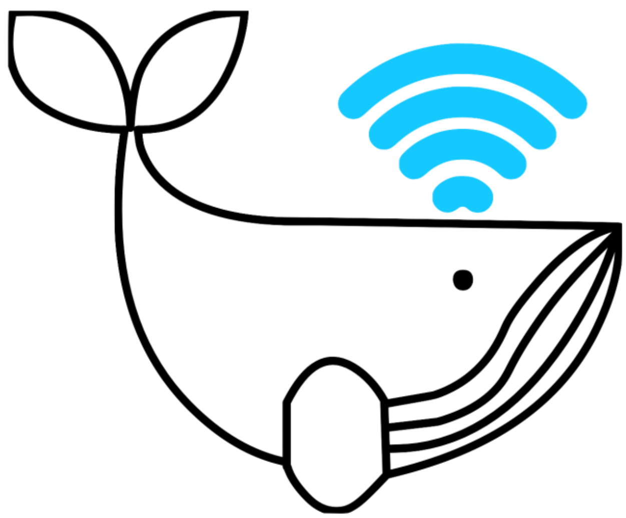
Modernizing Citizen Science: Improving Data Management for Effective Resource Protection

Rae Fuhrman, Sean Goral, Charlene Kormondy,
Jasmine Vazin, Molly Williams



FACULTY ADVISOR: DR. JAMES FREW

CLIENT: CHANNEL ISLANDS NATIONAL MARINE SANCTUARY

BREN SCHOOL OF ENVIRONMENTAL SCIENCE & MANAGEMENT

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

MARCH 22, 2019

Signature Page

Modernizing Citizen Science: Improving Data Management for Effective Resource Protection

As authors of this Group Project report, we archive this report on the Bren School's website such that the results of our research are available for all to read. Our signatures on the document signify our joint responsibility to fulfill the archiving standards set by the Bren School of Environmental Science & Management.

____
Rae Fuhrman

____
Sean Goral

____
Charlene Kormondy

____
Jasmine Vazin

____
Molly Williams

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principal of the School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions.

The Group Project is required of all students in the Master of Environmental Science and Management (MESM) Program. The project is a year-long activity in which small groups of students conduct focused, interdisciplinary research on the scientific, management, and policy dimensions of a specific environmental issue. This Group Project Final Report is authored by MESM students and has been reviewed and approved by:

____
Dr. James Frew, Faculty Advisor

____
Date

TABLE OF CONTENTS

Abstract

Citizen science can enable the collection of large, long-term datasets previously unobtainable through traditional methods. While citizens can effectively gather data like highly trained scientists, citizen science projects often lack proper data management, which limits data usability. The National Oceanic and Atmospheric Administration's (NOAA) Channel Islands National Marine Sanctuary (CINMS) has 20 years of marine mammal sighting data collected almost daily by citizen scientist volunteers in their Naturalist Corps program. The trained Naturalists record observations on smart phone apps aboard whale watching boats in the Santa Barbara Channel. This data has influenced local resource protection policy, but has greater potential to be used by researchers, conservation managers, and decision-makers. We outline CINMS' current citizen science data management system and identify improvements that will maximize the usability of this rich dataset. We created an open-source Python application to collate sightings data into a serverless SQL database that is structured to comply with international data quality standards and open repository requirements. We analyzed the quality differences between sightings collected by Naturalists and the general public, and found that trained users had a significantly lower percentage of potentially inaccurate sightings. We recommend that our client incorporate tools to filter these inaccuracies in the app interface, and follow data management best practices to maximize the impact of their citizen science program.

Executive Summary

This project uses the Channel Island National Marine Sanctuary (CINMS) as a case study of a long-term citizen science data collection program. Our objective is to assess the validity of app-based marine mammal sighting data and transform it to meet the Ocean Biogeographic Information System (OBIS) international repository's data standards. Using a data management system that incorporates cleaning and validating protocols will allow resource managers and researchers to use CINMS data more extensively to address their goals. We also advise CINMS on methods for structuring the app-collected data for integration into national repositories to facilitate broader use. Our analysis of and contributions to CINMS's data management should be applicable to other conservation and resource management activities where citizen science may be able to address critical knowledge gaps.

Background

Scientific investigations are often constrained by time, funding, and staffing. Citizen scientists—members of the public (usually volunteers without advanced scientific degrees) who contribute to research—can collect data beyond what is feasible with traditional research methods. Citizen science projects address diverse subjects, at a multitude of scales, and can be initiated by government agencies, universities, or private research institutions. The federally-run CINMS has been collecting marine mammal sighting data through their volunteer Channel Islands Naturalist Corps (CINC) program since the late 1990s. In 2013, CINMS transitioned their data collection to two smartphone apps: Whale Alert and Whale Spotter Pro, which are currently being combined into a single app called Ocean Alert. These apps have simplified and expanded CINMS's capability for data collection, and the data they collect has influenced CINMS's resource protection policy and strategies. In order to maximize the effectiveness of this wealth of data, thorough documentation and management are required. Usable and accessible data collected through these apps can strengthen the evidence base for marine policy and conservation management decision-making.

The data collected by the CINC volunteers aboard whale watching cruises is uploaded to Conserve.io (a data storage contractor hired by NOAA to manage the app outputs), where it can be accessed with login information provided by CINMS. Researchers and managers from government agencies (e.g. NOAA) and non-governmental organizations (e.g. Cascadia Research Collective, a private whale research institution) are using this data, but have needs addressable with improved data management. These include making complex data structures easily accessible in a single file, including weather and photos attached to other sighting information. A major component of this project was identifying best practices for effective citizen science and what appropriate data management looks like in practice. To accomplish this, we conducted an extensive literature review and interviewed the CINC program coordinator and volunteers. Our main deliverables include these findings, and we structured our project around aligning CINMS with proper data management practices, including generating and maintaining a database that is standardized, well-documented, and accessible.

Best Practices for Citizen Science and Data Management

CINMS is a federal agency mandated to protect marine ecosystems, facilitate the recovery of endangered species, track ecological changes, and monitor the effectiveness of management strategies. Addressing these goals requires high quality, long-term ecological monitoring (LTEM) data. The CINMS whale sighting data, collected by volunteers over two decades, is a rich LTEM resource with tens of thousands of observations. Resource managers at CINMS used this data to help justify moving the commercial shipping lanes in the Santa Barbara Channel, an action that faced significant opposition from the maritime shipping industry. Thus, data used to support arguments for resource

protection must be robust enough to withstand legal and political resistance, as well as to serve as evidence justifying conservation action.

Implementing effective data management can lead to improved quality and increased usability of data. One way to ensure quality and applicability of citizen science data is to create automated processes that format datasets according to standardized requirements, such as those enforced by international data repositories. One such repository is the Ocean Biogeographic Information System (OBIS), which is widely used by the global marine science community. OBIS has specific submission requirements that include standardization of species names and inclusion of detailed metadata. Formatting the app-collected data to be consistent with international marine biodiversity data repositories not only increases ease of use by CINMS, but also potentially by other users.

In addition to the standardization of data to the requirements of data repositories, other examples of quality assurance systems for citizen science project design and data management include:

- Using automated "sanity" filters to remove species "sighted" outside of their possible range, or in unrealistically large numbers;
- Following protocols established by a professional researcher;
- Educating, training, and testing for citizen scientist volunteers;
- Characterizing and verifying observer error;
- Asking for feedback from app users through questionnaires;
- Ensuring consistent usership (low volunteer attrition rates).

Results

Our exploration of data management best practices led us to develop an SQLite database for our client. SQLite implements a serverless SQL database in a single disk file. The file format is cross-platform and can be copied between 32-bit and 64-bit systems. We created virtual tables within the database that filter erroneous data and reformat the datasets to meet international data standards (OBIS). We also developed a Python script for constructing the database, which harvests data from the Conserve.io server on which the whale sighting data is stored.

Researchers and policy makers can now:

- Easily access and share the whale sighting data on various computing platforms;
- Visualize the data in a clear, well-documented format;
- Have confidence in the quality of the data because questionable data points are flagged and can be removed.

Recommendations

From our literature review on best practices and our experience working with CINMS data, we developed recommendations for app developers and data managers.

Recommendations for the new Ocean Alert app interface include:

- Ability to abort an incomplete marine mammal sighting. (Currently, users are unable to abort a sighting, and so resort to filling in sighting data with zeros, which can be misinterpreted as incorrect coordinates, counts, etc.);
- Sanity filters to check for correctness if unrealistic values are recorded;
- Location filters so users would not be able to record sightings on land;
- Use data formats (e.g., CSV) that match OBIS requirements.

Recommendations for CINMS data management include:

- Develop comprehensive, formal metadata for all data sets;
- Employ consistent user IDs to help managers identify users that may need assistance with data quality.

OBJECTIVES

1. Assess how Whale Spotter Pro and Whale Alert data is currently collected and stored, and determine how the data can be formatted for submission to a national repository (OBIS).
2. Determine the data needs and preferences of researchers/decision makers that use this type of data, and create a data management application that addresses their needs.
3. Provide recommendations to NOAA, CINC, and Ocean Alert app developers to improve citizen science data management practices, based on best practices in the literature and our findings during the data restructuring process.

SIGNIFICANCE

Scientific investigations often require data collection that extends beyond what trained scientists are capable of within time and budget constraints. One way to bolster datasets is to engage citizen scientist volunteers to collect data in coordination with professional researchers. Citizen collected data can address critical data gaps for resource management and can help engage, empower, and educate the public (Aceves-Bueno et al., 2015). However, providing clean, accessible citizen-collected data for use in research and decision-making is challenging. Organized and efficient data management is needed to make citizen science data more usable for science and policy. When citizen science projects are rigorously designed and documented, they can generate high-quality, sought-after data (McKinley et al., 2016). Improving citizen science data collection and management strategies can help researchers and decision-makers address environmental problems. Rigorous data management and documentation of methods in citizen science projects allows users to decide if the data sets are applicable for their specific needs, much like other types scientific data collection. Many citizen science initiatives lack stringent documentation and standard formatting protocols; this has led to underutilization of such data by science and policy communities. This project focuses on delineating methods and protocols for NOAA citizen science data to transform it into a format that can be easily applied to scientific and policy pursuits for whale monitoring in the Santa Barbara Channel. Our methods and final product can be applied to other citizen science projects to automatically format and clean data sets to enhance their use.

DELIVERABLES

1. A report which details the quality and unique characteristics of the CINMS citizen science data, including:
    a. The process for making citizen science data compatible with OBIS repository standards
    b. The methods for flagging data that may have been entered erroneously
    c. The effects of citizen science data collection process on the resulting data
2. A cleaned and organized database of all of CINMS citizen science data, suitable for upload into national repositories
3. Recommendations for app user interface design based on best practices researched in literature reviews

BACKGROUND

Citizen scientists are volunteers who collect data for scientific research (Silvertown 2009). Most are amateurs, rather than formally trained scientists, who want to get involved and learn about a subject or help document environmental problems (Cohn 2008). Modern citizen science is becoming widely accessible due to technology and grant stipulations (Silvertown 2009). The widespread availability of the internet and smartphones has allowed citizens to interact easily with professionally designed citizen science projects through websites and phone apps. Advancements in information science, graphical user interfaces, and geographic information system (GIS) web applications have also helped make data collection more accessible to non-specialists (Dickinson et al. 2012). There is also a growing recognition among scientists that citizens represent a powerful untapped source of labor and skills that can help generate scientific knowledge (Silvertown 2009). Research granting institutions, such as the National Science Foundation (NSF) in the U.S. and the Natural Environment Research Council the UK, often require science outreach in the projects they fund as well (Silvertown 2009).

Citizen science is particularly helpful in the environmental field because studying nature requires data across many locations and timespans (Bonney et al. 2009; Burgess et al. 2017; Theobald et al. 2015). Enlisting the public in data collection is a cost-effective method for gathering large quantities of data. Projects using citizen science have successfully advanced scientific knowledge on a wide variety of topics: from documenting bird breeding distributions, to monitoring water quality, to searching the sky for new galaxies (Bonney et al. 2009).

These types of programs also require participants to experience nature directly (Cohn 2008) and thus help them learn about the natural phenomena on which they are collecting data (Bonney et al. 2009). Citizen scientists learn scientific facts related to their projects and about how to use the scientific process (Trumbull et al. 2000; Cohn 2008). For example, after participating in a citizen project with the Cornell Lab of Ornithology (CLO), individuals were likely to increase the number of days bird-watching and to record information about these bird sightings (Bonney et al. 2009). Furthermore, analysis of 750 unsolicited letters from participants in the CLO's Seed Preference Test (SPT) project revealed that participation in the project triggered thinking that aligns with systematic, scientific inquiry. The participant letters contained detailed observations, described participants' efforts to attract birds if their experimental setup failed initially, and showed that participants had generated hypotheses to explain their observations (Trumbull et al. 2000). Citizen science combines research and public education in a way that could lead to broad societal impacts and public engagement with science (Dickinson et al. 2012).

Public engagement projects can have positive social and scientific outcomes as well (Bonney et al. 2014). For example, the Grupo Tortuguero, a sea turtle monitoring network, has citizen scientists collect information on diet, disease, and distribution of turtles in Mexico. This work influenced the creation of several marine protected areas and sustainable fisheries, which had a positive impact on the livelihoods of local people and protected turtles (Delgado and Nichols 2005). The West Oakland Environmental Indicators Project enabled individuals in low-income neighborhoods to collect data on air quality and health that helped document how environmental conditions were affecting residents ("Air Quality West Oakland Environmental Indicators Project" 2013). University College London scientists were able to work with illiterate individuals to use smartphones to capture data on illegal logging and poaching (Krisp 2013). These programs exemplify how citizen science can empower people around the world to use science to investigate and address issues in their local communities (Bonney et al. 2014).

Technical Approaches to Collecting and Using Citizen Science Data

Citizen science projects may be action-oriented projects, conservation projects, or investigative projects (Wiggins, 2011, Table 1). Action-oriented projects recruit citizens who typically design and implement the projects themselves in response to local concerns (Small, 1995). Conservation projects typically support environmental stewardship and natural resource management. Investigative citizen science projects involve collecting data from the environment and educating citizen scientists to follow specific protocols (Wiggins, 2011). The organization structure, project type, and methodology employed by these programs ultimately determines the overall usefulness of the data collected (McKinley et al., 2016).

The Cornell Laboratory of Ornithology (CLO) has a general model for developing a citizen science project:

1. Choose a scientific question
2. Form a scientist/educator/technologist/evaluator team
3. Develop, test, and refine protocols, data forms, and educational support material
4. Recruit participants
5. Train participants
6. Accept, edit, and display data Analyze and interpret data
7. Disseminate results
8. Measure outcomes

An interesting aspect of CLO's largest project, eBird, is their attempt to address and minimize error and bias within their datasets. One method involves steering the bird watcher towards providing more accurate identifications through education and training. CLO also employs automated filter limits that remove species spotted outside of their habitat range, or unusually large numbers of birds spotted within an area (Sullivan et al. 2009). Another mechanism to reduce data bias is called Avicaching. Since citizen science projects often have few restrictions, users tend to have a preferential bias about when, where, and how they collect data (Xue et al. 2016). Avicaching incentivizes eBird users to visit under-sampled areas by awarding them with points for each location they visit, which will enter them into a raffle to win free eBird hats and t-shirts (eBird, 2016). The point-based reward system is a promising technique to attempt to remove some eBird user bias. Overall, CLO understands that successful citizen science projects involve educating and training participants enough to feel confident in their ability to correctly identify species and follow established protocols, as well as finding methods to minimize user error and biases.

A related project to the CINMS Spotter Pro platform was conducted using a mobile app called Whale mAPP to collect marine mammal sighting data in Southeast Alaska. 39 citizen scientists used Whale mAPP to record 1261 marine mammals over a three-month period. Pre- and post-project questionnaires were given to the citizen scientists in order to evaluate the efficacy of a mobile app for scientific, educational, and engagement purposes. It was shown that citizen scientists improved their marine mammal identification skills and self-initiated further learning after two weeks of using the app. The study also attempted to measure and address spatial, user, and sample size biases. Initial enthusiasm for Whale mAPP was high, but usership dropped significantly without continued recruitment of citizen scientist volunteers. The study concludes that targeting self-selecting experts like naturalists and whale watching boat captains is a potential solution for establishing a sustainable citizen science engagement with mobile apps (Hann et al. 2018).

The organizational structures of citizen science initiatives also play a role in the development of projects and the methodologies they implement. *Consultative/functional governance* involves a central agency (e.g., CLO, Whale mAPP) that asks for data collection from the public. These are often referred to as top-down organizations. *Collaborative governance* is often incorporated in co-management plans representing as many stakeholders as possible in the decision-making process. Watershed monitoring often employs this type of governance (EPA, 1997). *Transformative governance* is an organization governed from the bottom-up (Conrad & Hilchey, 2011). These types of grassroots organizations often have no government support, which can be preferable because it shifts leadership to the people in the local community (Bradshaw, 2003). However, transformative governance projects have had many failures due to limited organization credibility and capacity given their lack of formal government backing (Bradshaw, 2003).

Table 1. Successful Citizen Science Projects, their Limitations, and Relevance to Spotter Pro

	Summary of Success	**Limitations/Issues**	**Relevance to Spotter Pro**
CLO eBird	-Largest biological citizen science project in the world -Gathers more than 100 million bird sightings a year -Experiences consistent usership	-Observer bias when recording sightings -Large number of users creates account for variability in accuracy	-Use education and dedicated users to foster sustainable usership and produce higher quality data -Collects photograph data
Whale mAPP	-Citizen scientists collected a relatively large cetacean dataset using a mobile app -Overlapped citizen science dataset with scientifically collected Stellar sea lion dataset to assess accuracy	-Only available on Android -Unable to sustain usership beyond initial citizen recruitment -Observer bias and accuracy uncertainty -Unable to maintain/update interface to operating system (OS) beyond initial grant funding	-Citizen scientists collecting cetacean sightings using a mobile app -Similar goal of uploading datasets to national and international data repositories -Overlapping citizen science datasets with scientifically collected cetacean datasets
Safecast	-Provides daily radioactivity maps in Japan using citizen collected data -Project is designed, implemented, and maintained by citizens	-Map accuracy is limited by the number and location of Safecast units -Lack of government support for funding and project expansion	-Ongoing environmental concerns contribute to its continued use -All code used to clean, analyze, and map the data is publicly available

The examples highlighted in Table 1 exemplify different approaches to crowd-sourced data collection. Safecast represents an open-source, bottom-up approach to citizen science that uses advancing technology and concerned citizens to provide real-time radiation mapping. Cornell Laboratory of Ornithology and Whale mAPP represent top-down approaches that institute specific methodologies tailored to answer specific research questions or reach conservation goals. In all of these cases, specific protocols increase the scientific usability of the data collected.

Sampling Bias in Opportunistic Data Collection

Citizen science datasets are often collected using opportunistic methods, non-systematic data recording methods, as opposed to standardized collection methods. Whale sightings gathered from whale watching boats are examples of opportunistic data, and are inherently biased towards intentionally locating whales. Opportunistic data collection can lead to sampling bias and detection errors related to survey locations if a measure of sampling effort is not incorporated. Often, citizen scientists have a bias towards recording rare species, which can result in an oversampling effect if effort is not accounted for, but also allows for the increased detection of population changes for these species (Snäll et al., 2011). Species distribution models (SDMs) created from a large quantity of opportunistic observations can be as accurate as designed surveys when all potential sources of bias are considered (Redfern et al. 2006, Sardà-Palomera et al. 2012). Literature comparing these methods has shown that with enough data points, opportunistic datasets often have the same sensitivity and capacity for statistical analysis as systematically-collected datasets (Munson et al. 2010).

Understanding species' geographic distribution is vital for effective conservation management, but distribution data is often severely lacking, particularly for marine species. SDMs identify regions where there is a high likelihood of species presence, and maximum entropy (MaxEnt) models can reduce sampling bias effects in opportunistic data (Fernandez et al. 2018; Elith et al. 2011). For instance, species observations from whale watching boats may result in several recorded sightings of the same animal. This oversampling effect can be mitigated by thinning the data through spatial and temporal filtering. Effort can be accounted for by mapping the path taken by the whale watching boat (Fernandez et al. 2018).

Standardization of Training and Approach

A meta-analysis of 71 citizen science-based species identification publications found that training and volunteer data-collecting responsibilities were both important factors in establishing consistency with professionally-collected data (Lewandowski and Specht 2015). Training volunteers extensively on the species of interest before the project began and having a professional present to facilitate data recordings were found to improve accuracy (Bell 2007; Lovell et al. 2009). Designing collection methods to better fit the scope of volunteer knowledge was also shown to be helpful; in some instances, volunteers were more accurate when recording presence-only data without any behavioral characteristics (Moyer-Horner et al. 2012). Conversely, if volunteers are given an option to describe species behavior, these descriptions can lead to definite species identification when reviewed by a professional (Cardamone et al. 2009). Volunteer input options can dramatically influence the reliability of the overall dataset and should be addressed on a case by case basis.

Citizen Science and Marine Management

A large dataset that can serve as an evidence base for decision making is critical for effective planning and management (Hyder et al., 2015). This is true for any biome, but data in marine ecosystems is often particularly lacking. Because data collectors are often unpaid volunteers operating across wide swaths of space and time, citizen science is a cost-effective strategy for obtaining increasingly large datasets to support evidence-based decision-making. Although smartphone apps have expanded the capability of citizen science to engage diverse volunteers across large landscapes, marine-based citizen science projects are still rare. Collaboration between experts and members of the public is crucial to decision-making as it helps build trust in scientific findings. However, challenges exist to the uptake of marine citizen science. A 2015 review found that 14% of marine citizen science projects deliver policy-relevant data, compared to 37% of citizen science projects overall. As referenced in the above sections, data quality is a main point of contention. Data used to support arguments for resource protection must be robust in the face of both legal and political opposition. Therefore, quality assurance systems including validation, verification, and characterization of observer error margins should be built into citizen science data collection frameworks. Creating transparent, open-access databases for this data is recommended to further the acceptance of citizen science generally (Hyder et al., 2015).

Channel Island Naturalist Corps

The Channel Island National Marine Sanctuary (CINMS) has been overseeing its own citizen science project since the 1990s. This initiative uses many of the successful methodologies and best practices discussed earlier in this report, and has helped create a uniquely massive dataset on marine life in the Santa Barbara Channel. Initially, cetacean sightings were recorded on paper logs. Since 2013, volunteers have entered data directly into the Whale Spotter Pro mobile application while aboard marine vessels, typically the *Condor Express* whale watching boat based in Santa Barbara, CA. Only trained CINC volunteers can access the Spotter Pro app, but another CINMS mobile application called Whale Alert allows the general public to record cetacean sightings. The CINC volunteers receive six weeks of training before they are given access to Spotter Pro. Each volunteer is provided with six informational videos that they must watch before each weekly meeting, where they discuss the information in the video with CINMS staff. Then they must pass a written test and shadow a certified volunteer twice before they can become certified themselves.

The CINC members communicate among themselves, through email, to stay in touch with the work the team as a whole is doing. Most volunteers write reports on the days that they go whale watching, to summarize what they saw for the other volunteers. There is also an official record of each day’s sightings on the Condor Express Captain’s public blog. Volunteers work with a dedicated volunteer program coordinator who leads the trainings and regularly communicates with the volunteers throughout their time on the project. The consistent interactions among volunteers and with the volunteer coordinator have led to extremely high retention rates in the volunteer group. There are currently 160 trained volunteers, many of whom have been participating in the program for 10-20 years. CINC volunteers serve a minimum of 8 hours each month and attend 3 hours per month of volunteer meetings in order to stay in the CINC program.

CINC Data Use in Policy and Science

In 2012, the CINC citizen science data was part of the evidence that NOAA used to successfully move shipping lanes in the Santa Barbara Channel one nautical mile away from whale feeding hotspots, to reduce ship strikes. The CINC dataset was determined to be the best available to policymakers for this decision, as it is one of the most extensive marine mammal sighting datasets available in this region. NOAA also uses CINC data to determine seasonal shipping speed reductions. Our client, NOAA, and another federal agency, the Bureau of Ocean and Energy Management (BOEM), want to see this data applied to cetacean research and protection projects on the national scale as well.

Cascadia Research Group currently uses CINC data for whale conservation research, using the photos taken through the Spotter Pro app to track the movements of specific whales. Benioff Ocean Initiative grant recipients are also interested in using this data to improve scientific understanding of whale migration patterns along the California Coast. Finally, Flukebook, a new collaborative initiative, is planning to use the CINC photos to create a database of individual cetaceans by analyzing pictures of their tails. All of these groups reported a need for data cleaning and increased ease of access in order to make best use of this dataset.

Data Management

As data collected by citizen scientists becomes increasingly common in various fields of research, the need to manage it becomes increasingly important. Federal agencies such as NOAA (and their subdivisions, including CINMS) are mandated to create policies that support ongoing protection of marine ecosystems and recovery of endangered species. In order to accomplish these goals, track ecological changes, and monitor the effectiveness of management strategies, decision makers and their research teams need access to high quality, long-term ecological monitoring (LTEM) data. However, many LTEM projects do not properly organize and document their data, so that its completeness, quality, and usability are clear to its users (Sutter et al., 2015).

Effective data management can help contribute to improved quality and usability of data (Sutter et al., 2015). An increasing number of publications address the importance of data management in conservation science, and effective management strategies are imperative to ensure usability of large long-term datasets. Decades-long collections must be continually updated and managed to remain relevant and usable (Addison et al. 2018). One way to ensure data quality is to create automated processes that format datasets according to standardized requirements (Costello et al. 2018). Our clients want the data collected through the Whale Spotter Pro and Whale Alert apps to be accessible and useful for their own research projects and management strategies, and to serve the needs of other conservation-oriented projects led by other agencies and NGOs. By formatting the app-collected data to be consistent with international marine biodiversity data repositories, specifically the OBIS, the data will be usable by both CINMS and other researchers and decision makers.

METHODS

Determining Data Flow

The first step in designing a data management process for the Spotter Pro and Whale Alert data was to determine the data flow from the citizen scientist to potential research and policy applications. To understand the details of the process, we observed citizen scientists as they used the apps and interviewed the individuals involved in each step.

Step 1: Citizen scientists collect data

The data originates at the citizen scientist when they record a cetacean sighting in Spotter Pro or Whale Alert. In October 2018, we observed CINC volunteers as they used Spotter Pro on a Condor Express whale watching trip. We also consulted Shauna Bingham, CINMS Citizen Science Coordinator, and read through the Spotter Pro application training materials that contain detailed instructions how the CINC volunteers should use the app to record sightings.

Step 2: Data is uploaded to Conserv.io

The recorded data is stored in the app until the citizen scientists return to a location with internet access. The data is then uploaded to Conserve.io, a data storage and software development contractor that NOAA has hired to manage the app outputs since 2013. Conserve.io stores all uploaded Spotter Pro and Whale Alert data for later use and distribution.

Step 3: Data is disseminated from Conserv.io to interested parties

The data is made available as a comma-separated value (CSV) file to individual researchers upon request. Organizations who currently use or are interested in using this data include NOAA, BOEM, Cascadia Research, the Bennioff Ocean Initiative, and WildMe Flukebook Photo-ID Matching Portal. We had conversations with representatives from each of these organizations to determine what would make the citizen science data more useful to them so it could be more easily used for research or policy applications.

After determining the data flow, our team developed a data management process for the Spotter Pro and Whale Alert data. Our goal was to convert the data into a format amenable to scientific research in an easy-to-query database, allowing researchers to pull out specific data (individual species, locations, etc.) We drew upon our conversations with current Spotter Pro data users to determine how we could format the data to best fit their needs. For example, the CSV downloaded from Conserv.io currently does not follow any national or international data quality standards.

Our team decided to reformat the data to OBIS data quality standards, an international data repository that is commonly used for marine organism data. Once in a standardized format, the data can be uploaded into a data repository, where it can be available for researchers to download. Also, based on input from researchers who currently use Spotter Pro data, we decided to link cetacean sighting photo IDs and vessel track-line data from Conserve.io to each trip ID to make the end product more amenable to effort tracking (which uses track-line data) and mark-and-recapture modeling (which uses photo IDs).

App data collection process

The trained Channel Island National Marine Sanctuary (CINMS) Channel Islands Naturalist Corps (CINC) volunteers record cetacean sightings in the Whale Spotter Pro app from marine vessels. Only trained CINC volunteers receive login information for Spotter Pro. Whale Alert is open to the general public to record cetacean sightings. The CINC volunteers often record sightings from CINMS Vessels, such as the Condor Express (https://condorexpress.com/), a whale watching vessel that is open to the general public and runs 5 to 7 days a week, depending on the time of year. There is at least one CINC volunteer aboard almost every Condor Express trip to record sightings in Spotter Pro and on board other participating whale watch operators, including Island Packers, Santa Barbara Sailing Center, Celebration Cruises, and Channel Islands Sportfishing Whale Watch. With over 160 volunteers, CINMS aims to provide public interpretation and collect opportunistic sightings data on all participating vessels (pers. comm. with Shauna Bingham).

A trained CINC volunteer can also log into the Spotter Pro app if they are on non-whale watching private vessels. When a trip is created in the app, the app prompts the user to record which CINMS Vessel or what “Other Vessel” they are on.

Spotter Pro app users manually record:

- the contact information of the observer
- species (selected from a drop-down menu)
- certainty (certain, possible, uncertain)
- total sighted (including calves)
- calves sighted
- approximate distance that the animals are from the boat (in yards <25, 25-250, >250)
- number of other vessels on the scene
- comments
- behaviors (selected from a drop-down menu, e.g. feeding, fluking, high-speed traveling)
- weather in terms of: visibility (in nautical miles); cloud cover (percent); Beaufort scale - an estimate of wind speed that ranges from 0 to 1, where 0 refers to sea smooth, mirror-like conditions (less than 1 knot) and 12 refers to hurricane-like conditions; and swell (feet)

Additionally, the Spotter Pro app automatically records:

- photo ID
- a unique trip ID
- the date and time that the trip starts (CINC volunteers are instructed to start trip as boat leaves dock)
- the time that the trip ends (no more editing can be done by the app user once trip is ended)
- a track-line showing the path that the boat took

There is an option for the CINC volunteer to review and edit sightings for the day, though there is no sighting delete button. This results in some of the results having sighting designations of zero. The Spotter Pro users are told “the time to enter the species is when the animal(s) are in the immediate vicinity of the boat and not when the captain announces he sees something ahead in the distance” but this method is not always adhered to, especially by Whale Alert Users. The data is stored on the device until the volunteers get back to land and have access to internet. The volunteers are instructed to immediately upload the data when they get back to shore.

Below are tables containing the original metadata for the data collected by Whale Alert and Spotter Pro as well as a table showing the transformation of attribute names to standardized OBIS attribute names. We added attributes to meet OBIS requirements, including scientific names, scientific names

ID from the World Register of Marine Species, and the basis of record (the nature of the data record; for example, HumanObservation).

Original CSV file metadata

Table 2. Whale Alert Metadata

Creator	Channel Islands National Marine Sanctuary
Subject	Citizen Science Collected Marine Mammal Sightings
Description	Whale Alert is an open access phone application that lets users log marine mammal sightings
Publisher	Channel Islands National Marine Sanctuary
Date	2013-2019
Type	Numerical, Point
Format	Comma-separated values file (.csv)
Identifier	Individual or Group Marine Mammal Sightings at single coordinate location
Data Columns	Project_ID, Trip_ID, Sighting_ID, Date, Lat, Long, Bearing, Species, Other_Species, Number_Sighted, Photo, Latitude, Longitude, User, Email, Phone, Notes, Time, Source (Username)
Source	Electronic (Mobile Application)
Language	English
Spatial Coverage	Anywhere the app is being used, but a majority of sightings are in the Santa Barbara Channel off the coast of Southern California, USA
Contact	Sean Hastings University of California Santa Barbara Ocean Science Education Building 514, MC 6155 Santa Barbara, CA 93106-6155 Phone: 805-893-6424 Email: sean.hastings@noaa.gov
File Name	WhaleAlertData.csv

Table 3. SpotterPro metadata

Creator	Channel Islands National Marine Sanctuary
Subject	Trained Volunteer Collected Marine Mammal Sightings
Description	SpotterPro is a login required phone application that lets trained volunteers log marine mammal sightings
Publisher	Channel Islands National Marine Sanctuary
Date	2013-2019
Type	Numerical, Point
Format	Comma-separated values file (.csv)
Identifier	Individual or Group Marine Mammal Sightings at single coordinate location
Data Columns	Project_ID, Trip_ID, Sighting_ID, Date, Lat, Long, Bearing (Angle of Observation), Species, Certainty, Number_Sighted (Total Mammals Sighted), Calves_Sighted, Vessels_on_scene, Other_Species, Comments, Time, Source
Source	Electronic (Mobile Application)
Language	English
Spatial Coverage	Anywhere volunteers use the app, but a majority of sightings are in the Santa Barbara Channel off the coast of Southern California, USA
Contact	Sean Hastings University of California Santa Barbara Ocean Science Education Building 514, MC 6155 Santa Barbara, CA 93106-6155 Phone: 805-893-6424 Email: sean.hastings@noaa.gov
File Name	SpotterProData.csv

Table 4. Standardized attribute names for OBIS

Whale Alert	**SpotterPro**	**OBIS**
id	id	*OccurenceID*
create_date	create_date	*EventDate*
species	Whale Alert Species & Whale Alert Other Species	*vernacularName*
Number_sighted & Total Sighted (Including Calves)	Number Sighted	*individualCount*
lat	lat	*DecimalLatitude*
lon	lon	*DecimalLongitude*
trip_id	trip_id	*collectionID*
Certainty	Animal Status	*occurenceStatus*
Additional attributes added to meet OBIS standards:		*scientificName; scientificNameID basisOfRecord*

Requirements for data management framework based on OBIS repository standards
OBIS has an online manual for OBIS nodes (localized data management groups in charge of ensuring data quality) and data providers, which specifies best management practices (BMPs) for data quality.

OBIS will accept data from any project, individual, organization, or consortium that wants to contribute data. OBIS Data Sources are the owners and custodians of the data; these are authors, editors, and/or organizations that have published one or more datasets through OBIS. OBIS however only harvests data from recognized "OBIS nodes". Until we identify the appropriate organization to serve as the OBIS node, the data management application will be stored at the Bren School of Environmental Science and Management at the University of California, Santa Barbara.

OBIS uses 3 standards: Darwin Core (species occurrence data), Ecological Metadata Language (dataset metadata), and Darwin Core Archive including OBIS-ENV-DATA (sampling events and facts, species occurrences and measurements). We have utilized the Darwin Core and Ecological Metadata Language, but have no appropriate data for OBIS-ENV-DATA. Darwin Core indicates the standardized attribute names for sharing biodiversity data. The Ecological Metadata Language is a standard for metadata expected with OBIS data records. The scientific species names must match a taxonomic register with a Life Sciences Identifier (LSID) from the World Register of Marine Species (WoRMS).

As an additional measure to protect data integrity, OBIS applies automated geographic and data format quality control for incoming data. These quality control procedures help data providers and data managers by checking completeness and detecting possible errors. Data that does not meet data format requirements are sent back to the submitting node. Additionally, quality flags are assigned to each available record for evaluation of fitness for purpose and use and to filter out records lacking completeness. OBIS will not ever change records in this scenario.

The Data Management Application
We created a Python application that automatically downloads app output data from Conserve.io into a cleanly formatted database. Structured Query Language (SQL) queries then perform filtering, aggregation, formatting to OBIS data standards, and export to OBIS. The basic methodology is outlined below:

1. **Download App Data**
   First, the Python application accesses the Conserve.io server and downloads the JSON files for each trip. When the application is run again, only new JSON files are added to the existing local folder and database.
2. **Extract Data from the JSON Files**
   Next, the application parses the species sightings, trip (vessel track), and weather data from each JSON into separate tables using the trip IDs as the unique identifier.
3. **Create an SQLite Database**
   The application creates an empty SQLite database. The sightings, trip, photo-log, and weather tables are then inserted.
4. **Filter the SQLite Database**
   Stored queries are used to view filters, aggregations, transformations, and joins of the datasets, as well as to format the attribute names into OBIS standards.
   a. To meet OBIS and best practice data management standards, data points missing critical information (latitude, longitude, species sighted, or number sighted) are removed.
   b. Impossible values are then removed, such as –999 for number sighted, over 10,000 for number sighted, and all marine mammals sighted on land. The on-shore sightings are removed by intersecting sightings with polygons representing the ocean. The result is a clean View (virtual table) with these incomplete and incorrect rows omitted. These Views are automatically updated when new sightings are recorded.
   c. Once formatting is complete, SQL queries flag potentially inaccurate data entries as a way to address accuracy. An additional SQL OBIS View with a Boolean flag column shows all trips with incomplete data. As an addition to the fully cleaned OBIS Views, this View preserves all data while still alerting users of the potential risks of incomplete sightings.
   d. A set of stock queries demonstrate how to extract useful information like specific species, trends, and probabilities from the database. These queries are stored in the database for other users to access and run.

RESULTS

Data Collection Methods

In all scientific endeavors, research methodology impacts usability within scientific conclusions and policy making. This is true for data collected by trained scientists as well as data collected by citizen science. Below, we report anomalies found in both app datasets and compare the data quality between sightings recorded on Spotter Pro and Whale Alert. Our cleaning and flagging processes highlighted unique effects that the different citizen science data collection schemes had on outputs of the two phone applications.

Spotter Pro

Currently, the Spotter Pro app is used by the CINC Naturalist Corps volunteers, which have been collecting marine mammal data almost daily since 2013. The app entries can be edited, and comments can be added to any sighting before it is uploaded to the server. CINC volunteers are encouraged to review each entry before uploading. Despite this, there were many sightings in the dataset that had no number of organisms sighted. This is because once a Spotter Pro user creates a sighting record, there is no way it can be deleted. Sometimes the captain of the ship spots a group of cetaceans in the distance and the volunteer creates a sighting record with the assumption that they will momentarily be able to count the animals, but then the animals disappear before the boat can catch up to them and the citizen scientist can record the sighted number in Spotter Pro. Since there is no way to delete a sighting ID once it has been created, the volunteers log the sighting as a zero, many times entering zero values for latitude and longitude as well.

We also flagged extremely high values in the total sighted column for review, and there were several anomalous sighting numbers in the data. For example, one record reported a sighting of 1976 dolphins. Dolphins often occur in large pods, so this number is plausible, but it would be impossible to record a sighting that large with that degree of accuracy. In conversations with CINC Volunteers on the Condor Express, Naturalists said that they sometimes input non-round values as estimates for large pods. Very few data points appear in this format—volunteers usually record round numbers when large numbers of cetaceans make accurate counts difficult. This is an interesting result that occurred when citizen scientists realized that they are unable to properly count individuals and have to make large estimations. This is an example of the unexpected data entries that can occur when citizen scientists collect data. These types of entries may suggest unlikely levels of accuracy if a data user is not aware of the rough estimation measures employed when logging large groups of cetaceans.

For example, dolphin pods can be huge and it is incredibly difficult to get an accurate count of total individuals or individual calf counts. CINC volunteers were instructed to count the animal by visual grid and then multiply by 6 or 7 to account for the dolphins that are under water. The CINC volunteers realize that this is a rough estimation, so the volunteers try to communicate with each other and get better at their estimations as they get more experience with the project. These numbers are useful to show the relative size of dolphin pods sighted from day to day, but it should be noted that large pods are very rough estimates. The explanation of these counting methods in the metadata for Spotter Pro can give insight into the applicability of these data points to specific projects based on users' own preferences and standards.

CINC volunteers were also instructed to focus on recording the target species (cetaceans) that the whale watching vessels are primarily looking for. Thus, non-target species (other marine fauna such as sharks, sea turtles, etc.), which are included as a sighting option in the drop-down menu in the app, are sometimes not recorded, according to CINC volunteers. Thus, non-target species are underrepresented in the data and their absence cannot indicate a true absence metric along the trip track lines.

Whale Alert

The Whale Alert application can be downloaded by any mobile phone user, and has also been in use since 2013. Sightings logged on this app only record species, location, time, user, and user comments. There are no track-line and weather metrics associated with these sighting data points. There are no testing or training aspects involved in using the platform, and users can enter as many sightings as they chose in any geographic location. So, while the number of sightings between the two apps is comparable, 19% of the Whale Alert data was flagged by our database for missing or questionable information (Table 6). Many of these flagged sightings were due to whales being recorded on land (16% of the total sightings).

Table 6. Comparative results of data flagging process. A Pearson's Chi-square test performed on the proportions of flagged sightings from Spotter Pro and Whale Alert revealed a significant difference between them ($p < 0.01$, χ^2 (1) = 1641.7). It should be noted that Spotter Pro has a large total organisms sighted value due to dolphin megapods that are frequently seen while volunteers are on whale watching cruises.

	Spotter Pro	**Whale Alert**
Dates in operation	2013-Present	2013-Present
Total sightings recorded	9,293	8,963
Flagged sightings	135 **(1.4%)**	1778 **(19%)**
Total organisms sighted	1,078,615	20,901
Total organisms sighted on land	45,157 **(4%)**	3,431 **(16%)**

SEAtizen Science Data Management Application

The primary product of this project is the SEAtizen Science open source software repository, which formats and stores the data collected from the Spotter Pro and Whale Alter mobile applications. Our application downloads data, creates a database, and allows direct queries on the data as a whole. Prior to our project, all of the output from these two apps was accessible only as CSV files downloadable from Conserve.io. These files did not include trackline, weather, and observer information.

Storing the data in a relational database enables querying using SQL. These queries can answer simple questions (e.g., species abundance during a particular time of year), or more complicated questions using geographic information. For example, the specific query in Figure 1 produces the five-year monthly averages for endangered whales sighted within the bounds of the international shipping lane that traverses through the Santa Barbara Channel. These values can be used to issue vessel speed reduction announcements, and could be used to further study the impacts of shipping lanes on these species. The database can also export data into geographic analysis platforms like ARCGIS or QGIS, which can conduct more complex analyses of the app information.

```sql
SELECT
    strftime('%m', eventDate) AS 'Month',
    ROUND(SUM(individualCount)/5) AS 'Endangered/Threatened Species in Shipping Lane'
FROM
    sp_wa_obis_ocean, ship_lane_2013
WHERE
    INTERSECTS(ship_lane_2013.geom, sp_wa_obis_ocean.geom) AND
    sp_wa_obis_ocean.vernacularName IN ('Humpback Whale', 'Gray Whale', 'Fin Whale', 'Blue Whale')
GROUP BY
    STRFTIME('%m', eventDate);
```

Grid view | Form view

Total rows loaded: 12

	Month	Endangered/Threatened Species in Shipping Lane
1	01	15
2	02	7
3	03	8
4	04	9
5	05	15
6	06	38
7	07	41
8	08	167
9	09	20
10	10	5
11	11	6
12	12	18

Figure 1. Sample Query of the SEAtizen Science Database: Average Monthly Endangered Whales in the International Shipping Lane 2013-2018.

In order for our data management process to be transferred effectively to our client, we encapsulated all components into a Docker container, a standardized unit of software. The Docker program combines all elements needed to run the code into a single package, which allows those without Python coding experience to effectively run the application in any computing environment. We wrote a Dockerfile that builds a Docker image, which contains all of the necessary Python scripts, required libraries and modules, the current database, and a shell script. The Docker image can then be pulled from the online SEAtizen Science Docker Hub by any user directly to their computer, where they can use Docker commands to create the database. The image is essentially a series of stacked layers of information that build off of each other using feedbacks from the database and the programs required to run the Python code (Figure 2). This container will update the database built into the Docker Image with new Spotter Pro and Whale Alert sightings whenever it is run. The updated database can then be copied from the Docker container to the local file system, where the user can perform queries to address their research questions. This process of containerization using Docker allows for seamless transition of our tool and increases overall usability of our code for the client.

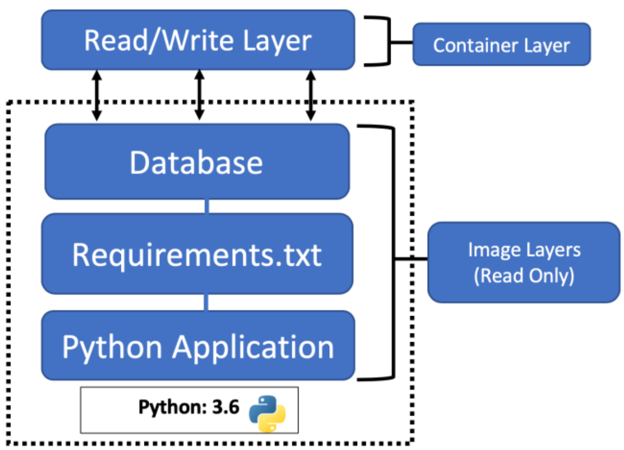


Figure 2. Simple diagram of the SEAtizen Science Docker image and container layers

Repository Contents

- **Readme**: Describes how to install and use the software. Also contains the list of authors, guidelines for contributing, license notice and background information for the project.
- **Requirements**: Lists the Python modules and libraries needed for a complete Python environment.
- **Python scripts**: The multi-script Python application.
- **Tests**: A testing unit focuses on one small section of functionality and proves it correct. Tests help ensure the code is functioning correctly, which will help current and future development.
- **Licensing**: Open-source license so that others are free to use, change, and distribute the software.
- **Dockerfile:** A Dockerfile is used to build the Docker image with all of the necessary Python scripts, all required libraries and modules within the Requirements.txt, the current database, and a shell script.

CONCLUSIONS AND RECOMMENDATIONS

The Value of Citizen Science

Citizen science endeavors can help answer science and policy questions, and also promote education, empowerment and engagement within communities. The several decades of marine mammal data collection by the Naturalist Corps has benefited the local Santa Barbara County community, as well as the policy and management goals of CINMS.

Beyond its qualitative value, the labor provided by these volunteers can be quantified in financial terms. CINMS performed a formal valuation of the contributions of the Naturalist Corps program for fiscal year 2018, which included all volunteer activities such as outreach and training, in addition to the hours spent in the field. The 164 volunteers contributed a total of 32,800 hours from October 2017 to September 2018, which translates to $954,152, or 15.76 full time employees[1]. We can put this value in context by comparing it to the number of sightings collected in this time period (1,230), and the cost of comparable standardized collection. At CINMS, this most often involves aerial flight surveys. Each flight costs $2,500, and happens once a month, resulting in a cost of $30k incurred each year for approximately 10-20 sightings. Aerial surveys provide a different metric of determining whale presence, but in terms of sighting volume, this method does not produce a fraction of the amount of data collected from boats by Naturalist Corps members.

Data from additional standardized marine mammal assessment surveys can also be compared to Spotter Pro and Whale Alert outputs. CINMS uses data from two government-funded recurring scientific cruises: NOAA-funded CalCurCEAS (US West Coast) and CalCOFI (California coast), a joint project between NOAA and the California Dept of Fish and Wildlife. These cruises produced intermittent data from 2008-2015, resulting in a total of 1997 sightings during this time period. The Spotter Pro dataset alone increases the number of unique observations by several orders of magnitude (Table 6).

Conservation Applications of Updated Database

Opportunistic sighting data collected by Naturalist Corps volunteers has been used in limited but significant ways by CINMS, such as the 2013 decision to move the shipping lanes that bisect the Santa Barbara Channel. Other groups have applied the whale sightings to their research, but have encountered issues with unwieldly data organization and lack of preservation of trackline and weather information. With our database and associated data management process, we are bridging the gap between users and data to address the needs of researchers and policymakers.

Our work enables users at CINMS and beyond to query the database in order to retrieve specific information that serves their needs. After creating filters that flagged potentially erroneous data points (e.g., whales sighted on land, or in unrealistically large numbers), we created virtual tables that omitted these points from the dataset, creating a clean and more accurate version. In addition to removing these questionable points, we converted the Spotter Pro and Whale Alert data to comply with OBIS standards by changing species names to scientific names and altering column headings to match required conventions. Further, we compiled metadata to go along with the datasets that adds additional contextual information about the data and how they were collected. Spotter Pro and Whale

[1] CINMS determined valuation data using the Independent Sector's annual calculation for the value of volunteer time. In 2017, the hourly value determined for California employees was $29.09.

Alert have produced thousands of observation points, which are now ready to be uploaded into OBIS for the first time. From this point, the dataset can be integrated with other similar observational data to serve research and management purposes in an open-source format. Future users can have confidence in the quality of the data and will be able to access it in a clear, well-documented format.

In addition to exporting this data to a globally-used repository, the relational database we created will be updated in real time as new observations are recorded in the apps, and can be easily be queried for details about the species sighted. Finally, the data management process we designed is potentially transferable to other citizen science projects, and may be used by other agencies and NGOs that are either starting or working to improve their programs.

Recommendations

Overall, the Channel Islands Naturalist Corps is a model citizen science program. Its intensive and extensive training regime, combined with volunteer retention and high volumes of data collection, aligns with the recommendations we found in the literature about what makes a citizen science project successful. Our contribution of a data management framework adds a component of control to the process and enhances the usability of this rich and continually expanding dataset. The following are detailed recommendations to address areas for improvement in the app interface and to establish an ongoing data management process.

Recommendations for App Developers

Comparing Whale Alert and Spotter Pro, there is a substantial difference in the proportion of "flagged" data points detected in each dataset. This indicates that the Naturalist Corps training is effective in producing more accurate data collectors. However, this system can be improved in a few critical ways. A few volunteers were responsible for several flagged sightings, so it would be useful to have a functionality within the app for providing regular feedback to volunteers on their data-recording accuracy. While the Whale Alert app is not as rigorous as the Spotter Pro app in the type of data collected, changing the app interface to try and match the methods implemented by Spotter Pro could greatly increase overall accuracy of the platform.

Additional recommendations for the app interface and data processing include:

- Functionality to abort an incomplete sighting (currently, users are unable to abort a recording, which results in a logged sighting at latitude/longitude 0,0 and potential zeroes or other inaccuracies in other sighting categories as well)
- Sanity filters – to check for correctness if unrealistic values are recorded
- Location filters – so users would not be able to record sightings on land
- Outputs organized into a CSV that automatically matches OBIS requirements

We hope that these changes will be incorporated in the combined app, Ocean Alert, in 2019. Naturalist Corps members should also be surveyed on how to improve the app. App users are able to identify confusing interface options or suggest more intuitive app workflows. For example, a volunteer suggested an improvement that could be made to the Species Sighted drop-down menu that could lead to more accurate sighting recordings. The volunteer suggested that "unidentified common dolphin" should go right under "long-beak common" and "short-beak common." This will make it clear that unidentified common dolphin is different from unidentified dolphin, which currently appear right next to each other on the drop-down menu and may confuse an app user.

Recommendations for Data Management

One of the main requirements of OBIS is to have metadata associated with any submitted data set. Ideally, comprehensive formal metadata that clearly explain the data collection process will be generated and updated regularly to enhance the usability of the data. This will help data users understand potential biases, anomalies, or unique characteristics of the data in order to make the best use of it in their research. We recommend that the following information be included in the metadata for Spotter Pro, based on observations of the CINC volunteers as they used the app to record sightings:

1. The estimation methods that CINC volunteers are taught, so that any data users are aware of the estimation methods used to count large groups of animals
2. The decision by CINMS leadership to have CINC volunteers focus on recording target species over non-target species.

Consistent user IDs would be helpful for managers to identify users that frequently have erroneous data inputs. Currently, single users are in the database with several different names, making it challenging to determine users that may need additional training, or to filter out inaccurate data by specific users. Assigning a specific user ID to each Naturalist Corps volunteer would mitigate this inconsistency.

References

Addison, P. F. E. *et al.* A new wave of marine evidence-based management: emerging challenges and solutions to transform monitoring, evaluating, and reporting. *ICES Journal of Marine Science* **75**, 941–952 (2018).

Aceves-Bueno, E., Adeleye, A. S., Bradley, D., Tyler Brandt, W., Callery, P., Feraud, M., … Tague, C. (2015). Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. *Ecosystems*, *18*(3), 493–506. https://doi.org/10.1007/s10021-015-9842-4

"Air Quality West Oakland Environmental Indicators Project."2013. http://www.woeip.org/airquality/

Bell, S., Marzano, M., Cent, J., Kobierska, H., Podjed, D., Deivida, V., … Muršič, R. (2008). What Counts? Volunteers and Their Organisations in the Recording and Monitoring of Biodiversity. *Biodiversity and Conservation*, *17*, 3443–3454. https://doi.org/10.1007/s10531-008-9357-9

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, *8*(6), e1000385. https://doi.org/10.1371/journal.pbio.1000385

Bonner, S. (2016, January 29). Safecast Air Prototypes. Retrieved May 7, 2018, from https://blog.safecast.org/2016/01/safecast-air-prototypes/

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next Steps for Citizen Science. *Science*, *343*(6178), 1436–1437. https://doi.org/10.1126/science.1251554

Bradshaw, B. (2003). Questioning the credibility and capacity of community-based resource management. *Canadian Geographer; Toronto*, *47*(2), 137–150.

Brown, A., Franken, P., Bonner, S., Dolezal, N., & Moross, J. (2016). Safecast: successful citizen-science for radiation measurement and communication after Fukushima. *Journal of Radiological Protection*, *36*(2), S82. https://doi.org/10.1088/0952-4746/36/2/S82

Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., … Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, *208*, 113–120. https://doi.org/10.1016/j.biocon.2016.05.014

Cohn, J. P. (2008). Citizen Science: Can Volunteers Do Real Research? *BioScience*, *58*(3), 192–197. https://doi.org/10.1641/B580303

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment; Dordrecht*, *176*(1–4), 273–291. http://dx.doi.org/10.1007/s10661-010-1582-5

Costello, M. J., Horton, T. & Kroh, A. Sustainable Biodiversity Databasing: International, Collaborative, Dynamic, Centralised. *Trends in Ecology & Evolution* **33**, 803–805 (2018).

Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, *10*(1), 117–128. https://doi.org/10.1007/s10530-007-9114-0

Delgado, S., & Nichols, W. J. (2005). Saving Sea Turtles from the Ground Up:, 16.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., … Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Dorazio, R. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2014) 23, 1472–https://doi.org/1484.10.1111/geb.12216

eBird Avicaching - eBird. (2016). Retrieved November 17, 2018, from https://ebird.org/ebird/science/applied-projects/avicaching/

Elith, J., Phillips S. J., Hastie T., Dudik M., Chee Y. E., Yates C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, (Diversity Distrib.) (2011) 17, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Environmental Protection Agency. (1997). *Volunteer Stream Monitoring: A Methods Manual* (No. EPA 841-B-97-003) (p. 227).

Fernandez, M., Yesson, C., Gannier A., Miller P.I., Azevedo J. M. N., (2018). *Marine Ecology Progress Series, Vol. 595: 217-231*
https://doi.org/10.3354/meps12551

Fletcher, S. (2007). Converting science to policy through stakeholder involvement: An analysis of the European Marine Strategy Directive. *Marine Pollution Bulletin*, *54*(12), 1881–1886. https://doi.org/10.1016/j.marpolbul.2007.08.004

Freedman, R., Herron, S., Byrd, M., Birney, K., Morten, J., Shafritz, B., … Hastings, S. (2017). The effectiveness of incentivized and non-incentivized vessel speed reduction programs: Case study in the Santa Barbara channel. *Ocean & Coastal Management*, *148*, 31–39. https://doi.org/10.1016/j.ocecoaman.2017.07.013

Hann, C. H., Stelle, L. L., Szabo, A., & Torres, L. G. (2018). Obstacles and Opportunities of Using a Mobile App for Marine Mammal Research. *ISPRS International Journal of Geo-Information*, *7*(5), 169. https://doi.org/10.3390/ijgi7050169

Hyder, K., Townhill, B., Anderson, L. G., Delany, J., & Pinnegar, J. K. (2015). Can citizen science contribute to the evidence-base that underpins marine policy? *Marine Policy*, *59*, 112–120. https://doi.org/10.1016/j.marpol.2015.04.022

International Atomic Energy Agency. (2015). *The Fukushima Daiichi accident*. Vienna. Retrieved from https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1710-ReportByTheDG-Web.pdf

Krisp, J. M. (Ed.). (2013). *Progress in Location-Based Services*. Berlin Heidelberg: Springer-Verlag. Retrieved from //www.springer.com/us/book/9783642342028

Levrel, H., Fontaine, B., Henry, P.-Y., Jiguet, F., Julliard, R., Kerbiriou, C., & Couvet, D. (2010). Balancing state and volunteer investment in biodiversity monitoring for the implementation of CBD indicators: A French example. *Ecological Economics*, *69*(7), 1580–1586. https://doi.org/10.1016/j.ecolecon.2010.03.001

Lewandowski, E., & Specht, H. (2015). Influence of volunteer and project characteristics on data quality of biological surveys. *Conservation Biology*, *29*(3), 713–723. https://doi.org/10.1111/cobi.12481

Lovell, S., Hamer, M., Slotow, R., & Herbert, D. (2009). An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation*, *18*(12), 3295. https://doi.org/10.1007/s10531-009-9642-2

Matteson, K. C., Taron, D. J., & Minor, E. S. (2012). Assessing Citizen Contributions to Butterfly Monitoring in Two Large Cities. *Conservation Biology*, *26*(3), 557–564. https://doi.org/10.1111/j.1523-1739.2012.01825.x

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., … Soukup, M. A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

Moyer-Horner, L., Smith, M. M., & Belt, J. (2012). Citizen science and observer variability during American pika surveys. *The Journal of Wildlife Management*, *76*(7), 1472–1479. https://doi.org/10.1002/jwmg.373

Mullen, K. A., Peterson, M. L., & Todd, S. K. (2013). Has designating and protecting critical habitat had an impact on endangered North Atlantic right whale ship strike mortality? *Marine Policy*, *42*, 293–304. https://doi.org/10.1016/j.marpol.2013.03.021

Munson, M. A., Caruana, R., Fink, D., Hochachka, W. M., Iliff, M., Rosenberg, K. V., … Kelling, S. (2010). A method for measuring the relative information content of data from different monitoring protocols. *Methods in Ecology and Evolution*, *1*(3), 263–273. https://doi.org/10.1111/j.2041-210X.2010.00035.x

Redfern, J., Ferguson, M., Becker, E., Hyrenbach, K., Good, C., Barlow, J., … Werner, F. (2006). Techniques for cetaceanhabitat modeling. *Marine Ecology Progress Series*, *310*, 271–295. https://doi.org/10.3354/meps310271

Riesch, H., & Potter, C. (2014). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, *23*(1), 107–120. https://doi.org/10.1177/0963662513497324

Roberson, L. A., Lagabrielle, E., Lombard, A. T., Sink, K., Livingstone, T., Grantham, H., & Harris, J. M. (2017). Pelagic bioregionalisation using open-access data for better planning of marine protected area networks. *Ocean & Coastal Management*, *148*, 214–230. https://doi.org/10.1016/j.ocecoaman.2017.08.017

Rockwood, R. C., Calambokidis, J., & Jahncke, J. (2017). High mortality of blue, humpback and fin whales from modeling of vessel collisions on the US West Coast suggests population impacts and insufficient protection. PloS one, 12(8), e0183052.

Safecast Tile Map - Data Processing Methodology. (n.d.). Retrieved May 7, 2018, from http://safecast.org/tilemap/methodology.html

Sardà-Palomera, F., Brotons, L., Villero, D., Sierdsema, H., Newson, S. E., & Jiguet, F. (2012). Mapping from heterogeneous biodiversity monitoring data sources. *Biodiversity and Conservation*, *21*(11), 2927–2948. https://doi.org/10.1007/s10531-012-0347-6

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Small, S. A. (1995). Action-Oriented Research: Models and Methods. *Journal of Marriage and Family*, *57*(4), 941–955. https://doi.org/10.2307/353414

Snäll, T., Kindvall, O., Nilsson, J., & Pärt, T. (2011). Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, *144*(2), 804–810. https://doi.org/10.1016/j.biocon.2010.11.010

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

1.

Sutter, R. D., Wainscott, S. B., Boetsch, J. R., Palmer, C. J. & Rugg, D. J. Practical guidance for integrating data management into long-term ecological monitoring projects: Practical Guidance for Data Management. *Wildlife Society Bulletin* **39**, 451–463 (2015).

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., … Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, *181*, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

Trumbull, D. J., Bonney, R., Bascom, D., & Cabral, A. (2000). Thinking scientifically during participation in a citizen-science project. *Science Education*, *84*(2), 265–275. https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<265::AID-SCE7>3.0.CO;2-5

Wiggins, A., & Crowston, K. (2011). From Conservation to Crowdsourcing: A Typology of Citizen Science (pp. 1–10). IEEE. https://doi.org/10.1109/HICSS.2011.207

Wiley, D. D. (2013). Marine Sanctuaries and Marine Planning, 6.

Xue, Y., Davies, I., Fink, D., Wood, C., & Gomes, C. P. (n.d.). Avicaching: A Two Stage Game for Bias Reduction in Citizen Science, 10.