Biological Conservation xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/bioc

The science of citizen science: Exploring barriers to use as a primary research tool

H.K. Burgess [a,*,1], L.B. DeBey [b,1], H.E. Froehlich [a,2], N. Schmidt [c], E.J. Theobald [b], A.K. Ettinger [b,3], J. HilleRisLambers [b], J. Tewksbury [b], J.K. Parrish [a]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington 98105, USA
[b] Biology Department, University of Washington, Seattle, Washington 98195, USA
[c] School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 16 April 2016
Accepted 23 May 2016
Available online xxxx

Keywords:
Biodiversity
Citizen science
Data quality
Outreach
Public participation in science
Research

ABSTRACT

Biodiversity citizen science projects are growing in number, size, and scope, and are gaining recognition as valuable data sources that build public engagement. Yet publication rates indicate that citizen science is still infrequently used as a primary tool for conservation research and the causes of this apparent disconnect have not been quantitatively evaluated. To uncover the barriers to the use of citizen science as a research tool, we surveyed professional biodiversity scientists ($n = 423$) and citizen science project managers ($n = 125$). We conducted three analyses using non-parametric recursive modeling (random forest), using questions that addressed: scientists' perceptions and preferences regarding citizen science, scientists' requirements for their own data, and the actual practices of citizen science projects. For all three analyses we identified the most important factors that influence the probability of publication using citizen science data. Four general barriers emerged: a narrow awareness among scientists of citizen science projects that match their needs; the fact that not all biodiversity science is well-suited for citizen science; inconsistency in data quality across citizen science projects; and bias among scientists for certain data sources (institutions and ages/education levels of data collectors). Notably, we find limited evidence to suggest a relationship between citizen science projects that satisfy scientists' biases and data quality or probability of publication. These results illuminate the need for greater visibility of citizen science practices with respect to the requirements of biodiversity science and show that addressing bias among scientists could improve application of citizen science in conservation.

1. Introduction

Public participation in scientific research, or citizen science, is a growing practice that could be a powerful addition to the conservation toolbox (Cooper et al., 2007; Danielsen et al., 2010; Cosquer et al., 2012; Theobald et al., 2015). From documenting climate change impacts (e.g., Boyle and Sigel, 2015) to informing land management (e.g., Martin, 2015), applications of citizen science data are broad across subjects and at scales relevant to today's conservation issues (Theobald et al., 2015). Many authors have argued that both ecology and conservation would benefit from greater use of citizen science due to its ability to provide data at the broad spatiotemporal scales and fine grain resolution needed to address global-scale conservation questions (Jiguet et al., 2005; Couvet et al., 2008; Schmeller et al., 2009; Devictor et al., 2010; Magurran et al., 2010; Loss et al., 2015). This argument may be particularly strong for biodiversity science, because abundance and/or density of taxa is the focus of a large number of citizen science projects (Theobald et al., 2015).

Despite these arguments, citizen science has yet to be fully embraced by either the ecological or conservation communities (Silvertown, 2009; Riesch and Potter, 2013; Tulloch et al., 2013; Cooper et al., 2014; Bonney et al., 2014). Citizen science projects within these fields generally report only modest peer-reviewed publication rates (Theobald et al., 2015), and they have rarely generated well-known, highly cited data (Jiguet et al., 2005; Zuckerberg et al., 2009; Silvertown et al., 2011, but see Devictor et al., 2010 and Edgar et al., 2014 for exceptions).

One important obstacle to the scientific use of citizen science data may be the perceptions of scientists. For example, efforts to incorporate citizen-generated data into conservation are sometimes met with concerns regarding rigor of data collection and, ultimately, data quality. Specific critiques include lack of attention to study design (Newman et al., 2003; Krasny and Bonney, 2005), inconsistent or suboptimal training (Conrad and Hilchey, 2011), absent or problematic standardization and verification methods (Cohn, 2008; Dickinson et al., 2010; Bonter and

* Corresponding author.
E-mail address: hkb10@uw.edu (H.K. Burgess).
[1] Burgess and DeBey contributed equally to the work.
[2] National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California 93106, USA.
[3] Arnold Arboretum of Harvard University, Boston, Massachusetts 02131, USA.

http://dx.doi.org/10.1016/j.biocon.2016.05.014
0006-3207/© 2016 Published by Elsevier Ltd.

Cooper, 2012), and observer or sampling biases (Galloway et al., 2006; Delaney et al., 2008; Dickinson et al., 2010). These criticisms are often countered by specific examples of citizen science projects producing data comparable in quality to that collected by professionals (e.g. Elbroch et al., 2011; Forrester et al., 2015; Lewandowski and Specht, 2015; Lin et al., 2015). This debate highlights the potential influence of perceptions in shaping the use of citizen science (Conrad and Hilchey, 2011; Henderson, 2012). For example, Riesch and Potter (2013) found that scientists' perceptions that citizen science data would not be well received by peers in the scientific community contributed to lack of use.

In this paper, we quantitatively and systematically examine factors related to scientific use of citizen science data, assessed as publication in the peer-reviewed literature. Our goals are to evaluate the influence of perception on use, and elucidate barriers to the use of citizen science data in conservation and ecology. We use multivariate analysis to explore survey results from 423 biodiversity scientists and 125 managers of biodiversity citizen science projects, with respect to three categories: scientists' perceptions of citizen science, scientists' requirements for their own data, and actualities of biodiversity citizen science projects. Our analyses focus on the influence of data source, quality, and visibility. Our results point to specific opportunities for expanded integration of citizen science into the fields of ecology and conservation science.

2. Methods

2.1. Surveys

2.1.1. Scientists

To determine factors that influence whether biodiversity scientists use citizen science data, we created an online survey targeting biodiversity scientists (IRB approval 43,438) to assess: (1) the extent to which citizen science data are presently used in their research, (2) perceptions of citizen science and its resultant data, and (3) requirements for methods and data (e.g. standardization, procedures, measure of error) when conducting their own biodiversity research (see Appendix A Table A1 for full text of survey questions and response options). We defined biodiversity citizen science in the survey introduction as “programs collecting taxon-specific information at known locations and date/times.” The survey contained 25 multi-part questions with binomial (yes/no), multiple choice (inclusive and exclusive), and Likert scale (e.g., strongly agree to strongly disagree) answers, as well as free responses to select prompts. Survey respondents were free to skip any question, resulting in a variable sample size for each question.

We identified publishing biodiversity scientists using a Web of Science search, restricted to natural science, of corresponding authors of papers containing the word “biodiversity” in the title, abstract, or keywords; this yielded a pool of 3148 scientists as potential survey respondents. We contacted the corresponding author of each publication by email; 423 scientists completed the survey (13.4% response rate; see Table A1 for respondent demographics). It is possible that scientists who responded were more interested in, and/or aware of, citizen science compared to those who did not respond, and we consider this when interpreting our results.

2.1.2. Citizen science

In order to compare scientists to citizen science programs, we surveyed citizen science project managers (IRB approval 43,438) regarding (1) project goals and details of project administration, (2) data collection protocols, and (3) participant demographics (see Table A2 for full text of survey questions and response options). We defined biodiversity citizen science in the survey introduction as “programs collecting taxon-specific information at known locations and date/times.” The survey contained 32 multi-part questions with similar types of questions as the scientist survey (i.e., binomial, multiple choice, Likert scale, and free response); where applicable, we asked identical questions of both scientists and citizen science projects. Again, respondents were free to skip any question, resulting in a variable sample size for each question.

We identified potential respondents from a database of biodiversity citizen science projects that aggregates projects from seven publicly available databases (see Theobald et al., 2015 for database details). Of these 388 projects, 329 were extant and had contact information that enabled our communication with project managers via email at the time of survey administration. We received a total of 125 responses (38% response rate). Respondent projects were predominantly housed in North America (66.4%) followed by 9%, Europe, 2.5% Asia, 2.5% multiple, 1.6% Oceana with 18% Unknown.

2.2. Analysis

2.2.1. Factors influencing publication of citizen science data

While peer-reviewed scientific publication is not the only outlet for citizen science research outcomes, patterns of publication of citizen science data in the peer-reviewed literature represent a measure of the current extent to which citizen science reaches professional scientists. We used a non-parametric modeling approach known as Random Forest analysis to derive explanatory patterns between the probability of publication using citizen science data and survey responses, for both professional scientists and citizen science project managers.

Random Forest analysis (RF) is a statistical technique that uses recursive and “out-of-bag” bootstrap sampling (i.e., predicting data not in the bootstrap sample) to construct binary partitions of predictor variables, fitting regression trees ($n = 1000$) to the dataset, and ultimately combining the predictions from all trees (Breiman, 2001). Predictors are ranked by mean squared error (Breiman, 2001; Cutler et al., 2007); the order reflects the influence of each predictor on the response variable. We conducted three separate RF analyses: two using explanatory variables from scientists' survey responses (Table A1), and one using explanatory variables from citizen science project manager responses (Table A2). See Appendix A Supplemental Methods for additional details.

We distinguished two categories of variables a priori to explore via RF in association with scientists' potential engagement with citizen science projects (perceptions/preferences and requirements), and one set of variables for the citizen science managers. “Perceptions/preferences” captures opinions regarding the purpose of citizen science, the quality of citizen science data, and the degree of trust in various data sources. “Requirements” consists of awareness of citizen science projects that matched their research area and factors that are required to successfully conduct their particular research (e.g., specific methods, protocols, or data attributes). We assumed that for citizen science data to be used for research purposes by our respondent scientists these latter factors must be satisfied. We performed two separate RF analyses on these mutually exclusive sets of variables: scientists' perceptions/preferences (398 respondents, 29 predictor variables: 27 numeric, 3 binary), and scientists' requirements (388 respondents, 27 predictor variables: 23 binary, 2 factors, one 1 numeric), each with the binary response variables of either “have published using citizen science generated data” or “have not” (1 or 0, respectively; see Table A1, Question 21). We conducted a third RF analysis on citizen science survey responses (118 respondents, 49 predictor variables: 27 numeric, 22 binary) to predict the probability of whether a project reported having one or more peer-reviewed articles using project-generated data (1; see Table A2, Question 18), or no publications associated with that project's data (0).

The full RF models incorporated all possible respective variables, which we reduced to the best-fit model based on a subset of those predictors. The full RF models provided an initial ranked order of all predictor variables associated with publication for each dataset. In a stepwise elimination, starting from least to most influential, predictor variables were removed, without replacement, from the model and variance explained was determined at each step. We then compared all models, selecting the best-fit as the model that explained the greatest amount