). This can be done through input from expert verifiers, or by using secondary metadata such as historical data recorded through the scheme or external datasets. These data can then be used to cross-reference the metadata with each record (Terry, Roy, and August 2020). If schemes have large volumes of data across many species and records with varying amounts of information, a hierarchy of approaches could be implemented. This

allows the bulk of records to be verified by automated and community approaches, and then flagged records undergo additional levels of verification (see [Figure 3](#)).

Automated verification approaches are flexible and—resources for implementation permitting—could be used more widely across citizen science schemes to verify large quantities of data efficiently. Automation can be implemented within schemes that already have large quantities of historic data, as these can be used to inform algorithms and develop filters for the datasets (Kelling et al. 2011). To account for verification metrics for the species, environmental context, and recorder expertise (see [Figure 3](#)), automated approaches can incorporate record-level information and secondary metadata (Terry, Roy, and August 2020), as well as expert knowledge (Kelling et al. 2011). For automated approaches to account for environmental factors, location, date, and time are required, as well as prior knowledge of the species' geographical and temporal range (Sutherland, Roy, and Amano 2015). Using contextual information is most useful for schemes that focus on monitoring species' phenology,

or when there are no photos or recordings submitted with a record. However, it is associated with the risk that sightings could be rejected if the species displays novel activity patterns or range shifts. To account for recorder expertise, individual recorders require a unique ID. It is important to consider that as individuals submit more records, their accuracy when identifying species may improve. When accounting for environmental context or recorder expertise in automated verification approaches, it is essential to retain flexibility, with rules being dynamically updated as unexpected sightings accumulate or as recorder expertise improves.

Another approach that can be used as the first level of verification is community consensus (see [Figure 3](#)). This approach is less widely applicable than automated verification and typically requires an online platform that connects recorders and verifiers, and large enough numbers of volunteers to verify the volume of records (Silvertown et al. 2015; Siddharthan et al. 2016; Swanson et al. 2016; Hsing et al. 2018). Community consensus approaches are more suitable for species that are more widely recognised by the public and where there is photographic evidence with each record (Swanson et al. 2016), as this means that the record can be verified based on visual attributes of the species, and no prior knowledge of the environmental context is required.

If automated and community approaches cannot verify records with an appropriate level of certainty, experts can provide additional levels of verification (see [Figure 3](#)). It is important, therefore, for schemes to decide on their required level of certainty, which may vary depending on the species and the purpose for which the data will be used. For most schemes, a proportion of the data will ultimately need to be referred to experts for verification. A key aim of automated approaches is to minimise the proportion of the data that require expert verification. This additional verification is likely to be required for species that have not been recorded before through the scheme, for rarer species, for invasive species for which pinpointing the exact location of individuals is necessary (Lagoze 2014), and for species that are recorded beyond their typical range or habitat. If a scheme is focusing exclusively on these kinds of species, expert verification may be the most appropriate approach. Expert insight can also be used to inform automated verification approaches, by providing information on the species and environmental context that can be accounted for in data filters. Furthermore, if a scheme is considering recorder expertise when verifying data, expert insight could also be beneficial to identify trusted recorders, allowing their submissions to be used in place of a gold standard when verifying and analysing data.

CONCLUSIONS

We reviewed approaches to data verification across ecological citizen science datasets, and assessed factors that appear to influence the choice of verification approach. Alongside this, we highlighted that the verification approaches of many citizen science schemes are not readily available to the public. We recommend how citizen science schemes can approach verification and make appropriate choices to ensure data quality. Citizen science plays an important role in data collection at a geographical and temporal scale unmatched by other data collection methods, and is a valuable means of engaging the public in scientific endeavours. By developing improved verification approaches and using the full range of information available, issues of data quality within citizen science datasets can be addressed, thereby increasing trust in citizen science approaches and strengthening the place of citizen science within ecological research.

SUPPLEMENTARY FILES

The supplementary files for this article can be found as follows:

- **Supplemental File 1.** Literature Search References. DOI: <https://doi.org/10.5334/cstp.351.s1>
- **Supplemental File 2.** Citizen Science Scheme Data. DOI: <https://doi.org/10.5334/cstp.351.s2>

ACKNOWLEDGEMENTS

EB was supported by the Natural Environment Research Council's IAPETUS2 Doctoral Training Partnership award number NE/S007431/1. DR was supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Emily Baker  orcid.org/0000-0003-3571-6888
Conservation Ecology Group, Department of Biosciences, Durham University, GB

Jonathan P. Drury  orcid.org/0000-0002-0262-8761
Conservation Ecology Group, Department of Biosciences, Durham University, GB

Johanna Judge
National Biodiversity Network Trust, Nottingham, GB

David B. Roy  orcid.org/0000-0002-5147-0331
Biological Records Centre, UK Centre for Ecology & Hydrology, Wallingford, GB

Graham C. Smith  orcid.org/0000-0002-9897-6794
National Wildlife Management Centre, Animal and Plant Health Agency, York, GB

Philip A. Stephens  orcid.org/0000-0001-5849-788X
Conservation Ecology Group, Department of Biosciences, Durham University, GB

REFERENCES

- August, T, Fox, R, Roy, DB and Pocock, MJO.** 2020. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Scientific Reports*, 10(1): 11009. DOI: <https://doi.org/10.1038/s41598-020-67658-3>
- Barton, K.** 2020. MuMIn: Multi-Model Inference. R package version 1.43.17. <https://CRAN.R-project.org/package=MuMIn>.
- Bates, AJ, Lakeman Fraser, P, Robinson, L, Tweddle, JC, Sadler, JP, West, SE, Norman, S, Batson, M and Davies, L.** 2015. The OPAL bugs count survey: exploring the effects of urbanisation and habitat characteristics using citizen science. *Urban Ecosystems*, 18(4): 1477–1497. DOI: <https://doi.org/10.1007/s11252-015-0470-8>
- Bone, J, Archer, M, Barraclough, D, Eggleton, P, Flight, D, Head, M, Jones, DT, Scheib, C and Voulvoulis, N.** 2012. Public participation in soil surveys: Lessons from a pilot study in England. *Environmental Science and Technology*, 46(7): 3687–3696. DOI: <https://doi.org/10.1021/es203880p>
- Bonter, DN and Cooper, CB.** 2012. Data validation in citizen science: a case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, 10(6): 305–307. DOI: <https://doi.org/10.1890/110273>
- Borden, KA, Kapadia, A, Smith, A and Whyte, L.** 2013. Educational Exploration of the Zooniverse: Tools for Formal and Informal Audience Engagement. *ASP Conference Series*, 473: 101–107.
- Chase, SK and Levine, A.** 2016. A framework for evaluating and designing citizen science programs for natural resources monitoring. *Conservation Biology*, 30(3): 456–466. DOI: <https://doi.org/10.1111/cobi.12697>
- Collaboration for Environmental Evidence.** 2018. *Guidelines and Standards for Evidence synthesis in Environmental Management*. Version 5.0. In: Pullin, AS, Frampton, GK, Livoreil, B and Petrokofsky, G (Eds.). Available at: <http://www.environmentalevidence.org/information-for-authors> (Last accessed: 2 January 2020).
- Crall, AW, Newman, GJ, Stohlgren, TJ, Holfelder, KA, Graham, J and Waller, DM.** 2011. Assessing citizen science data quality: an invasive species case study. *Conservation Letters*, 4(6): 433–442. DOI: <https://doi.org/10.1111/j.1755-263X.2011.00196.x>
- Dennis, EB, Morgan, BJT, Freeman, SN, Brereton, TM and Roy, DB.** 2016. A generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4): 1305–1314. DOI: <https://doi.org/10.1111/biom.12506>
- Devictor, V, Whittaker, RJ and Beltrame, C.** 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, 16(3): 354–362. DOI: <https://doi.org/10.1111/j.1472-4642.2009.00615.x>
- Dickinson, JL, Zuckerberg, B and Bonter, DN.** 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1): 149–172. DOI: <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- Donnelly, A, Crowe, O, Regan, E, Begley, S and Caffarra, A.** 2014. The role of citizen science in monitoring biodiversity in Ireland. *International Journal of Biometeorology*, 58(6): 1237–1249. DOI: <https://doi.org/10.1007/s00484-013-0717-0>
- Flaherty, M and Lawton, C.** 2019. The regional demise of a non-native invasive species: the decline of grey squirrels in Ireland. *Biological Invasions*, 21(7): 2401–2416. DOI: <https://doi.org/10.1007/s10530-019-01987-x>
- Flower, E, Jones, D and Bernede, L.** 2016. Can Citizen Science Assist in Determining Koala (*Phascolarctos Cinereus*) Presence in a Declining Population? *Animals*, 6(7): 42. DOI: <https://doi.org/10.3390/ani6070042>
- Gardner, E.** 2019. Make the Adder Count: population trends from a citizen science survey of UK adders. *Herpetological Journal*, 29(1): 57–70. DOI: <https://doi.org/10.33256/hj29.1.5770>
- Gilfedder, M, Robinson, CJ, Watson, JEM, Campbell, TG, Sullivan, BL and Possingham, HP.** 2019. Brokering Trust in Citizen Science. *Society and Natural Resources*, 32(3): 292–302. DOI: <https://doi.org/10.1080/08941920.2018.1518507>
- Green, SE, Rees, JP, Stephens, PA, Hill, RA and Giordano, AJ.** 2020. Innovations in Camera Trapping Technology and Approaches: The Integration of Citizen Science and Artificial Intelligence. *Animals*, 10(1): 132. DOI: <https://doi.org/10.3390/ani10010132>
- Hof, AR and Bright, PW.** 2016. Quantifying the long-term decline of the West European hedgehog in England by subsampling citizen-science datasets. *European Journal of Wildlife Research*, 62(4): 407–413. DOI: <https://doi.org/10.1007/s10344-016-1013-1>
- Hsing, P, Bradley, S, Kent, VT, Hill, RA, Smith, GC, Whittingham, MJ, Cokill, J, Crawley, D, MammalWeb Volunteers and Stephens, PA.** 2018. Economical crowdsourcing for camera trap image classification. *Remote Sensing in Ecology and*

- Conservation*, 4(4): 361–374. DOI: <https://doi.org/10.1002/rse2.84>
- Huber, B, Barnidge, M, Gil de Zúñiga, H and Liu, J.** 2019. Fostering public trust in science: The role of social media. *Public understanding of science (Bristol, England)*, 28(7): 759–777. DOI: <https://doi.org/10.1177/0963662519869097>
- Isaac, NJB, van Strien, AJ, August, TA, de Zeeuw, MP and Roy, DB.** 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10): 1052–1060. DOI: <https://doi.org/10.1111/2041-210X.12254>
- Iyengar, S and Massey, DS.** 2019. Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16): 7656–7661. DOI: <https://doi.org/10.1073/pnas.1805868115>
- Jones, FM, Allen, C, Arteta, C, Arthur, J, Black, C, Emmerson, LM, Freeman, R, Hines, G, Lintott, CJ, Macháčková, Z, Miller, G, Simpson, R, Southwell, C, Torsey, HR, Zisserman, A and Hart, T.** 2018. Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, 5(1): 180124. DOI: <https://doi.org/10.1038/sdata.2018.124>
- Kabat, AP, Scott, R, Kabat, TJ and Barrett, G.** 2012. 2011 Great Cockey Count: Population estimates and identification of roost sites for the Carnaby's Cockatoo (*Calyptorhynchus latirostris*). Perth WA.
- Kamp, J, Opper, S, Heldbjerg, H, Nyegaard, T and Donald, PF.** 2016. Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, 22(10): 1024–1035. DOI: <https://doi.org/10.1111/ddi.12463>
- Kelling, S, Yu, J, Gerbracht, J and Wong, WK.** 2011. Emergent filters: Automated data verification in a large-scale citizen science project. *Proceedings – 7th IEEE International Conference on e-Science Workshops, eScienceW 2011*. IEEE, Stockholm on 5–8 December 2011, 20–27. DOI: <https://doi.org/10.1109/eScienceW.2011.13>
- Kimura, AH and Kinchy, A.** 2016. Citizen Science: Probing the Virtues and Contexts of Participatory Research. *Engaging Science, Technology, and Society*, 2: 331. DOI: <https://doi.org/10.17351/ests2016.99>
- Kosmala, M, Crall, A, Cheng, R, Hufkens, K, Henderson, S and Richardson, A.** 2016b. Season Spotter: Using Citizen Science to Validate and Scale Plant Phenology from Near-Surface Remote Sensing. *Remote Sensing*, 8(9): 726. DOI: <https://doi.org/10.3390/rs8090726>
- Kosmala, M, Wiggins, A, Swanson, A and Simmons, B.** 2016a. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10): 551–560. DOI: <https://doi.org/10.1002/fee.1436>
- Kullenberg, C and Kasperowski, D.** 2016. What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE*, 11(1): e0147152. DOI: <https://doi.org/10.1371/journal.pone.0147152>
- Křeček, J, Palán, L, Pažourková, E and Stuchlík, E.** 2019. Water-quality genesis in a mountain catchment affected by acidification and forestry practices. *Freshwater Science*, 38(2): 257–269. DOI: <https://doi.org/10.1086/698533>
- Lagoze, C.** 2014. eBird: Curating Citizen Science Data for Use by Diverse Communities. *International Journal of Digital Curation*, 9(1): 71–82. DOI: <https://doi.org/10.2218/ijdc.v9i1.302>
- Lyon, JP, Bird, TJ, Kearns, J, Nicol, S, Tonkin, Z, Todd, CR, O'Mahony, J, Hackett, G, Raymond, S, Lieschke, J, Kitchingman, A and Bradshaw, CJA.** 2019. Increased population size of fish in a lowland river following restoration of structural habitat. *Ecological Applications*, 29(4): e01882. DOI: <https://doi.org/10.1002/eap.1882>
- Mason, CE and Garbarino, J.** 2016. The Power of Engaging Citizen Scientists for Scientific Progress. *Journal of Microbiology & Biology Education*, 17(1): 7–12. DOI: <https://doi.org/10.1128/jmbe.v17i1.1052>
- McBride, MF, Fidler, F and Burgman, MA.** 2012. Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research. *Diversity and Distributions*, 18(8): 782–794. DOI: <https://doi.org/10.1111/j.1472-4642.2012.00884.x>
- Meehan, TD, Michel, NL and Rue, H.** 2019. Spatial modeling of Audubon Christmas Bird Counts reveals fine-scale patterns and drivers of relative abundance trends. *Ecosphere*, 10(4): e02707. DOI: <https://doi.org/10.1002/ecs2.2707>
- Miller-Rushing, A, Primack, R and Bonney, R.** 2012. The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, 10(6): 285–290. DOI: <https://doi.org/10.1890/110278>
- National Coordinating Centre for Public Engagement (NCCPE).** 2016. *Examples of successful public engagement : additional evidence submitted to the Science and Technology Committee (Science Communication) by the National Coordinating Centre for Public Engagement.*
- Newman, G, Wiggins, A, Crall, A, Graham, E, Newman, S and Crowston, K.** 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6): 298–304. DOI: <https://doi.org/10.1890/110294>
- Pace, ML, Hampton, SE, Limburg, KE, Bennett, EM, Cook, EM, Davis, AE, Grove, JM, Kaneshiro, KY, LaDeau, SL, Likens, GE, McKnight, DM, Richardson, DC and Strayer, DL.** 2010. Communicating with the public: opportunities and rewards for individual ecologists. *Frontiers in Ecology and the Environment*, 8(6): 292–298. DOI: <https://doi.org/10.1890/090168>
- Pescott, OL, Walker, KJ, Pocock, MJO, Jitlal, M, Outhwaite, CL, Cheffings, CM, Harris, F and Roy, DB.** 2015. Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland. *Biological Journal of the Linnean Society*, 115(3): 505–521. DOI: <https://doi.org/10.1111/bij.12581>

- Pocock, MJO, Chapman, DS, Sheppard, LJ and Roy, HE.** 2014. *Choosing and using citizen science: a guide to when and how to use citizen science to monitor biodiversity and the environment – NERC Open Research Archive.* Centre for Ecology and Hydrology, Wallingford.
- Pocock, MJO, Roy, HE, Preston, CD and Roy, DB.** 2015. The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnean Society*, 115(3): 475–493. DOI: <https://doi.org/10.1111/bij.12548>
- Pocock, MJO, Tweddle, JC, Savage, J, Robinson, LD and Roy, HE.** 2017. The diversity and evolution of ecological and environmental citizen science. *PLOS ONE*, 12(4): e0172579. DOI: <https://doi.org/10.1371/journal.pone.0172579>
- Pusceddu, M, Floris, I, Mannu, R, Cocco, A and Satta, A.** 2019. Using verified citizen science as a tool for monitoring the European hornet (*Vespa crabro*) in the island of Sardinia (Italy). *NeoBiota*, 50: 97–108. DOI: <https://doi.org/10.3897/neobiota.50.37587>
- Rotman, D, Hammock, J, Preece, J, Hansen, D, Boston, C, Bowser, A and He, Y.** 2014. Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries. *iConference 2014 Proceedings*, 110–124. DOI: <https://doi.org/10.9776/14054>
- Roy, DB and Sparks, TH.** 2000. Phenology of British butterflies and climate change. *Global Change Biology*, 6(4), 407–416. DOI: <https://doi.org/10.1046/j.1365-2486.2000.00322.x>
- Roy, HE, Pocock, MJO, Preston, CD, Roy, DB, Savage, J, Tweddle, JC and Robinson, LD.** 2012. *Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.* NERC/Centre for Ecology & Hydrology, Wallingford.
- Salomon, AK, Lertzman, K, Brown, K, Wilson, KB, Secord, D and McKechnie, I.** 2018. Democratizing conservation science and practice. *Ecology and Society*. DOI: <https://doi.org/10.5751/ES-09980-230144>
- Sewell, D, Beebee, TJC and Griffiths, RA.** 2010. Optimising biodiversity assessments by volunteers: The application of occupancy modelling to large-scale amphibian surveys. *Biological Conservation*, 143(9): 2102–2110. DOI: <https://doi.org/10.1016/j.biocon.2010.05.019>
- Siddharthan, A, Lambin, C, Robinson, AM, Sharma, N, Comont, R, O'mahony, E, Mellish, C and Van Der Wal, R.** 2016. Crowdsourcing without a crowd: Reliable online species identification using Bayesian models to minimize crowd size. *ACM Transactions on Intelligent Systems and Technology*, 7(4). DOI: <https://doi.org/10.1145/2776896>
- Silvertown, J.** 2009. A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9): 467–471. DOI: <https://doi.org/10.1016/j.tree.2009.03.017>
- Silvertown, J, Harvey, M, Greenwood, R, Dodd, M, Rosewell, J, Rebelo, T, Ansine, J and McConway, K.** 2015. Crowdsourcing the identification of organisms: A case-study of iSpot. *ZooKeys*, 480(480): 125–146. DOI: <https://doi.org/10.3897/zookeys.480.8803>
- Smale, DA, Epstein, G, Parry, M and Attrill, MJ.** 2019. Spatiotemporal variability in the structure of seagrass meadows and associated macrofaunal assemblages in southwest England (UK): Using citizen science to benchmark ecological pattern. *Ecology and Evolution*, 9(7): 3958–3972. DOI: <https://doi.org/10.1002/ece3.5025>
- Sutherland, WJ, Roy, DB and Amano, T.** 2015. An agenda for the future of biological recording for ecological monitoring and citizen science. *Biological Journal of the Linnean Society*, 115(3): 779–784. DOI: <https://doi.org/10.1111/bij.12576>
- Swanson, A, Kosmala, M, Lintott, C and Packer, C.** 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, 30(3): 520–531. DOI: <https://doi.org/10.1111/cobi.12695>
- Swanson, A, Kosmala, M, Lintott, C, Simpson, R, Smith, A and Packer, C.** 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, 2(1). DOI: <https://doi.org/10.1038/sdata.2015.26>
- Terry, JCD, Roy, HE and August, TA.** 2020. Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, 11(2), 303–315. DOI: <https://doi.org/10.1111/2041-210X.13335>
- Torney, CJ, Lloyd-Jones, DJ, Chevallier, M, Moyer, DC, Maliti, HT, Mwitia, M, Kohi, EM and Hopcraft, GC.** 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, 10(6): 779–787. DOI: <https://doi.org/10.1111/2041-210X.13165>
- Tweddle, JC, Robinson, D, Pocock, MJO and Roy, HE.** 2012. *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK – NERC Open Research Archive.* Wallingford: Centre for Ecology and Hydrology.
- van der Wal, R, Sharma, N, Mellish, C, Robinson, A and Siddharthan, A.** 2016. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology*, 30(3): 550–561. DOI: <https://doi.org/10.1111/cobi.12705>
- Venables, WN and Ripley, BD.** 2002. *Modern Applied Statistics with S.* Fourth Edition. New York: Springer. ISBN 0-387-95457-0. DOI: https://doi.org/10.1007/978-0-387-21706-2_14
- Waetjen, DP and Shilling, FM.** 2017. Large Extent Volunteer Roadkill and Wildlife Observation Systems as Sources of Reliable Data. *Frontiers in Ecology and Evolution*, 5(89). DOI: <https://doi.org/10.3389/fevo.2017.00089>
- Wehrmann, J, de Boer, F, Benjumea, R, Cavallès, S, Engelen, D, Jansen, J, Verhelst, B and Vansteelant, WMG.** 2019. Batumi

Raptor Count: autumn raptor migration count data from the Batumi bottleneck, Republic of Georgia. *ZooKeys*, 836, 135–157. DOI: <https://doi.org/10.3897/zookeys.836.29252>

West, P. 2012. *FeralScan: web-based community reporting, education and extension tool for landholders and community groups*. Orange NSW.

Wiggins, A, Newman, G, Stevenson, RD and Crowston, K. 2011. Mechanisms for data quality and validation in citizen science. *Proceedings – 7th IEEE International Conference on e-Science Workshops, eScienceW 2011*, Stockholm on 5–8 December 2011, 14–19. DOI: <https://doi.org/10.1109/eScienceW.2011.27>

Woolley, JP, McGowan, ML, Teare, HJA, Coathup, V, Fishman, JR, Settersten, RA, Sterckx, S, Kaye, J and Juengst, ET. 2016.

Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives. *BMC Medical Ethics*, 17(1). DOI: <https://doi.org/10.1186/s12910-016-0117-1>

Yu, J, Kelling, S, Gerbracht, J and Wong, WK. 2012. Automated data verification in a large-scale citizen science project: A case study. *2012 IEEE 8th International Conference on E-Science, e-Science 2012*. Chicago, IL on 8–12 October 2012. DOI: <https://doi.org/10.1109/eScience.2012.6404472>

Yu, J, Wong, WK and Hutchinson, RA. 2010. Modeling Experts and Novices in Citizen Science Data for Species Distribution Modeling. *IEEE*, Sydney, NSW on 13–17 December 2010, 1157–1162. DOI: <https://doi.org/10.1109/ICDM.2010.103>

TO CITE THIS ARTICLE:

Baker, E, Drury, JP, Judge, J, Roy, DB, Smith, GC and Stephens, PA. 2021. The Verification of Ecological Citizen Science Data: Current Approaches and Future Possibilities. *Citizen Science: Theory and Practice*, 6(1): 12, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.351>

Submitted: 24 July 2020 Accepted: 12 February 2021 Published: 13 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.

