Recommendations

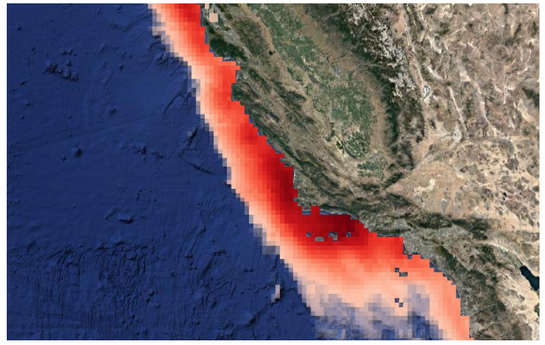
CINMS Data Management

- Use a Data Management Guide for a cleaned, organized, standardized, and more easily accessible CINMS database
- Develop comprehensive, formal metadata for all data sets
- Employ consistent user IDs to help managers identify users that may need assistance to improve observation quality

Ocean Alert App Development

- Add ability to abort an incomplete marine mammal sighting
- Reject unrealistic recorded values
- Reject sightings on land or outside of known geographic ranges
- Make outputs adhere to OBIS requirements

Science and Policy Applications



Research Applications

- Model distributions of multiple cetacean species in the Santa Barbara Channel
- Use photo IDs to track individual whales
- Assess trends in whale migratory patterns

Policy Applications

- Scale to projects at the national level through increased standardization and accessibility
- Help set time frames for the CINMS Vessel Speed Reduction program
- Determine whether to move shipping lanes to reduce whale ship strikes

Acknowledgements

We would like to thank our advisors, Dr. James Frew, Niklas Griessbaum, and Dr. Mark Buntaine, and our clients, Sean Hastings and Shauna Bingham from the Channel Islands National Marine Sanctuary. We would also like to express gratitude to Jacob Levenson, Greg Sanders, John Calambokidis, Julie Hower, Virgil Zetterlind, Courtney Hann, and the Channel Islands Naturalist Corps Volunteers for their valuable input.

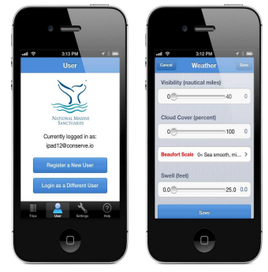
Improving Citizen Science Data Management for Resource Protection

Rae Fuhrman, Sean Goral, Charlene Kormondy, Jasmine Vazin, Molly Williams
Faculty Advisor: James Frew, PhD Student Advisor: Niklas Griessbaum

Visit our website to learn more: seatizenscience.org | Group Project Brief | Spring 2019

Background

Scientific investigations often lack time, funding, and staffing. Citizen scientists—members of the public who contribute to research—can collect data beyond what researchers are able to collect. Our client, the Channel Island National Marine Sanctuary (CINMS) has been collecting marine mammal sighting data through their volunteer Channel Islands Naturalist Corps (CINC) program in the Santa Barbara Channel since the late 1990s. In 2013, CINMS began collecting data with two smartphone apps: Whale Alert, which can be used by anyone, and Whale Spotter Pro, which can only be used by trained CINC volunteers. These two apps will be replaced by a single app called Ocean Alert.



Motivation

Improving data management processes could maximize the usefulness of CINMS citizen science data to scientist and policymakers.

Assessing problems in the data collection of the current apps could inform the design of the new Ocean Alert app.

What does good data management look like?

Clean **Organized** **Standardized** **Accessible**

We gleaned these data management best practices from the scientific literature and applied them to a data management system for the CINMS citizen science data.

Clean

Methods

- Flag unreasonable data points (for example, a sighting of 1 million whales)
- Filter data by location to identify marine mammal sightings recorded on land

Results

	Spotter Pro	Whale Alert
Total sightings	9,293	8,963
Flagged sightings	135 **(1.4%)**	1,778 **(19%)**

Our data forensics process revealed a significant difference in flagged sightings between the two apps. 19% of the Whale Alert data was flagged by our database for missing or questionable data, while 1.4% of Spotter Pro data was flagged.

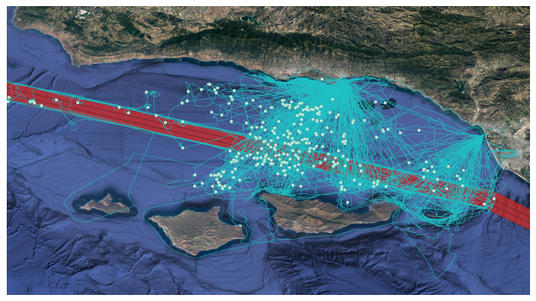
A common data entry error was organism sightings with geographic coordinates on land, or with erroneous zero values. Many sightings showed up along the prime meridian, at 0° longitude.

Organized

Methods

- Preserve all data collected by apps
- Organize data in an easy-to-use database structure

Results



This map of a shipping lane overlayed with 2018 sightings and whale watching boat paths was made possible because our data management process preserved boat paths, which were previously lost during data extraction. Additionally, our database preserves weather data and observer name.

```
SELECT SUM (spotter_pro_obis.individualCount) /
           (SUM (julianday (trips.end_date) -
                 julianday (trips.start_date) ) *24)
AS         "Average Number Sighted/Hour"
FROM       spotter_pro_obis INNER JOIN trips
ON         spotter_pro_obis.collectionID =
           trips.trip_id
WHERE      spotter_pro_obis.vernacularName =
           "Humpback Whale";
```

0.38
Humpbacks Per Hour

What is the average number of Humpback whales sighted per hour?

Standardized

Methods

- Format data according to the standards of a reputable international data repository, the Ocean Biogeographic Information System (OBIS)
- Include comprehensive metadata
- Use consistent column headings
- Use consistent scientific naming conventions for all species

Results

Variable Names

Previous Variable Name	→	Standardized Variable Name
Number_sighted	→	individualCount
species	→	vernacularName

Whale Alert and Spotter Pro use inconsistent variable names, preventing their data from being merged. Our database converts all variable names to OBIS standards.

User Identification

App User ID Entry Examples	→	Standardized Entry Example
Jane Doe jane D. janeD	→	UserID = 55555 OR jane_doe

App users sometimes enter their names inconsistently, making it difficult to track observers. We recommend using a standardized naming convention (e.g. jane_doe) or assigning each user a unique numeric identifier.

Accessible

Methods

- Upload data to a well-known repository (OBIS)
- Make code accessible online in a Github repository
- Package all code and data in a Docker container that can be run on any computer operating system

Results

Data Repository

Now that the data has been formatted to meet OBIS standards, it can be uploaded to OBIS and downloaded by scientists and policymakers around the world. Sensitive user information (e.g. username) would be redacted before uploading to a repository.

OBIS
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

Data Management Guide

Tools used to access, operate, and view the database.

- python — Programming language used to create and update the database
- GitHub — Code repository for ongoing development of the database management application
- docker — Download and run code and data in a single package
- SQLite — Single-file database (with SpatiaLite extension) for storing and querying the data