# Progress Report Enhancement of Project AEDES

# Summary

This document summarizes the progress of Project AEDES Enhancement as of February 4, 2022, vis-à-vis the project’s objectives/expected improvements to the platform.

|  |  |
| --- | --- |
| Improvements | Status |
| Automation of data gathering from various sources. | A working Python package is now available.  It allows automated data collection covering the entire data stack. It also provides the capability of visualizing all the points of interest with their proper labels using one line of code. |
| Addition of new weather, satellite, geospatial and socioeconomic data to enrich dataset | Data stack has been enriched with new datasets. Refer to Data Stack section. |
| Open API | Data Management System (CKAN) has been created and being tested to include datasets generated from the Python package. |
| Enhancing the predictive modeling by adding additional ML algorithms to improve model fitting performance applicable to Dengue Forecasting and Hotspot Detection | Ongoing Machine Learning (ML) Model Training. Refer to Models and Frameworks section for more details. |
| Incorporating the INFORM Epidemic Risk Framework with data gathered by AEDES teams to generate location-based risk maps, and advise policy interventions to mitigate the impacts of dengue | Ongoing INFORM Risk model testing utilizing current datasets. |
| Improvement of User Interface to make it feel more like a consumer utility. | UI/UX mockup currently being developed. |

### 

# Data Gathering Automation

## Previous Process

The following diagrams show the existing data gathering processes for each data category.

#### Social Listening Data

Diagram

Description automatically generated

#### Weather Data

A picture containing diagram

Description automatically generated

#### Health Data

Diagram

Description automatically generated

#### Geospatial Data

Diagram

Description automatically generated

## Improved Process

Data collection and processing will now be done in a Python package, effectively eliminating significant manual processes presented previously.

**Python Package Index (PyPI):** <https://pypi.org/project/aedes/>

**Repository**: <https://github.com/xmpuspus/aedes>

Graphical user interface, text, application

Description automatically generated

It utilizes algorithms that pull data from various resources thereby generating the datasets being utilized by the platform.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

### Sample Web Application Data Input

At a user level, the only input required to generate/collect necessary data and features for modelling is a bounding box geojson or a polygon.

Graphical user interface, map

Description automatically generated

# Data Stack

## Previous Datasets

* Google Search Trends
* PAGASA Precipitation
* PAGASA Temperature
* DOH EpiBureau Cases and Deaths
* Sentinel FAPAR
* Sentinel NDVI
* Sentinel NDWI

## Enriched Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Data Usage Type** | **Modelling Use** | **INFORM Component** |
| longitude | Geolocation | Spatial | NONE |
| latitude | Geolocation | Spatial | NONE |
| NDVI | Satellite Remote Sensing | Spatiotemporal | VULNERABILITIES |
| NDBI | Satellite Remote Sensing | Spatiotemporal | VULNERABILITIES |
| NDWI | Satellite Remote Sensing | Spatiotemporal | HAZARDS |
| NDMI | Satellite Remote Sensing | Spatiotemporal | VULNERABILITIES |
| Aerosol Index | Satellite Remote Sensing | Spatiotemporal | VULNERABILITIES |
| FAPAR | Satellite Remote Sensing | Spatiotemporal | VULNERABILITIES |
| Surface Temperature | Satellite Remote Sensing | Spatiotemporal | HAZARDS |
| Precipitation Rate | Satellite Remote Sensing | Spatiotemporal | HAZARDS |
| Relative Humidity | Satellite Remote Sensing | Spatiotemporal | HAZARDS |
| Search Trends | Social Listening | Temporal | VULNERABILITIES |
| Dengue Case Count | Government Data | Spatiotemporal | HAZARDS |
| Distance from n nearest clinics | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Distance from n nearest hospitals | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Distance from n nearest doctors | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Count of clinics within d km | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Count of hospitals within d km | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Count of doctors within d km | Geolocation | Spatial | LACK OF COPING CAPACITY |
| Distance from n nearest school | Geolocation | Spatial | VULNERABILITIES |
| Distance from n nearest college | Geolocation | Spatial | VULNERABILITIES |
| Distance from n nearest university | Geolocation | Spatial | VULNERABILITIES |
| Distance from n nearest kindergarten | Geolocation | Spatial | VULNERABILITIES |
| Count of school within d km | Geolocation | Spatial | VULNERABILITIES |
| Count of college within d km | Geolocation | Spatial | VULNERABILITIES |
| Count of university within d km | Geolocation | Spatial | VULNERABILITIES |
| Count of kindergarten within d km | Geolocation | Spatial | VULNERABILITIES |
| Distance from n nearest toilets | Geolocation | Spatial | HAZARDS |
| Distance from n nearest water points | Geolocation | Spatial | HAZARDS |
| Count of toilets within d km | Geolocation | Spatial | HAZARDS |
| Count of water points within d km | Geolocation | Spatial | HAZARDS |
| Distance from n nearest sanitary dump station | Geolocation | Spatial | HAZARDS |
| Distance from n nearest waste disposal | Geolocation | Spatial | HAZARDS |
| Distance from n nearest waste transfer station | Geolocation | Spatial | HAZARDS |
| Count of sanitary dump stations within d km | Geolocation | Spatial | HAZARDS |
| Count of waste disposals within d km | Geolocation | Spatial | HAZARDS |
| Count of waste transfer stations within d km | Geolocation | Spatial | HAZARDS |

# Models and Frameworks

Model testing is ongoing. The datasets per geolocation are feature-engineered and fed through to a machine learning model. The inference inputs are classified into two: Spatial (longitude, latitude, ISO geotag) and Temporal (year, month, day).

Text, whiteboard

Description automatically generated

INFORM Risk model is being explored based on the datasets presented in the Enriched Datasets table. The exploration includes envisioning the kind of policies we want to influence with the given data and how to make the model effective on any granularity.

# Data Management System

A more user-friendly way to access the data is through a data management system (CKAN), where the datasets can be retrieved through API.

Working URL: <https://aedes-datacatalogue-beta.xyz/>

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Graphical user interface

Description automatically generated

*\*\*\*End of this Document\*\*\**