Using open geospatial data and technology to support malaria risk modelling

Malaria Atlas Project: Data Engineering Team

Who we are

Malaria Atlas Project

A geospatial disease modelling research group. Focussed primarily on malaria

Why malaria?

- Responsible for an estimated 640,000 deaths in 2019
- Disproportionally affects sub-Saharan Africa and kills children under 5

Mission

- Improve understanding of global malaria risk and the impact of interventions
- Disseminate free, accurate and up-to-date geographical information on malaria
- Provide policy makers with information on trends and on which interventions are most effective
- Reduce malaria morbidity and mortality by facilitating targeting of resources

Who we are

Data Engineering Team

Support modelling work by building and using software tools across three core areas:

1. **Data acquisition**

   Identification and ingestion of new datasets relevant to malaria

2. **Data processing**

   Extraction of useful information into a consistent format

3. **Data dissemination**

   Publication of input and output datasets on the web

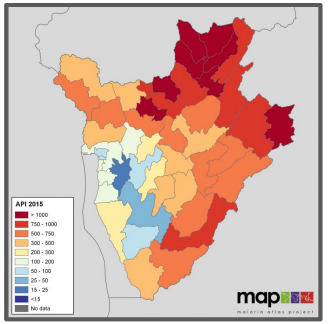
Data Sources

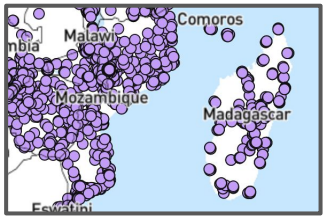
Administrative unit boundaries:
- GADM, GAUL, country ministries of health

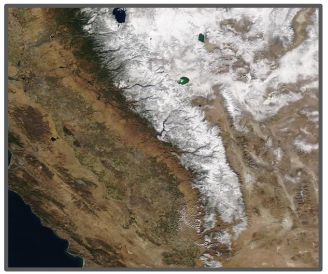
Routine surveillance: data associated with administrative units (polygons)
- MoH reports/websites, WHO reports, privately shared MoH reports

Parasite Rate and Interventions: data associated with survey locations (points)
- Surveys from academic lit. reviews, DHS Surveys, MICS surveys, SPA Surveys

Satellite imagery: environmental data
- NASA MODIS data; other publicly available satellite data

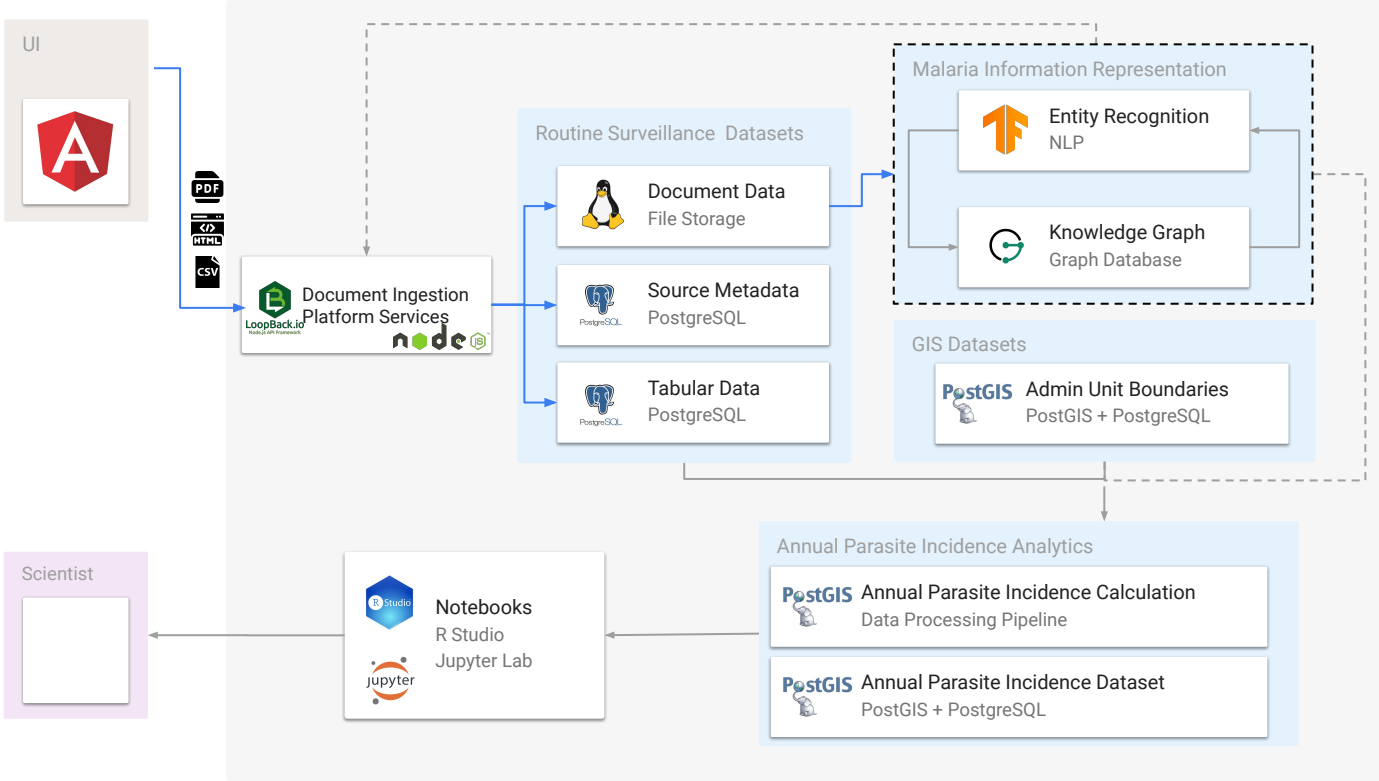






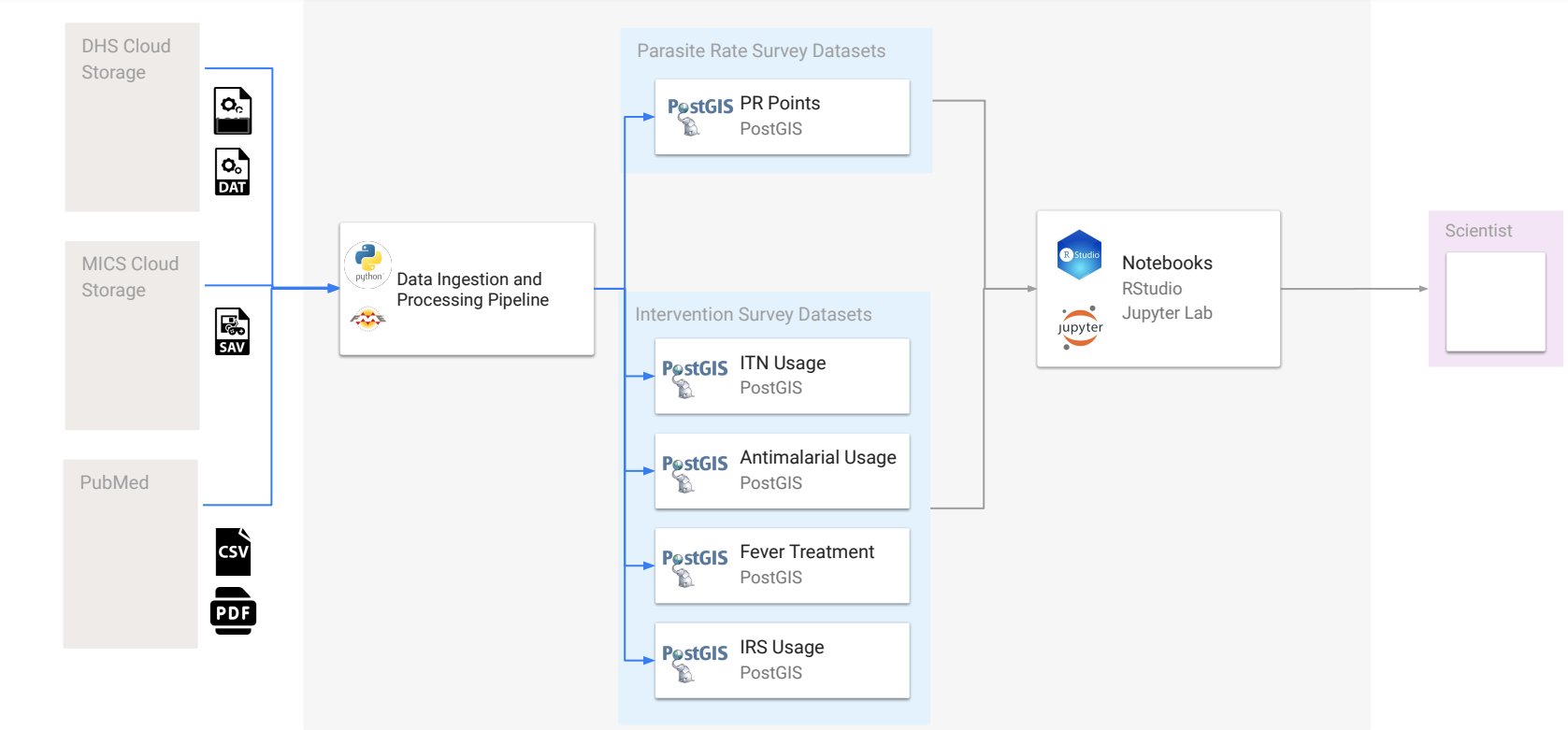
Routine Surveillance Data

Architecture: Surveillance Data

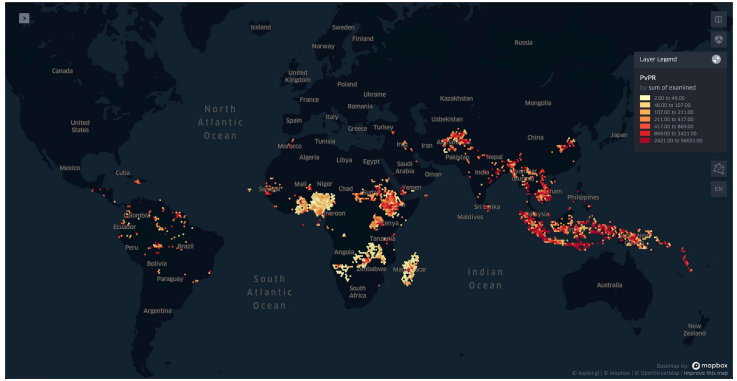


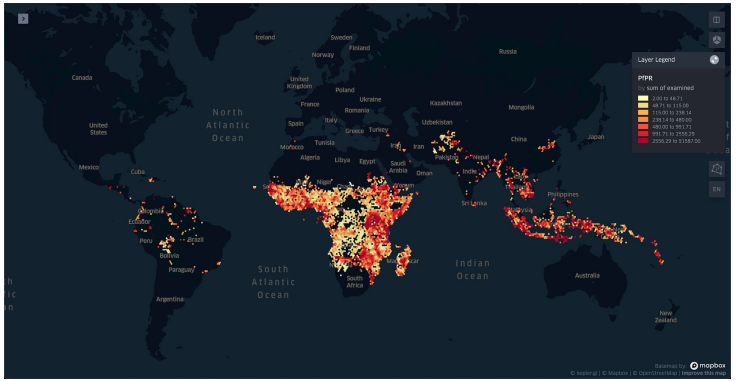
Parasite Rate and Intervention Survey Data

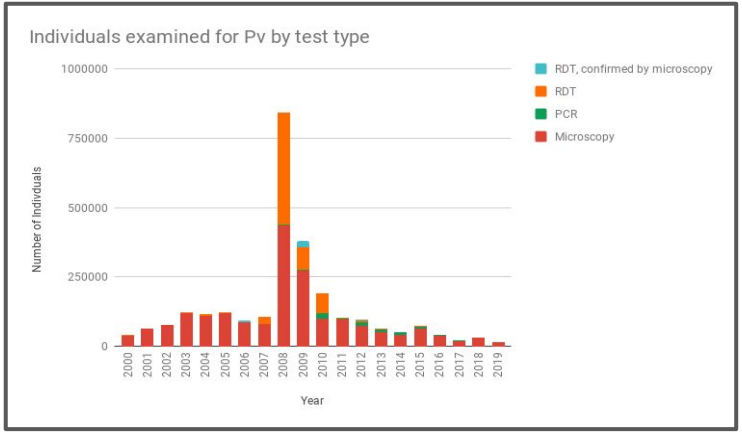
Architecture: Survey Data

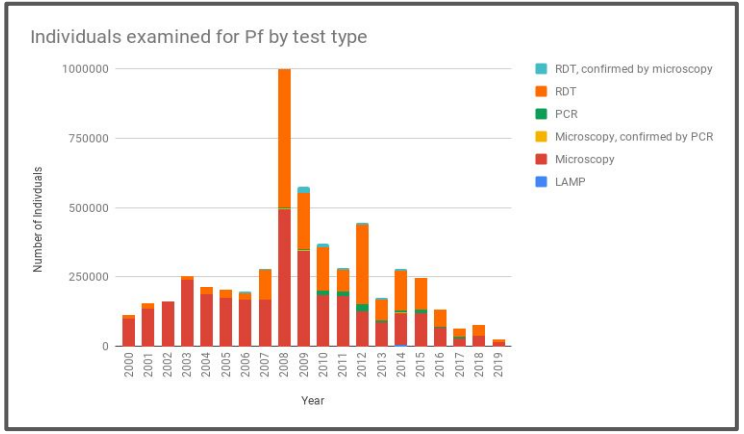


Parasite Rate Survey Data



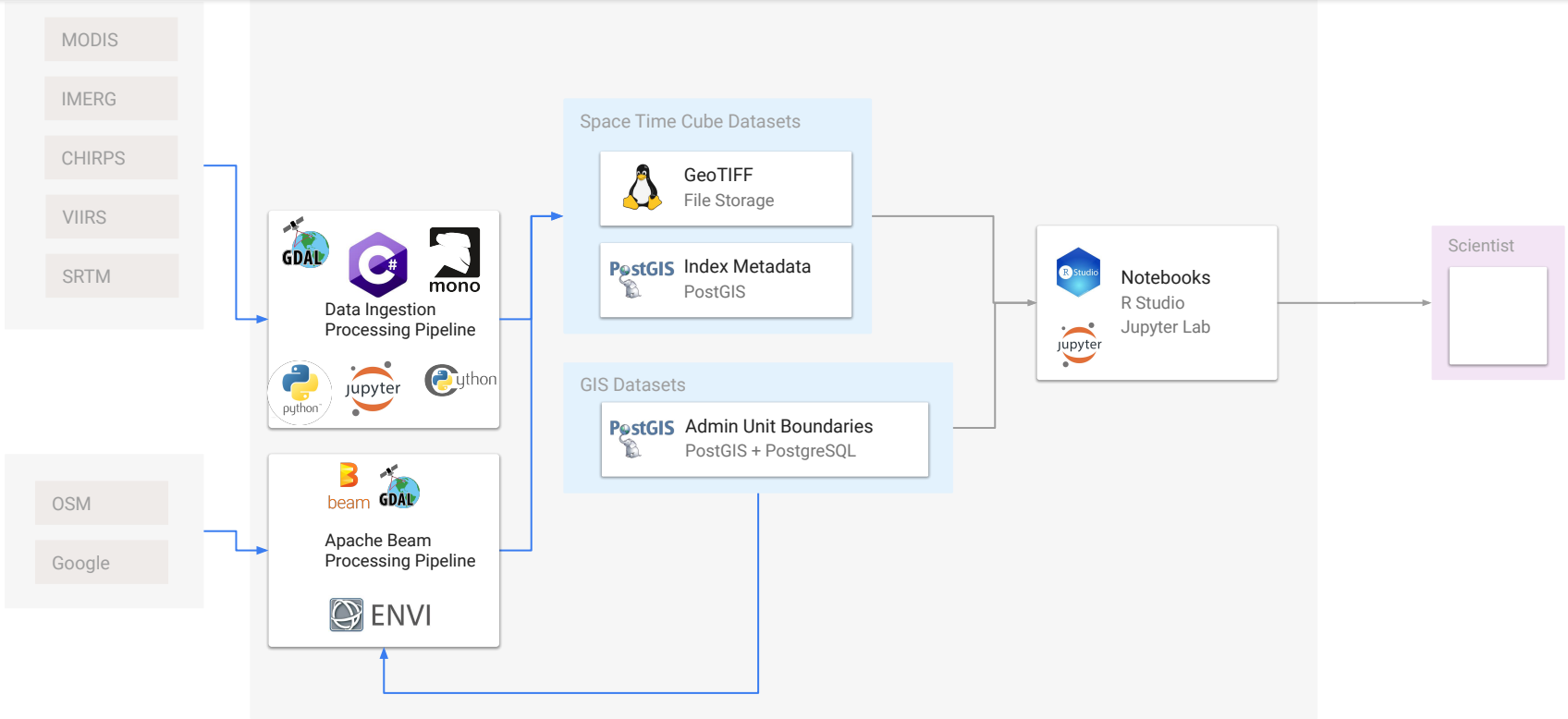




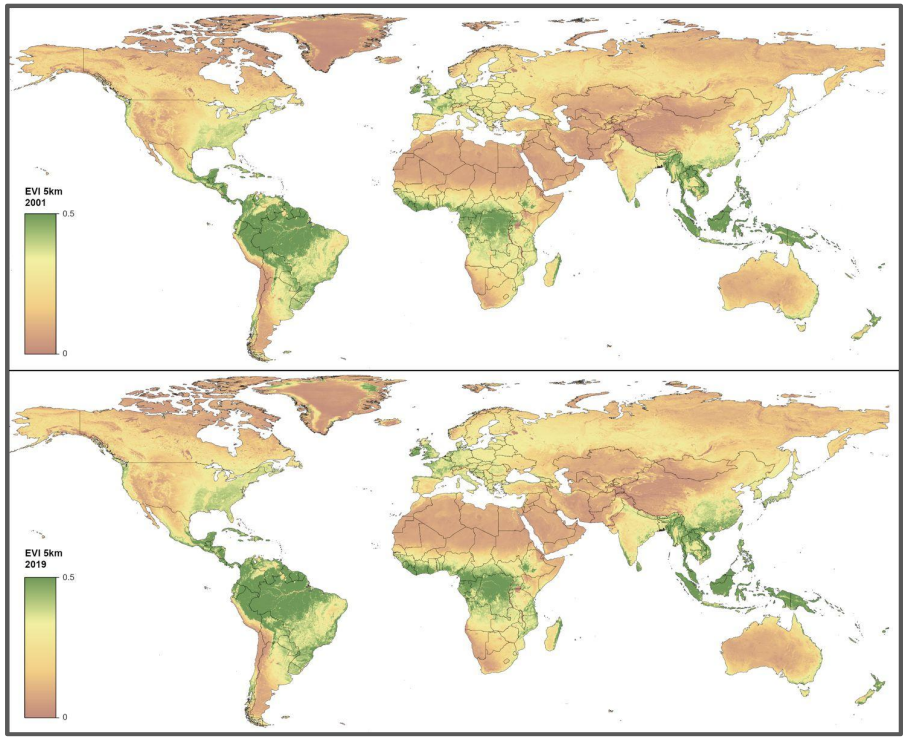


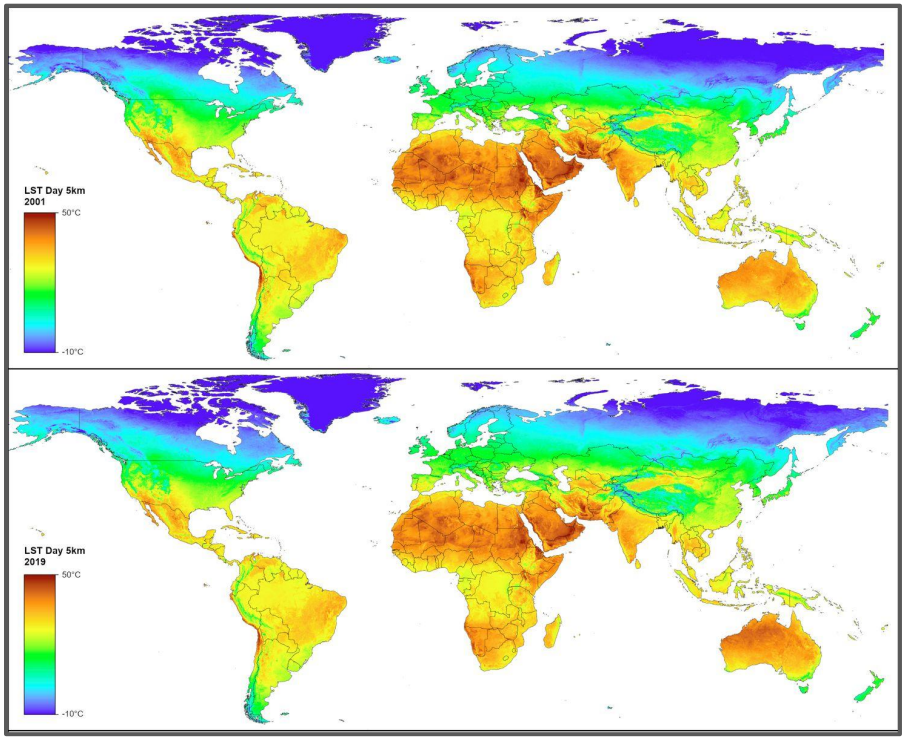
Covariate Data Processing

Architecture: Covariates Data



Covariate Products: EVI and LST Day





Geostatistical Modelling

Modelling Pipeline: Global Cubes

Covariates
GeoTiff Raster Files

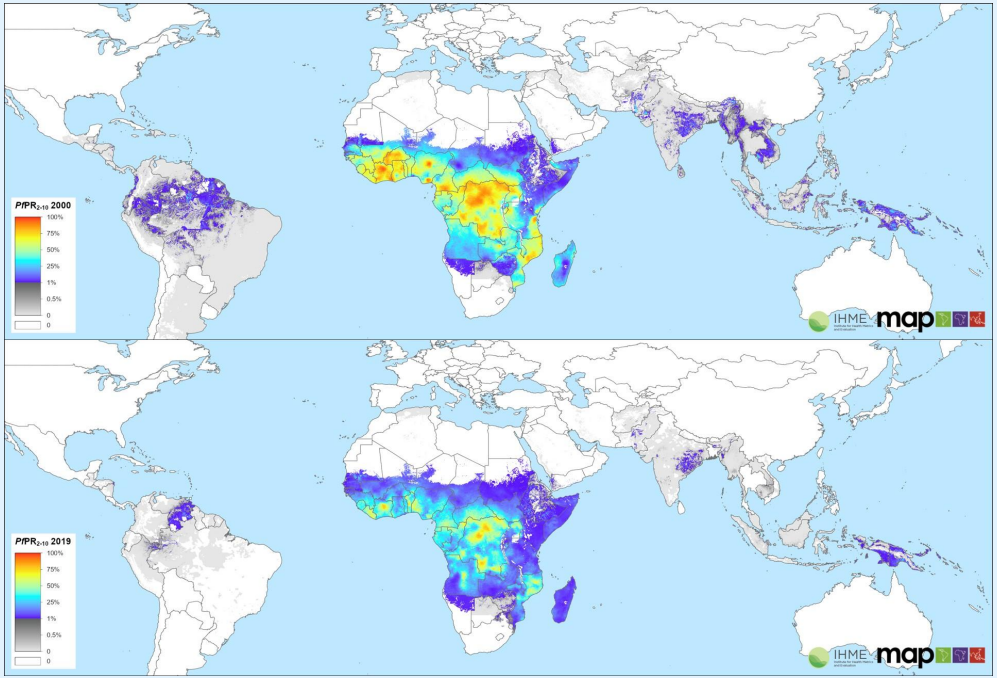
Administrative Unit Boundaries
PostGIS + PostgreSQL

Parasite Rate Points
PostGIS + PostgreSQL

Intervention Datasets
PostGIS + PostgreSQL

Annual Parasite Incidence
PostGIS + PostgreSQL

jupyter
python
GDAL
PyTorch
ENVI
RStudio
R



Geostatistical Modelling

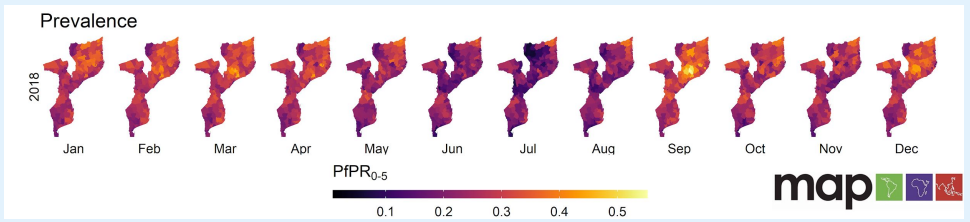
Modelling Pipeline: High-Resolution Country-Level Time-Series

- **Covariates**
  GeoTiff Raster Files
- **Administrative Unit Boundaries**
  PostGIS + PostgreSQL
- **Parasite Rate Points**
  PostGIS + PostgreSQL
- **Intervention Datasets**
  PostGIS + PostgreSQL
- **Annual Parasite Incidence**
  PostGIS + PostgreSQL

→

jupyter
python
GDAL
PyTorch
ENVI
RStudio
R

→



Geostatistical Modelling

Modelling Pipeline: Intervention Coverage Cubes

Covariates
GeoTiff Raster Files

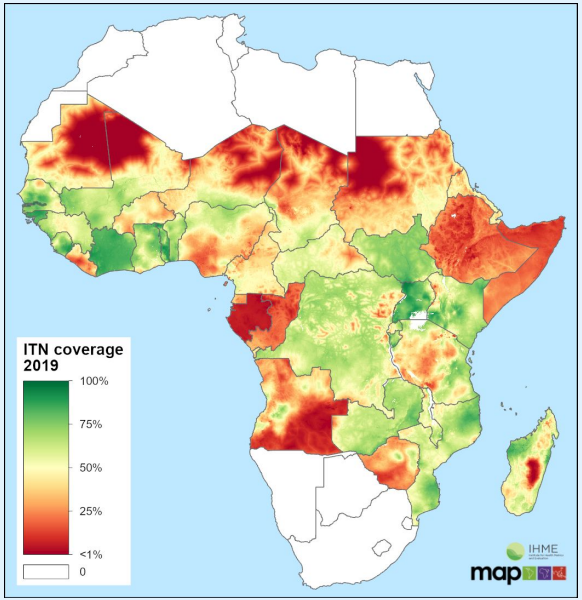
Administrative Unit Boundaries
PostGIS + PostgreSQL

Parasite Rate Points
PostGIS + PostgreSQL

Intervention Datasets
PostGIS + PostgreSQL

Annual Parasite Incidence
PostGIS + PostgreSQL

jupyter
python
GDAL
PyTorch
ENVI
RStudio
R



Dissemination: Data

https://malariaatlas.org/api-docs

Simplified Architecture: MAP Web Portal

MAP Explorer
Angular + OpenLayers

Trends Explorer
Angular

malariaAtlas
R Package

Third-Party Applications

WMS | WFS | WCS | WMTS | WPS | Vector Tiles | JSON-stat

OGC Web Services
Geoserver

Tiling Server
GeoWebCache

RESTful Web Service
Spring Framework

RESTful Web Service
NodeJS + Koa2

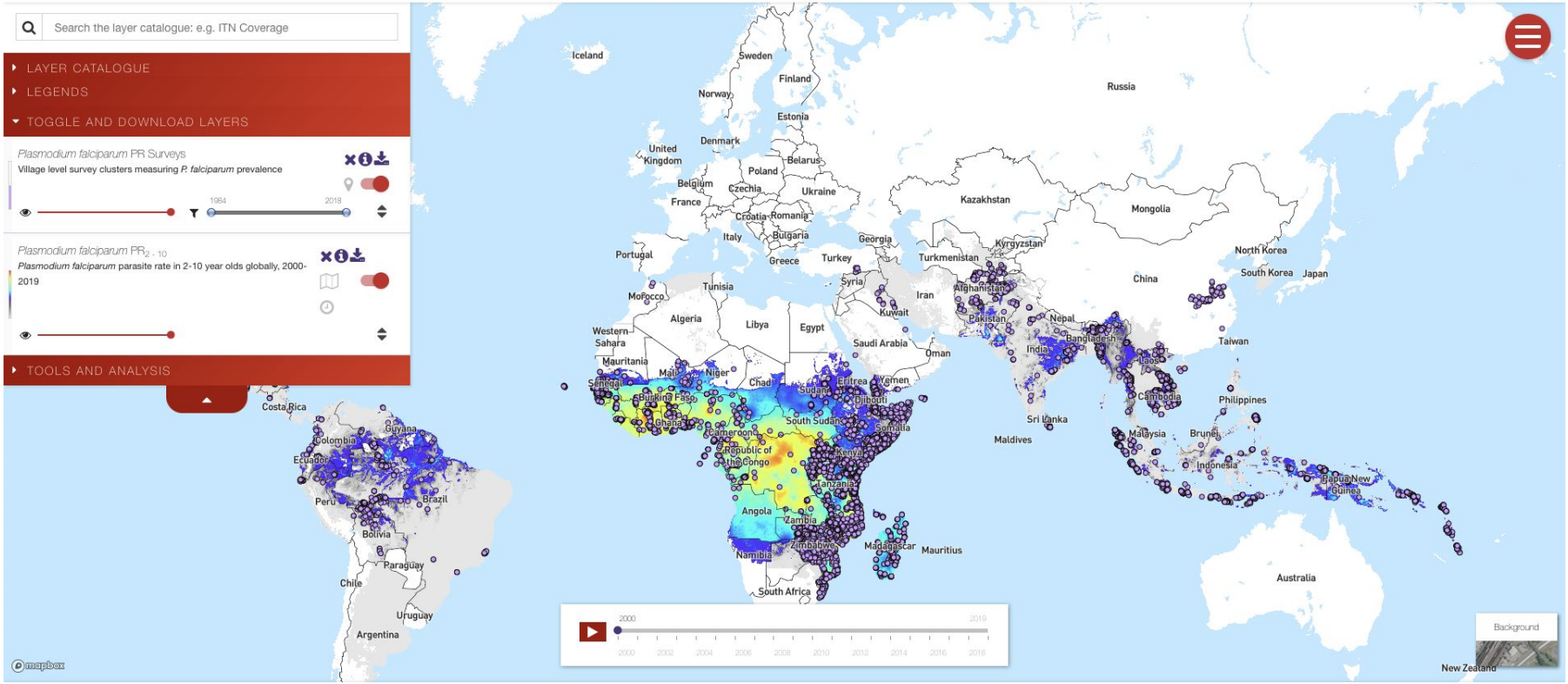
Pixel Level Estimates for Malaria related metrics
GeoTIFF files in File Storage

Admin Units Boundaries
PostGIS + PostgreSQL

Population-weighted metrics per administrative unit
PostgreSQL

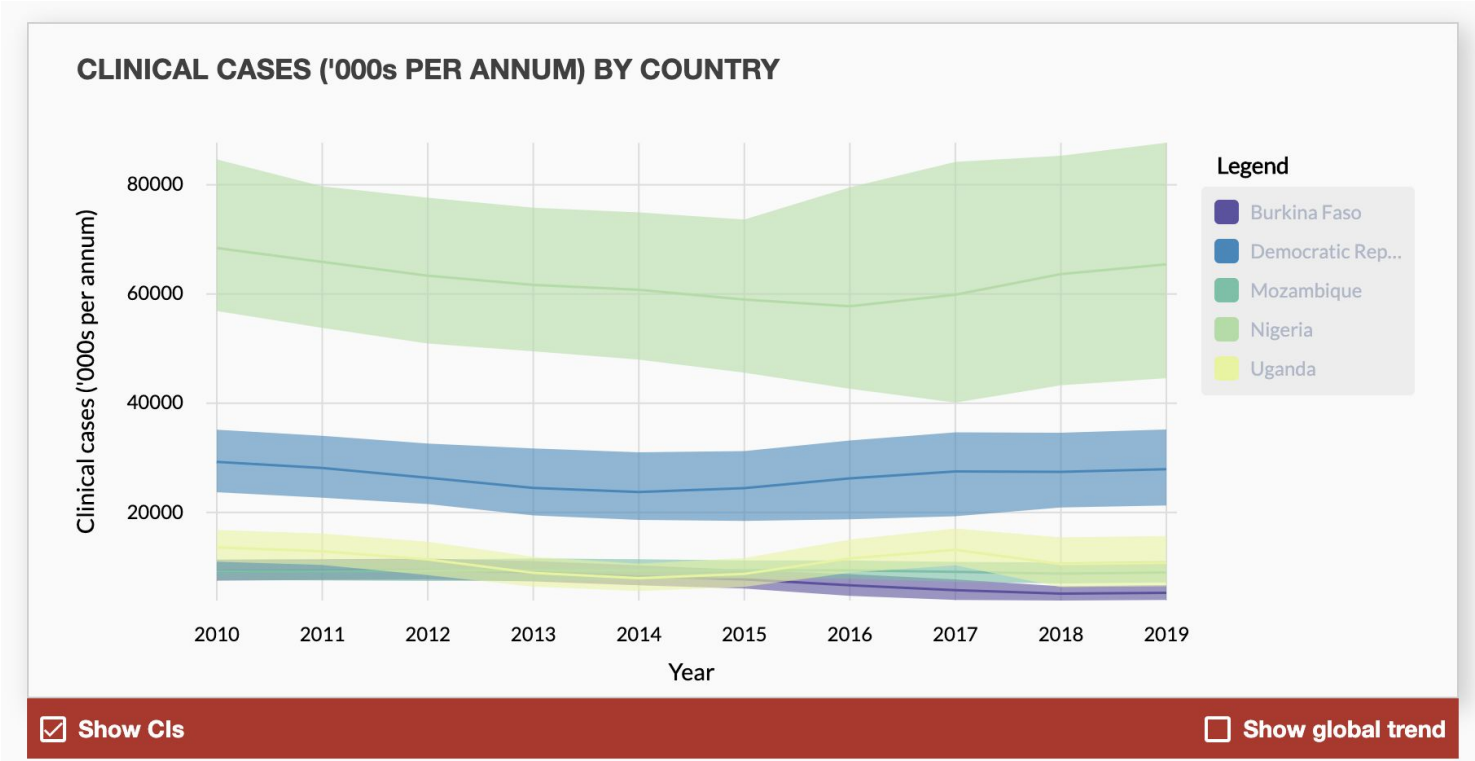
Dissemination: Web Mapping

https://malariaatlas.org/explorer



Dissemination: Data Visualisation

https://malariaatlas.org/trends



Acknowledgements

BILL & MELINDA GATES foundation