Advantages and Drawbacks of Open-Ended, Use-Agnostic Citizen Science Data Collection: A Case Study

CASE STUDIES

]u[ubiquity press

YOLANDA F. WIERSMA TOM CLENCHE MARDON ERBLAND GISELA WACHINGER ROMAN LUKYANENKO JEFFREY PARSONS

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

ABSTRACT

Citizen science projects that collect natural history observations often do not have an underlying research question in mind. Thus, data generated from such projects can be considered "use-aqnostic." Nevertheless, such projects can yield important insights about species distributions. Many of these projects use a class-based data schema, whereby contributors must supply a species identification. This can limit participation if contributors are not confident in their identifications, and can introduce data quality issues if species identification is incorrect. Some projects, such as iNaturalist, circumvent this with crowdsourced species identifications based on contributed photographs, or by grading confidence in the data based on attributes of the sighting and/or contributor. An alternative to a class-based data schema is an open-ended (instance-based) one, where contributors are free to identify their sighting at whatever taxonomic resolution they are most confident, and/or describe the sighting based on attributes. This can increase participation (data completeness) and have the benefit of adding additional (and sometimes unexpected) information. The regionally-focused citizen science website NLNature.com was designed to experimentally examine how class-based versus instance-based schema affected contributions and data quality. Here, we show that the instance-based schema yielded not only more contributions, but also several of ecological importance. Thus, allowing contributors to supply natural history information at a level familiar to them increases data completeness and facilitates unanticipated contributions.

CORRESPONDING AUTHOR: Yolanda F. Wiersma

Department of Biology, Memorial University, St. John's, Newfoundland and Labrador, CA ywiersma@mun.ca

KEYWORDS:

citizen science; data quality; information systems; natural history; species distributions; taxonomy; wildlife

TO CITE THIS ARTICLE:

Wiersma, YF, Clenche, T, Erbland, M, Wachinger, G, Lukyanenko, R and Parsons, J. 2024. Advantages and Drawbacks of Open-Ended, Use-Agnostic Citizen Science Data Collection: A Case Study. *Citizen Science: Theory and Practice*, 9(1): 5, pp. 1–13. DOI: https:// doi.org/10.5334/cstp.676

INTRODUCTION

The earth's biodiversity is vast, and scientists estimate that over 5 million species are still unidentified (Moura and Jetz 2021), many of which may disappear due to human activities before they are identified (Liu et al. 2022; Martin et al. 2023). Scientific funding for biodiversity inventories is limited. Harnessing citizen science to collect observations of species occurrence can be a cost-effective and viable solution (Mesaglio et al. 2023). Citizen scientists are more spatially distributed than research scientists, and have more opportunities to observe nature; particularly those individuals who spend time in the outdoors as part of their livelihood (e.g., fishers, adventure guides, farmers) or hobbies (e.g., bird watchers, hikers, paddlers, foragers). There is also strength in numbers: Virtually anyone can become a citizen scientist. Contrast this with biologists, who, for example, in the United States, represent about 0.018% of the total population, according to the US Bureau of Labor Statistics (US Bureau of Labor Statistics 2022).

The widespread use of smart phones and the ability to upload data and photos from mobile devices facilitated new ways for citizen scientists to contribute to biodiversity inventories over the past two decades. For example, the app iNaturalist, launched in 2008, harnesses crowd knowledge by allowing observers to upload photos and using the knowledge of the online community to identify the species. iNaturalist grew rapidly concurrent with widespread adoption of smart phones, approximately doubling the contributed observations per annum in 2012, with over 185 million at last count in 2023 (https://www.inaturalist. ora/stats). Recent assessments of iNaturalist data have shown that accurate species identification as well as completeness of contribution was possible to a high degree for some (usually more charismatic) species (Mesaglio et al. 2023) and that it improves with screening and training of volunteers (Wittmann, Girman, and Crocker 2019). Data from iNaturalist can be useful to track changes in species distributions, especially when used in combination with museum records and scientific surveys (Gaier and Resasco 2023; O'Neill et al. 2023).

Other similar platforms, such as eBird.org, have comparably high levels of participation and broad utility for research. For example, eBird data have been used to assess phenology (Campioni, Madeiro, and Becciu 2023) and the impacts of the COVID-19 pandemic on global eco-tourism (Qiao et al. 2023). Other studies have compared or combined eBird data with other monitoring data to estimate population sizes (Stillman et al. 2023). However, these large databases cannot always substitute for systematic data collection. Shen, Ding, and Tsai (2023) cautioned that eBird data can produce biased estimates, and Kelling et al. (2019) emphasized that eBird data were more useful for monitoring when coupled with details on observation methods (e.g., search effort).

Citizen science data have been critiqued as being spatially and temporally biased (Dimson and Gillespie 2023; but see Geurts et al. 2023), and not all scientists see value in engaging with citizen science (Golumbic et al. 2017). While one main concern from skeptics is data quality, numerous strategies exist to address this (e.g., Freitag, Meyer, and Whiteman 2016; Baker et al. 2021). Because some projects are open ended, they do not generate data equivalent to systematic surveys and thus may not be useful for targeted research and monitoring. However, such anecdotal citizen science sightings may have utility (Wiersma 2010; Parsons, Lukyanenko, and Wiersma 2011). We aim here to provide a narrative overview of the history and development of a regionally focused digital citizen science project. We briefly summarize the original research goals of the site (which focused on data quality), along with key results from past work. Our main focus is to illustrate how our project has also contributed both unanticipated natural history findings and intangible benefits through transference of impacts (Lynch-O'Brien et al. 2021).

BRIEF HISTORY OF NLNATURE.com

In 2009, we founded a citizen science website, NLNature. com, focused specifically on the flora and fauna of the province of Newfoundland and Labrador, Canada (hereafter NL). The initial purpose of NLNature was to examine ways to harness citizen science participation in online data collection (Lukyanenko, Parsons, and Wiersma 2011; Parsons et al. 2011). The province was chosen as a test case as part of a larger Canada-wide study (Tudge et al. 2012); its vast size (405,212 km²) but sparse population (~500,000 people) provided an interesting venue in which to investigate issues of participation, digital divides, and motivations. Approximately half of the population lives outside of the greater capital region, in small, rural (mostly coastal) towns. Many citizens hunt, fish, and forage to supplement their diets. Furthermore, the province of NL is a popular tourist destination; the Department of Tourism estimates 180,000 people visit per year (Government of Newfoundland and Labrador 2021). One attraction for visitors is the unique natural beauty of the province. Thus, while there is a small population of potential citizen scientists, a large proportion of these people spend a great deal of time outdoors and are keen observers of their surroundings. The site experienced its peak activity between 2011 and 2015. We took the site down in early 2022, owing to a lack of resources to keep it consistent with current web standards.

DATA QUALITY DIMENSIONS

The main purpose of NLNature.com was to serve as a vehicle to explore how to design data collection for an effective citizen science project. Although the founder is an ecologist, the intent of the site was never for rigorous data collection for use in ecological studies. As such, the data are "use agnostic" (Lukyanenko, Parsons, and Wiersma 2014). Instead of trying to collect data for a specific use, we designed the site to investigate whether and how we could motivate citizen science participation across a broad geographic region, and specifically, to conduct experiments to test questions about data quality, usability and engagement.

Data quality is one of the primary concerns with citizen science data (Baker et al. 2021), and has multiple dimensions (Wand and Wang 1996; Mesaglio et al. 2023). The common data quality dimensions include accuracy, precision, completeness, and timeliness. Our theoretical framework borrows from the field of information systems (IS), particularly platforms that rely on user-generated content (UGC), which can include product reviews, discussion boards, and citizen science. We focused specifically on the data quality dimensions of accuracy and completeness. Generally, we understood accuracy as the congruence between the labels and descriptions provided by the citizens and the standards for identification that represent state-of-the-art scientific knowledge in biology. Completeness was assessed as the depth and breadth of the descriptions of objects observed, as well as the fraction of objects captured in the database compared with the observations contributors came in contact with (the latter was naturally difficult to quantify, hence we employed estimations and used laboratory experiments with the live NLNature interface to simulate real-world sightings). For example, in a preliminary study, we asked non-biology students to identify images of plants and animals found in the province and to describe their attributes. We found that participants were able to accurately identify the images but only at class levels higher than species (e.g., they could identify something as a bird, but not what species of bird). Detailed results of the preliminary study are reported in Lukyanenko et al. (2014). We used this to inform design choices and rationale in the design of the online experiment (Lukyanenko, Parsons, and Wiersma 2016; Lukyanenko et al. 2017; Lukyanenko and Parsons 2020a; Lukyanenko and Parsons 2020b), which is summarized briefly below. The bulk of the work leading up to this writing made contributions to theories of software development, design science research, conceptual modeling, and data management.

UNANTICIPATED CONTRIBUTIONS

We aim here to investigate what ecological value we realized from the site beyond the contributions to the IS field, as detailed elsewhere (Lukyanenko et al. 2014; Lukyanenko et al. 2016; Lukyanenko et al. 2019). Because we did not design the site with an ecological research question in mind, and we made no attempts to facilitate standardized data collection, the site represents a type of "surveillance monitoring" as opposed to "targeted monitoring" (Wiersma 2010; Wintle, Runge, and Bekessy 2010; Callaghan et al. 2018). Nonetheless, the site yielded some valuable ecological data, which are detailed below. We use these observations to discuss the pros and cons of use-agnostic online citizen science projects.

Beyond the generation of data, citizen science projects can bring intangible benefits to participants. This can include gaining knowledge (e.g., Crall et al. 2012), feeling connected to a community (e.g., Evans et al. 2005; Overdevest et al. 2004), and contributing to a collective challenge (McKinley et al. 2017). Lynch-O'Brien et al. (2021) discuss a theory for transference of impacts and describe five stages through which citizen scientists themselves transfer the impacts of their participation. Generally, individuals who can have impacts beyond a given citizen science project begin with a long-term interest in nature (stage 1), then join a citizen science project (stage 2). Over time, their participation in citizen science causes them to be attributed with expertise by peers (stage 3); they eventually acquire the role of an expert (stage 4) and may in turn influence change in others (stage 5) (Lynch-O'Brien et al. 2021). Through a summary of interviews and focus groups we conducted with NLNature participants, and reflections of the authors, we examine whether and how participants in the NLNature project realized some of these intangible benefits.

METHODS

WEBSITE AND EXPERIMENTAL DESIGN

We designed NLNature.com to test how data quality (specifically accuracy and completeness) varied with different data entry constraints on user contribution. Participants contributed a sighting by clicking a point on a map that corresponded with the position where their sighting was made; this then opened up an interface where they were asked to contribute the date of the sighting and additional sighting information. During initial account setup, participants were randomly assigned to one of two treatments. The online interface for each treatment was nearly identical; the single difference was whether users were constrained to a limited list of species names when identifying their sighting (the class-based interface), or if the field for entering their observation was open ended (the instance-based interface). In the class-based interface, users had to classify their observation based on a predefined list of 343 species (contributed by the first author and updated based on contributions from contributors) known to be extant in the province, or click "unknown." In the instance-based interface, users could describe their sighting any way they wished. We describe the interface design in more detail in Lukyanenko and Parsons (2020a).

We promoted awareness of the site through extensive public outreach and media appearances between 2009 and 2010. We also further promoted the site via speaking tour across the province in 2015 2012, which targeted tourism businesses and interest groups (e.g., photography clubs, outdoors clubs, and naturalist groups).

To assess accuracy, we carried out a laboratory experiment similar to our initial studies, where we showed images of species from the province on the screen. Instead of pencil-and-paper responses, these users had to create an account on NLNature, and identify the species shown on the screen as best as possible using the online interface. The participants in the study were all non-biology majors. On a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree), participants in the experiments gave themselves a mean rank of 2.2 (s.d. = 1.36) in response to the statement "I consider myself an expert in local (NL) flora, fauna and wildlife," and 4.0 (s.d. = 1.77) in response to the statement "I am familiar with Newfoundland and Labrador's plants and animals." Additional information about the design of this experiment is detailed in Lukyanenko et al. (2019).

DATA ANALYSIS

We analyzed the number of contributions made by participants by summarizing all observations contributed in each treatment (class-based and instance-based interfaces) after removing any non-species observations (e.g., "iceberg") over a six-month period beginning May 1, 2013. We tested for significant difference in number of observations and number of novel contributions (i.e., unanticipated species that were not in the initial drop-down list of species expected to be present in the province) in the instance- versus class-based conditions using an exact permutation test. We assessed accuracy in the classroom experiment using MANCOVA to test for experimental differences between experimental groupings (degree of familiarity of species) and conditions (instance- versus class-based). See Lukyanenko et al. (2019) for details.

FOCUS GROUPS

In 2015, we conducted individual interviews (n = 8) and focus group discussions (n = 2) with contributors to NLNature. We ran one focus group (5 participants) in the same city as the university campus, and one (2 participants) in a small town 2 hours outside the city. Of the 15 total participants, 5 identified as female and 10 as male. When

asked about profession/occupation, 8 reported that they worked outside of biology/ecology; 5 said they worked in biology/ecology, and 3 did not provide information on occupation. The purpose of this was to conduct a qualitative analysis of responses to determine contributors' motivations for participation, their perceptions of data quality on the site, and the applications to which they thought the data on NLNature could be used. A copy of the interview instrument is provided in the Supplemental File 1: Appendix A, and further details on interview methods and interviews were recorded and transcribed, and we qualitatively analyzed the text of the responses to assess for patterns as well as differences in responses.

RESULTS

Between 2009 and 2022, just over 4,000 people signed on to NLNature. However, not all who joined the site contributed a sighting. By January 20, 2022, participants had contributed 10,675 sightings in total. As expected, and consistent with other citizen science data, the contributions were long-tailed (Figure 1), with only a few participants contributing the majority of the observations (Szabo, Forti, and Callaghan 2023). The top five participants in NLNature contributed 60% of all observations, with the top three participants (which includes the second and third authors), contributing 53% of all observations. The second author (TC) contributed the highest number of sightings (2,763 total), including most of the ones described here.

DATA QUALITY

During the six-month online experiment on data quality (May-October 2013), there were 42 participants in the class-based treatment and 39 in the instance-based treatment. Our analysis of data quality showed that when forced to choose a species from a dropdown list, there were significantly fewer observations (87 in the classbased and 390 in the instance-based, p = 0.033) and fewer unanticipated observations (7 in the class-based and 119 in the instance-based, p = 0.007) (Figure 2). In at least six instances, users abandoned a post when they system did not accept their initial entry (analyzed via data logs). This suggests that contributors in the instance-based case might have been less confident in their species identification and defaulted to unknown or simply did not contribute rather than provide a guess that might be erroneous. However, we cannot verify this across all sightings, as not all contributors appended a photograph, and even if they did, not all species can be identified from a photograph. Thus, we rely on the lab experiment to assess the data quality dimension of

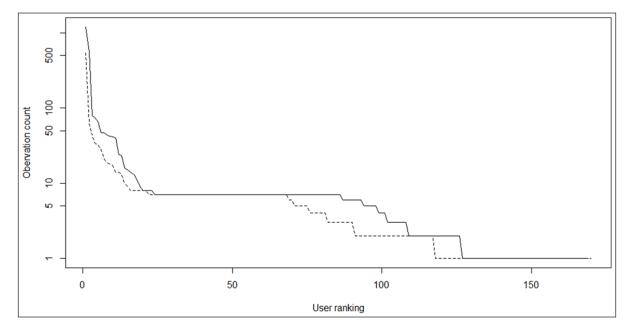


Figure 1 Count of observations (y-axis) to NLNature.com per user, ranked by user (x-axis). Solid line is observations from contributions from observers in the instance-based condition, and dashed line is contributions from observers in the class-based condition. The long-tailed distribution is typical for most citizen science projects.

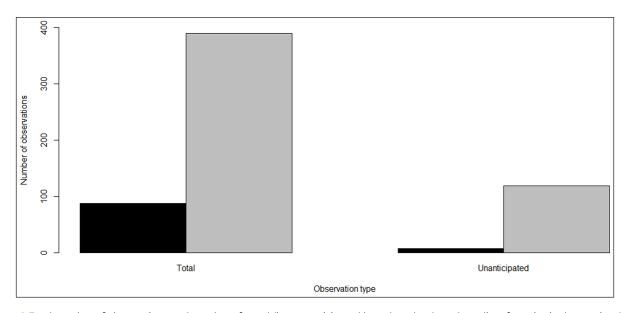


Figure 2 Total number of observations and number of novel (i.e., unanticipated based on the drop-down list of species in the province) observations contributed to NLNature by users in the class-based condition (*black bars*) and instance-based condition (*grey bars*). An exact permutation test showed that these were significantly fewer observations in the class-based condition (p = 0.033 for total number of observations and p = 0.007 for novel observations).

accuracy and compare this between instance- and classbased data collection methods.

In the lab experiment, for which we knew the correct identity of the species, there was a significant difference in accuracy; the instance-based condition had higher accuracy than the class-based one (F = 156.6, p < 0.001), but significantly lower precision (F = 53.9, p < 0.001). This is because in the instance-based condition participants

could identify at the taxonomic level they felt confident (thus we coded responses such as "gull" as correct for the image of the herring gull), but in the class-based condition, participants had to select from a drop-down species list that included many gulls (and other birds), forcing them to guess if they were not confident of the species identity. Further details on these experiments are provided in Lukyanenko et al. (2019). Despite the limitations, the site did yield several valuable ecological observations. These include a record of a new species to the province (Fielden et al. 2015); sightings that expand the known range of two species in the province, including a species ranked as "imperiled" (S-2) (Clenche 2020); and a possible new species to science. We summarize these findings in more detail below.

UNANTICIPATED CONTRIBUTIONS

Record of a new species to the province

The third author (ME) joined NLNature in 2013. In September 2013, he noticed a mosquito with unusual leqband pattern in his backyard, about 1.4 km outside the capital city. He managed to take a photograph of it on his arm (Figure 3). It was the species Aedes japonicas japonicas, which researchers had first detected in Canada in 2001, but had not yet found in Newfoundland and Labrador. The species is native to eastern Asia and is a vector for West Nile virus in other parts of North America. ME asked the administrators of NLNature if the mosquito he had photographed was a potential disease vector. We passed the information to colleagues with mosquito expertise who were able to trap more individuals in the capital region and verify its presence. ME, along with entomology experts at Memorial University of Newfoundland, published a note describing this new species record for the province (Fielden et al. 2015).

Expansions of species distribution

In 2018, the second author (TC, member since 2013) was walking along the side of a favourite swimming location (48.115000 N, -53.751667 W) in rural Newfoundland and



Figure 3 Aedes japonicas, a mosquito species native to eastern Asia, and discovered for the first time on the island of Newfoundland by a participant on the citizen science website NLNature.com. Photo by Mardon Erbland taken 24 September 2013 in the vicinity of Devereaux Lane, Outer Cove, NL.

noticed an interesting-looking aquatic plant. He keyed it out as *Nymphoides cordata* (Little Floating-heart), and had the identification confirmed by a plant expert at the botanical gardens in St. John's (Figure 4). This is an S-2 ranked ("imperiled") species that was previously known from only eight localities on the island of Newfoundland, the nearest of which was 73.5 km from where TC discovered it. This greatly expands the known distribution of a rare species in the province. Since the initial discovery, TC found it in four other locations within the same watershed (Clenche 2020). His notation of such a significant finding led the lead author to notice it in a pond in a different watershed ~12 km (48.150556 N, -53.895000 W) from where TC discovered it.

In August 2016, TC was on a fishing trip in Labrador, and found an unusual freshwater clam in a river downstream from Igloo Lake (53.051038 N, -58.732001 W). He posted to NLNature, and alerted a local mollusk expert, who identified it as the Newfoundland floater clam (Pyganodon fragilis); this was the first record for this species anywhere within the Eagle River Watershed in Labrador. NatureServe lists it as S-2 ("imperiled") in Labrador and "apparently stable" (S-4) on the island of Newfoundland. In August 2020, he found the same species in a pond on the island of Newfoundland (Shoal Pond, Carmanville; 49.368803 N, -54.164032 W) (Figure 5); again, the first record for that waterbody. He was with his father-in-law at the time, also an avid outdoorsperson. This led his father-inlaw to find it in a different pond (Weirs Pond, Bonavista North; 49.218276 N, -54.382719W), also the first known occurrence of this species in this area. Thus, the sightings facilitated by NLNature.com have expanded our knowledge of the distribution of this species, which is currently confounded by uncertainty over its taxonomy (Martel et al. 2010) and a lack of survey effort. This species is likely quite common in the Atlantic region, but its distinctive features are noticeable only to someone with a keen eye and an interest in collecting bivalves. Thus citizen science observations could yield specimens useful to resolve the taxonomic uncertainty (Martel et al. 2010). Interestingly, another outdoorsperson, William Larkham Jr., also found P. fragilis in Labrador in 2020 and noted that it looked different from other freshwater bivalves. Larkham details his finding in two YouTube videos; the comments in the videos suggest that others think they have this species in their local watersheds as well.

Possible new species to science

TC also posted a wasp sighting that was unlike anything he had seen before. An entomologist in Europe, who believed it could be a new species of cuckoo wasp, contacted him. However, scientists cannot verify whether this is a new species without a specimen in hand. Despite keeping a



Figure 4 (a) The aquatic plant *Nymphoides cordata* and **(b)** its habitat as found 8 August 2018. The S-2 ranked species was only known from eight localities in the province until a participant on NLNature.com found it in several new locations. Photos by Tom Clenche.

close look out, and even setting insect traps, no one has reported seeing this organism again, and hence it remains an intriguing mystery.

IMPACTS BEYOND NLNature

Lynch-O'Brien et al. (2021) proposed a theory of transference of citizen science impacts by which participants can influence those around them and raise awareness of a particular issue. They documented its presence in an entomology citizen science project (Lynch-O'Brien et al. 2021). There is evidence that NLNature achieved transference of impacts in several cases. For example, TC was able to show his father-in-law the floater clam he found in Labrador, who then noted it in another pond; and TC's documentation of the Little Floating-heart led two others to notice its presence in two different watersheds. This is an example of citizen scientists becoming recognized as experts by their peers (stage 3 in Lynch-O'Brien et al.'s

2021 framework). Both professional entomologists and the natural history museum's curator emeritus have contacted TC about his findings, illustrating that he is acquiring the role of a local expert (stage 4 in Lynch-O'Brien et al. 2021). In a similar vein, ME is also recognized as a local expert. He notes that two of his observations introduced him to the world of amphipods and cyclopid copepods (stages 1 and 2 in Lynch-O'Brien et al. 2021). This spurred his current goal of finding DNA Barcodes for a sampling of freshwater crustaceans on the Northeast Avalon, including ostracods, anomopods, amphipods, and copepods. He recently collected amphipods for a doctoral candidate (stages 3 and 4 in Lynch-O'Brien et al. 2021) at Miami University in Ohio; DNA sequencing showed that two of his specimens were Hyalella wellborni, likely the first record of this species in the province.

The analysis of the 15 respondents from the focus groups and interviews provided evidence that many of the citizen scientists who interacted with NLNature passed through at least a few of the 5 stages for transference of impacts identified by Lynch-O'Brien et al. (2021). For example, when asked why they were motivated to participate, two spoke specifically of a long-term interest in the natural world, which Lynch-O'Brien et al. (2021) see as the first stage in their theory (Meldt 2017). Eight expressed that enjoyment of outdoor activities prompted them to interact with NLNature. Six expressed a motivation to join the citizen science project (stage 2 in Lynch-O'Brien et al.'s 2021 framework) in order to learn more about the natural world around them. Specifically, participants spoke about wanting to learn to identify species they did not know (Drechsel 2017; Lukyanenko et al. 2017; Meldt 2017), sharing their own observations with other people, and learning about the biodiversity of their home province (Meldt 2017). As well, they felt their data could be useful to track changes in the environment, especially unanticipated ones (Lukyanenko et al. 2017). Participants also noted that they felt that through active participation, they had earned the label "citizen scientist," (stage 3 in Lynch-O'Brien et al. 2021), even if they had not identified themselves as such initially (Meldt 2017). Interestingly, at least three interview subjects expressed that did not feel they were necessarily credentialed as experts in the eyes of their peers on the site, but did see others on the site as experts, sometimes pointing to specific users as key participants they watched and learned from (Meldt 2017). Only two felt that they had acquired the roles of an expert (stage 4 in Lynch-O'Brien et al 2021), although some acknowledged that they had professional expertise in some domains of natural history (e.g., birds) but enjoyed participating on the site to learn about other taxa in which they did not have formal training (Drechsel 2017). In terms of using involvement in NLNature

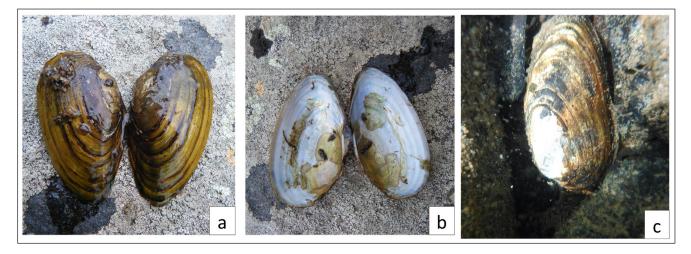


Figure 5 (a) Outer and (b) inner surface of a specimen of the Newfoundland floater clam *Pyganodon fragilis* found downstream from Igloo Lake, Labrador, NL, 23 August 2016. (c) *Pyganodon fragilis* found in Shoal Pond 23 August 2020. Photos by Tom Clenche.

to influence change in others (stage 5 in Lynch-O'Brien et al. 2021), the main motivation participants had was to interact with others (both on the site and with experts who might be looking at the sightings contributed by citizens) to influences changes. Respondents expressed a high degree of motivation to contribute to research and to protect the natural beauty of the province (Meldt 2017).

DISCUSSION

Our previous publications summarize in more detail the value of a use-agnostic approach to a citizen science project from the perspective of data quality. These findings are similar to those of Serret et al. (2019), who compared contributions from two different citizen science pollinator monitoring projects and found that there were more contributions per person in the more open-ended condition. Unlike Serret et al.'s project (2019), NLNature was not designed with an ecological question in mind. Despite this, the site did yield some valuable ecological observations. Not surprisingly, the most active contributors made the most significant findings; those who participated only a handful of times generally contributed only highly familiar species, or common species that they wanted the community's help identifying. This illustrates that serendipitous natural history discoveries usually require a keen observer who spends a great deal of time outdoors and who thus can recognize when something is unusual or notable. Moreover, most citizen scientists are most familiar with their local area—as an example, the majority of TC's observations are within a 15 km radius of his home, and ME found the mosquito in his backyard. Thus, citizen scientists can contribute highly accurate and extensive data over time, but usually more limited in space than academic or government scientists, who have the resources and personnel to coordinate large-scale surveys.

While writing this paper, we discovered the YouTube videos documenting the floater clam. The comments on the videos show there is interest in these species (mostly motivated by whether they are edible or not). Such comments on amateur naturalist videos could be a useful source of data, but could be difficult to extract into a usable format, although perhaps AI tools could be harnessed (e.g., Wood et al. 2022). As well, there is uncertainty about spatial and temporal references, and what can be inferred from generic comments like "we have those around here," and about species identification (for example, comments suggest two incorrect species identities in Larkham's first video). Nonetheless, systematic, detailed videos that clearly show anatomical structure and habitat, and have several specimens of different sizes in hand, such as in Larkham's videos, have the potential to be another form of citizen science data useful for documenting species occurrences.

The data from NLNature are use-agnostic. However, by 2015, the site was active and high enough in profile in the provincial natural history community that we were approached by a research scientist at the federal Department of Fisheries and Oceans (DFO) (C. Bourne, pers. comm.) to see if the users of the site could be rallied to contribute observations of the capelin spawning events. Capelin are a small pelagic fish that spawn on beaches in NL annually sometime between June and August. In small coastal communities, this is a major community event, as the fish is valued as both a food source and for garden fertilizer. We sent out several email messages to all NLNature members asking them to contribute capelin spawning sightings in the spring/summer of 2016. This was successful enough that the following year, DFO created its own citizen science site to collect capelin data (ecapelin.ca). Thus, we were able to harness our use-agnostic project for a purpose-driven targeted monitoring program.

Novel and unanticipated sightings of scientific phenomena have value to research, and ours is not the only citizen science project to yield such findings (Straub 2016). Our initial experiment showed that allowing citizen scientists to contribute data at the level of taxonomic resolution with which they felt comfortable yielded better data quality in terms of accuracy and completeness than when they were forced to classify to species level. Our unanticipated contributions all came from participants in the instance-based treatment. However, use-agnostic citizen science data, such as iNaturalist and NLNature, have disadvantages as well. As we saw with our findings here, sightings tend to be biased to localities close to the observer's home, or to heavily used recreation areas (Di Cecco et al. 2021; although see Geurts et al. 2023). Thus, sampling may be too uneven to be useful for applications like species distribution modelling (Drew, Wiersma, and Huettmann 2011; although see Feng and Lougheed 2023; Redolfi De Zan et al. 2023; and Serniak et al. 2023 for examples of species distribution models developed with citizen science data). As well, the regional focus of NLNature proved in the end to be somewhat of a limitation. TC notes that when he posted an unknown sighting, he did not easily get help with species identification unless he specifically targeted an expert within the province. Since joining iNaturalist, he has discovered that he is linked to a global community of natural history experts and enthusiasts, and has received assistance with species identification from Canada, the United States, Russia, and Australia. The individuals he is interacting with on iNaturalist would not be on NLNature, thus the wider scope has advantages. As coordinators of the project, the small population size of the province made it difficult to generate high numbers of sightings. Some focus group and interview participants also saw the restricted focus to the province as a disadvantage, while also acknowledging the value of local knowledge over a long period of time as contributing to citizen science, particularly around climate change and other anthropogenic stressors (Drechsel 2017).

Ultimately, the site was not yielding sufficiently useful data to ecologists to warrant its upkeep, and we did not have the resources to keep the website up to date with current web standards. When we first launched NLNature, mobile technology was not ubiquitous, and rural Newfoundland had limited cellular cover (parts of it still do). While setting up a website to collect anecdotal data appears simple and low cost, we learned from the 14 years that NLNature was active that there are costs, both

to building and maintaining the site and to promoting its use and keeping users engaged. The expansion of the much larger and well-resourced iNaturalist platform, along with high-quality cameras and GPS on most mobile phones, led us to take the NLNature site down and encourage its contributors to post their sightings on iNaturalist. Both TC and ME continue to contribute natural history observations to iNaturalist (further evidence of the transference of impact of NLNature, sensu Lynch-O'Brien et al. 2021). For example, one of TC's videos of sandlance spawning contributed on iNaturalist created excitement in the fisheries community, as this is a rarely observed phenomenon. Both TC and ME are in the top ten contributors on two iNaturalist projects: the Biodiversity of Newfoundland and Labrador, and Nature Newfoundland and Labrador. In addition, ME contributes to a specialized project on the Arachnids and Myriapods of North America (where he is ranked 129 of 6,145 in terms of contributions as of June 28, 2023), and TC contributes to a Canadawide invasive species detection project (where he is ranked 133 of 911 as of June 28, 2023).

Our case study shows that, when designing open-ended citizen science projects (i.e., those not motivated by a specific research question), there are sampling and design issues realized post hoc that a targeted program might have wished to address at the outset. For example, with NLNature, if we had active observers equally distributed geographically in the project, there would likely be similar novel and unanticipated discoveries made elsewhere in addition to those documented here. In another example, the BirdNET project allowed a broad suite of citizens to collect bird songs on their phones and use AI to evaluate the calls. However, the frequencies that the AI could evaluate were limited to bird song. In contrast, the Dawn Chorus project (dawn-chorus.org), while designed to capture bird song, was more open ended and allowed for recording of other sounds (e.g., insects, amphibians), even though it was designed for a project focused on birds. Thus, open-endedness in sampling design (in Dawn Chorus, not limiting the frequency of recordings that could be analyzed) yielded more unanticipated information, consistent with our hypothesis that use-agnostic projects may yield more unanticipated insights.

CONCLUSIONS

Our case study of the citizen science project NLNature suggests two design characteristics made it a pioneering project in citizen science. These include its explicit useagnostic design, which facilitated research outcomes that could not be foreseen at the project outset, and its emphasis on providing a range of input options (photos, attributes, instances, classification at any level). These attributes have been adapted in at least two new citizen science projects of which we are aware (Dawn Chorus and Bio-O-Ton). In our opinion, the largest value of NLNature was the creation of an online community of people interested in natural history who were excited about sharing sightings and learning to identify things. Other case studies of novel/unanticipated discoveries by citizen scientists point to the creation of an online community, for example via discussion forums, as instrumental (Straub 2016). Some of these values have continued, through participants from NLNature contributing to other citizen science projects. In sum, the NLNature citizen science project yielded contributions towards education, research, and peer-to-peer and peerto-expert networking.

SUPPLEMENTARY FILE

The Supplementary file for this article can be found as follows:

• **Supplemental File 1.** Appendix A, Interviews and Focus Groups Protocol Details. DOI: https://doi.org/10.5334/cstp.676.s1

ETHICS AND CONSENT

This research was carried out under the authorization of Memorial University's Interdisciplinary Committee on Ethics in Human Research (certificates 2013-0257-BA and 2018-0994 YW).

ACKNOWLEDGEMENTS

We thank Daria Lukyanenko for her assistance in conducting the field and laboratory experiments. We are especially grateful to the hundreds of citizens who contributed sightings to NLNature.com over the years and to the organizations and individuals who helped to promote the site.

FUNDING INFORMATION

Funding was supported by grants from the Natural Sciences and Engineering Research Council of Canada to YFW and JP, and NSERC Doctoral Fellowship to RL, a grant from the Social Sciences and Humanities Research Council of Canada to JP, funding from the GEOIDE Network of Centres of Excellence Canada to YFW, and from the Institute for Data Valorisation (IVADO), and Memorial University's Harris Centre to RL.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

RL built, maintained, and promoted NLNature.com with advisement from JP and YFW. TC and ME contributed the unanticipated sightings reported here. GW and YFW carried out the interviews and focus groups; GW supervised the qualitative analysis. YFW wrote the manuscript and all authors reviewed the content and provided input on the writing.

AUTHOR AFFILIATIONS

Yolanda F. Wiersma D orcid.org/0000-0003-4604-9240 Department of Biology, Memorial University, St. John's, Newfoundland and Labrador, CA

Tom Clenche

Petley, Random Island, Newfoundland and Labrador, CA

Mardon Erbland Logy Bay, Newfoundland and Labrador, CA

Gisela Wachinger

DIALOGIK gemeinnützige Gesellschaft für Kommunikations- und Kooperationsforschung mbH, Stuttgart, DE

Roman Lukyanenko crcid.org/0000-0002-2776-9640 Faculty of Business Administration, Memorial University, St. John's, Newfoundland and Labrador, CA; McIntire School of Commerce, University of Virginia, US

Jeffrey Parsons D orcid.org/0000-0002-4819-2801 Faculty of Business Administration, Memorial University, St. John's, Newfoundland and Labrador, CA

REFERENCES

- 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. DOI: https://doi.org/10.5334/cstp.351
- Callaghan, CT, Martin, JM, Major, RE and Kingsford, RT. 2018. Avian monitoring – comparing structured and unstructured citizen science. *Wildlife Research*, 45(2): 176–184. DOI: https://doi.org/10.1071/WR17141
- Campioni, L, Madeiros, J and Becciu, P. 2023. Citizen science data reveal possible multi-decadal phenological changes in

the arrival time of a migratory tropical seabird species at the breeding ground. *Marine Biology*, 170(10): 124. DOI: https://doi.org/10.1007/s00227-023-04268-3

- **Clenche, T.** 2020. Little floating-heart (*Nymphoides cordata*) on Random Island. *Sarracenia*, 23(2): 13–16.
- Crall, AW, Jordan, R, Holfelder, K, Newman, GJ, Graham, J and Waller, DM. 2012. The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Public Understanding of Science*, 22: 745–764. DOI: https://doi.org/10.1177/0963662511434894
- Di Cecco, GJ, Barve, V, Belitz, MW, Stucky, BJ, Guralnick, RP and Hurlbert, AH. 2021. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, 71(11): 1179–1188. DOI: https://doi.org/10.1093/biosci/biab093
- Dimson, M and Gillespie, TW. 2023. Who, where, when: Observer behavior influences spatial and temporal patterns of iNaturalist participation. *Applied Geography*, 153: 102916. DOI: https://doi.org/10.1016/j.apgeog.2023.102916
- Drechsel, S. 2017. Wie kann Citizen Science für die Beteiligten zufriedenstellend funktionieren? [trans: How can Citizen Science work satisfactorily for those involved?] – Unpublished thesis (PhD), University of Stuttgart, Stuttgart, Germany.
- Drew, CA, Wiersma, YF and Huettmann, F. 2011. Predictive species and habitat modeling in landscape ecology: Concepts and applications. New York: Springer. DOI: https://doi. org/10.1007/978-1-4419-7390-0
- Evans, C, Abrams, E, Reitsma, R, Roux, K, Salmonsen, L and Marra, PP. 2005. The neighborhood nestwatch program: particpatn outcomes of a citzen-science ecolgoical research project. *Conservation Biology*, 19: 589–594. DOI: https://doi. org/10.1111/j.1523-1739.2005.00s01.x
- Feng, W and Lougheed, SC. 2023. Integrating eDNA and citizen science observations to model distribution of a temperate freshwater turtle near its northern range limit. *PeerJ*, 11. DOI: https://doi.org/10.7717/peerj.15120
- Fielden, MA, Chaulk, AC, Bassett, K, Wiersma, YF, Erbland, M, Whitney, H and Chapman, TW. 2015. Aedes japonicus japonicus (Diptera: Culicidae) arrives at the most easterly point in North America. Canadian Entomologist, 147(6): 737– 740. DOI: https://doi.org/10.4039/tce.2015.5
- Freitag, A, Meyer, R and Whiteman, L. 2016. Strategies employed by citizen science programs to increase the credibility of their data. *Citizen Science: Theory and Practice*, 1(1): 2. DOI: https://doi.org/10.5334/cstp.6
- Gaier, AG and Resasco, J. 2023. Does adding community science observations to museum records improve distribution modeling of a rare endemic plant? *Ecosphere*, 14(3): e4419. DOI: https://doi.org/10.1002/ecs2.4419
- Geurts, EM, Reynolds, JD and Starzomski, BM. 2023. Not all who wander are lost: Trail bias in community science. *PLoS*

ONE, 18(6): 0287150. DOI: https://doi.org/10.1371/journal. pone.0287150

- Golumbic, YN, Orr, D, Baram-Tsabari, A and Fishbain, B. 2017. Between vision and reality: a study of scientists' views on citizen science. *Citizen Science: Theory and Practice*, 2(1): 6. DOI: https://doi.org/10.5334/cstp.53
- Government of Newfoundland and Labrador. 2021. Tourism snapshot. Accessed from https://www.gov.nl.ca/tcar/ files/Tourism-Monthly-Snapshot_Infographic-YTD-DECEMBER-2021_FINAL.pdf.
- Kelling, S, Johnston, A, Bonn, A, Fink, D, Ruiz-Gutierrez, V, Bonney, R, Fernandez, M, Hochachka, WM, Julliard, R, Kraemer, R and Guralnick, R. 2019. Using semistructured surveys to improve citizen science data for monitoring biodiversity. *BioScience*, 69(3): 170–179. DOI: https://doi. org/10.1093/biosci/biz010
- Liu, J, Slik, F, Zheng, S and Lindenmayer, DB. 2022. Undescribed species have higher extinction risk than known species. *Conservation Letters*, 15(3): e12876. DOI: https://doi. org/10.1111/conl.12876
- Lukyanenko, R and Parsons, J. 2020a. Design theory indeterminacy: what is it, how can it be reduced, and why did the polar bear drown? *Journal of the Association for Information Systems*, 21(5): 1. DOI: https://doi. org/10.17705/1jais.00639
- Lukyanenko, R and Parsons, J. 2020b. Easier crowdsourcing is better: designing crowdsourcing systems to increase information quality and user participation. In: vom Brocke, J, Hevner, A, Maedche, A (eds.), *Design Science Research Cases*. Cham: Springer. pp. 43–72. DOI: https://doi.org/10.1007/978-3-030-46781-4_3
- Lukyanenko, R, Parsons, J and Wiersma, Y. 2011. Citizen science 2.0: Data management principles to harness the power of the crowd, *Lecture Notes in Computer Science*, 6629: 465–473. DOI: https://doi.org/10.1007/978-3-642-20633-7_34
- Lukyanenko, R, Parsons, J and Wiersma, YF. 2014. The IQ of the crowd: Understanding and improving information quality in structured user-generated content. *Information Systems Research*, 25(4): 669–689. DOI: https://doi.org/10.1287/ isre.2014.0537
- Lukyanenko, R, Parsons, J and Wiersma, YF. 2016. Emerging problems of data quality in citizen science. *Conservation Biology*, 30(3): 447–449. DOI: https://doi.org/10.1111/ cobi.12706
- Lukyanenko, R, Parsons, J, Wiersma, YF and Maddah, M. 2019. Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced usergenerated content. *MIS Quarterly: Management Information Systems*, 43(2): 623–647. DOI: https://doi.org/10.25300/ MISQ/2019/14439

 Lukyanenko, R, Parsons, J, Wiersma, YF, Sieber, R and Maddah,
 M. 2016. Participatory design for user-generated content: Understanding the challenges and moving forward.
 Scandinavian Journal of Information Systems, 28(1): 2.

Lukyanenko, R, Wiersma, Y, Huber, B, Parsons, J, Wachinger, G and Meldt, R. 2017. Representing crowd knowledge: Guidelines for conceptual modeling of user-generated content. *Journal of the Association of Information Systems*, 18(4): 2. DOI: https://doi.org/10.17705/1jais.00456

Lynch-O'Brien, LI, Babchuk, WA, Dauer, JM, Heng-Moss, T and Golick, D. 2021. Transference of citizen science program impacts: A theory grounded in public participation in scientific research. *Diversity*, 13(8): 339. DOI: https://doi. org/10.3390/d13080339

- Martel, AL, McApline, DF, Madill, JB, Sabine, DL, Paquet, A, Pulsifer, MD and Elderkin, MF. 2010. Freshwater mussels (Bivalvia: Margaritiferidae, Unionidae) of the Atlantic Maritime Ecozone. In McAlpine, DF and Smith, IM (eds.), Assessment of Species Diversity in the Atlantic Maritime Ecozone. Ottawa: NRC Research Press. pp. 551–598.
- Martin, TE, Bennett, GC, Fairbairn, A and Mooers, AO. 2023. "Lost" taxa and their conservation implications. *Animal Conservation*, 26(1): 14–24. DOI: https://doi.org/10.1111/ acv.12788
- McKinley, DC, Miller-Rushing, AJ, Ballard, HL, Bonney, R, Brown, H, Cook-Patton, SC, Evans, DM, French, RA, Parrish, JK, Phillips, TB, Ryan, SF, Shanley, LA, Shirk, JL, Stepenuck, KF, Weltzin, JF, Wiggins, A, Boyle, OD, Briggs, RD, Chapin, SF, Hewitt, DA, Preuss, PW and Soukup, MA. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, 208: 15–28. DOI: https://doi.org/10.1016/j. biocon.2016.05.015
- Meldt, R. 2017. Motive von Citizen Scientists: eine qualitative Untersuchung von Interviews mit Nutzern von NLNature [transl: Motivatins of citizen scientists: a qualitative analysis of interviews with NLNature users] – Unpublished thesis (PhD), University of Stuttgart, Stuttgart, Germany.
- Mesaglio, T, Callaghan, CT, Samonte, F, Gorta, SB and
 Cornwell, WK. 2023. Recognition and completeness: two key metrics for judging the utility of citizen science data.
 Frontiers in Ecology and the Environment, 21(4): 167–174.
 DOI: https://doi.org/10.1002/fee.2604
- Moura, MR and Jetz, W. 2021. Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nature Ecology and Evolution*, 5(5): 631–639. DOI: https://doi.org/10.1038/ s41559-021-01411-5
- O'Neill, D, Häkkinen, H, Neumann, J, Shaffrey, L, Cheffings, C, Norris, K and Pettorelli, N. 2023. Investigating the potential of social media and citizen science data to track changes in species' distributions. *Ecology and Evolution*, 13(5): e10063. DOI: https://doi.org/10.1002/ece3.10063

- **Overdevest, C, Orr, CH** and **Stepenuck, K.** 2004. Volunteer stream monitoring and local participation in natural resource issues. *Human Ecological Reviews*, 11: 177–185.
- Parsons, J, Lukyanenko, R and Wiersma, Y. 2011. Easier citizen science is better. *Nature*, 471(7336), 37. DOI: https://doi.org/10.1038/471037a
- Qiao, H, Orr, M, Yang, Q, Zhan, X, Lei, F and Hughes,
 AC. 2023. Global birdwatching data reveal uneven consequences of the COVID-19 pandemic. *Biological Conservation*, 288: 110351. DOI: https://doi.org/10.1016/j. biocon.2023.110351
- Redolfi De Zan, L, Rossi de Gasperis, S, Andriani, V, Bardiani, M, Campanaro, A, Gisondi, S, Hardersen, S, Maurizi, E, Mosconi, F, Nardi, G, Zapponi, L, Rombolà, P and Romiti, F. 2023. The big five: species distribution models from citizen science data as tool for preserving the largest protected saproxylic beetles in Italy. *Diversity*, 15(1): 96. DOI: https:// doi.org/10.3390/d15010096
- Serniak, LT, Chan, SS and Lajtha, K. 2023. Predicting habitat suitability for Amynthas spp. in the United States: a retrospective analysis using citizen science data from iNaturalist. *Biological Invasions*, 25(3): 817–825. DOI: https://doi.org/10.1007/s10530-022-02947-8
- Serret, H, Deguines, N, Jang, Y, Lois, G and Juillard, R. 2019. Data quality and participant engagement in citizen science: comparing two approaches for monitoring pollinators in France and South Korea. *Citizen Science: Theory and Practice*, 4(1): 22. DOI: https://doi.org/10.5334/ cstp.200
- Shen, F-Y, Ding, T-S and Tsai, J-S. 2023. Comparing avian species richness estimates from structured and semi-structured citizen science data. *Scientific Reports*, 13(1): 1214. DOI: https://doi.org/10.1038/s41598-023-28064-7
- Stillman, AN, Howell, PE, Zimmerman, GS, Bjerre, ER, Millsap, BA, Robinson, OJ, Fink, D, Stuber, EF and Ruiz-Gutierrez, V. 2023. Leveraging the strengths of citizen science and structured surveys to achieve scalable inference on population size. *Journal of Applied Ecology*, 60(11): 2389–2399. DOI: https://doi.org/10.1111/1365-2664.14512
- Straub, M. 2016. Giving citizen scientists a chance: a study of volunteer-led scientific discovery. *Citizen Science: Theory and Practice*, 1(1): 5. DOI: https://doi.org/10.5334/ cstp.40
- Szabo, JK, Forti, LR and Callaghan, CT. 2023. Large biodiversity datasets conform to Benford's law: Implications for assessing sampling heterogeneity. *Biological Conservation*, 280: 109982. DOI: https://doi.org/10.1016/j.biocon.2023.109982
- Tudge, P, Sieber, R, Wiersma, YF, Corbett, J, Chung, S, Allen, P and Robinson, P. 2012. Collaborating towards innovation: lessons from the participatory GeoWeb GEOIDE network. In Chrisman, N and Wachowicz, M (eds.), The Added Value of Scientific Networks. GEOIDE: Quebec City. pp. 103–118.

- US Bureau of Labor Statistics. 2022. Accessed from https://www. bls.gov/ces/data/.
- Wand, Y and Wang, RY. 1996. Anchoring data quality dimensions in ontological foundations. *Communications* of the ACM, 39(11): 86–95. DOI: https://doi. org/10.1145/240455.240479
- Wiersma, YF. 2010. Birding 2.0: Citizen science and effective monitoring in the web 2.0 world. Avian Conservation and Ecology, 5(2): 1–9. DOI: https://doi.org/10.5751/ACE-00427-050213
- Wintle, DA, Runge, MC and Bekessy, SA. 2010. Allocating monitoring effort in the face of unknown

unknowns. *Ecology Letters*, 13: 1325–1337. DOI: https://doi. org/10.1111/j.1461-0248.2010.01514.x

- Wittmann, J, Girman, D and Crocker, D. 2019. Using iNaturalist in a coverboard protocol to measure data quality: suggestions for project design. *Citizen Science: Theory and Practice*, 4(1): 21. DOI: https://doi.org/10.5334/ cstp.131
- Wood, CM, Kahl, S, Rahaman, A and Klinck, H. 2022. The machine learning-powered BirdNET app reduces barriers to global bird research by enabling citizen science participation. *PLoS Biology*, 20(6): e3001670. DOI: https://doi.org/10.1371/ journal.pbio.3001670

TO CITE THIS ARTICLE:

Wiersma, YF, Clenche, T, Erbland, M, Wachinger, G, Lukyanenko, R and Parsons, J. 2024. Advantages and Drawbacks of Open-Ended, Use-Agnostic Citizen Science Data Collection: A Case Study. *Citizen Science: Theory and Practice*, 9(1): 5, pp. 1–13. DOI: https://doi.org/10.5334/ cstp.676

Submitted: 28 August 2023 Accepted: 17 January 2024 Published: 22 February 2024

COPYRIGHT:

© 2024 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.

