

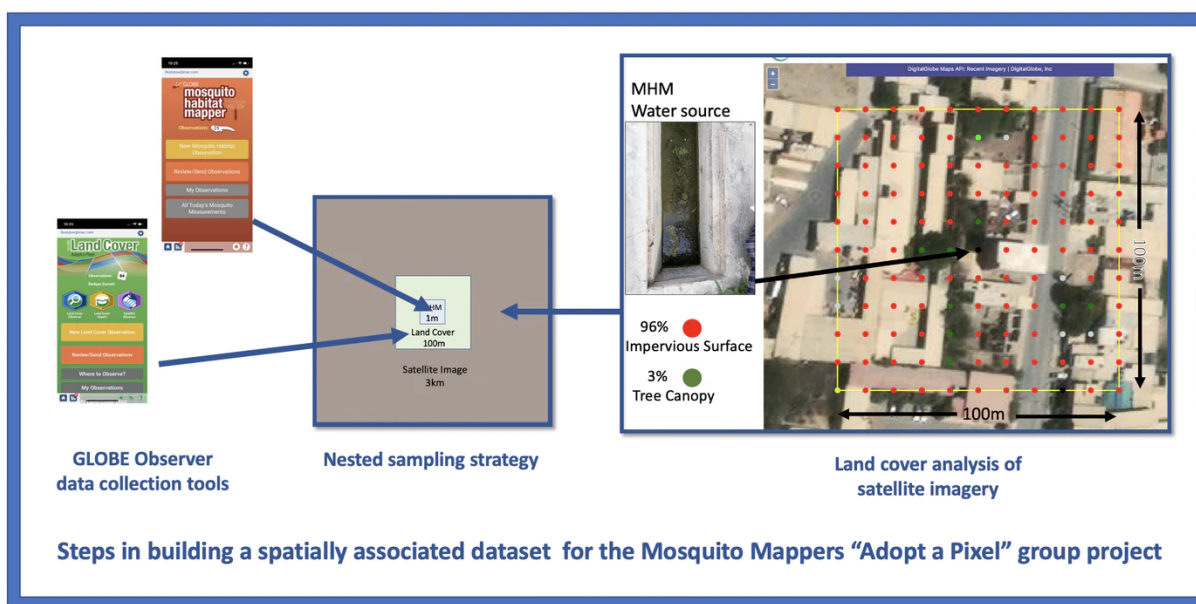
# Citizen Scientist Contributions to Data Quality: Developing Algorithms to Improve Science Outcomes

Authors: Matteo Kimura, Independence High School, Frisco, TX; Pratham Babaria, Harmony School of Endeavor, Austin TX; Prachi Ingle, Westwood High School, Austin TX; Andrew Clark, Institute for Global Environmental Strategies, Arlington VA; Russanne Low PhD, Institute for Global Environmental Strategies, Arlington VA; Peder Nelson, Oregon State University, Corvallis OR

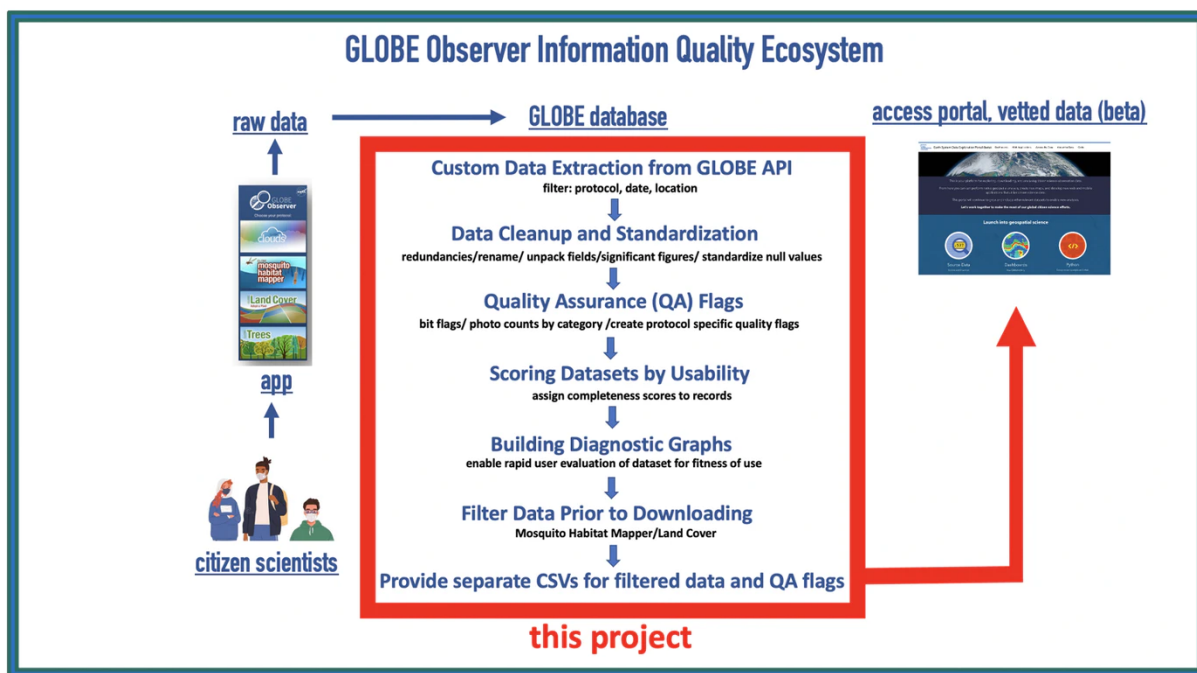
**Abstract:** *Raw data obtained through the GLOBE database requires formatting and a number of preprocessing steps in order to obtain reliable outcomes when using web-based data analysis tools and programs. In this project we automated several data processing steps that enable citizen scientists or professional researchers access to analysis-ready data, downloadable as a CSV file. The steps completed so far include algorithms that enable custom data extraction, data clean up and standardization, quality assurance flagging, and applying an extensible photographic image naming convention. These steps have been applied to Mosquito Habitat Mapper and Land Cover datasets, and provide a model that could be applied to other data products. This*

*project was developed as part of the 2020 STEM Enhancement in the Earth Sciences Mosquito Mappers high school summer research internship, and completed through a spring 2021 data science internship offered by IGES by the first three authors of this poster.*

The Mosquito Mappers team addressed two barriers to the use and analysis of citizen science data. The first project was related analytical challenges related to the opportunistic sampling strategy employed by GLOBE Observer.



The second barrier to using GLOBE Observer citizen science data in research is related to how long it takes to clean and organize the data in the GLOBE database so that it is analysis ready. Our project domain is identified in red, below:



Data preprocessing routines we established as part of this project make it easier for researchers to obtain a subset of data that meets the unique fitness for use requirements for their project. Here is an example output of one of these routines, that enables scientists to quickly visualize and evaluate their data for its intended use. Note that by filtering RAW MHM Data, we reduce the number of records in our curated dataset, but all records in the data return have been evaluated for completeness and data quality.

Since data collection is voluntary, Citizen Science Data is not always complete. So, we created a Completeness Score that indicates the percentage of data fields filled out by the User. Below: Larger, darker circles represent more complete entries.



Data products from our project are currently available through the Earth System Data Exploration portal (beta), developed by IGES to support SEES research interns ([geospatial.strategies.org](https://geospatial.strategies.org)). The SEES 2021 high school internship team, Earth Explorers, will have access to this portal to extract and use preprocessed data using algorithms we created as part of this project.

INSTITUTE  
GLOBAL  
ENVIRONMENTAL  
STRATEGIES

Earth System Data Exploration Portal (beta) Dashboards Web Applications Access the Data About the Data Code

This is your platform for exploring, downloading, and analyzing citizen science observation data.

From here you can perform native geospatial analysis, create new maps, and develop new web and mobile applications that utilize citizen science data.

This portal will continue to grow and include other relevant datasets to enable new analyses.

**Let's work together to make the most of our global citizen science efforts.**

**Launch into geospatial science**

**Source Data**  
Explore and Download

**Dashboards**  
View Global Activity

**Python**  
Data processing scripts on GitHub

Please view our complete poster here: <https://youtu.be/wJZoRr-9IGE>

### About the authors:



Matteo Kimura is a GLOBE Observer citizen scientist. He is a rising high school senior at Independence High School, Frisco, Texas and is currently planning to study computer science. "This project was an incredible learning experience as we got to apply data science techniques to enhance the usability and accessibility of Citizen Science data."





Pratham Babaria is a GLOBE Observer citizen scientist. He is currently a senior at Harmony School of Endeavor, Austin, TX. He is interested in studying computer science and astrophysics and plans to earn

a PhD in computational physics. "My work as a SEES Intern strengthened my skills as a scientist/researcher and taught me the importance of employing the scientific method when engineering solutions to the environmental problems that face us."



Prachi Ingle is a GLOBE citizen scientist. She attends Westwood High School, Austin, TX. Next year she will attend the University of Texas, Austin, where she will major in Computer Science under the Turing

Scholars and the Dean's Scholars Honors Programs. "The NASA SEES Program and this research project helped me to discover my passion for using scientific computing to investigate global issues. I am so thankful for this opportunity to grow as a scientist and as an individual."

Boulder, CO · Updated Wed, April 21



**Rusty Low**