



Hiding in Plain Sight: Secondary Analysis of Data Records as a Method for Learning about Citizen Science Projects and Volunteers' Skills

METHOD

KAREN PETERMAN

VERONICA DEL BIANCO

ANDREA GROVER

CATHLYN DAVIS

HOLLY ROSSER

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

]u[ubiquity press

ABSTRACT

This paper is the culmination of several facilitated exercises and meetings between external researchers and five citizen science (CS) project teams who analyzed existing data records to understand CS volunteers' accuracy and skills. CS teams identified a wide range of skill variables that were "hiding in plain sight" in their data records, and that could be explored as part of a secondary analysis, which we define here as analyses based on data already possessed by the project. Each team identified a small number of evaluation questions to explore with their existing data. Analyses focused on accurate data collection and all teams chose to add complementary records that documented volunteers' project engagement or the data collection context to their analysis. Most analyses were conducted as planned, and included a range of approaches from correlation analyses to general additive models. Importantly, the results from these analyses were then used to inform the design of both existing and new CS projects, and to inform the field more broadly through a range of dissemination strategies. We conclude by sharing ways that others might consider pursuing their own secondary analysis to help fill gaps in our current understanding related to volunteer skills.

CORRESPONDING AUTHOR:

Karen Peterman

Catalyst Consulting Group, US
karenpetermanphd@gmail.com

KEYWORDS:

citizen science; data validation;
secondary analysis; volunteers;
skill

TO CITE THIS ARTICLE:

Peterman, K, Del Bianco, V,
Grover, A, Davis, C and Rosser,
H. 2022. Hiding in Plain Sight:
Secondary Analysis of Data
Records as a Method for Learning
about Citizen Science Projects
and Volunteers' Skills. *Citizen
Science: Theory and Practice*,
7(1): 35, pp. 1–12. DOI: [https://
doi.org/10.5334/cstp.476](https://doi.org/10.5334/cstp.476)

INTRODUCTION

Recent studies have noted the gap between the number of citizen science (CS) projects that require volunteers to use science inquiry skills to collect high-quality data, and the extent to which these projects assess volunteers' abilities to demonstrate those same skills (Bowser et al. 2020; Burgess et al. 2017; Phillips et al. 2018; Styliniski et al. 2020). For those who work in CS, there is little doubt that volunteers need robust science inquiry skills to ensure the collection of high-quality data and to meet education outcomes. Even so, scientists outside of CS continue to question the validity of volunteer-collected data (Burgess et al. 2017). The CS field must ensure that volunteer skill prowess is evaluated to support continued improvement of programs and studied to continue to document the veracity of volunteer-collected data (Becker-Klein, Peterman, and Styliniski 2016). While few projects evaluate volunteers' skills directly, researchers have devoted significant attention to CS data quality (Kosmala et al. 2016) using a variety of data validation and verification strategies that are often specific to project goals and reflect relevant standards of scientific practice for different types of data (Parrish et al. 2018; Stevenson et al. 2021). Though collected for other purposes, these kinds of data records have the potential to serve as a key source of information about volunteer skill if reframed to use the volunteer as the unit of analysis. This study explores the possibility that the data needed to answer some of the field's questions about volunteers' data collection skills are "hiding in plain sight" in these data records and could be explored through a *posteriori* analysis.

Several studies have scanned the field to learn whether and how CS project leaders validate volunteer data (Baker et al. 2021; Bowser et al. 2020; Freitag, Meyer, and Whiteman 2016). Strategies employed by citizen science programs to increase the credibility of their data. *Citizen Science: Theory and Practice*, 1(1)2016; Wiggins et al. 2011). At least half of the project leaders in each study reported implementing data validation methods, indicating that there is broad potential for conducting secondary analyses to study volunteers. In the only known example of secondary analysis in the literature, Kelling et al. (2015) analyzed volunteers' submissions to the eBird project to identify latent indicators of skill in volunteers' existing data. Most volunteers' rates of detection increased with ongoing participation, suggesting that secondary analysis of data records can generate new learning about volunteers' abilities. The current study was designed to explore this potential through a series of case studies with established projects in the CS field.

Using data records in this way has the potential to fill a current gap in our understanding of volunteer skill development and the relationship between volunteer skills

and high-quality CS data. Evaluation is defined as "the systematic collection of data to determine strengths and weaknesses of programs, policies, or products, so as to improve their overall effectiveness" (Phillips et al. 2014). As noted earlier, direct evaluation of volunteers' skills is infrequent. For example, only 4 projects out of 36 surveyed by Bowser et al. (2020) conducted any skill assessment. A recent study by Styliniski et al. (2020) confirmed the lack of rigorous evaluation methods to understand volunteer skill prowess and identified a number of barriers: limited time and limited staff to devote to evaluation efforts, lack of evaluation expertise among CS team members, and lack of supporting resources. In other cases, CS project leaders have concerns that evaluating volunteer skills will create a barrier to participating in the project itself. Direct evaluation of volunteers often requires human subject review (e.g., IRB or similar), which may be overlooked or avoided by scientists who are less familiar with these procedures (Resnik 2019). Many of these challenges might be resolved if CS projects could use existing data records to understand more about volunteer skill in ways that promote the continued development of the CS field through data-driven decision-making. For example, the results from such analyses could enhance understanding of volunteers' competencies, provide valuable data for CS project leaders to continue to refine their projects, and further support using volunteer-collected data for scientific purposes. Given that current evaluation practice in CS includes few studies that focus on volunteer skill, taking advantage of existing data sources to answer these questions has the potential to catalyze learning about volunteers' skills, as well as their contributions to both CS projects and the scientific enterprise.

With funding from the National Science Foundation (DRL# 1713424), our team examined volunteer skill assessment processes and impacts within CS and is publishing a series of articles based on the results. A commonality across the papers is the use of embedded assessment, or activities integrated into the learning experience that allow learners to demonstrate competencies for evaluation purposes. Davis et al. (2022) examined how results from such evaluation efforts informed 15 CS projects and the broader field. Becker-Klein et al. (in review) shares a development process, including challenges and opportunities, for creating performance-based embedded assessments that can be shared across citizen science projects. Here, we conducted a phenomenological study with five CS teams (Saldana 2011). The phenomenon was defined as the process of using a *posteriori* analysis of previously collected data to study volunteer skill. Several questions were of interest:

1. What kinds of information about volunteer skills are "hiding in plain sight" in data records?

2. What kinds of information do CS projects use to make decisions about conducting a secondary analysis of data records?
3. How can conducting secondary analyses of data records answer questions about volunteer skills to inform CS projects and the field?
4. What were the challenges in conducting secondary analyses of data records to understand more about CS projects and volunteers?

METHOD

STUDY CONTEXT

We gathered information from 13 candidate CS project leaders who involved volunteers in observation-based data collection and expressed interest in learning more about their volunteers’ skill proficiency. Candidates were identified through our work on a prior project (Stylinski et al. 2020), and through a snowball sampling method. From these, we selected five based on four predetermined commonalities that were verified during the interview process: (1) databases containing repeated submission of data by individual (unique, known) volunteers, (2) existing procedures for validating volunteer data, (3) scientific investigation as a primary goal (see Wiggins and Crowston 2011), and (4) primarily adult volunteers. Beyond their data validation practices, one of the five had conducted prior studies of volunteer skills prior to this project (Parrish et al. 2019) and another used machine learning based on skill levels, but had not studied skills directly (Zevin et al. 2017).

Our five partners reflect the range of CS projects and included field monitoring, species identification, and image classification projects that utilized a range of data collection strategies: hands-on; out-of-doors repeated sampling; and online, crowdsourcing projects. The number of volunteers ranged from a few hundred to several thousand, and available data ranged from hundreds to over one million data points (See Table 1). Each team included a CS project director (who received a stipend) and an analyst (who received payment for their work).

INSTRUMENTS AND PROCEDURES

Four semi-structured interviews were conducted with each team via Zoom using a standard protocol designed to capture information on secondary analysis processes. Baseline interviews were conducted in spring 2018 to plan the secondary analysis. Questions focused on the tasks and science inquiry skills volunteers needed to fulfill the citizen science project work, the common mistakes that volunteers might make, and how the project’s data validation process worked to catch those mistakes. Using the results from these interviews, our team continued to support planning efforts through the summer of 2018 via facilitated exercises both prior to and during a two-day in-person meeting.

Additional facilitated exercises were used to document and reflect on the process each partner team used to conduct their secondary analysis. The first of these was a Skill Ranking Worksheet that includes nine questions that helped prioritize the list of variables identified in the first interview based on data availability and the kinds of results that were likely to have the greatest influence on the project. We also shared a Skill Hierarchy Worksheet that was used to identify the higher-level goal or outcome of interest for the analysis (e.g., accurate data collection), as well as the underlying tasks, skills, and sub skills that must be accomplished to achieve the overall goal. Copies of both worksheets can be found at the project’s website (<https://sites.google.com/umces.edu/embeddedassessment/welcome-to-eas>).

A midpoint interview was conducted between fall 2018 and winter 2019; for most teams, these occurred after data had been prepared for analysis. Questions in the second interview centered on each team’s progress toward conducting their secondary analysis with a focus on the steps that were both easier and more challenging than originally expected. The third and fourth interviews were conducted after the CS team had completed their final analyses. These began in late 2019 and continued through early 2021; the third interview was conducted soon after each team completed their analysis and final interviews were conducted several months later. Interview questions for the third interview documented the key findings of the analysis and the value of the process to the project. The

PROJECT NAME	PRIMARY TASK FOR VOLUNTEERS	YEAR FOUNDED	# VOLUNTEERS	# “SURVEYS” COMPLETED
Alliance for Aquatic Resource Monitoring (ALLARM)	Environmental measurements	2010	390	5,500
Coastal Observation and Seabird Survey Team (COASST)	Species identification	2000	4,000	27,000
Colorado Pika Project	Environmental measurements	2012	136	360
Gravity Spy on Zooniverse	Image classification	2016	14,000	1,200,000
Reef Environmental Education Foundation (REEF)	Species identification	1993	15,000	236,000

Table 1 Description of our five project partners at the beginning of our project.

final interview focused on ways teams had used the results of their analysis to inform their work and the CS field.

This research was approved by the University of Maryland Institutional Review Board (IRB #1072528). A sample set of interview questions are included in the supplemental documents.

CODING

To analyze the interview data, three researchers used a six-step collaborative qualitative analysis process to develop two coding schemes (Richards and Hemphill 2018). A deductive coding scheme focused on the steps CS teams took to prepare for and conduct their secondary analysis. An inductive scheme documented the reasons behind CS teams’ decision-making throughout their secondary analysis. A coding scheme from the evaluation literature was also used (Alkin and King 2016; Bundi, Frey, and Witmer 2021). The four-part scheme with CS-based examples is described fully in Davis et al. (2022). For the purposes of this analysis, data related to two codes were examined in greater detail for five projects and in the context of secondary analysis. For a full list of codes and definitions, see the supplemental documents.

Interview data were coded and analyzed using NVivo 12 with consensus coding. Two researchers coded each document independently and then argued to consensus on disagreements. The themes from the interviews are presented below.

RESULTS

This study was designed to describe the essential opportunities and challenges that CS teams encountered when conducting a secondary analysis of data records to understand volunteer skill. The phenomenon being studied—that is, the process of conducting the secondary analysis—was situated within five project cases that reflect the range of projects in the field. Given that we used a case-oriented rather than a variable-oriented approach, the results are presented by project (Miles, Huberman, and Saldana 2020). This section is organized by our research questions, and follows the process used by our teams to conduct their secondary analysis.

WHAT KINDS OF INFORMATION ABOUT VOLUNTEER SKILLS ARE “HIDING IN PLAIN SIGHT” IN DATA RECORDS?

To begin this work, each team reviewed their data records to identify the specific data points that could serve as indicators of a volunteer’s skill. These were considered the skills that teams could investigate as part of a secondary analysis; teams identified between 5 and 20 skills as part of this process. Table 2 presents the skills, using the language shared by each team. Most of the skills named were related to specific data collection tasks such as navigating to the data collection site or taking measurements. A smaller number were broad measures related to data collection,

	PROJECT A	PROJECT B	PROJECT C	PROJECT D	PROJECT E
Skill, as indicated by specific data collection activities	<ul style="list-style-type: none"> - Habitat choice - Identify non-charismatic habitat - Identify nocturnal species - Location accuracy - Species accumulation 	<ul style="list-style-type: none"> - Dual calibration - Duplicate sampling 	<ul style="list-style-type: none"> - Accurate species identification - Beach data collection - Body measurements - Carcass data collection - Family level identification - Foot type accuracy - Photos - Plumage status - Species un/verifiable - Verified identification types - Wing key accuracy 	<ul style="list-style-type: none"> - Geo validated navigation - Identify land cover types - Identify signs of pika - Photo validation in talus 	<ul style="list-style-type: none"> - Accurate pattern matching - Matrix comparisons
Skill, as indicated by broad data-related activities	<ul style="list-style-type: none"> - Duplicate survey submissions - Error reporting/flags - Experience level - “Groaner” abundance - Rarity reporting/flags - Species-only surveys - Types of edits to surveys 	<ul style="list-style-type: none"> - Complete data - Time to sample - Zero observations 	<ul style="list-style-type: none"> - Complete data entry - Error conditions - Exception species identification - First identification accuracy - Frequency of finds - Number of species identified - Outliers - Zeros versus nulls 	<ul style="list-style-type: none"> - Complete data 	<ul style="list-style-type: none"> - Sophistication - Species accumulation
Skill, as indicated by data interpretation		<ul style="list-style-type: none"> - Data interpretation 			

Table 2 Skills Skills in data records, by project.

such as submitting complete data or reporting zero observations. Only one project included data interpretation as a skill that was included in their data records.

Some skills were named by multiple teams, including accuracy of identification tasks, submitting complete data, location accuracy, accurately recording habitat features, the frequency of reporting outliers, the rate at which volunteers confused commonly mistaken objects, data interpretation skills, and species accumulation.

The range in the types of variables listed also hints at the value of this analytic method in that the variables span the entire inquiry process. Some variables relate to verifying the necessary conditions for high-quality data before the collection process begins (e.g., accurate location, dual calibration), while others are related to data collection itself (e.g., identifying land cover, using the key accurately). Examples of skills related to later stages of the inquiry process include submitting complete data and data interpretation.

WHAT KINDS OF INFORMATION DO CS PROJECTS USE TO MAKE DECISIONS ABOUT CONDUCTING A SECONDARY ANALYSIS OF DATA RECORDS?

Once they had identified a full list of skills that could be investigated through secondary analysis, teams prioritized their list using the *Skill Ranking Worksheet* and through meetings with our research team. Each interview also included details related to the decisions that contributed to each team’s analysis. Two themes were common. The first focused on ways the team hoped to use the results of their analysis. The most common reason for wanting to do (or not do) a secondary analysis was intellectual interest. As exemplified in the quotes provided in [Table 3](#),

leaders were interested in learning more about the experiences they were providing for volunteers and the ways volunteers engaged with the support provided by the project. Four teams each made decisions based on the potential to adaptively manage their program and to guide future evaluation efforts. In these cases, teams expected the results to provide data that could be used to improve the training provided to volunteers and ways to study volunteers and their contributions to the project. Three teams made decisions based on the potential to affect the CS field. In these cases, teams chose an analysis that they believed would have broad applicability across their sector of CS (e.g., water quality monitoring projects).

The second focal point for decision-making was whether data were in a format that was easy to access and use. Decisions were more often based on the data that were available, though all five teams narrowed the focus of their analysis by discussing both the data they could and could not access in meaningful ways in their existing data records.

In the examples below, data were available but the effort required to make those data usable for secondary analysis was considered beyond the scope of the current project.

It would be a combination of [using the online feature], data mining, as well as linking proficient users who talk about images in a specific way, and link it up with their particular evolution of how well they classify over time along with...the way they talk about the images evolve over time...Theoretically, that data quote-unquote exists, but it doesn't exist in a way that we'd have the time to study or dig into it.

Intellectual interest (all 5 teams)	<i>What we're interested in interpretation-wise is how effective are these trips, as opposed to how effective is it when [someone] goes out there for the 20th time? It could be interesting to see how much they use all the information available to them. And if over time once you reach level five you do that less and less.</i>
Program improvement (4 teams)	<i>We have really only used our data validation to do the basic quality data that we can use in our research...So it would be helpful to know whether volunteers are able to do some of the things we're asking them to do, in thinking about that new protocol...whether there are tasks we're kind of assuming the volunteers are able to do, that they're actually not very skilled at. Are there particular things that folks struggle with more than others that we could tune our training and materials to address, so certain species or certain things that we ask of them that we could focus our communications around.</i>
Changing evaluation practices (4 teams)	<i>We could learn more about our volunteers...So we're hoping this project will force us to actually think through that, and go through all of those steps, and then hopefully we can use that to improve the way we're assessing volunteers moving forward. It just goes back to the research questions we had of what processes are involved in catapulting users to like the most helpful contributors of information, and then that'll help us set up best practices for our research teams.</i>
Informing the CS field (3 teams)	<i>How can this case study on [our project] be extrapolated to something that at least the 300 coordinating projects could take away and incorporate as well? Part of my motivation was choosing [this project] was providing the opportunity to choose [a project] that is pushing ahead in really interesting ways, and I want that to inform your study. Because I think that's where citizen science should be headed, where it makes sense to do these.</i>

Table 3 Intentions for using the results of the secondary analysis, as described by CS teams.

We have all of the datasheets archived, and all the images archived...We've talked about...going backwards in time and putting accuracy in [older records where it is missing]. But that literally requires somebody going back to the datasheets and then looking to see what's in the database, and then putting in the accuracy scores...Yeah, it would be hundreds and hundreds of hours.

An important factor in these decisions was the time required to accomplish secondary analysis tasks. The importance of time in determining evaluation effort is consistent with the prior literature on CS evaluation (Stylinski et al. 2020). All partners had to balance their interests with the time and budget available to prepare for and conduct the analysis.

HOW CAN CONDUCTING SECONDARY ANALYSES OF DATA RECORDS ANSWER QUESTIONS ABOUT VOLUNTEER SKILLS TO INFORM CS PROJECTS AND THE FIELD?

Table 4 presents the questions investigated by CS teams and a summary of their results. Despite the range of skills listed, all teams chose to focus their secondary analysis on a topic related to accurate data collection. Even with this shared focus, the range of skills that teams selected remained broad: (1) accurate site navigation, (2)

submitting complete data as a measure of accuracy, (3) relationship between measurement context and accuracy, (4) accumulated knowledge of volunteers based on the number of identified species, and (5) online user behavior in response to an inaccurate image classification.

Table 4 summarizes the evaluation questions that were answered by each team's secondary analysis, as well as the overall results for each CS project. As shown in the italicized text, all CS teams included at least one independent variable in their secondary analysis to explore whether and how variability in program or data collection context related to volunteers' skill. For example, McNeill and Vastine (2019) shared a correlation analysis that explored the relation between ALLARM staff support and volunteers submitting complete data. Jackson et al. (2020) used a combination of nonparametric tests and mixed-effects logistic regression to understand volunteer engagement with different Gravity Spy resources. Simonis and Pattengill-Semmens (unpublished manuscript) used generalized additive models to explore the growth in volunteers' ability to identify new species over time and in relation to individual versus group-based data collection contexts. Existing data records were used to support a wide range of analyses that included correlation analysis, analysis of variance (ANOVA), general linear models, generalized additive models, and sequence analyses. The detailed methods and results of

PROJECT	EVALUATION QUESTIONS BY PROJECT TEAM	RESULTS
Project A	Which <i>online project resources</i> , if any, are associated with increased performance with classifying images?	During the early stages of project involvement, the authoritative resources such as field guides are used most often by volunteers when they make an incorrect classification. Over time, the social features and interactions, such as chat and instant messaging, seem to be what volunteers find most useful to bridge their learning gaps after receiving feedback about a mistake. However, the majority of users do not consult with any resources and instead just continue classifying.
Project B	What is the relation between the amount of <i>volunteer engagement with the project team</i> and whether they pass quality assurance/quality control measures? Are volunteers who <i>pass quality assurance/quality control</i> measures more likely to submit a complete data set ?	The vast majority of volunteers pass quality assurance/quality control measures, and so this was not a useful predictor for contributing complete data. The lack of variability also meant that the relation between volunteer engagement and quality assurance/quality control could not be explored.
Project C	Are there differences in accurate navigation to historical pika nesting sites based on the <i>difficulty of the site location</i> ?	Volunteers were less accurate when navigating to the more difficult sites, and so the characteristics of the site mattered.
Project D	Does accuracy vary with <i>body condition</i> of the bird carcass found, based on body parts present?	Of three different body parts measured, the foot measurement was significantly less precise (more variable) than either the wing or bill measurements. However, foot measurements were rarely needed for species identification. When only feet were present, measurement precision was significantly greater.
Project E	Is a volunteers' rate of species acquisition accelerated if they <i>collected data as part of a group versus only on their own</i> ?	All else being equal, group size is positively correlated with species accumulation. Volunteers' growth in their ability to identify a wider range of different species is faster if they go and observe with others.

Table 4 Sample evaluation questions and summary of results.

Bold text is used to indicate the dependent variable (i.e., volunteer skills being studied). *Italicized text* is used to indicate additional independent variables that were explored.

this research is beyond the scope of the current study, and can be found instead in the work cited above.

As noted, we expanded on evaluation use findings in Davis et al. (2022) to examine in more detail how CS teams used the results from their secondary analysis. At the time of their final interview, four of five teams had used their results to make changes to their CS project, and all five teams had shared their results in at least one way to inform the CS field (see Table 5).

Four of five projects made programmatic changes to volunteer training. In multiple cases, for example, the results revealed that volunteers needed additional content or skill training on a specific topic; three projects responded by adding content to their initial training materials, two updated their website or online support systems, and two projects integrated additional training content into their follow-up communication and training for volunteers.

Three teams each used the results from their secondary analysis to make changes to their data collection procedures. In all cases, these teams applied what they had learned from their results to design the data collection procedures for similar, new CS projects that were being initiated. Three teams also used the results from their analysis to change their data validation processes (e.g., to ensure that volunteers were collecting field data at the intended location). Regarding volunteer management, two teams realized that their project would benefit from collecting additional information about their volunteers; both teams added new registration processes to collect this information so that it could inform their work moving forward.

All teams used the results to attempt to persuade others (e.g., to gain board member support for project expansion, to retain current and obtain new funding, and to recruit new collaborators). Three teams shared their results through conference presentations or publications to inform the field

(Jackson et al. 2020; McNeill and Vastine 2019; Simonis and Pattengill-Semmens, unpublished manuscript). Two teams shared the results of their secondary analysis with volunteers to provide feedback and be transparent about their secondary analysis process.

WHAT WERE THE CHALLENGES IN CONDUCTING SECONDARY ANALYSES OF DATA RECORDS TO UNDERSTAND MORE ABOUT CS PROJECTS AND VOLUNTEERS?

Throughout the process of planning for and conducting secondary analyses, we and our CS partner teams reflected on the expected and unexpected challenges of this work that might be used by others to decide whether to use this approach within the context of their own project. Identifying skills was one such challenge. A number of skills were eliminated from consideration during the baseline conversations. See Table 6 for a full list, by project. A total of 84 potential skills were named in the baseline interview discussions, and 34 were eliminated. Twelve were eliminated as independent variables related to the project, rather than dependent skill variables. Some of these focused on project engagement (e.g., when and how volunteers were trained, the frequency of project participation), whereas others focused on the data collection context (e.g., whether volunteers used available project supports or characteristics of the physical environment where data were collected).

An additional 22 skills, half of which were considered vital to successful data collection, were eliminated because they were not captured in existing data records. These included contextual factors that affected data collection (e.g., whether and how visual feedback was used during an observation, whether group members helped with an identification), and a volunteer’s ability to detect the organisms or habitats of interest to the project (e.g., level of effort given to search for

	PROJECT A	PROJECT B	PROJECT C	PROJECT D	PROJECT E	TOTAL PROJECTS
Programmatic use						
Volunteer training	—	✓	✓	✓	✓	4
Data collection procedures	—	✓	—	✓	✓	3
Data validation process	—	✓	—	—	✓	3
Volunteer management	—	✓	—	—	✓	2
Dissemination use						
Share results to persuade	✓	✓	✓	✓	✓	5
Share results to advance the field	✓	✓	—	✓	—	3
Share results to inform volunteers	✓	—	—	✓	—	2

Table 5 Ways that results were used to inform CS project implementation and the field.

	PROJECT A	PROJECT B	PROJECT C	PROJECT D	PROJECT E
Independent variables	<ul style="list-style-type: none"> - Consistent surveying - 1st survey w/ a group 	<ul style="list-style-type: none"> - Data entry frequency - Trainer effects 	<ul style="list-style-type: none"> - Cumulative surveys/hrs - # Carcasses ID'd - Personal background - Re-finds - Time since training 0 # - Intact birds ID'd - # uses of wing key 		<ul style="list-style-type: none"> - Leveling down
Data not included in data records	<ul style="list-style-type: none"> - Taxonomic updates - Community ID help - Photos - Can ID to family - Uses ID references - Mystery fish - Records survey # 	<ul style="list-style-type: none"> - Replicates with flags - Using visual feedback 	<ul style="list-style-type: none"> - Beach search effort - Transcription accuracy - Not-birds - Data entry into Db - Detection rate - Uses guide to ID - Project terminology 	<ul style="list-style-type: none"> - Zero observations - Identifying habitat - Measure in meters 	<ul style="list-style-type: none"> - Targeted samples - Phase II project building - Focus on X-Y axes

Table 6 Skills that were independent variables or unavailable in data records.

organisms to record). Data entry skills and recording zero observations were additional skills that were of interest but that could not be verified using data records.

The baseline conversations also illuminated a limitation of using secondary analysis of data records to investigate skill gains. Some variables were eliminated because data were not stored to track mistakes as evidence of learning. For example, online data entry forms use several strategies to help prevent common errors and typically record the final response submitted without recording the foregoing occurrence or types of mistakes. Similarly, some projects overwrite volunteer data as part of the validation process to correct for errors. It is logical that the “correct” answer would be the only one recorded, as not all projects require an audit trail for validation. However, recording only the final answer eliminates the potential traces of learning in action. These competing priorities are described best by our project partners:

They're too smart [the data collection filters]. They let people know when they've made mistakes [but] we don't get to see our volunteers self-correct. We don't get to see them, over time, refine the technique.

In terms of database design and the way that we track information, there's kind of two goals that don't really easily converge and that's tracking and preserving what participants do so that we can evaluate it and having the very best possible data set. Without having a complete duplicate—where you maintain the original and then you also have the version that we've finessed over time because we know where errors happened and we're making improvements—we don't have a perfect record.

The results above demonstrate that identifying skills is not a straightforward process. Preparing participant

engagement data was also a challenge for four of the five teams. In three cases, these data were available to teams but not in a format that was easy to access and use. Volunteer management data were not stored in a database for two teams and thus had to be entered and merged with the scientific data records prior to analysis. In another example, online user behavior data were captured, but not in a format that was immediately useful to the analyst; this team spent time querying the database to isolate the data needed. When reflecting on preparing volunteer management data, these teams shared the following:

So the easy side was compiling the volunteer participation, I would say, so training days, do they follow-up in a follow-up meeting, or conference calls, and [with] the quality control data. That was really easy to pull together. And then the thing that is really time-consuming...is taking [our data records] and then trying to bring that into a format where that all lines up.

We had quite a bit of work to do, kind of going back and reconstructing some of the stuff that we needed on the volunteer end of things. That might be one thing that was kind of harder than we anticipated. We've done a really nice job of keeping track of the stuff that we're doing for data quality control. But these other measures of how many trainings a volunteer attended, those records were really messy... So we've gone back through and reconstructed all that now and so now we have a good spreadsheet with all those things.

Using the Python API to the Google Analytics, I roughly can reproduce the clickstream to some degree for that individual at that time...The biggest

issue for me is that the Google API docs is super haphazard, especially for Python. So it's just kind of a learning curve type of thing.

Four teams experienced challenges related to sample size. Two had too much data and spent time determining the appropriate data subset to use based on two constraints: the time needed to clean the data, and the time it takes to run complex higher-level statistical models. Other teams had the opposite challenge and did not have enough data to conduct multivariate analyses to understand relationships between participant engagement and skill levels. Challenges related to sample size were described as follows:

Probably the hardest [thing] is going to be finding our way around this very giant database and getting the export stuff streamlined so that [the analyst is] getting what they need and nothing that they don't. And I guess another thing would be for me, making sure that the information about [participation] is in a way that makes sense and is easy to incorporate into the analysis.

I think that sample size sort of prevented us from looking at skill development over time too. We were just looking at whether or not they were good at getting to the sites, not so much did their skills at doing that improve over time, and that was just an issue with not having enough returning volunteers to be able to do that.

DISCUSSION AND CONCLUSIONS

The results of this phenomenological study of five CS teams provide evidence to support the utility of secondary analyses as a method for filling the current gap in the literature related to evaluating volunteer skill. As shown in Davis et al. (2022), the results from these analyses were used to hone the implementation of CS projects and to inform the field. Here, we highlight specific uses among the five projects, which included changes to data validation processes, the addition of new volunteer management practices, and the sharing of results to persuade and inform others.

Few CS projects evaluate volunteer skill currently (Bowser et al. 2020; Burgess et al. 2017; Phillips et al. 2018; Stylinski et al. 2020), but many do have processes in place to assess data quality (Kosmala et al. 2016; Stevenson et al. 2021). Our results indicate that records of data quality may often include valuable data for understanding CS volunteers' skills in ways that can inform practice. Given that the challenges associated with measuring skills are common across many informal learning contexts (Bell et al. 2009; Fenichel and Schweingruber 2010), secondary analyses of

CS data records have the potential to make significant and broader contributions. Two of our partners, for example, were motivated to conduct a specific analysis because they anticipated their findings would generalize across their sector of CS. The results from these kinds of analyses also demonstrate the level of attention and detail needed to support skill learning in informal learning contexts.

Identifying the range of volunteer skills represented in existing data records was a critical first step in the process. Although seemingly simple, a number of skills were eliminated from consideration because they were project-rather than skill-related or because a concrete indicator of the skill was not available. Though our initial conversations focused on a wide range of skills related to a CS project, each team ultimately focused on skills related to accuracy. This commonality is not surprising given that the requirement to use existing data records meant that data quality was the most prominent outcome to consider. For others considering a secondary analysis, focusing on accuracy-related skills from the beginning may streamline their process of identifying possible skills. The Skill Hierarchy Worksheet may be useful as a way to map the full complement of tasks, skills, and sub-skills needed to collect accurate data (see our project website for an example, <https://sites.google.com/umces.edu/embeddedassessment/welcome-to-eas>). Given our experiences to date, it is likely that some tasks, skills, and sub-skills will already be included explicitly in a project's existing training materials while others will not. We recommend that projects begin their secondary analysis by focusing on the skills that are included in their training, while concurrently adding training for the tasks, skills, and subskills that are missing.

We were surprised that few teams chose to pursue an analysis that included learning trajectories (i.e., skill gains at the individual level). Three of the five teams spent some time considering this analysis plan; only one pursued this approach. We believe using data validation records is an ideal way to explore learning trajectories. This kind of analysis may be more feasible for systems designed with this analytic plan in mind, rather than via secondary analysis. Ideal systems would allow researchers to query data by individual volunteer, and be organized to compare validated data records for that individual over time.

Our partners found value in working with our team and with one another as they navigated the opportunities and challenges of their secondary analysis (see Davis et al. 2022 for a full description). Project leaders who are considering a secondary analysis might benefit from identifying others in their network who are similarly interested and who might support one another through collaborative learning strategies. Working together to generate lists of skills that might be found in data records, verifying skills by distinguishing them from independent variables, and then using the Skill Rating and Skill Hierarchy worksheets

are all steps that benefit from collaboration. Small projects with limited staff and projects that are unable to work with professional evaluators are likely to see particular benefits from this type of collaborative learning.

This work also highlighted potential limitations of conducting a secondary analysis of CS data records, as well as some related opportunities. Online data entry systems are often designed to prioritize accurate data entry and to capture final responses rather than mistakes. CS teams might consider whether and how to capture data related to mistakes for tracking changes in proficiency over time. Given the high costs associated with making changes to online data collection systems, this approach might be most feasible for CS projects that have funding to study and optimize their technical infrastructure regularly, and especially those teams that include learning scientists. Platforms like Zooniverse, iNaturalist, and citsci.org have the potential to support these types of analyses by beginning to capture online behavior that might help document skill, as well as engagement data that could be used to conduct similar analyses to those chosen by our partner teams. Engagement data, whether captured about live or online project participants, may be of particular interest to project organizers who want to learn more about their volunteers but do not have expertise in human subject research requirements. Many IRBs consider this type of analysis to be exempt from human subject review. Teams with optimized systems may not find the secondary analysis method useful, and may encounter diminishing returns if their only goal is further optimization. Optimized projects with teams interested in applying their learning more broadly may still find value in secondary analysis of their data records.

This study was designed to be broad and somewhat exploratory as it considered the phenomenon of conducting a secondary analysis of data records to understand volunteer skills within the context of five existing CS projects. Although secondary analyses cannot substitute for direct project evaluation, conducting a secondary analysis of data records holds promise as one solution to begin to fill the gap in our current understanding of volunteer skill development and the relationship between volunteer skills and high-quality CS data and thus has the potential to make significant and broader contributions. We hope that this study provides practical considerations for those who might consider this approach within the unique context of their own project.

DATA ACCESSIBILITY STATEMENTS

The data used in the research project has not been made available, in accordance with our Internal Review Board's determination about the best way to protect confidentiality.

SUPPLEMENTARY FILES

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

- **Supplementary File 1.** Sample Interview Questions. DOI: <https://doi.org/10.5334/cstp.476.s1>
- **Supplementary File 2.** Code Book. DOI: <https://doi.org/10.5334/cstp.476.s2>

ETHICS AND CONSENT

As noted in the text, this study was approved by the University of Maryland Institutional Review Board (IRB #1072528).

ACKNOWLEDGEMENTS

We would like to thank the following partners for their participation in this research project: Brad Schrom, Christy Pattengill-Semmens, Corey Jackson, Erica Garrouette, Hillary K. Burgess, Julia Parrish, Julie Vastine, Juniper Simonis, Laura Trouille, Lisie Lohre, Megan Mueller, Natalie McNeill, Paul Millhouser, Scott B Coughlin, Suzanne Hartley and Timothy Jones.

FUNDING INFORMATION

This material is based upon work supported by the National Science Foundation under Grant No. DRL-171342.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Karen Peterman  orcid.org/0000-0003-4388-9412

Catalyst Consulting Group, US

Veronica Del Bianco  orcid.org/0000-0001-9088-9488

University of Maryland Center for Environmental Science, US

Andrea Grover  orcid.org/0000-0003-4082-4138

University of Nebraska, US

Cathlyn Davis  orcid.org/0000-0002-0968-4336

University of Maryland Center for Environmental Science, US

Holly Rosser  orcid.org/0000-0003-0773-0930

University of Nebraska, US

REFERENCES

- Alkin, MC** and **King, JA**. 2016. The historical development of evaluation use. *American Journal of Evaluation*, 37(4): 568–579. DOI: <https://doi.org/10.1177/1098214016665164>
- 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>
- Becker-Klein, R, Davis, C, Phillips, T, DelBianco, V, Grack Nelson, A** and **Christian Ronning, E**. in review. Using a shared embedded assessment tool to understand participant skills: Processes and lessons learned. *Citizen Science: Theory and Practice*.
- Becker-Klein, R, Peterman, K** and **Stylinski, C**. 2016. Embedded assessment as an essential method for understanding public engagement in citizen science. *Citizen Science: Theory and Practice*, 1(1). DOI: <https://doi.org/10.5334/cstp.15>
- Bell, P, Lewenstein, B, Shouse, AW** and **Feder, MA**. (eds.) 2009. *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, DC: National Academies Press.
- Bowser, A, Cooper, C, De Sherbinin, A, Wiggins, A, Brenton, P, Chuang, TR, Faustman, E, Haklay, M** and **Meloche, M**. 2020. Still in need of norms: The state of the data in citizen science. *Citizen Science: Theory and Practice*, 5(1). DOI: <https://doi.org/10.5334/cstp.303>
- Bundi, P, Frey, K** and **Widmer, T**. 2021. Does evaluation quality enhance evaluation use? *Evidence & Policy: A Journal of Research, Debate and Practice*, 17(4): 661–687. DOI: <https://doi.org/10.1332/174426421X16141794148067>
- Burgess, HK, DeBeyb, LB, Froehlich, HE, Schmidt, N, Theobald, EJ, Ettinger, AK, HilleRisLambers, J, Tewksbury, H** and **Parrish, JK**. 2017. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, 208: 113–120. DOI: <https://doi.org/10.1016/j.biocon.2016.05.014>
- Davis, C, Del Bianco, V, Peterman, K, Grover, A, Phillips, T** and **Becker-Klein, R**. 2022. Diverse and important ways evaluation can support and advance citizen science. *Citizen Science: Theory and Practice*, 7(1): 30. DOI: <https://doi.org/10.5334/cstp.482>
- Fenichel, M** and **Schweingruber, HA**. 2010. Surrounded by science: Learning science in informal environments. *Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.
- 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). DOI: <https://doi.org/10.5334/cstp.6>
- Jackson, CB, Østerlund, C, Crowston, K, Harandi, M** and **Trouille, L**. 2020. Shifting forms of engagement: volunteer learning in online citizen science. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1): 1–19. DOI: <https://doi.org/10.1145/3392841>
- Kelling, S, Johnston, A, Hochachka, WM, Iliff, M, Fink, D, Gerbracht, J, Lagoze, C, La Sorte, FA, Moore, T, Wiggins, A** and **Wong, WK**. 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? *PLoS One*, 10(10): e0139600. DOI: <https://doi.org/10.1371/journal.pone.0139600>
- Kosmala, M, Wiggins, A, Swanson, A** and **Simmons, B**. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10): 551–560. DOI: <https://doi.org/10.1002/fee.1436>
- McNeill, N** and **Vastine, J**. 2019. Discovering unexpected outcomes mid-stream: Lessons learned from data interpretation affiliation: Alliance for Aquatic Resource Monitoring (ALLARM). *A panel presented at the biannual meeting of the Citizen Science Association*. Atlanta, GA.
- Miles, MB, Huberman, MA** and **Saldana, J**. 2020. *Qualitative Data Analysis: A Methods Sourcebook*. Los Angeles: SAGE Publications.
- Parrish, JK, Burgess, H, Weltzin, JF, Fortson, L, Wiggins, A** and **Simmons, B**. 2018. Exposing the science in citizen science: Fitness to purpose and intentional design. *Integrative and Comparative Biology*, 58(1): 150–160. DOI: <https://doi.org/10.1093/icb/icy032>
- Parrish, JK, Jones, T, Burgess, HK, He, Y, Fortson, L** and **Cavalier, D**. 2019. Hoping for optimality or designing for inclusion: Persistence, learning, and the social network of citizen science. *Proceedings of the National Academy of Sciences*, 116(6): 1894–1901. DOI: <https://doi.org/10.1073/pnas.1807186115>
- Phillips, TB, Ferguson, M, Minarchek, M, Porticella, N** and **Bonney, R**. 2014. *User's Guide for Evaluating Learning Outcomes in Citizen Science*. Ithaca, NY: Cornell Lab of Ornithology.
- Phillips, T, Porticella, N, Conostas, M** and **Bonney, R**. 2018. A framework for articulating and measuring individual learning outcomes from participation in citizen science. *Citizen Science: Theory and Practice*, 3(2): 3. DOI: <https://doi.org/10.5334/cstp.126>
- Resnik, DB**. 2019. Citizen scientists as human subjects: Ethical issues. *Citizen Science: Theory and Practice*, 4(1): 11, 1–7. DOI: <https://doi.org/10.5334/cstp.150>
- Richards, KAR** and **Hemphill, MA**. 2018. A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical Education*, 37(2): 225–231. DOI: <https://doi.org/10.1123/jtpe.2017-0084>
- Saldana, J**. 2011. *Fundamentals of Qualitative Research*. Oxford: Oxford University Press.
- Simonis, JL** and **Pattengill-Semmens, CV**. In prep. Quantifying and accounting for variable observer learning in volunteer-based biodiversity surveys.
- Stevenson, RD, Suomela, T, Kim, H** and **He, Y**. 2021. Seven primary data types in citizen science determine data quality

requirements and methods. *Frontiers in Climate*, 3: 645120. DOI: <https://doi.org/10.3389/fclim.2021.645120>

Stylinski, CD, Peterman, K, Phillips, T, Linhart, J and Becker-Klein, R. 2020. Assessing science inquiry skills of citizen science volunteers: A snapshot of the field. *International Journal of Science Education, Part B*, 10(1): 77–92. DOI: <https://doi.org/10.1080/21548455.2020.1719288>

Wiggins, A and Crowston, K. 2011, January. From conservation to crowdsourcing: A typology of citizen science. In *2011 44th Hawaii international conference on system sciences* (pp. 1–10). IEEE. DOI: <https://doi.org/10.1109/HICSS.2011.207>

Wiggins, A, Newman, G, Stevenson, RD and Crowston, K. 2011, December. Mechanisms for data quality and validation in citizen science. In *2011 IEEE Seventh international conference on e-Science Workshops* (pp. 14–19). IEEE. DOI: <https://doi.org/10.1109/eScienceW.2011.27>

Zevin, M, Coughlin, S, Bahaadini, S, Besler, E, Rohani, N, Allen, S, Cabero, M, Crowston, K, Katsaggelos, AK, Larson, SL and Lee, TK. 2017. Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, 34(6): 064003. DOI: <https://doi.org/10.1088/1361-6382/aa5cea>

TO CITE THIS ARTICLE:

Peterman, K, Del Bianco, V, Grover, A, Davis, C and Rosser, H. 2022. Hiding in Plain Sight: Secondary Analysis of Data Records as a Method for Learning about Citizen Science Projects and Volunteers' Skills. *Citizen Science: Theory and Practice*, 7(1): 35, pp. 1–12. DOI: <https://doi.org/10.5334/cstp.476>

Submitted: 20 November 2021 **Accepted:** 24 August 2022 **Published:** 11 November 2022

COPYRIGHT:

© 2022 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.