

# CITIZEN SCIENCE **CONFERENCE**

14.–15.10.2020

Knowledge for Change:  
A decade of Citizen Science (2020–2030)  
in support of the Sustainable Development  
Goals



CitSci Earth

## Team Sarjom, USA

Winner EU-WeObserve Open Data Challenge 2020

Presentation by Turam Purty  
Lead UX Researcher, [www.citsci.earth](http://www.citsci.earth)

# Team Sarjom for WeObserve ODC Challenge



## Vision

Create sustainable and  
collaborative technologies  
for the environment

## Mission

Drive community action for  
climate change



**Turam Purty**  
UX Research & Design  
Founder Sarjom



**Kiranmayi KLC**  
Tech Lead, AI/ML Developer  
Open-Source Contributor



**Anindya Pandey**  
Backend Engineering  
Open-Source Contributor



**Vighnesh Misal**  
Frontend Engineering  
Open-Source Contributor



**Ashish Anand**  
Product Manager  
Sustainability & Business Strategy

## AGENDA

In this presentation, we will go over the following topics,

### **Problem Statement**

Citizen Science & WeObserve  
Datasets

### **Concept & Solution**

Wireframes and Demo

### **Recommendations**

Market, Sustainability & Future  
Developments



## PROBLEM STATEMENT

How can we make citizen science datasets accessible and easy to use across a diverse set of projects?

## AREAS OF INTERVENTION

- ① Engaging Young People on open data and climate: Education
- ② Open-source infrastructure for data inter-operability across different projects
- ③ Citizen science volunteers have a choice to get attributions whenever their datasets are published

## UNITED NATION SUSTAINABLE DEVELOPMENT GOALS(SDGS)

Our solution targets the following goals



Image Source: United Nations Sustainable Goals Website

## ODC CHALLENGE THEMATIC AREAS

- ① Engaging Young People on open data and climate: Education
- ② Open-source infrastructure for data inter-operability across different projects
- ③ Citizen science volunteers have a choice to get attributions whenever their datasets are published

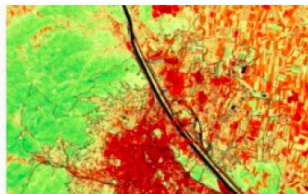
## WEOBSERVE DATASETS

### Issues with Datasets

- 1 Varying data structures
- 2 Varying formats
- 3 Varying sizes

### Possible Solutions

- 1 Normalize data structures
- 2 Process different data formats
- 3 Limit file size for quick analysis

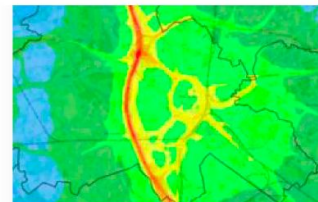


#### LandSense

6 Different Projects

File Formats: csv(16),png(8),geojson(6)

File size Range: 1kb - 3GB



#### GroundTruth 2.0

3 Different Projects

File Formats: csv(4),kml(2)

File size Range: 1kb - 345kb



#### GROW

2 Different Projects

File Formats: csv(16),excel(8),accdb(6)

File size Range: 1kb - 15GB



#### SCENT

1 Single Project

File Formats: Supported by OGC WMS & WFS

File size Range: 13.5 GB

Image Source: WeObserve ODC 2020 Website

## CONCEPT

### **CitSciData-Manager**

An open source tool to effectively manage datafiles and metadata of citizen science projects.

- ❶ CSV File - frequently used
- ❷ GROW BigData File - 15GB CSV
- ❸ SELECT Metadata columns
- ❹ JOIN Metadata columns

### **CitSci Earth - Sarjom**

An open source platform that hosts open source tools, datasets and technologies for environment.

- ❶ Collect Datasets from citizen scientists
- ❷ Publish DataStory - Map/Timeseries
- ❸ Volunteers get attributions
- ❹ Organisations can host samples of their datasets for quick analysis



## TECHNOLOGY ARCHITECTURE

### Schema less approach

The tool is primarily developed using Python, Flask and MongoDB

The diversity of the projects and datasets led us to use a document based database

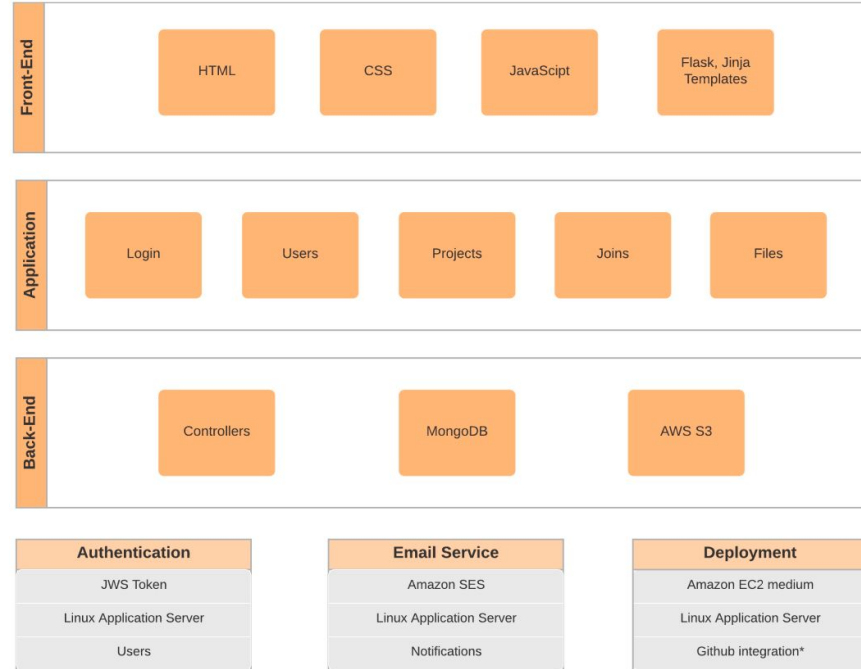
Additional Services were used from Amazon Web Service(AWS)

DataBricks & ApacheSpark BigData Processing tools were used to process 15GB csv file

The repository is currently licensed with open-source GNU Public Version 3 license

### CitSci Earth - WeObserve DevOps

Branch Name - MVP-Merged | Python-Flask-MongoDB Application



\*Github to EC2 Deployment is pending. It will be completed once other features are developed, tested and deployed. WeObserve-ODC branch(mvp-merged) will freeze development and the rest of the development will continue on a separate branch. Documentation of use will be available on the github page of the branch.

## TOOL DEMO

### CitSci-Manager



#### Play Demo Video

In the Demo - CitSciManager works with a GROW BigData CSV file (15GBs) and gets combined with external weather dataset picked up from National Oceanic and Atmospheric Administration(NOAA).

This concept was developed in approximately 1 month

2 weeks for data analysis and research

2 weeks for programming and implementations

#### Current Features

Upload upto 2 CSV Files

Limit file size to 2GB

## STAKEHOLDERS

### Who benefits?

- ➊ Researchers - Climate Change/Environment
- ➋ Citizen Science Volunteers
- ➌ Government and Policy Makers
- ➍ Amateur Enthusiasts
- ➎ Student Community across the world

IMPACT

## Why the solution matters?

- ➊ Saves crucial time and effort lost in data cleaning
- ➋ Researchers can spend more time in data analysis
- ➌ Makes datasets across projects easily accessible
- ➍ Opens up avenues for collaborations and civic engagements
- ➎ Attributions can be provided to data-contributors in compliance with GDPR/CAA
- ➏ Trains the next generation of students about climate change

## MARKET

Sarjom is a not-for-profit organization that collaborates with research agencies and policy organizations to create open-source projects/datasets that university students participate in state, national and international competitions.

### Current Partners/Grants in USA

- ❶ National Science Foundation(NSF)-Innovation Corps Customer Discovery Grant
- ❷ Seattle Local Chapter Partners - COASST project, Vashon Nature Center(VNC)
- ❸ Anecdata.org - An online citizen science data aggregator with 200+ projects, 7000+users.

### Potential Grants

- ❹ National Geographic Society & Microsoft Earth - AI for Earth Innovation Grant
- ❺ National Science Foundation(NSF) - Civic Innovation Grant in collaboration with Anecdata.org
- ❻ Mozilla Open Source Support(MOSS) Grant

### Targets 2021

- ❷ Potential Collaborations with Universities/Research labs in European Union(EU), India & China



### **Vision**

Create sustainable and collaborative technologies for the environment

### **Mission**

Drive community action for climate change

### **Contact**

[hello@citsci.earth](mailto:hello@citsci.earth)

THANK YOU

### **Credits**



Open Data Challenge

