Enhancement of Project AEDES

Inception Report for

UNICEF Philippines and the Digital Public Goods Alliance

Prepared By

Cirrolytix Research Services

January 2022



Applied Analytics for Social Impact

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

[Introduction 4](#_Toc96088766)

[Overview 4](#_Toc96088767)

[About Project AEDES 4](#_Toc96088768)

[Current State 4](#_Toc96088769)

[Objectives of the Project 4](#_Toc96088770)

[Enhanced AEDES Platform 5](#_Toc96088771)

[Solutions and Enhancements 5](#_Toc96088772)

[Technology Stack 5](#_Toc96088773)

[Data Architecture 8](#_Toc96088774)

[Data Sources 8](#_Toc96088775)

[Licensing 8](#_Toc96088776)

[Data Model 9](#_Toc96088777)

[High-Level Data Architecture 9](#_Toc96088778)

[Datasets 10](#_Toc96088779)

[Data Collection 11](#_Toc96088780)

[Beneficiaries 11](#_Toc96088781)

[Risk-based Framework 12](#_Toc96088782)

[About INFORM Epidemic GRI Model 12](#_Toc96088783)

[Use-Cases 12](#_Toc96088784)

[INFORM AEDES Data 13](#_Toc96088785)

[Project Plan 13](#_Toc96088786)

[Expected Outcomes 13](#_Toc96088787)

[Risks 14](#_Toc96088788)

[Project Activities 14](#_Toc96088789)

[Project Artefacts 15](#_Toc96088790)

[Project Design 15](#_Toc96088791)

[Project Timeline 16](#_Toc96088792)

[Organizational Profile 18](#_Toc96088793)

# Introduction

## Overview

In 2019, the Philippine government declared a national epidemic as it struggled to contain its worst dengue outbreak since 2012. Total cases reached 420,453 and deaths at 1,565, 78% increase in infections and 33% in deaths from 2018. 42% of deaths are children between 5 and 9 years old (Relief Web, 2019). In 2020 and 2021, a total of 90,135 and 79,872 dengue cases were reported, respectively (WHO, 2022).

The battle against dengue epidemic is faced with the following problems:

LACK OF PROACTIVE CASE DETECTION – There is no existing standardized tool for forecasting case counts and identifying possible vector-borne disease outbreak locations.

LACK OF DATA AVAILABILITY – Dengue data (vector-borne disease data, in general) is not easily accessible, while environmental data (climate, remote sensing) are scattered resources.

LACK OF AVAILABILITY OF OPEN-SOURCE TOOLS – There is a wide net of resources that needs to be collected to support epidemic initiatives, most tools/scripts are behind paywalls and entail steep learning curves.

With manual reporting, release of data is delayed which hampers the health sector’s ability to effectively deal with the threat. Prioritizing prevention is essential to mitigate the associated risks of dengue outbreak.

## About Project AEDES

Project AEDES was developed in 2019 as a big data early warning and surveillance system for dengue. The system intends to nowcast probable dengue cases and dengue-related deaths using Google Search Trends, precipitation, and temperature readings from climate data. Google Search Trends represent the public interest and panics related to dengue. Moreover, to detect potential mosquito hotspots, the system utilizes FAPAR, NDVI, and NDWI readings from remote sensing. Information derived from such data would help prioritize interventions and resource allocation.

### Current State

Since the platform was developed under the NASA International Space Apps Challenge, there are several known limitations and opportunity for enhancements, which includes the following:

* Manual data gathering and preparation
* Unavailability of current data
* Lack of user-friendly features
* Unclear policy usage of the information
* Limited scope of solution (i.e. dengue)

### Objectives of the Project

Following the limitations stated above, the key improvements to the AEDES functionality are:

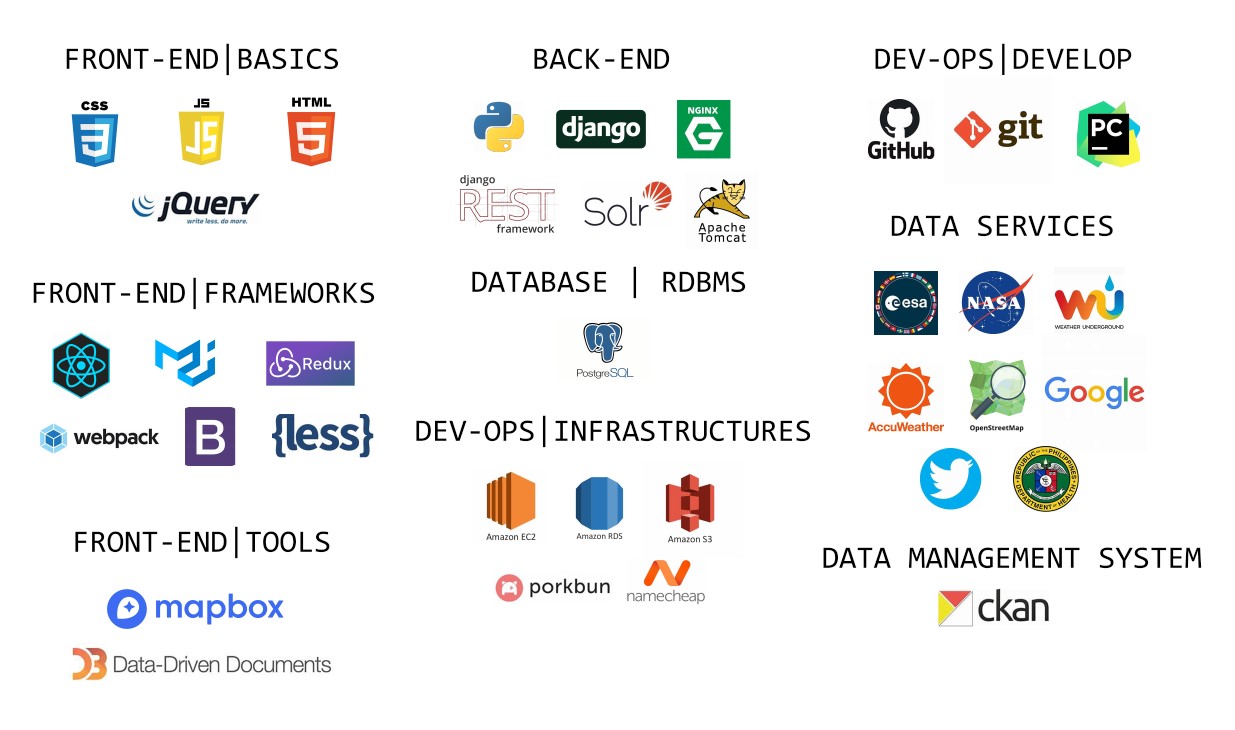
* Automation of data gathering from various sources, especially weather data which was sourced from an offline source (DOST-PAGASA)
* Addition of new weather, satellite, geospatial and socioeconomic data to enrich dataset
* Enhancing the predictive modeling by adding additional ML algorithms to improve model fitting performance
* Integrate AEDES data with INFORM Epidemic Risk Framework to generate location-based risk maps, and advise policy interventions to mitigate the impacts of dengue
* Improvement of User Interface to make it feel more like a consumer utility (e.g. Waze)

# Enhanced AEDES Platform

## Solutions and Enhancements

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FORECASTING |  | HOTSPOT DETECTION |  | INFORM RISK MAPPING |  | OPEN API |  | PYTHON PACKAGE |
| •Predict future number of cases/deaths of vector-borne diseases | •Identify locations of possible hotspots for outbreaks | •Map out risk framework using  data representing  human  settlement situation (hazard and exposure,  vulnerability, coping capacity) | •Publicly open pre-processed  satellite, weather, socioeconomic, and health datasets | •Open-source tools used for data collection, feature engineering and automated machine learning. |

## Technology Stack



|  |  |
| --- | --- |
| **SECTION** | **DESCRIPTION** |
| **Front-end | Basics** | **HTML** - Hypertext Markup Language, a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on World Wide Web pages.  **JavaScript** - an object-oriented computer programming language commonly used to create interactive effects within web browsers.  **CSS** - Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.  **Jquery** - is a JavaScript library that allows web developers to add extra functionality to their websites. |
| **Front-end | Frameworks** | **ReactJS** - React makes it painless to create interactive UIs. Design simple views for each state in your application and React will efficiently update and render just the right components when your data changes.  **Material UI** - Material-UI is an open-source project that features React components that implement Google’s Material Design.  **Redux** - is an open-source JavaScript library for managing the application state.  **Webpack** - is a module bundler for modern Java script applications.  **Bootstrap** - is the most popular CSS Framework for developing responsive and mobile-first websites.  **LessJS** - is a backwards-compatible language extension for CSS |
| **Front-end | Tools** | **Mapbox** - provides powerful routing engines, accurate, traffic-powered travel times, and intuitive turn-by-turn directions to help you build engaging navigation experiences.  **D3 (Data-Drive Documents)** - is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a datadriven approach to DOM manipulation. |
| **Back-end** | **Python** -is a popular programming language. Python can be used on a server to create web applications.  **Django** - Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design.  **Nginx** - accelerates content and application delivery, improves security, and facilitates availability and scalability for the busiest websites on the Internet.  **Django-REST** - Django REST framework is a powerful and flexible toolkit for building Web APIs.  **Solr** - is the popular, blazing-fast, open-source enterprise search platform.  **Tomcat** - is a free and open-source implementation of the Jakarta Servlet, Jakarta Expression Language, and WebSocket technologies. |
| **Database | RDBMS** | **Postgres** - also known as Postgres, is a free and open-source relational database management system (RDBMS) emphasizing extensibility and SQL compliance. |
| **Dev-Ops | Infrastructures** | **Amazon EC2** - Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) Cloud.  **Amazon RDS** - Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud.  **Amazon S3** - Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance.  **Porkbun** - is an amazingly awesome ICANN accredited domain name registrar based out of the Pacific Northwest.  **Namecheap** - is a leading ICANN Accredited Domain Registrar, serving over 2 million customers with over 13 million domains worldwide. |
| **Dev-Ops | Develop** | **Github** - is a web-based version-control and collaboration platform for software developers.  **Git** - is a distributed revision control and source code management system with an emphasis on speed.  **Pycharm** - is a hybrid-platform developed by JetBrains as an IDE for Python. It is commonly used for Python application development. |
| **Data Management System** | **CKAN** - is a data management software with a combination of open-source codebase (including API). It is a tool for making open data websites that helps in managing and publishing collections of data. |

## Data Architecture

### Data Sources

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Satellite and Map Data |  | Social Listening |  | Health-related/Other Secondary Data |  | Climate Data |
| •**NASA GIBS WMS** -  dynamically expose and visualize various satellite layers and indicators  •**Sentinel HUB** access to and visualization of Sentinel satellite data  •**OpenStreetMap** -  access to clinics and hospitals around an area as well as their distances from points of interest | •**Google Trends** access and extraction of public search interest  •**Twitter** - access to dengue-related tweets | •**DOH Epi Bureau** provision data on dengue cases and deaths | •**Wunderground** access to historical weather data  •**Accuweather** access to live weather data |

### ***Licensing***

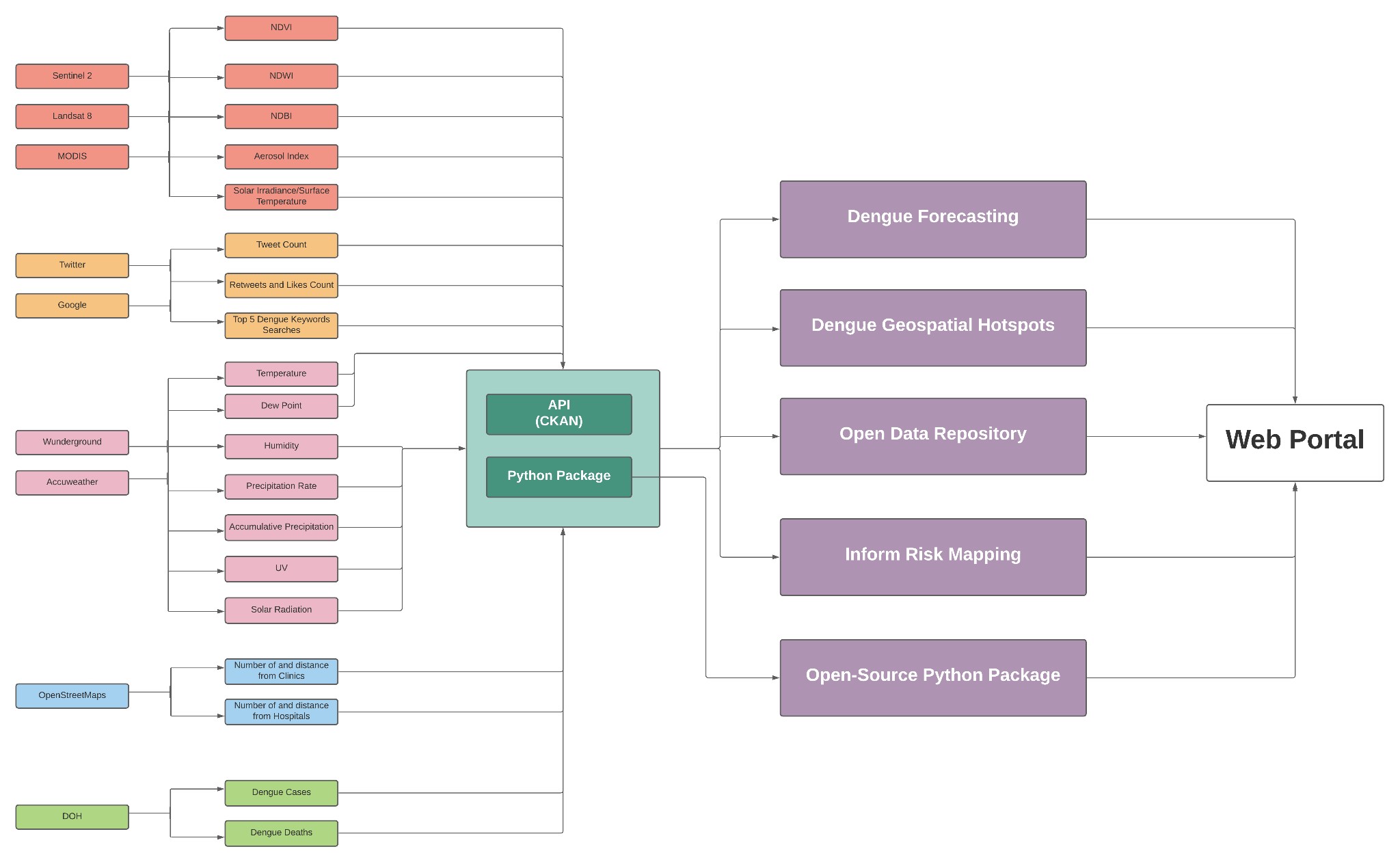
Open Data

* Satellite Data - Landsat 8 (NASA) and Sentinel-2 Copernicus (ESA)
* Wunderground Weather Data - Wunderground
* Google Search Trends - Google https://policies.google.com/terms
* Twitter Data - Twitter
* OpenStreetmap Data - OpenStreetmap
* Disease Surveillance Data - Department of Health (via Freedom of Information)

Freemium

* Accuweather Weather data - Accuweather

### Data Model



### High-Level Data Architecture

Sentinel 2 and Landsat 8 (NASA GIBS WMS/Sentinel HUB)

•Normalized Difference Water Index (NDWI), Fraction of Absorbed Photosynthetic Active Radiation

(FAPAR), and Normalized Difference Vegetation Index (NDVI) extracted using QGIS

•Overlap of indicators provides data for mosquito hotspots which are ingested into Mapbox for visualization

Google Trends and Twitter Data

•Search interest on ‘dengue’, ‘dengue symptoms’, ‘dengue medicine’, and ‘dengue cure’

•Tweets on dengue

•Top search trends and tweet sentiments and engagements are visualized using ChartJS

DOH Epibureau

•Provision data on dengue case counts and deaths by region

### Datasets

Project AEDES relies on the following datasets as input data to the dashboard:

**Google Trends Data**

The raw values, ranging between 0-100, represent search interest as the proportion of all searches of the inputted term for a given period of time and location. A value of 100 is the peak popularity of the term, 50 represents half of the popularity, and 0 means inadequate data.

Search data of related dengue words from Google Trends are collected at specific areas of interest using dengue top 5 dengue keywords as dictated by Google search trends.

**Disease Surveillance Data**

Monthly reported dengue cases and deaths of the selected regions from 2015 to 2018 were obtained from the public records of disease surveillance released by the Department of Health on their website.

**Weather Data**

Time-series data of multiple weather attributes such as temperature, precipitation rate, accumulation of precipitation, wind, solar radiation, etc. are collected from Wunderground and Accuweather.

**Satellite Data**

Process Sentinel2, Landsat, and MODIS band satellite data using python to determine stagnant water locations (NDW), high vegetation (NDVI), high built-up index (NDBI), air quality index (AEROSOL) and solar irradiance/surface temperature in the specified area of interest. The generated dataframe coinciding specific pairs of longitude and latitude are used to determine hotspots of dengue outbreak.

**Twitter Data**

Scrape dengue-related twitter data in order to capture time-series tweet count, engagements such as retweets and likes, and tweet sentiment such as positive, negative, or neutral.

**OpenStreetMap Data**

Availability of clinics and hospitals in an area are available through OpenStreetMap. Data include distance and count from points of interest.

### Data Collection

Data are collected through Python, except for health-related data (e.g. dengue):

|  |  |
| --- | --- |
| **Data** | **Source** |
| Time-series Temperature | https://www.accuweather.com/ |
| Dew Point | https://www.wunderground.com/ |
| Humidity | https://www.wunderground.com/ |
| Wind Speed | https://www.wunderground.com/ |
| Gust | https://www.wunderground.com/ |
| Pressure | https://www.wunderground.com/ |
| Precipitation Rate | https://www.wunderground.com/ |
| Accumulative Precipitation | https://www.wunderground.com/ |
| UV | https://www.wunderground.com/ |
| Solar Radiation | https://www.wunderground.com/ |
| time-series NDVI | https://modis.gsfc.nasa.gov/ |
| time-series EVI | https://modis.gsfc.nasa.gov/ |
| Tweet Count | twitter.com |
| Tweet Engagements (retweets, etc) | twitter.com |
| Tweet Sentiment (positive, negative, neutral) | twitter.com |
| Google trends | google.com |
| geospatial NDWI | https://landsat.gsfc.nasa.gov/ |
| geospatial NDBI | https://landsat.gsfc.nasa.gov/ |
| geospatial NDVI | https://landsat.gsfc.nasa.gov/ |
| geospatial Aerosol Index | https://landsat.gsfc.nasa.gov/ |
| Surface Temperature | https://landsat.gsfc.nasa.gov/ |
| AOI Polygon Geojson | https://boundingbox.klokantech.com/ |
| Hospitals and Clinics | https://www.openstreetmap.org |
| Dengue case count | https://doh.gov.ph/ |
| Dengue deaths | https://doh.gov.ph/ |

## Beneficiaries

The following LGUs will be targeted as the initial beneficiaries of the platform, who will help validate its impact and relevance.

|  |  |
| --- | --- |
| **LGU Coverage** | **Current Status** |
| San Fernando, Pampanga | Inactive MOU – via PSPHP, need to revisit |
| CALABARZON | Existing engagement with DOST R4 |
| Tacloban, Eastern Visayas | No relationship |
| Iloilo, Western Visayas | Initial linkage via DAP, Project SPARTA |
| Cotabato, BARMM | Initial linkage via TAF, UNDP, UNICEF Country Office |

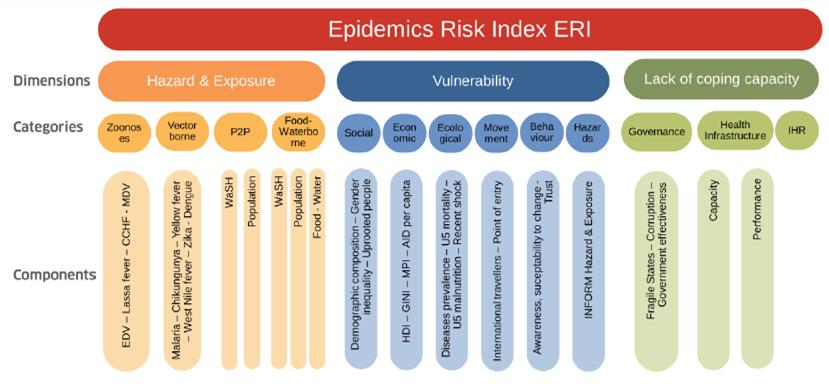
## Risk-based Framework

To further strengthen the relevance of AEDES, one of the major enhancements is to integrate a risk assessment framework for dengue in regions and provinces of the Philippines using the INFORM Epidemic GRI Model. It has the potential to be adapted for evaluating dengue risk at both the national and local levels given that the model is applicable to several types of risks and epidemics.

### About INFORM Epidemic GRI Model

The Index for Risk Management (INFORM) was developed by the JRC and endorsed by INFORM partners. It is a composite indicator that combines various indicators into three dimensions or risk: (2) Hazards, which captures the events that could occur and exposure to them; (2) Vulnerability, which shows the susceptibility of communities to the identified hazards; and (3) the Lack of Coping Capacity, which represents the lack of resources to lessen the impact.

Below figure shows the Epidemic Risk Index Conceptual Framework:



Given that the model follows a consensus-based methodology, it is an appropriate framework to follow for analyzing crisis risk at all levels (global, regional, or national).

### Use-Cases

Following the INFORM model, the indicators shall be modified to suit the platform’s purpose of being and early warning and surveillance system of epidemic outbreaks. It will be designed to achieve the following use-cases:

* **Hazards:** Monitor progress of epidemic, generate alerts
* **Vulnerabilities:** Prioritize areas with vulnerable groups, suggest demographic and geographic determinants of risk
* **Coping Capacity:** Prioritize areas for emergency aid, recommend infrastructure investment

### INFORM AEDES Data

The following datasets per dimension are aimed to be used in the risk-based assessment, subject to data availability and sustainability.

|  |  |  |
| --- | --- | --- |
| Hazards | Vulnerabilities | Coping Capacity |
| Dengue Case incidence  Flood Occurrence  Temperature  Precipitation  COVID-19 Incidence  Access to water  Access to sanitation | Population ages 0-20  Poverty Index  Population affected by natural disasters  Population previously infected by dengue  Mortality  Land-use types  Social listening  Primary and secondary schools  PhilHealth coverage  Human mobility | Presence of health centers  Presence of hospitals  Number of health workers  Health expenditure  Vaccination coverage |

# Project Plan

## Expected Outcomes

|  |  |
| --- | --- |
| **OUTCOME** | **DESCRIPTION** |
| **Data Collection and Processing** | Database Management and Automated Data Ingest for Search Trends, Climate, Satellite, and Health Data which entails continuous research on alternate global open data sources. |
| **Data Analysis** | Incorporate Socio-Economic indicators using Dengue RISK INFORM in the predictive modeling and deploy to all regions. Enhancement of Dengue Case and Deaths Nowcasting. |
| **Product Development** | Redesign AEDES interface and functionalities which include information portals, publicly accessible APIs, and near-real-time daily updates.  This will entail Dengue Trends Overview, Outbreak/Epidemic Monitoring (time-series projections, dengue hotspot map visualizations), At-Risk Community Assessment (risk ranking of regions and provinces, risk maps), and Actions and Recommendations. |
| **Beneficiary Engagement** | Validate impact and relevance through partnerships with local government units as initial beneficiaries. Engage with them through the formation of public health campaigns, target control activities, and on-site validation. |

## Risks

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk (description)** | **Level of risk (high, low)** | **Probability (high, low)** | **Action planned to mitigate** |
| Data  (Timeliness, Accuracy) | High | Low | Reliance on credible 3rd parties (NASA, Landsat, ESA) |
| LGU Buy-in | Medium | Medium | Lobbying local government partnerships through UN system (UNDP, WHO, UNICEF) |
| Dedicated staffing due  to reliance on volunteers | Medium | Low | Dedicated staffing costs to hire full time dedicated staff as part of the project (50% of funding) |

## Project Activities

Pertinent documentations and codebase can be accessed publicly through the project’s designated GitHub repository: <https://github.com/Cirrolytix/aedes_unicef_2022>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Duration** | **Feature details** | **Status** |  | **Result** |
| 1 month | API Development - Satellite (NDWI,  NDVI, NDBI, Aerosol Index, Solar Irradiance/Surface Temperature) from Sentinel HUB, Landsat, Euro  Data Cube | Refining | Automate data collection, repository, and feeder into application for remote sensing and dengue case count nowcasting/forecasting | |
| Socio-economic and weather data (hospitals, population, temperature, humidity, solar radiation, precipitation rate, cumulative precipitation) |
| 1 month | Risk INFORM Framework and Model Integration | Developing | Risk modeling incorporating hazard, vulnerability, and resilience indicators | |
| 0.5 months | Social Listening - Google Trends, Twitter count, engagements, sentiment | Refinement | Automated data gathering of social indicators | |
| 0.5 months | Dengue Trends Overview,  Outbreak/Epidemic  Monitoring, At-Risk Community  Assessment, Action and  Recommendations | Refinement | Descriptive, Predictive, and  Prescriptive  Modules | |
| 0.5 months | Dengue case and deaths nowcast enhancement | Refinement | Show results of prediction including more regions. | |
| 1.5 months | UI/UX Enhancement | Refinement | Mobile Responsive Web-App | |

## Project Artefacts

* **Python Package**: 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.

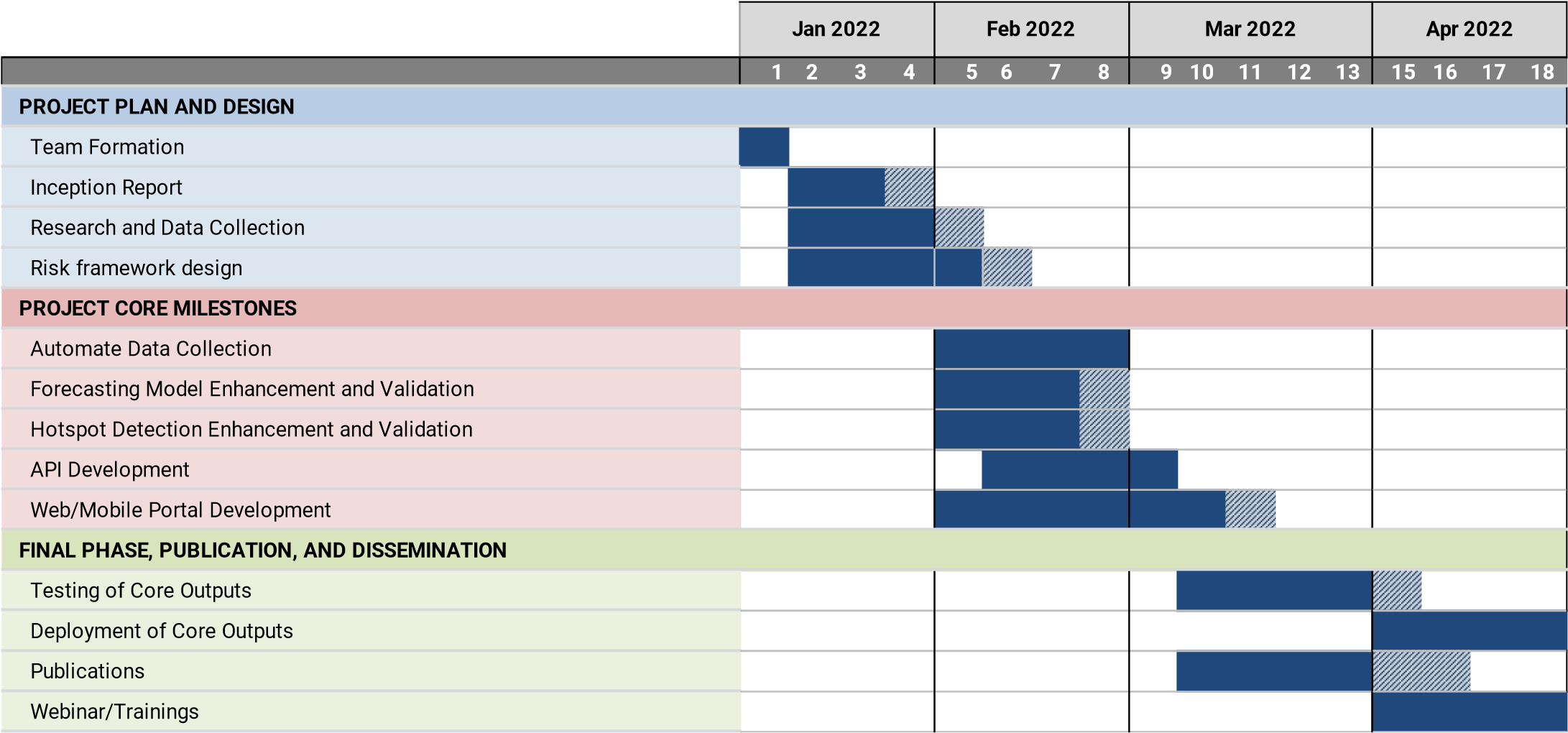
It can be accessed by public through the designated Python Package Index repository: <https://pypi.org/project/aedes/>

* **Data Management System**: Allows for management, retrieval, and publishing of the generated data from the Python package. It will be developed using CKAN (<https://github.com/ckan/ckan>) – an open-source data management system.
* **Open API**: API will operate under [CKAN API](https://docs.ckan.org/en/ckan-2.2.3/api.html).
* **Web Portal** (<https://www.aedesproject.org/>)**:** Main platform which exhibits the AEDES dashboard and the consolidated artefacts.

## Project Design

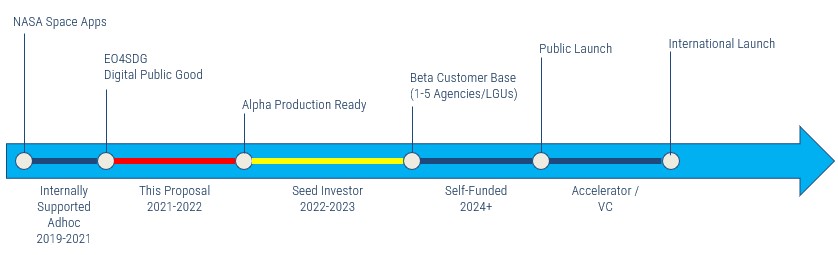
|  |  |  |
| --- | --- | --- |
| **Milestone** |  | **Details** |
| Team Formation and Kickoff | 1. | Finalize team skill requirements |
|  | 2. | Role design |
|  | 3. | Promote job openings |
|  | 4. | Recruit team members |
|  | 5. | Onboarding and debriefing |
| Automate data gathering | 1. | Finalize automation requirements and python package design |
|  | 2. | Automation design   * Remote Sensing * Google Trends * Weather |
|  | 3. | Data ingestion design - dengue cases and deaths |
|  | 4. | Implement automation   * Remote Sensing * Google Trends * Weather |
|  | 5. | Implement data ingestion - dengue cases and deaths |
|  | 6. | Automation testing and evaluation, and finalizing data gathering python package |
| Enhance nowcasting models | 1. | Nowcasting design |
|  | 2. | Data gathering   * dengue cases and deaths * google trends, twitter data * weather and satellite data |
|  | 3. | Nowcasting modeling   * correlations and linear regression * time-series forecasting * Automated Machine Learning |
|  | 4. | Nowcasting testing and evaluation |
|  | 5. | Nowcasting model deployment |
| Enhance mosquito hotspot detection model | 1. 2. | Finalize hotspot detection requirements Hotspot detection design |
|  | 3. | Data gathering   * remote sensing * ground observations |
|  | 4. | Hotspot modeling through geospatial clustering with exogenous data |
|  | 5. | Hotspot model testing and evaluation |
|  | 6. | Hotspot model deployment |
| Web / mobile portal development and enhancement | 1. 2. | Finalize web and mobile portal requirements Design and wireframes |
|  | 3. | Portal development |
|  | 4. | Testing and evaluation |
|  | 5. | Portal deployment to production |
| Implement INFORM Risk Framework | 1. | Finalize risk framework design |
|  | 2. | INFORM inputs data gathering |
|  | 3. | Implementation |
|  | 4. | Testing and evaluation |
|  | 5. | Model deployment |
| Publication and dissemination | 1. | Finalize publication requirements |
|  | 2. | Publication plan and design |
|  | 3. | Manuscript drafting and development |
|  | 4. | Editing and feedback |
|  | 5. | Publication |
|  | 6. | Webinar and training |

## Project Timeline



Development Roadmap

To appreciate the potential of Enhanced AEDES platform, below figure shows an initial development roadmap.



# Organizational Profile

The following resources are involved in this project.

|  |  |  |
| --- | --- | --- |
| **NAME** | **RESPONSIBILITIES** | **EMAIL** |
| **EMILY JO VIZMONTE,**  Project Coordinator and  Research Lead | Manage and oversee the project implementation and provide oversight and direction to project activities. | emilyjo.vizmonte@cirrolytix.com |
| **XAVIER PUSPUS,**  Data Science and  Machine Learning Lead | Prepare data for processing and analysis, leverage machine learning techniques for interpretation and insight generation | xavier.puspus@cirrolytix.com |
| **MARK NEIL PASCUAL,** Platform Development and Engineering Lead | Develop the product design of the AEDES platform and manage the API infrastructure. | markneil.pascual@cirrolytix.com |

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