

HARMONIZE

Earth Observation Data Cubes tuned for Health Response Systems

Integrating Earth Observation Data for Enhanced Health Response Systems:

The EODCtHRS component of HARMONIZE Project

Date: 2024/12/06



This work was supported by the Wellcome Trust grant number 224694/Z/21/Z.

HARMONIZE Brazil Team



Miguel Monteiro



Rachel Lowe



Christovam Barcellos



Lubia Vinhas



Karine Ferreira



Gilberto Queiroz



Isabel Escada



Sidnei Sant'Anna



Cláudia Codeço



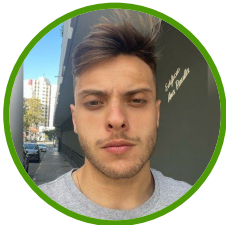
Diego Xavier



Adeline Maciel



Marcos Rodrigues



Yuri Domaradzki



Luana Becker



Ana Paula Dal'Asta



Gabriel Sansigolo



Ana Rorato



Raphael Costa



Rennan Marujo



Vanderlei Matos



Raphael Saldanha



Fabiana Zioti



Abner Anjos



Alessandra Gomes

What is Brazil Data Cube?

Brazil Data Cube is a research and development project that goal is he production visualization and analysis of large volumes of remote sensing images modeled as multidimensional data cubes for the entire Brazilian territory.

Source: <http://www.brazildatacube.org/en/home-page-2>



Brazil Data Cube



SpatioTemporal Asset Catalog

- Common language to describe geospatial information;
- It standardizes the way geospatial asset metadata is structured and queried;
- Created by several organizations collaborating to improve the interoperable search for satellite imagery;
- **BDC STAC:** Image collections and data cubes produced in the BDC project can be accessed through STAC API;

STAC Technologies

- STAC uses JSON format to wrap around any geospatial data.
- Data can be accessed using API packages in R and Python:

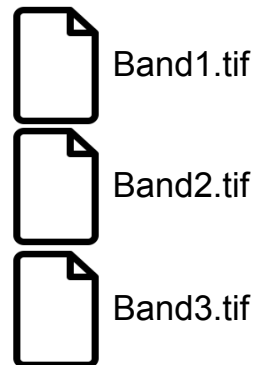
STAC Catalog



STAC Item



STAC Asset



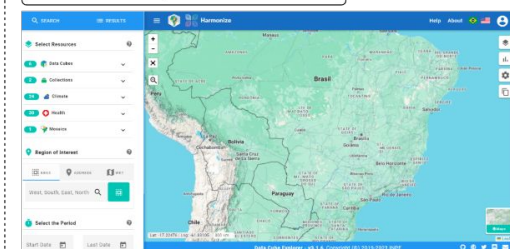
HARMONIZE Instance Overview

A technical-scientific concept for data visualization and analysis based on the Brazil Data Cube (BDC) project platform. It comprises a web portal (HARMONIZE Explorer) and a Data Science environment (BDC Lab) that provide mechanisms for manipulating Earth Observation data (Cubes), drone data, health and climate data (indicators).

HARMONIZE INSTANCE

Front-end Infrastructure

HARMONIZE Explorer - WEB Platform



HARMONIZE DSE - Data Science Environment



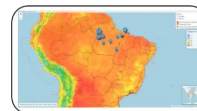
Back-end Infrastructure

Request data via STAC service and GeoServer

HARMONIZE Collection



Image collections and data cube of Earth Observation, Climatic, Weather, maps and time series with Socio-economic and Health Indicators.

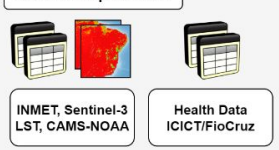


Data cataloged following the STAC and GeoServer specifications

Data harmonization

Query the indicators/variables in external data sources

External Repositories



External Applications

Instance Modules

The EODCtHRS component of HARMONIZE is composed of six modules:

Module 1 - Drone Data

Module 2 - Health Data

Module 3 - Climate Data

Module 4 - BDC-Lab

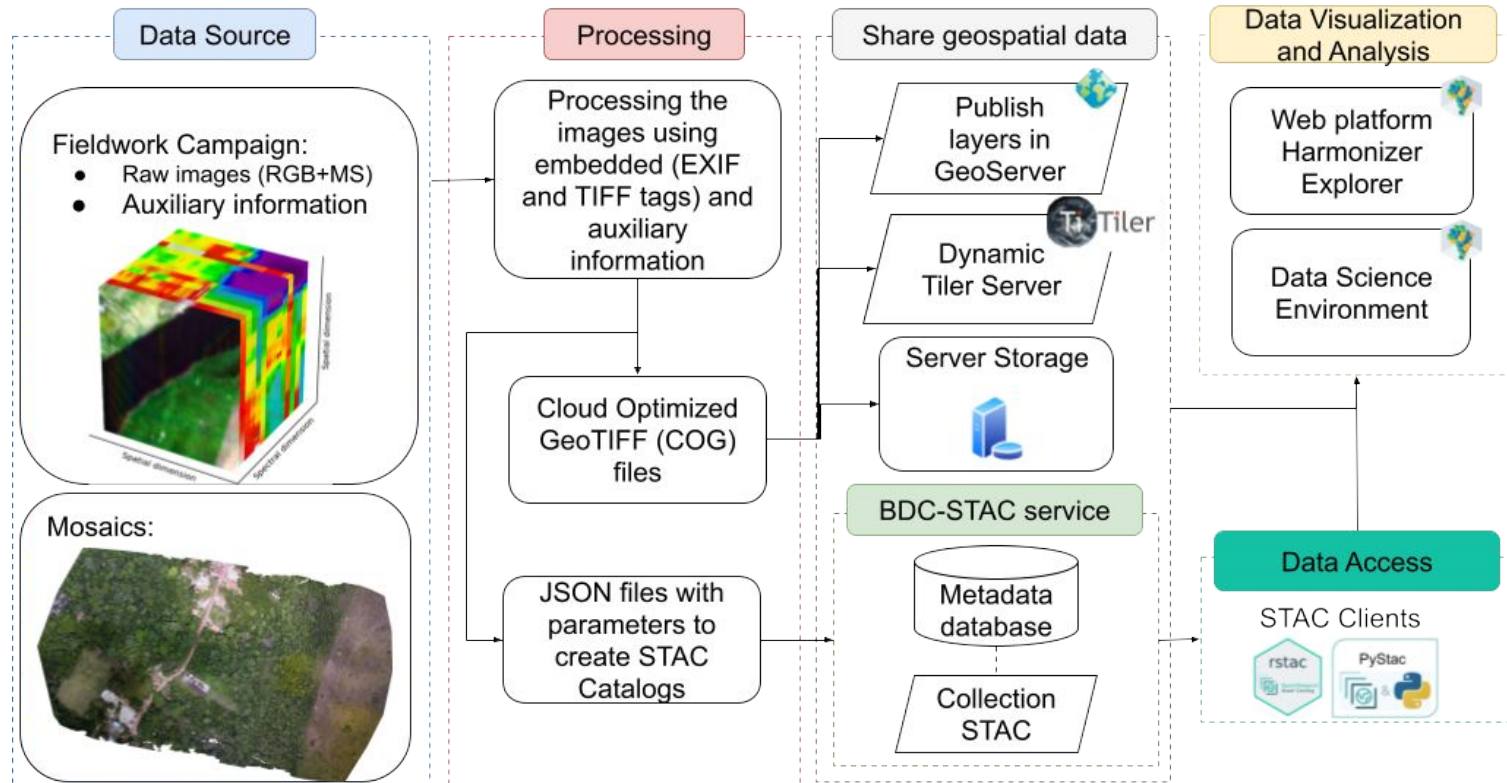
Module 5 - HARMONIZE Explorer

Module 6 - EO Data Cubes

Module 1 - Drone

Module 1 - Infrastructure

Prototype summarizing the main steps for processing raw images and mosaics from drones to produce COG files and metadata, which are available for visualization and analysis based on a BDC-STAC service.



Module 1 - Infrastructure

Drone Images Flow - Input and Output Formats

Input Format

Data format:

- The images must be organized following the protocol established for the data organization after the fieldwork campaign, accompanied by auxiliary files (JSON) with information about the sensor used and flight height



Interface Python

Output Protocol: Raster (COG)



Input parameters:

- Directory with files organized according to the protocol
- Templates files (JSON) defined for each drone and flight height using STAC specification format for collections of images

EODCtHRS Drone Image Processing Algorithm

Output Format

Drone Images Projected

Output Protocol:



Raster files

Third's application (OpenDroneMap)

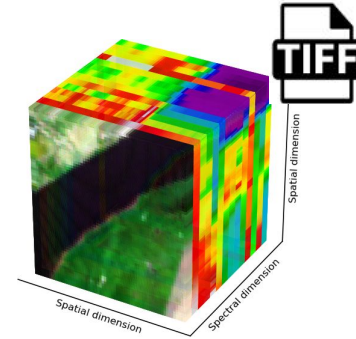


- The images are processed using an open-source application to produce mosaics

Output Protocol:



Raster files



Multispectral Images (Near Infrared, Red Edge, Red and Green) and NDVI



RGB Images



RGB and Multispectral mosaics

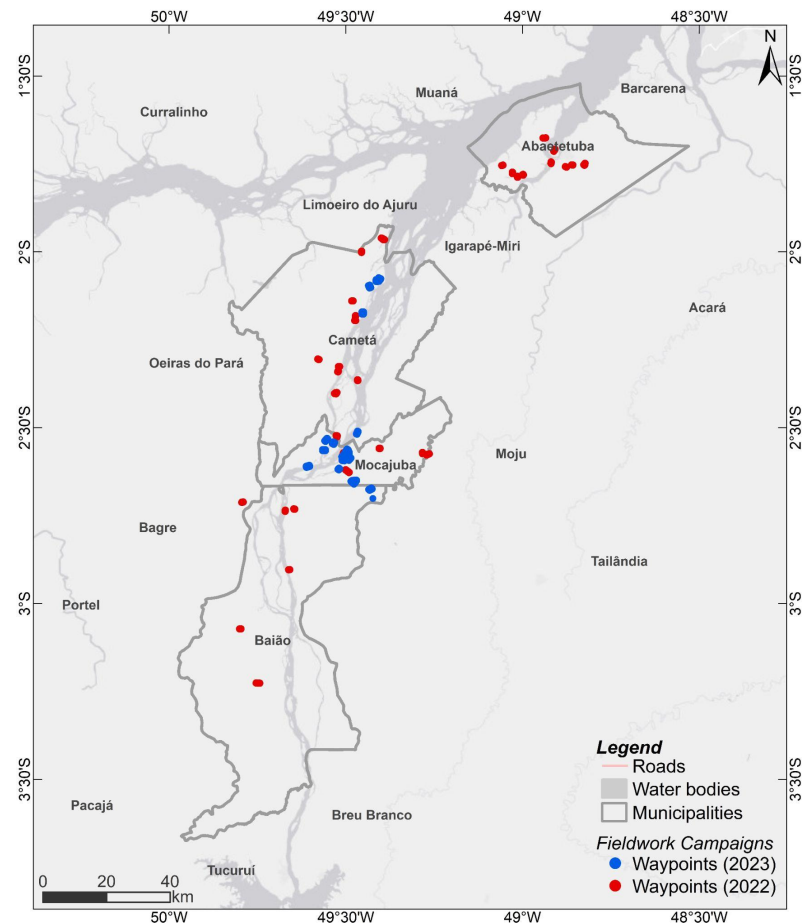
Module 1 - ALPHA R

2022

- 46 flights (RGB mosaics):
 - Flight Height 120 m
- 3044 waypoints (RGB images):
 - Front Overlap Ratio (FOR) = 80%
 - Side Overlap Ratio (SOR) = 70%
- ~15.6 GB of data (raw data)
- Temporal extension:
 - Start - 2022/11/14
 - End - 2022/11/23 (10 days)

2023

- 23 flights (RGB mosaics):
 - Flight Height 120 m
- 10681 waypoints (RGB, NIR, RE, R, G, and NDVI images):
 - Front Overlap Ratio (FOR) = 80%
 - Side Overlap Ratio (SOR) = 70%
- ~600 GB of data (raw data)
- Temporal extension:
 - Start - 2023/11/07
 - End - 2023/11/14



Spatial distribution of ALPHA R version of drone data

Module 1 - ALPHA R at Harmonize Explorer

Q

SEARCH

≡

RESULTS

Select Resources

?

06

Data Cubes

▼

03

Collections

▲

Mavic3M_120m_RGB

i

Mavic3M_120m_MS

i

Phantom3Adv_120m_RGB

i

44

Climate

▼

48

+

Health

▼

02

Mosaics

▼

Region of Interest

?

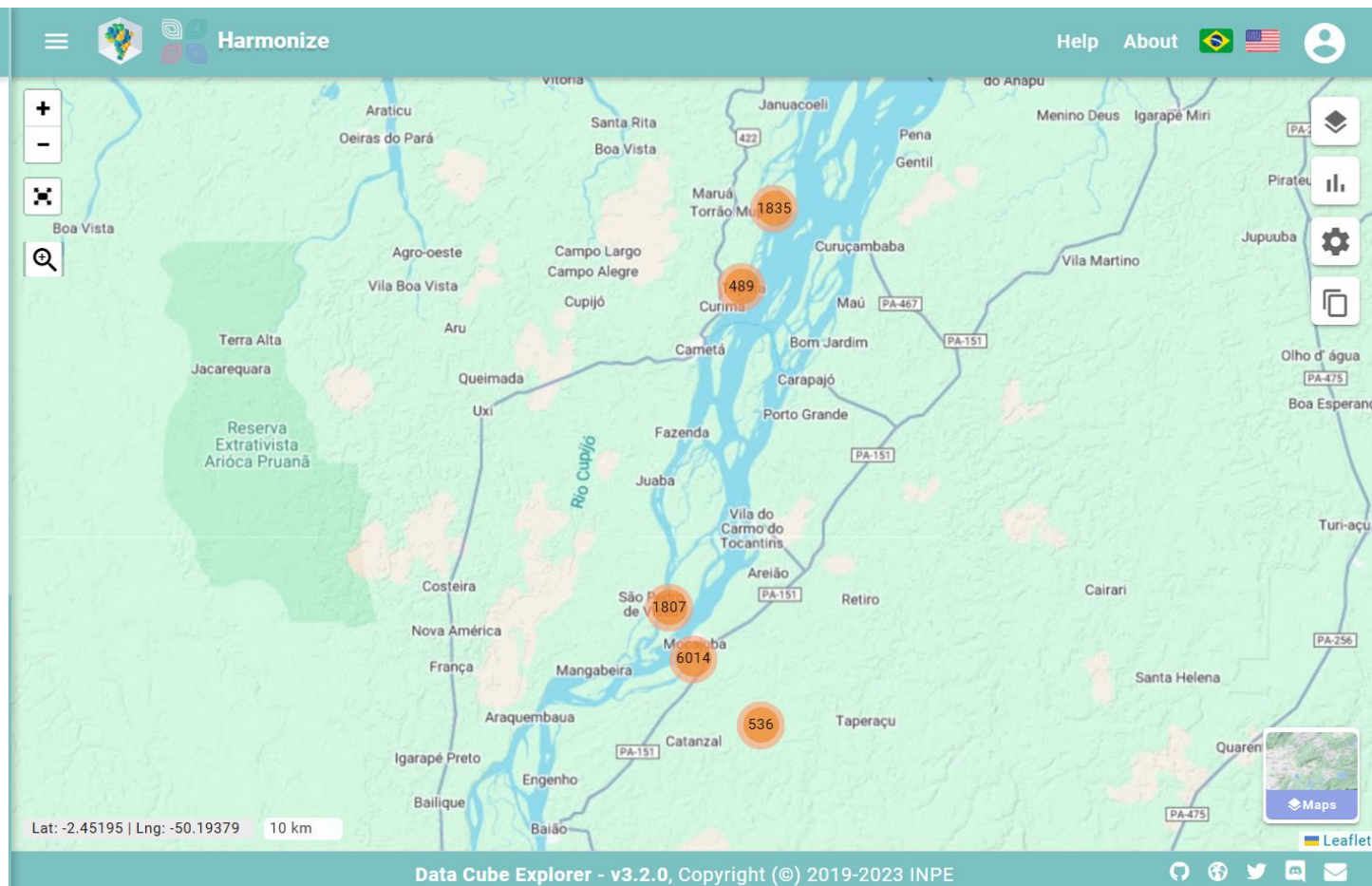
BBOX

ADDRESS

WKT

West, South, East, North

Q



Module 1 - ALPHA R at Harmonize Explorer

The screenshot displays the Harmonize Explorer web application interface. The top navigation bar features the Harmonize logo, a search icon, and links for Help and About, along with flags for Brazil and the United States. The main map area shows an aerial view of a coastal region with houses, trees, and a body of water. A timeline at the top of the map indicates the date 11/14/2023. The left sidebar contains a search bar, a results section, and a list of tiles with their respective timestamps and toggle switches. The bottom status bar displays the coordinates (Lat: -2.17696 | Lng: -49.45073) and the map scale (10 m). The bottom right corner shows the Leaflet logo and the map data source (Mavic3M_FlightHeight120m).

Search **RESULTS**

Period: 11/07/2023 - 11/15/2023
Region: -49.4498,-2.1771,-49.4497,-2.1769

13 Mavic3M_120m_RG ...

Type
Tile ID

2023-11-14 18:36:45
2023-11-14 18:36:43
2023-11-14 18:36:41
2023-11-14 18:36:39
2023-11-14 18:34:12

11/14/2023 11/14/2023

Lat: -2.17696 | Lng: -49.45073 10 m

Data Cube Explorer - v3.2.0, Copyright (©) 2019-2023 INPE

Leaflet | Mavic3M_FlightHeight120m

The portal with RGB image from Mavic 3M collected at Cametá Tapera locality (Cametá-PA).

Module 1 - ALPHA R at Harmonize Explorer (Mosaic)

The screenshot displays the Harmonize Explorer (Mosaic) interface. The top navigation bar includes a search icon, the word "SEARCH", a menu icon, "RESULTS", the "Harmonize" logo, and links for "Help", "About", and flags for Brazil and the USA. The main content area shows a 3D aerial view of a city grid. A timeline at the top of the view area indicates the date "11/07/2023". On the left, a sidebar contains a "Period" filter set to "11/07/2023 - 11/07/2023" and a "Region" filter set to "All features". Below this, a list of features is shown, including "Mavic3M_120m_RG ..." with a toggle switch. A "VIEWING" section shows the time range "2023-11-07 00:00:00 - 2023-11-07 00:00:00". The bottom of the interface shows the coordinates "Lat: -2.57271 | Lng: -49.50033" and a scale bar for "100 m". The bottom right corner features a "Maps" button and a footer with the text "Data Cube Explorer - v3.1.6, Copyright (©) 2019-2023 INPE" and social media icons.

SEARCH RESULTS

Period: 11/07/2023 - 11/07/2023
Region: All features

1 Mavic3M_120m_RG ...

2023-11-07 00:00:00 - 2023-11-07 00:00:00

VIEWING
2023-11-07 00:00:00 - 2023-11-07 00:00:00

Harmonize

11/07/2023 11/07/2023

Radio Studio Mix

a cosméticos variedades
ja de cosmético

mbleia De Deus
Vida & Luz

Tv. Betel

Lat: -2.57271 | Lng: -49.50033 100 m

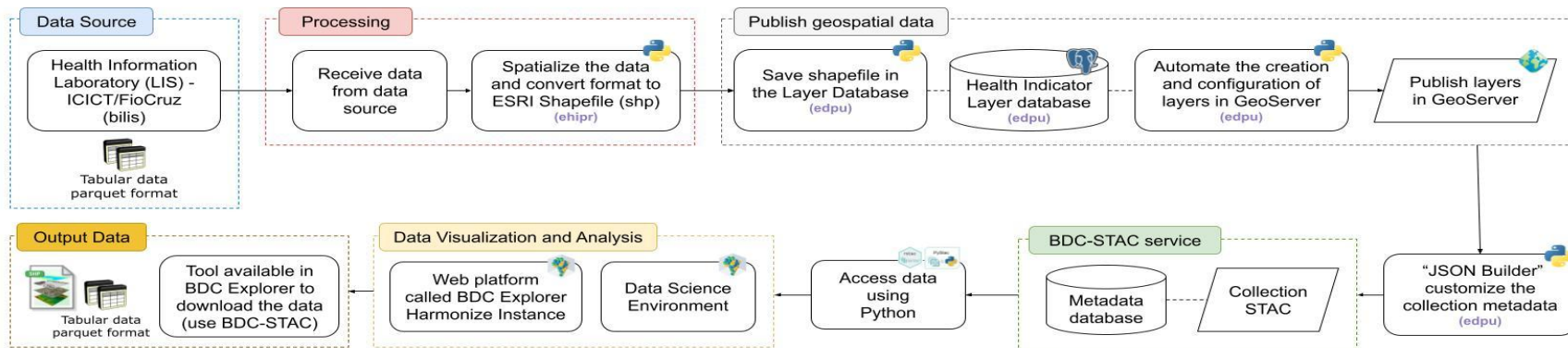
Leaflet | Mavic3M_FlightHeight120m_Mosaic

Data Cube Explorer - v3.1.6, Copyright (©) 2019-2023 INPE

Module 2 - Health Data

Module 2 - Infrastructure

- Architecture of health data integration and dissemination.
- The tasks within the Data Source and Processing blocks were tested using the ehpr package.
- The tasks in the Share of geospatial data and the BDC-STAC service blocks were facilitated through the edpu application.
- In summary, the tasks within the Data Source, Processing, geospatial sharing data, and Catalog service blocks are completed, while the Data Visualization and Analysis block is currently in progress.



Developed Python Packages

ehpr - EODctHRS Health Indicator Processing Package
edpu - EODctHRS Data Publisher Package

Python package to process health data

EODCtHRS Health Indicator Processing Package (ehipr) developed by the project's fellows using the Python programming language. This package provides a set of functions aimed at acquiring health indicators, separate it by combining spatial and temporal aggregations, add the spatial component to the data and publish it as a layer in GeoServer and its metadata in STAC using the **EODCtHRS Data Publisher Package (edpu)**.

The screenshot shows the GitHub repository for 'ehipr' by YuriDomaradzki. The repository is private and has 2 branches and 0 tags. The main branch is 2 commits ahead of the main branch. The repository contains 16 files, including ehipr.py, examples, img, settings, shp_malhas, .gitignore, INSTALL.md, LICENSE, Makefile, README.md, USAGE.md, docker-compose.yml, requirements.txt, and setup.py. The repository has 1 star and 0 forks. The 'About' section describes the package as scripts to manipulate health indicators data produced by Laboratório de Informação em Saúde from Iciit (LIS). The 'Releases' section shows no releases published. The 'Packages' section shows no packages published. The 'Languages' section shows Python at 93.2% and Makefile at 6.8%. The 'Suggested workflows' section is based on the user's tech stack.

Health Data Flow - Input and Output Formats

Input Format



Data format:

- Expected fields are **value** and **spatial indexer** (an integer that represents the index and must correspond to the grid index of the municipality)

Interface Python Package

Output Protocol: Shapefile data

EODCtHRS Health Indicator Processing Package (ehipr)

Input parameters:

- List of indicator ids;
- Directory with the parquet files;
- GitHub user token to download data from Brazil;

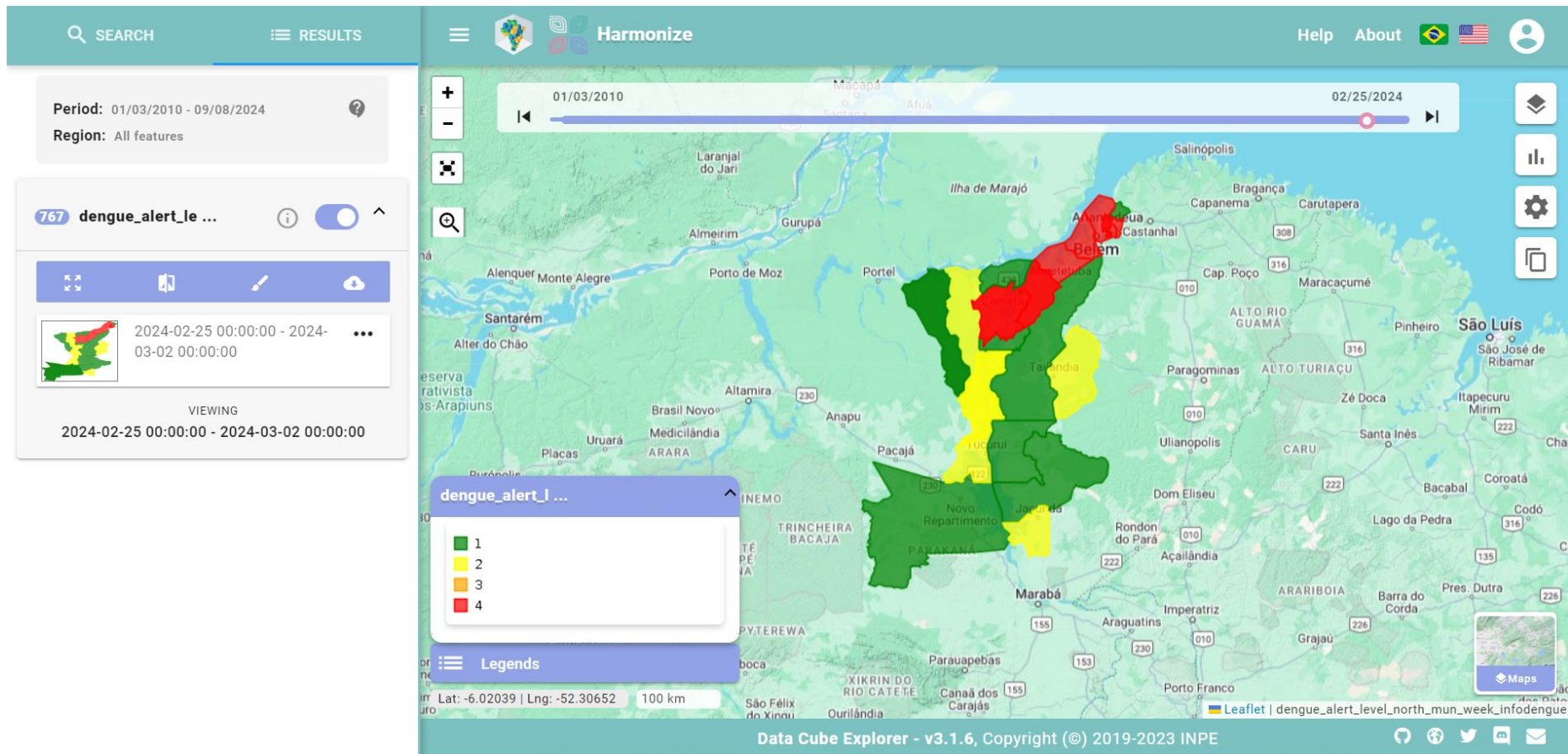
Output Format

Health as Vector Layers (GeoServer)

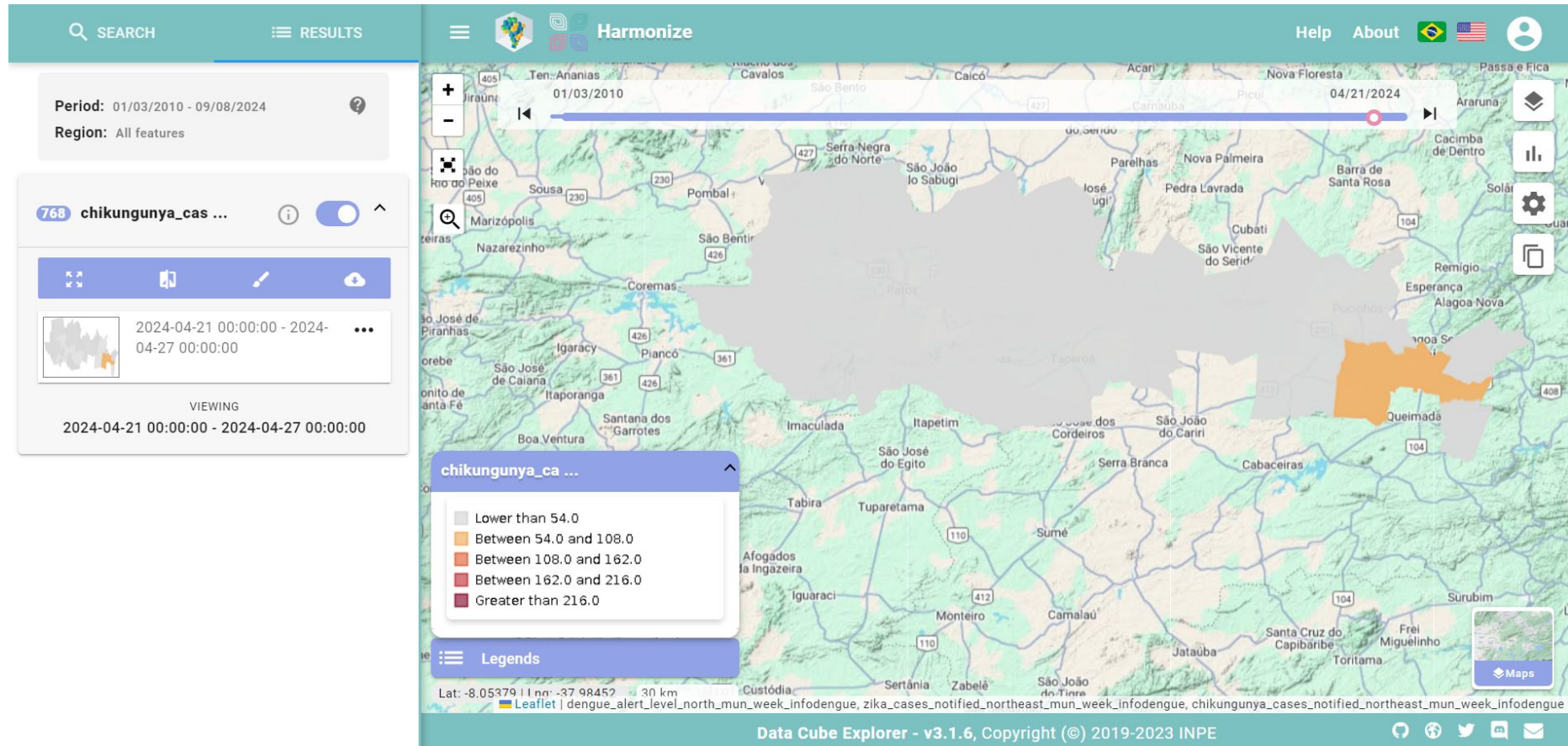
Output Protocol:



Module 2 - Explorer with Health Data - Dengue Alert Level



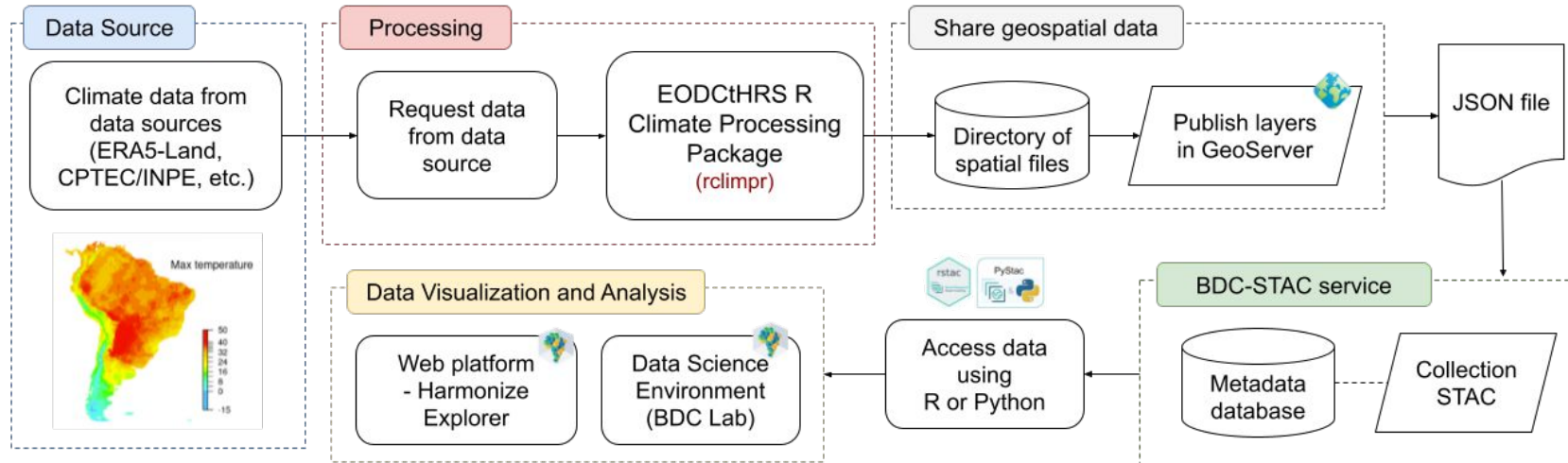
Module 2 - Explorer with Health Data - Chikungunya Notified Cases



Module 3 - Climate Data

Module 3 - Infrastructure

- Architecture of climate data integration and dissemination.
- The tasks within the Data Source and Processing blocks were tested using the rclimpr R package.
- The tasks in the Share of geospatial data and the BDC-STAC service blocks were facilitated through the edpu application.
- In summary, the tasks within the Data Source, Processing, Sharing geospatial data, and Catalog service blocks are completed, while the Data Visualization and Analysis block is currently in progress.



Module 3 - Infrastructure

EODCtHRS **R Climate Processing Package (rclimpr)**, developed using the R programming language, processes climate indicators from different input data formats, including netCDF, GRIB2, and CSV. It offers a set of functions specifically designed to generate climate indicators.

The screenshot shows the GitHub repository for **rclimpr** (Private). The repository is owned by **Harmonize-Brazil**. The file structure includes:

- R**: Update 2024-10-15. Add relative humidity indicator, color, and adjust ... 3 weeks ago
- data**: Update 2024-02-18. Add new folder 9 months ago
- inst**: Update 2024-10-15. Add relative humidity indicator, color, a... 3 weeks ago
- man**: Update 2024-10-15. Add relative humidity indicator, color, a... 3 weeks ago
- tests**: Update 2024-02-14. Update R package, add scripts, funcio... 9 months ago
- _Rbuildignore**: Update 2024-02-14. Update R package, add scripts, funcio... 9 months ago
- .gitignore**: Update 2024-02-20. Improve functions and variable names 9 months ago
- DESCRIPTION**: Update 2024-08-22. Add geobr option as interesting ares to... 3 months ago
- LICENSE**: Initial commit 10 months ago
- NAMESPACE**: Update 2024-10-15. Add relative humidity indicator, color, a... 3 weeks ago
- README.Rmd**: Update 2024-08-22. Add geobr option as interesting ares to... 3 months ago
- README.md**: Update 2024-08-22. Add geobr option as interesting ares to... 3 months ago
- rclimpr.Rproj**: Update 2024-02-14. Update R package, add scripts, funcio... 9 months ago

The repository also shows a commit history table with columns for commit message, commit hash, and time ago. The repository is licensed under MIT and has 0 stars, 2 watchers, and 0 forks.

Climate Data Flow - Input and Output Formats

Input Format



Data format:

- A netCDF or GRIB2 format file with an internal variable representing data aggregated on a daily. Accompanied by auxiliary files such as epidemiological week dates (CSV) and municipality spatial information (Shapefile).

Interface R Package

Output Protocol: Raster (**COG**) and Vector (Shapefile or GeoJSON)

EODCtHRS R
Climate Processing
Package
(**rclimpr**)

Input parameters:

- Directory with an input format files
- Variable
- Temporal aggregation (months or epidemiological week)
- Spatial aggregation (interesting area defined by a Shapefile or municipality code)

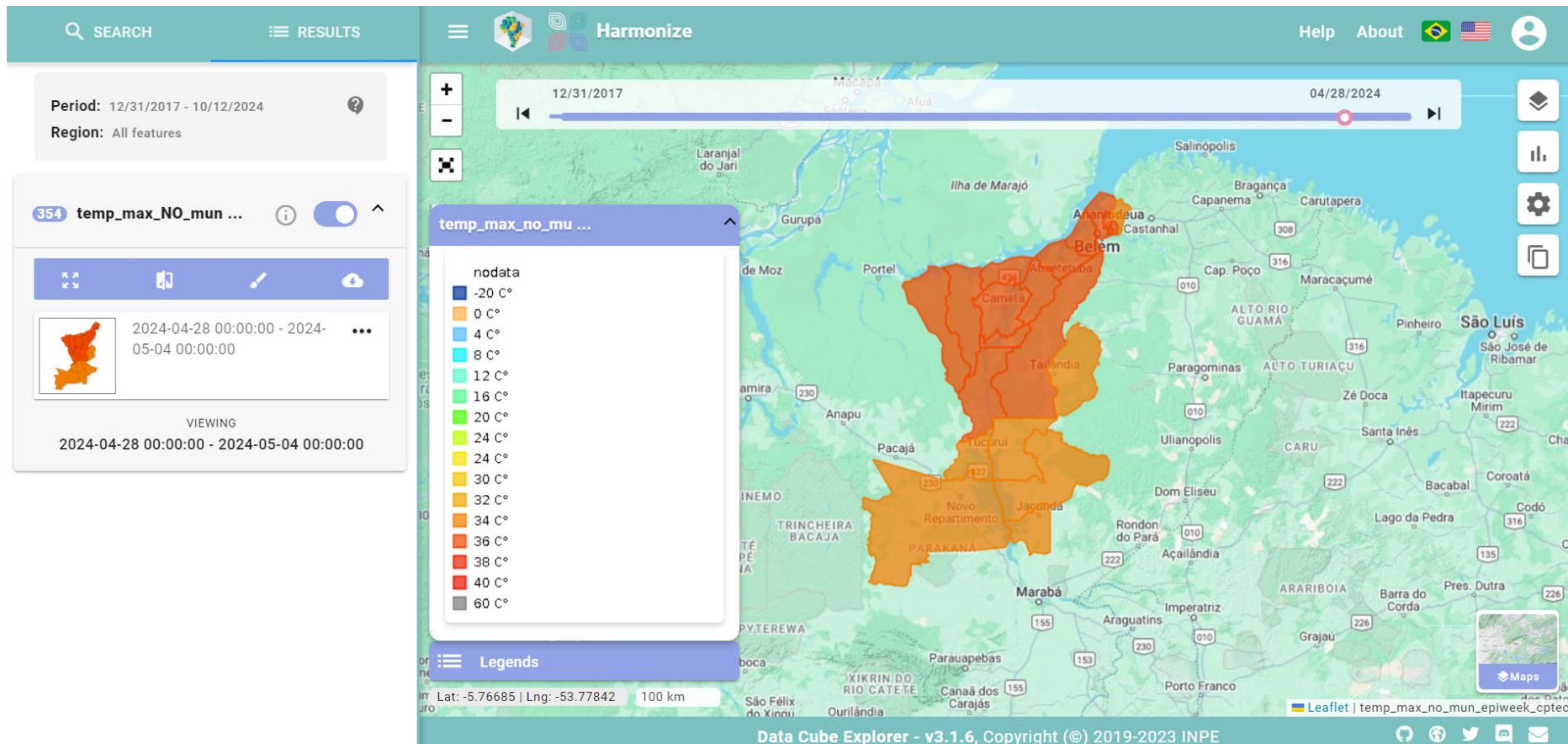
Output Format

Climate indicators as
Raster and Vector
files

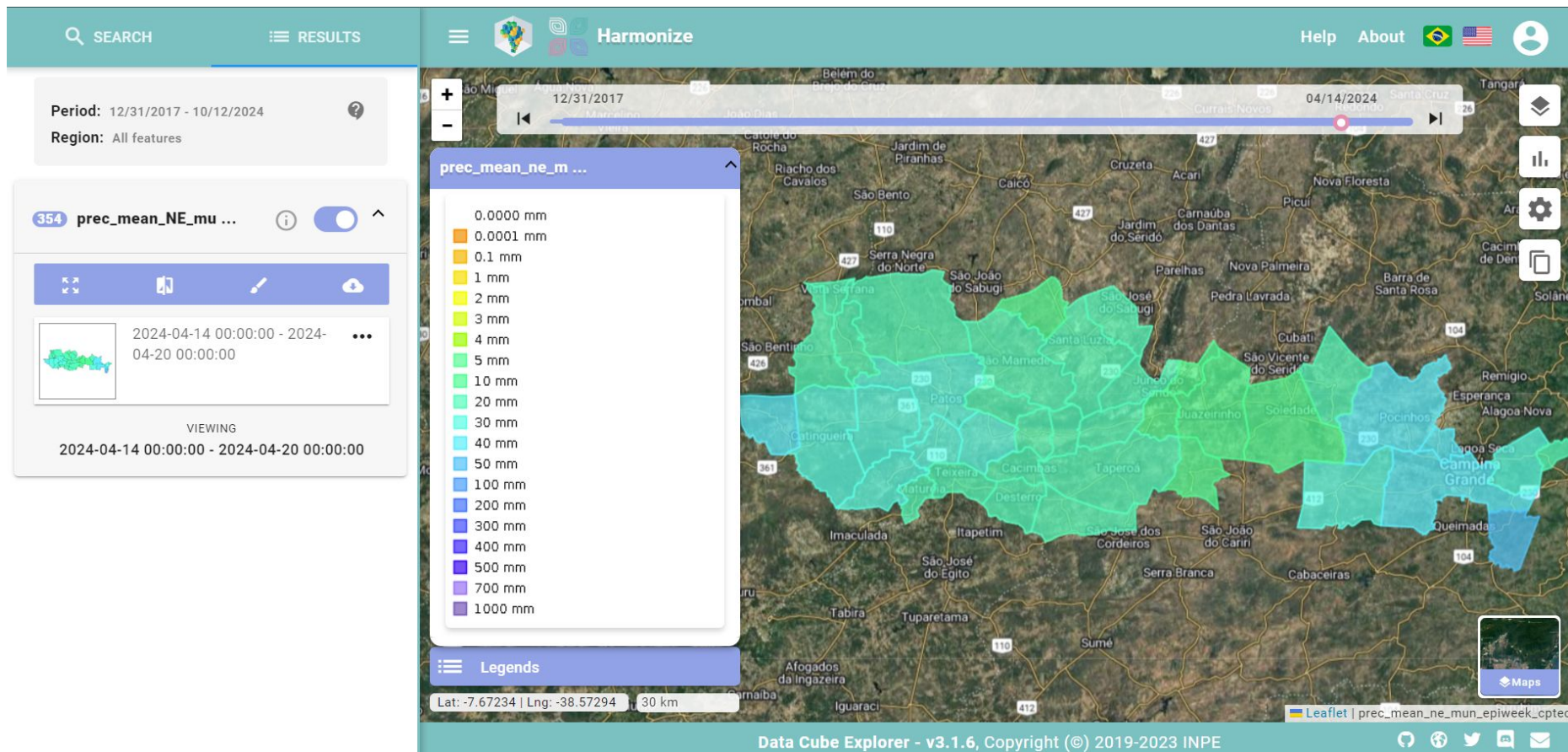
Output Protocol:



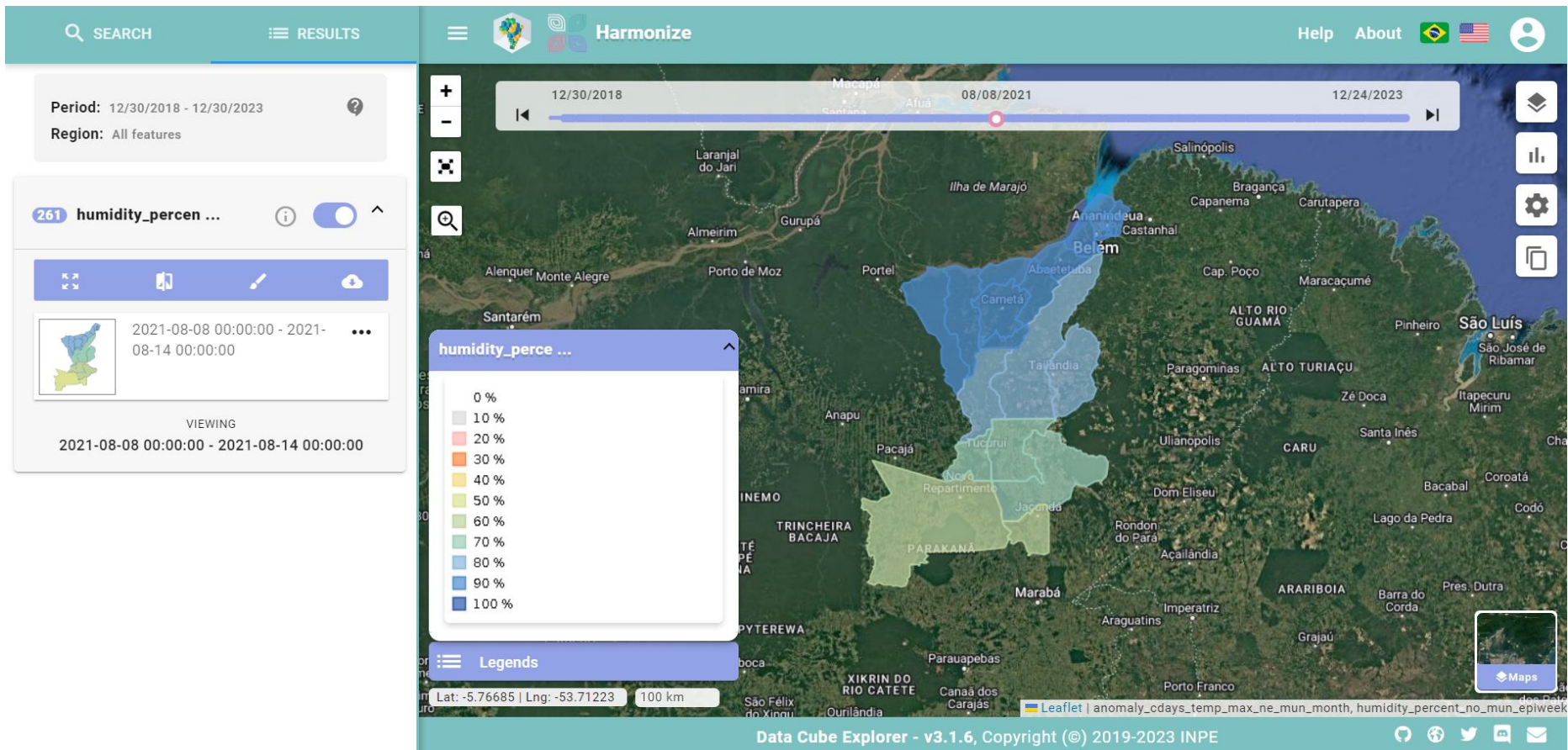
Module 3 - Explorer with Climate Data - Temperature



Module 3 - Explorer with Climate Data - Precipitation

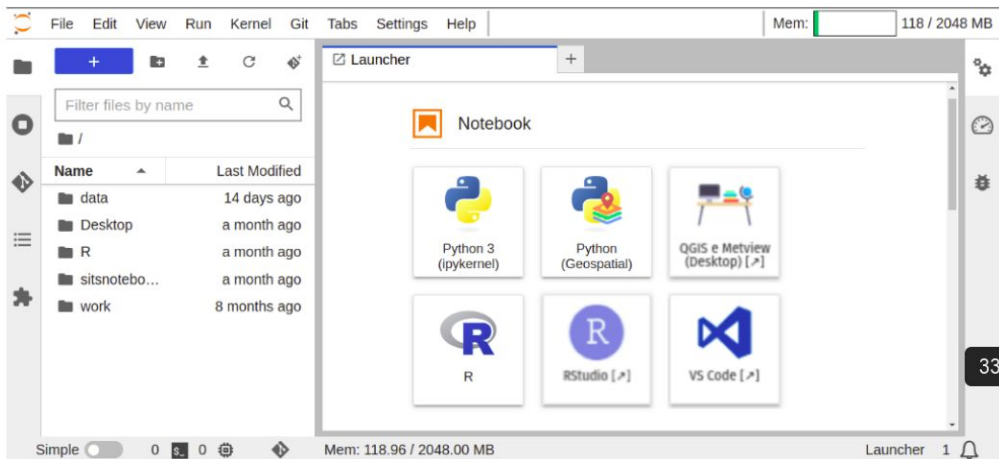


Module 3 - Explorer with Climate Data - Relative Humidity



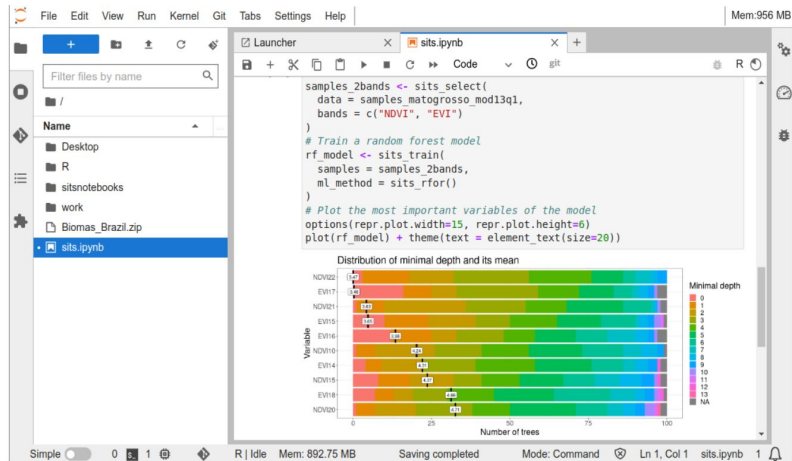
Module 4 - BDC-Lab

Data science tutorials



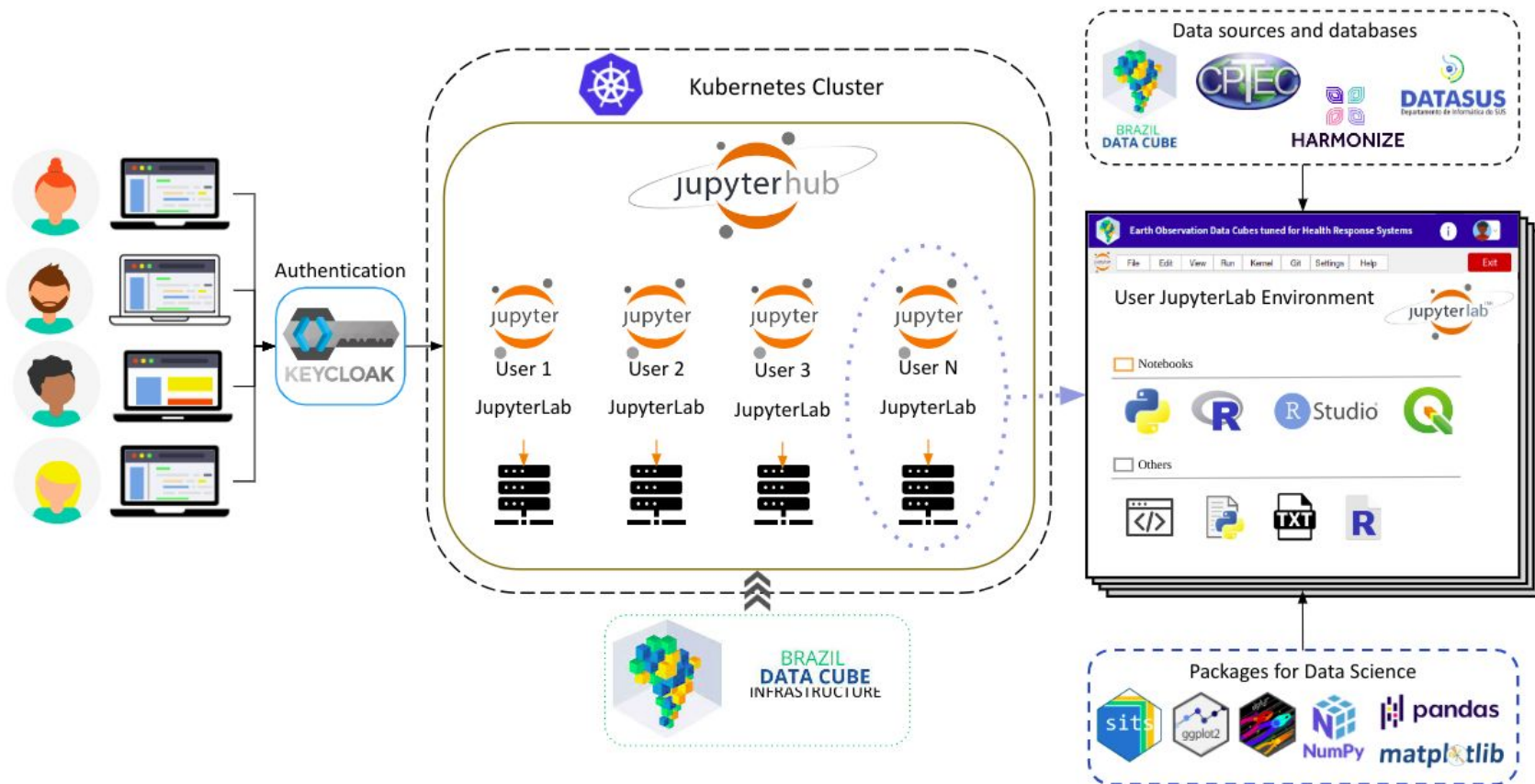
BDC-Lab is being used to develop
Harmonize data science tutorials using
Jupyter Notebooks.

The objective of these tutorials is to guide users on
how to access drone, health and climate data and
perform data science analyses.



Data science tutorials

BDC-Lab infrastructure and flow



Data science tutorials

BDC-Lab - updates

Since May 2024, the main tasks and results achieved:

- Release of versions **1.2.1** and **1.2.2**:
 - Interface improvements;
 - Security updates;
 - New processing packages and applications (VSCode and Metview);
- Ongoing Harmonize tutorials for accessing health, drone and climate data;
- Experimental platform being tested by users of INPE projects and partner institutions;

NOTE: BDC-Lab is not a HARMONIZE package/deliverable. It requires a computational infrastructure hosted by INPE.

The contributions made to the cross-cutting themes:

- **Topics related to Harmonize:** BDC-Lab has been used to access EODCtHRS data (tutorials are in progress);
- **Topics related to other projects/institutions:** It has been a tool for generating products by INPE and partner institutions (below)



Data science tutorials

BDC-Lab user session



INPE Website



BIG Website



Catálogo Integrado de Imagens



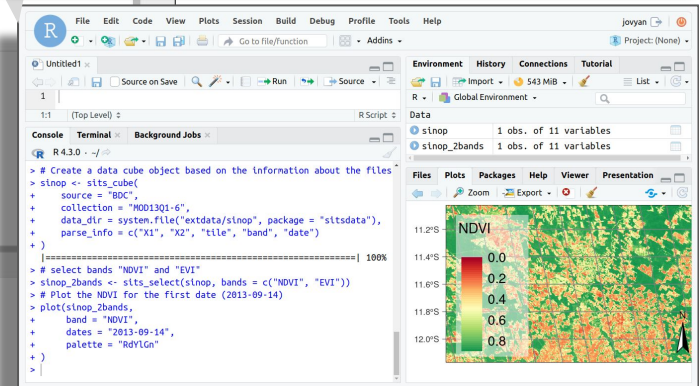
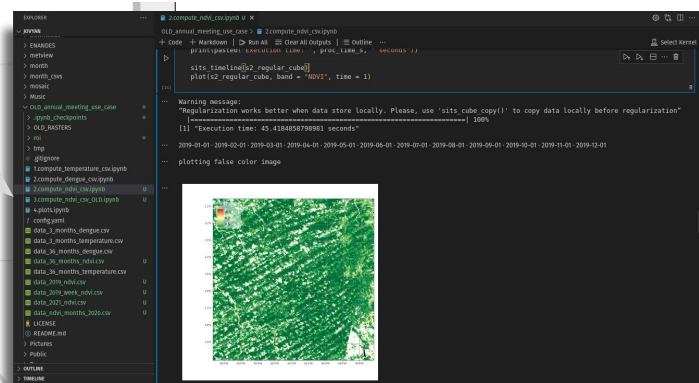
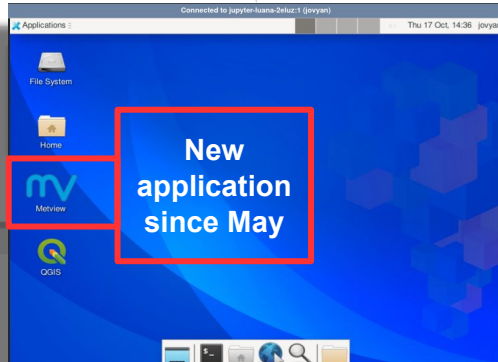
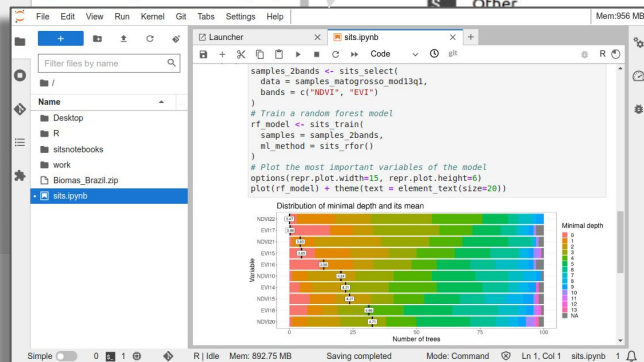
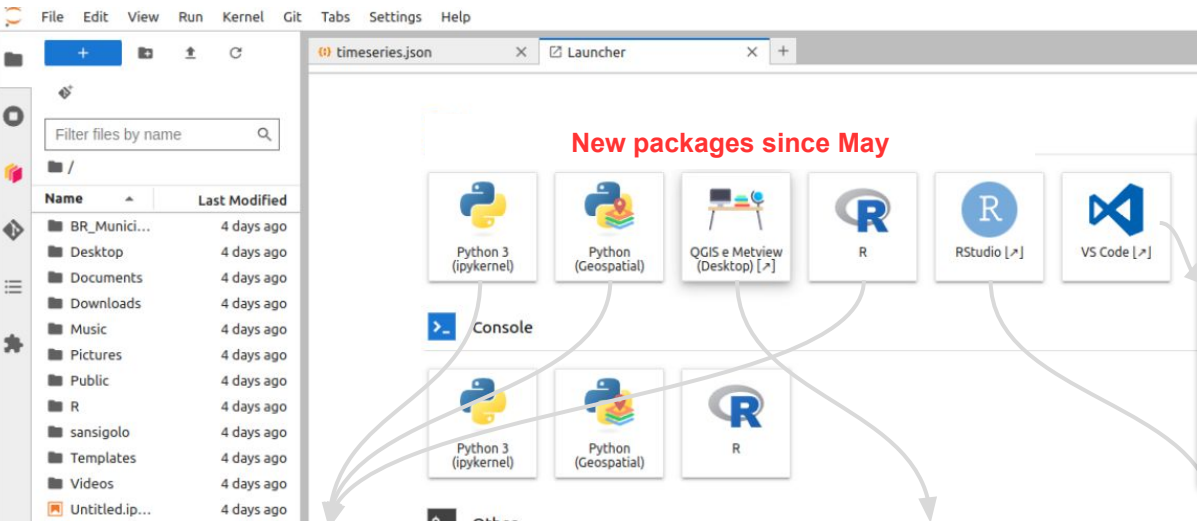
Data Cube Explorer



Catálogo de Imagens CBERS e Amazonia-1

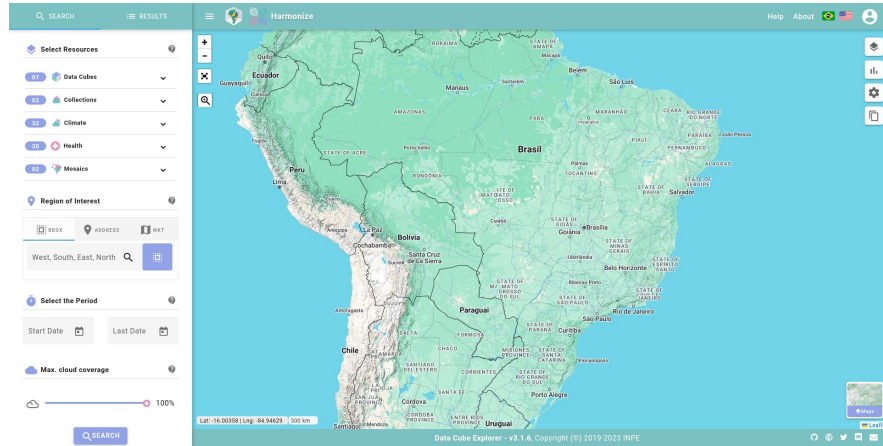


Visualização Interativa de Imagens do GOES-16



Module 5 - HARMONIZE Explorer

Harmonize Explorer



HARMONIZE Explorer is the web portal for the EODCtHRS component,

It provides features for visualizing data cubes, drone image collections, health data, climate data and drone mosaics.

Q SEARCH

RESULTS

Select Resources

07

Data Cubes

02

Collections

32

Climate

30

Health

02

Mosaics



Region of Interest

BBOX

ADDRESS

WKT

West, South, East, North



Select the Period

Start Date



Last Date



Max. cloud coverage



100%

Q SEARCH



Harmonize

Help About



SEARCH

RESULTS



Harmonize

Help About



Period: 11/07/2023 - 11/07/2023

Region: All features



1 Mavic3M_120m_RG ...



2023-11-07 00:00:00 -
2023-11-07 00:00:00



VIEWING

2023-11-07 00:00:00 - 2023-11-07 00:00:00



11/07/2023



11/07/2023



Lat: -2.57150 | Lng: -49.49774

50 m



Maps

Leaflet | Mavic3M_FlightHeight120m_Mosaic

Data Cube Explorer - v3.1.6, Copyright (©) 2019-2023 INPE



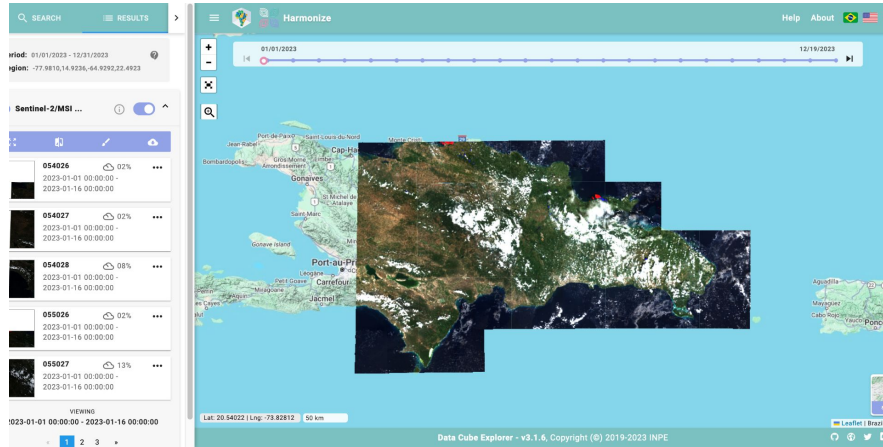
Module 6 - EO Data Cubes

EO Data Cubes - Dominican Republic

One of the activities was the publication of Sentinel-2 data cubes for the Dominican Republic. Unlike the Sentinel-2 and landsat-8 data cubes for Colombia and Peru, external support was needed to generate the grid for the Dominican Republic.

Process:

- Download of ESA Sentinel-2 images
- Creation of S2 cubes for Dominican Republic using Grids.
- Regularization
- Visualization and testing of cubes (private)
- Opening of Data
- Addition to STAC Harmonize catalog.
- Testing in Harmonize Explorer.



SEARCH RESULTS

Period: 01/01/2023 - 12/31/2023
Region: -77.9810,14.9236,-64.9292,22.4923

Sentinel-2/MSI ...



054026 02%

2023-01-01 00:00:00 -
2023-01-16 00:00:00

054027 02%

2023-01-01 00:00:00 -
2023-01-16 00:00:00

054028 08%

2023-01-01 00:00:00 -
2023-01-16 00:00:00

055026 02%

2023-01-01 00:00:00 -
2023-01-16 00:00:00

055027 13%

2023-01-01 00:00:00 -
2023-01-16 00:00:00

VIEWING

2023-01-01 00:00:00 - 2023-01-16 00:00:00

1 2 3

Harmonize

Help About



01/01/2023

12/19/2023



Lat: 20.54022 | Lng: -73.82812 50 km

Data Cube Explorer - v3.1.6, Copyright (©) 2019-2023 INPE



SEARCH RESULTS

Period: 01/01/2023 - 12/31/2023
Region: -77.9810,14.9236,-64.9292,22.4923

Sentinel-2/MSI ...

054026 02%
2023-01-01 00:00:00 -
2023-01-16 00:00:00

054027 02%
2023-01-01 00:00:00 -
2023-01-16 00:00:00

054028 08%
2023-01-01 00:00:00 -
2023-01-16 00:00:00

055026 02%
2023-01-01 00:00:00 -
2023-01-16 00:00:00

055027 13%
2023-01-01 00:00:00 -
2023-01-16 00:00:00

VIEWING
2023-01-01 00:00:00 - 2023-01-16 00:00:00

1 2 3

Harmonize

Help About



Lat: 18.78542 | Lng: -71.30659 1 km

Data Cube Explorer - v3.1.6, Copyright (©) 2019-2023 INPE



Thank you!

Email: gabriel.sansigolo@inpe.br

For more details, visit: <https://www.harmonize-tools.org/>