



Co-funded by the
Erasmus+ Programme
of the European Union

Open data landscape

In today's data-driven world, open data platforms play a critical role in advancing research, supporting the development of AI models, and providing valuable resources for environmental and societal studies. Notable platforms include *Kaggle*, *Hugging Face*, and *Code for Africa*, which offer vast repositories of datasets relevant to data science, AI model training, and analysis of global issues, especially in the context of environmental impact.

Kaggle is a well-established platform where users can access over 100,000 datasets spanning multiple domains, such as environmental science, healthcare, agriculture, and finance. Datasets like *Water Quality Classification*, *Satellite Images to Predict Poverty*, and *Wildfire Smoke Datasets* directly address global and local environmental concerns. Kaggle provides both free and paid datasets, with the opportunity for collaboration and data sharing, fostering innovation across industries.

Hugging Face specializes in machine learning models and datasets, primarily in the field of natural language processing (NLP) and computer vision. Its models and datasets are highly accessible, and tools like the *Transformers library* have democratized AI model integration. Researchers and developers can use datasets to train models on issues such as *sentiment analysis*, *fake news detection*, and *environmental sensor telemetry data*, contributing to the global knowledge base.

Code for Africa focuses on civic technology and open data to solve societal challenges. Their repositories offer datasets and tools related to governance, transparency, and environmental sustainability, such as the *openAFRICA* repository, which is the largest independent collection of open data on the continent. Through projects like *Climate Change Projections* and *AfricanSpending*, they contribute to critical insights into climate change, governance, and public policy.

In addition to these platforms, local data from **Zanzibar KIST University** is invaluable for projects involving meteorological and air quality monitoring, as well as sensor data integration through the *OpenAQ* platform. This collaboration allows for real-time data sharing and analysis, contributing to the global conversation on air pollution and climate change.

AP_TERRA is the platform for all environmental sensor, deployed by AP and its international partners. All the data is free to access through an API. This is expected to be finished in June 2025.

Code for Africa Datasets

Dataset/Tool Name	Description	Link
HURUmap	an interactive web platform that leverages datasets for versioning data, such as national censuses, allowing for updates with new releases over time.	https://github.com/CodeForAfrica/HURUmap
COVID-19 in Africa	a dataset providing daily updates on COVID-19 statistics across African countries, aiding in tracking and analysis of the pandemic's impact on the continent.	https://github.com/CodeForAfrica/covid19-in-africa
openAFRICA	aims to be the largest independent repository of open data on the African continent, facilitating access to a wide range of datasets for research and	https://github.com/CodeForAfrica/openAFRICA

	development purposes.	
Climate Change Projections	an embeddable map that displays climate change projections, utilizing vector files created from raster datasets to visualize potential environmental changes.	https://github.com/CodeForAfrica/ClimateChangeProjections
AfricanSpending	a project focused on transparency and accountability in government spending across African nations, providing datasets related to public expenditures.	https://github.com/CodeForAfrica/AfricanSpending

Kaggle datasets

Name	Description	Hyperlink
Water Quality Classification	this dataset contains information on various water quality parameters to determine the potability of water samples.	https://www.kaggle.com/datasets/adityakadiwal/water-potability
Satellite Images to Predict Poverty	a collection of satellite images from Africa aimed at researching the correlation between visual data and poverty levels.	https://www.kaggle.com/datasets/sandeshbhat/satellite-images-to-predict-povertyafrica
Agricultural Crop Images - Classification of Crops	this dataset offers images of 30 different agricultural products, including maize, rice, and sugarcane, for crop classification tasks.	https://www.kaggle.com/datasets/mdwaquarazam/agricultural-crops-image-classification
Wildfire Smoke Dataset	a compilation of satellite images from Canadian regions that have experienced wildfires, useful for studying wildfire patterns and effects.	https://www.kaggle.com/datasets/abdelghaniaaba/wildfire-prediction-dataset

10 Monkey Species	a dataset featuring images of ten different monkey species, designed for image classification and recognition projects.	https://www.kaggle.com/datasets/slothkong/10-monkey-species
Corn Leaf Infection Dataset	this dataset includes images of corn leaves affected by various diseases, facilitating research in plant pathology and disease detection.	https://www.kaggle.com/datasets/saroz014/corn-leaf-infection-dataset
Fake and Real News	a collection of news articles labeled as fake or real, intended for training models in fake news detection and analysis.	https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset
User Complaint Database	this dataset comprises user complaints across various sectors, useful for sentiment analysis and customer service improvement studies.	https://www.kaggle.com/datasets/prosperchuks/consumer-complaints-dataset
Sentiment Analysis on Financial Tweets	a dataset containing financial-related tweets labeled for sentiment,	https://www.kaggle.com/datasets/yash612/stockmarket-sentiment-dataset

	aiding in market sentiment analysis and financial forecasting.	
Hate Speech Identification	a collection of social media posts labeled for hate speech, supporting research in content moderation and hate speech detection.	https://www.kaggle.com/datasets/mrmorj/hate-speech-and-offensive-language-dataset
Patient Survival Prediction	this dataset includes medical records aimed at predicting patient survival rates, valuable for healthcare analytics and predictive modeling.	https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data
Brain Tumor Image Dataset	a collection of MRI images of brain tumors, facilitating research in medical imaging and tumor detection.	https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset
Heart Disease Health Indicators Dataset	this dataset contains health indicators related to heart disease, aiding in cardiovascular research and risk assessment.	https://www.kaggle.com/datasets/khoongweihao/heart-disease-health-indicators-dataset
Glaucoma Detection	a dataset comprising	https://www.kaggle.com/datasets/andrewmvd/glaucoma-detection

	retinal images for the detection of glaucoma, supporting ophthalmological research and diagnostics.	
Malaria Cell Images Dataset	this dataset includes microscopic images of cells infected with malaria, useful for developing diagnostic tools and research in parasitology.	https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria
Credit Card Fraud Detection Dataset	this dataset includes transactional data labeled for fraudulent activities, supporting research in fraud detection and prevention.	https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
Environmental Sensor Telemetry Data	a collection of telemetry data from environmental sensors, useful for monitoring and analyzing environmental conditions.	https://www.kaggle.com/datasets/sootersaalu/telemetry-data-set
Customer Personality Analysis	this dataset contains customer data aimed at analyzing personality traits for targeted marketing and customer segmentation.	https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

Predicting Student Performance	a dataset that includes student academic records, facilitating research in educational analytics and performance prediction.	https://www.kaggle.com/datasets/whenamancodes/student-performance
Household Electric Power Consumption	this dataset comprises measurements of electric power consumption in households, useful for energy consumption analysis and forecasting.	https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set

HuggingFace

Hugging Face is a leading platform for sharing and deploying state-of-the-art machine learning models and datasets. Its primary purpose is to make advanced AI tools accessible to researchers, developers, and organizations by offering a collaborative ecosystem. The platform hosts a wide array of models for natural language processing, computer vision, and other AI applications, along with high-quality datasets to support training and evaluation. Hugging Face emphasizes usability, providing tools like the Transformers library to integrate these models into various projects seamlessly. This fosters innovation and democratizes access to AI, enabling users to solve real-world problems efficiently.

AP_TERRA

AP_TERRA collects data from all the environmental sensors deployed by AP and its international partners. This includes air quality parameters in Zanzibar, Dar es Salaam and Antwerp, water quality in Lier - Belgium, Mbarara – Uganda (fish ponds), Zanzibar (seaweed farming). The system is ready end of June 2025, integration of all projects before the end of the AHUMAIN project.

The URL for now: <https://apterra.be/>

Some screenshots:



AP_TERRA

- Dashboard
- Devices
- API
- Notifications
- Settings



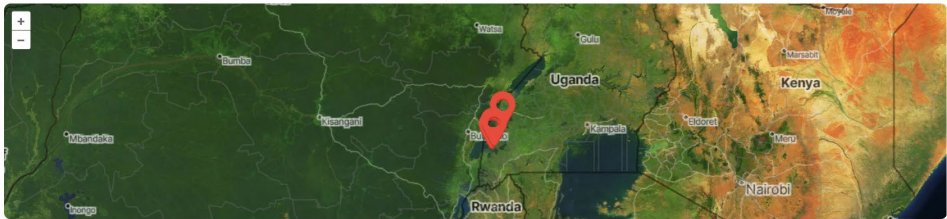
Live Device Status

Name	Status	Last message	
POL-01 Wifi or LTE device using MQTT	Offline	33 minutes, 55 seconds ago	Details
ROB-CAM-01 Camera	Running	3 minutes, 52 seconds ago	Details
Router-ROB-01 Wifi or LTE device using MQTT	Offline	12 hours, 13 minutes, 27 seconds ago	Details
WIM-01 Wifi or LTE device using MQTT	Running	4 minutes, 57 seconds ago	Details
WIM-02 Wifi or LTE device using MQTT	Running	3 minutes, 7 seconds ago	Details
WIM-03 Wifi or LTE device using MQTT	Running	3 minutes, 9 seconds ago	Details
WIM-04 Wifi or LTE device using MQTT	Running	2 minutes, 56 seconds ago	Details



AP_TERRA

- Dashboard
- Devices
- API
- Notifications
- Settings



WIFI OR LTE DEVICE USING MQTT
POL-01
Obscure level gauge at Kilembe bridge.
LAT 0.197939 LONG 30.012687
 Xander Van...
Created: 05/06/2025 [Details](#)



CAMERA
ROB-CAM-01
ROB camera at Kilembe bridge
LAT 0.197681 LONG 30.012916
 Xander Van...
Created: 30/05/2025 [Details](#)



WIFI OR LTE DEVICE USING MQTT
Router-ROB-01
Microtik router of ROB at Kilembe bridge.
LAT 0.197681 LONG 30.012916
 Xander Van...
Created: 05/06/2025 [Details](#)



WIFI OR LTE DEVICE USING MQTT
WIM-01
WIM module at serengeti power plant
LAT 0.232202 LONG 29.981443
 Xander Van...
Created: 28/05/2025 [Details](#)



WIFI OR LTE DEVICE USING MQTT
WIM-02
WIM module at nursery school (upper)
LAT 0.147767 LONG 29.930988
 Xander Van...
Created: 29/05/2025 [Details](#)



WIFI OR LTE DEVICE USING MQTT
WIM-03
WIM module at kyarumba primary school
LAT 0.125984 LONG 29.948568
 Xander Van...
Created: 29/05/2025 [Details](#)



WIFI OR LTE DEVICE USING MQTT
WIM-04
WIM module at Kilembe bridge
LAT 0.197681 LONG 30.012916



WIFI OR LTE DEVICE USING MQTT
WIM-05
WIM module at MMU campus Fort Portal
LAT 0.6879956 LONG 30.2428796



Co-funded by the
Erasmus+ Programme
of the European Union

30/6/2025 12:00:03 (5 MB)



30/6/2025 11:00:02 (5 MB)



30/6/2025 10:00:02 (6 MB)



30/6/2025 09:00:02 (5 MB)



30/6/2025 08:00:02 (5 MB)



30/6/2025 07:00:02 (5 MB)



30/6/2025 06:00:02 (5 MB)



29/6/2025 17:00:02 (4 MB)



29/6/2025 16:00:03 (5 MB)



29/6/2025 15:00:02 (5 MB)



29/6/2025 14:00:03 (4 MB)



29/6/2025 13:00:02 (5 MB)





Co-funded by the
Erasmus+ Programme
of the European Union

© European Union, 2024

The information and views set out in this document are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein
Reproduction is authorized provided the source is acknowledged