



BRAZIL DATA CUBE

brazildatacube.org

Part of the Environmental Monitoring of
Brazilian Biomes project (2019 – 2023)

Karine R. Ferreira

Gilberto R. de Queiroz

Earth Observation and Geoinformatics Division (DIOTG)

National Institute for Space Research (INPE)

**FUNDO
AMAZONIA**

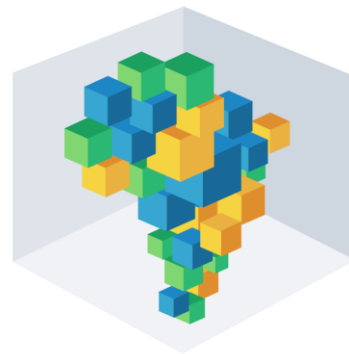


MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÃO



**Big data of remote sensing
images modeled as
multidimensional data cubes**

**Land use and cover
mapping**



**BRAZIL
DATA CUBE**

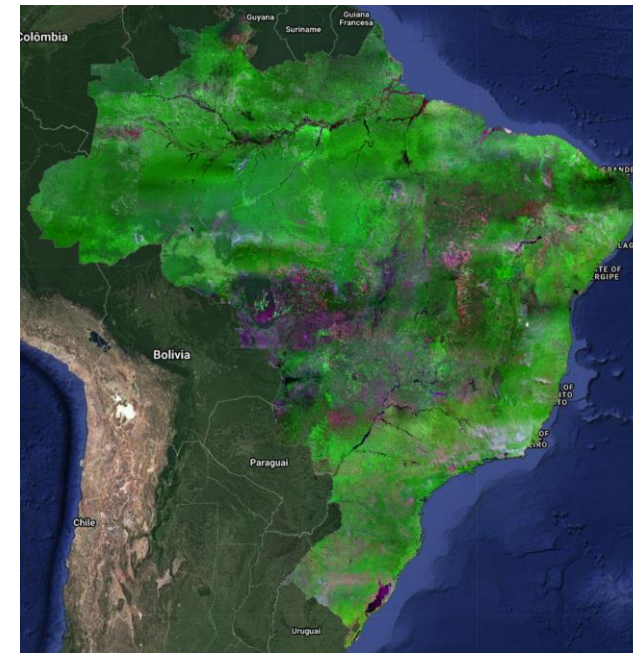
**Image time series
analysis**

**Big data technologies and
machine learning**

Motivation

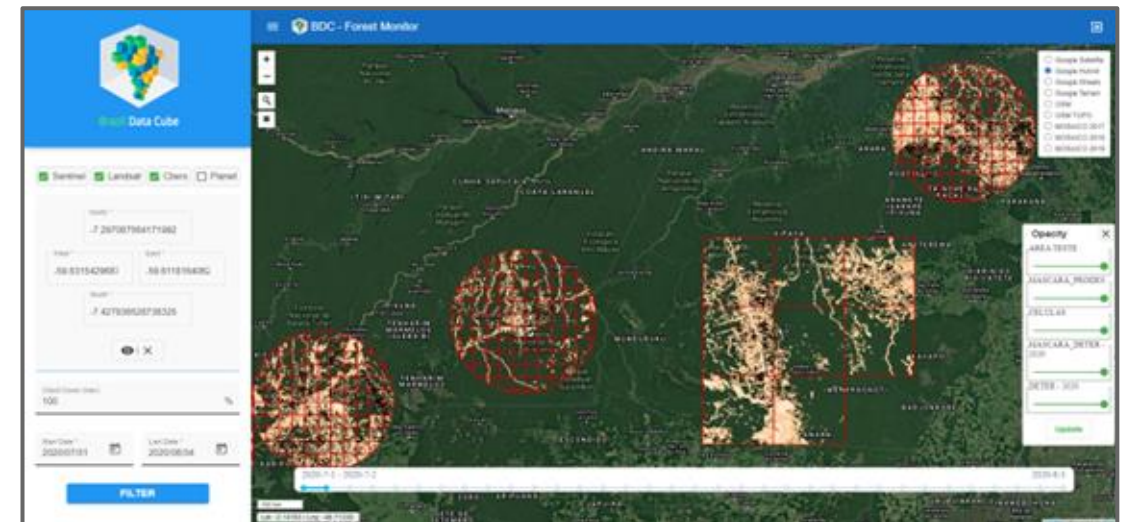
Technological innovation for the environmental monitoring projects of INPE

TerraClass Cerrado 2020 (December, 2022) - BDC



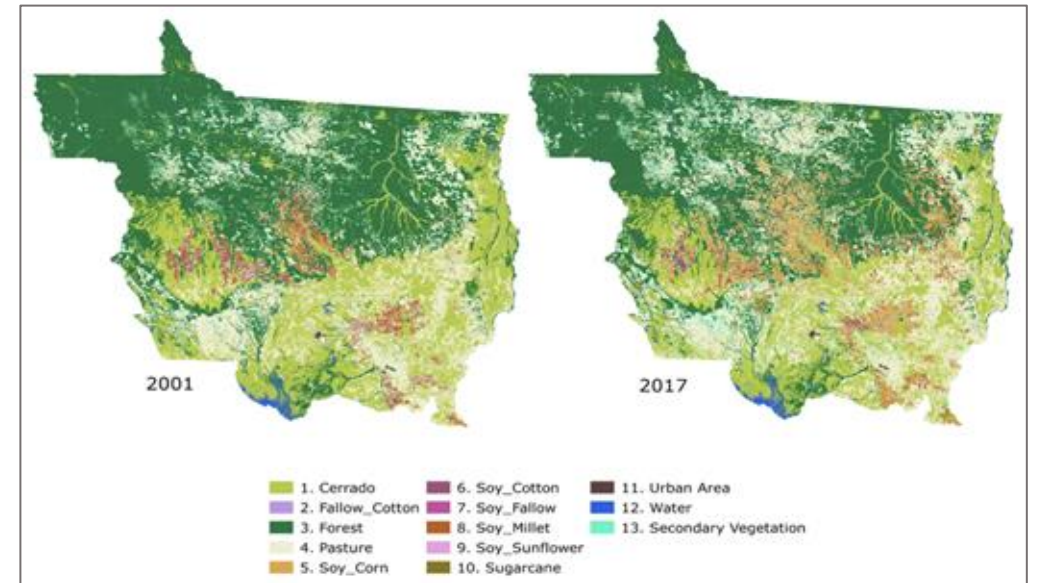
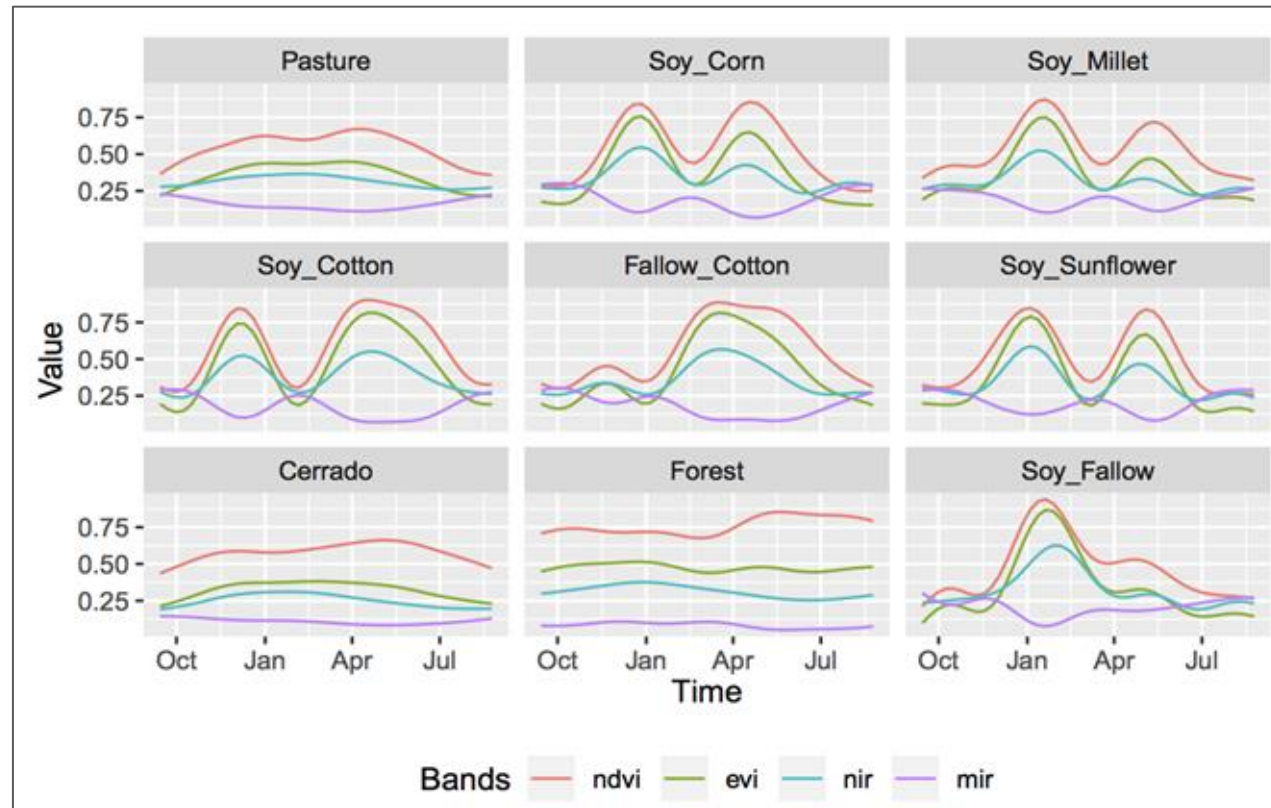
Mosaics – selection of the best pixels (free of clouds or cloud shadow) for periods.

Forest Monitor - DETER Intenso
Service to visualize big Earth observation data on AWS



Motivation

Image time series analysis and machine learning to produce land use and cover information from big Earth observation data



Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017, Scientific Data, 2020 (Simoes et al., 2020)

Image time series NDVI, EVI, NIR, MIR - agriculture year
MODIS – MOD13Q1 Product / Method – SVM (Support Vector Machine)

Motivation

Image time series analysis to extract vegetation phenological metrics.

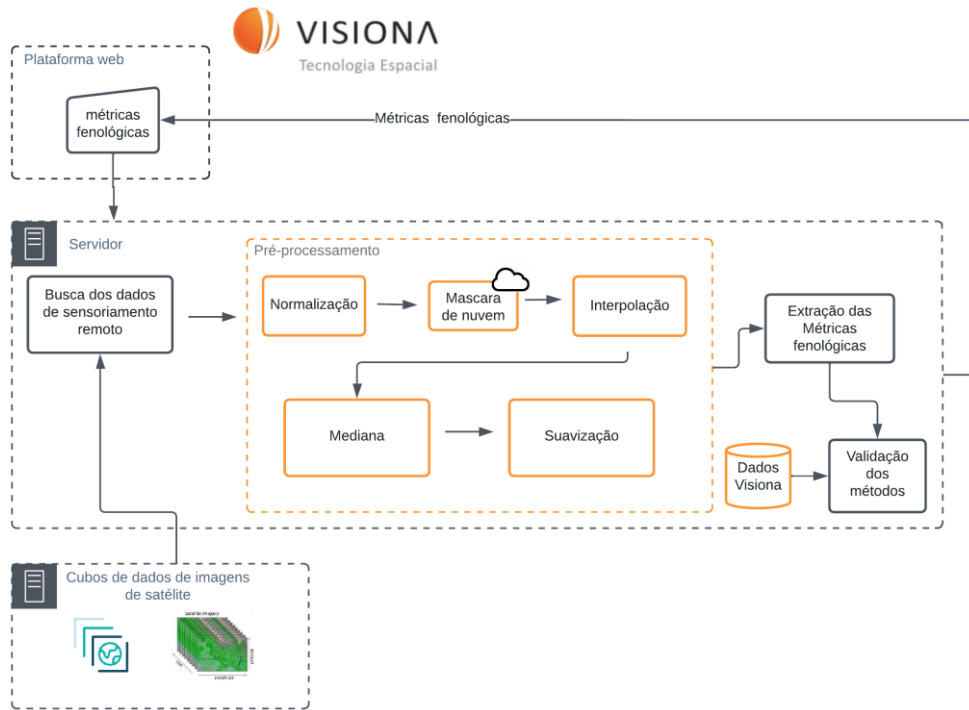
A review of vegetation phenological metrics extraction using time-series, multispectral satellite data, Remote Sensing of Environment, 2020 (Zeng et al., 2020)

Table 4
Summary of main phenological metrics extraction methods for species-specific vegetation types from satellite imagery.

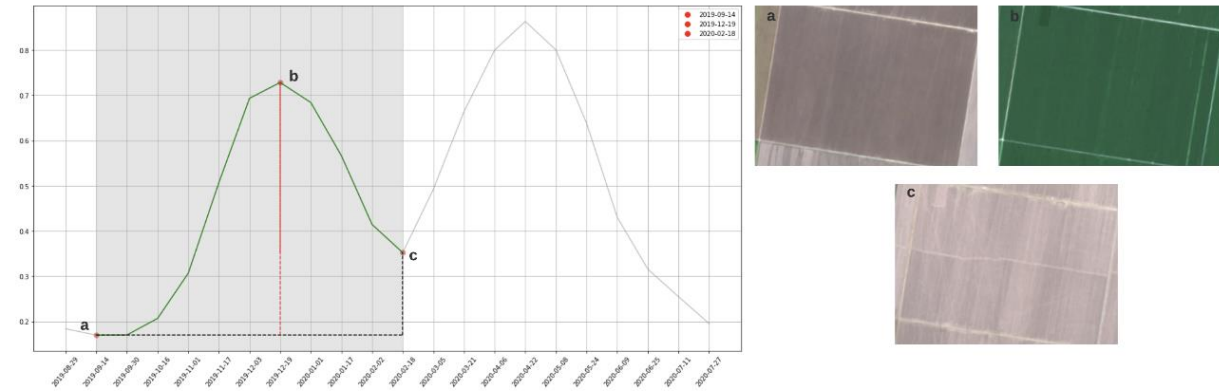
Methods	Vegetation types	Sensors	Stage classification	Specific Stages	Method Classification	Reference
Line segment fitted parameters and statistics	Quercus petraea, Fagus ylvatica L	AVHRR	Physiological-based phenological stages	Budburst, senescence	Empirical Statistics method	Duchemin et al. (1999)
Inflection points determined by derivative	Rice	MODIS	Physiological-based phenological stages	Planting, heading, and harvesting	Empirical method	Sakamoto et al. (2005)
Based on the parameters derived from the best fitted polynomial curve	Potato	MODIS	General phenological stages	12 metrics for potato	Empirical method	Islam and Bala (2008)
Use TIMESAT software to detect rice phenological stages	Rice	MODIS	General phenological stages	Start, peak and end of season	Empirical method	Boschetti et al. (2009)
Derive phenological dates based on the optimum scaling parameters and shape model.	Corn and soybeans	MODIS	Physiological-based phenological stages	8 stages for corn and soybeans respectively	Phenology matching	Sakamoto et al. (2010)
Regress the ground measure degree days and VI values	Sugarcane	ASTER	Physiological-based phenological stages	6 stages	Simulation	Mobasheri et al. (2010)

Motivation

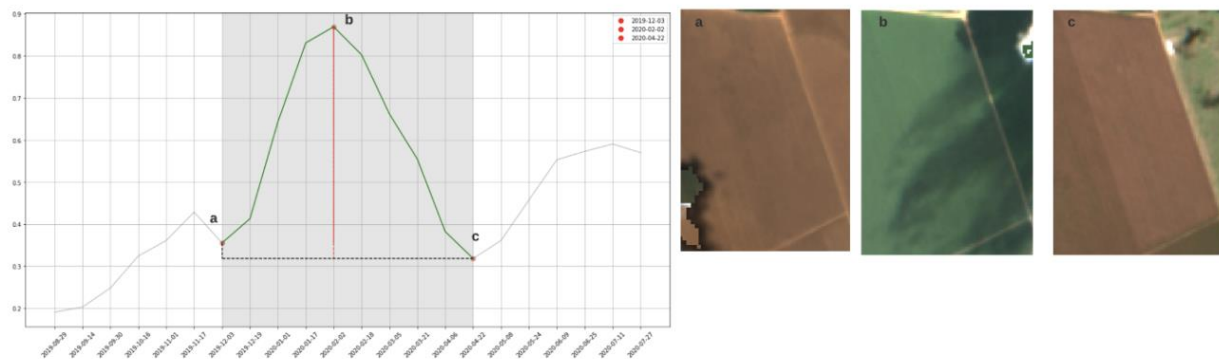
Image time series analysis to extract vegetation phenological metrics.



(a)



(b)



(c)

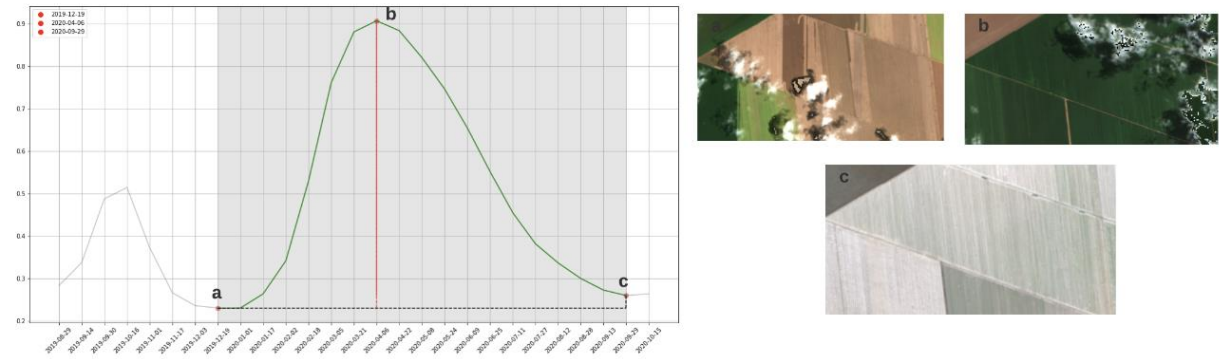


Figura 4. Métricas fenológicas de início, fim e máximo vigor vegetativo de plantio extraídas para soja (a), milho de primeira safra (b) e algodão (c) utilizando séries Sentinel-2A/B de NDVI a cada 16 dias.

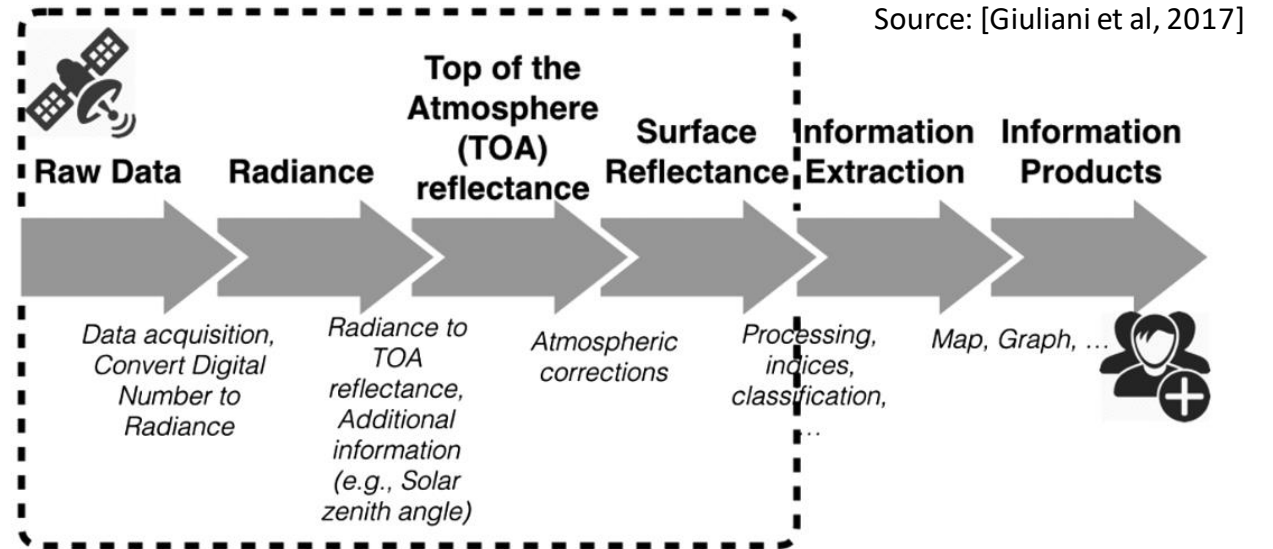
Objectives

(1) Analysis-Ready Data (ARD) of medium-resolution satellite images for Brazil: CBERS-4 Landsat 8/9 Sentinel 2.

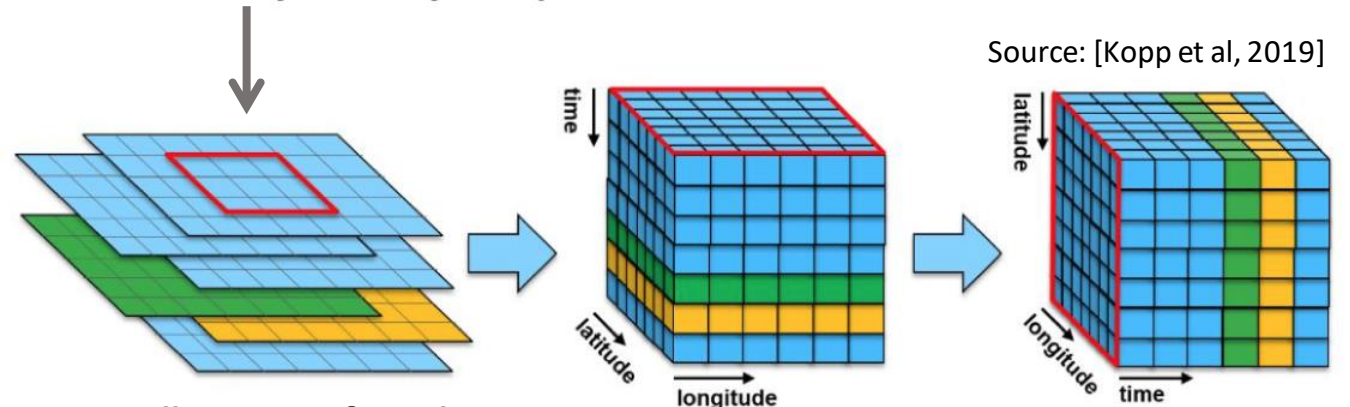
(2) Multidimensional data cubes.



CEOS Analysis Ready Data for Land:
<https://ceos.org/ard/index.html>



Analysis Ready Data production



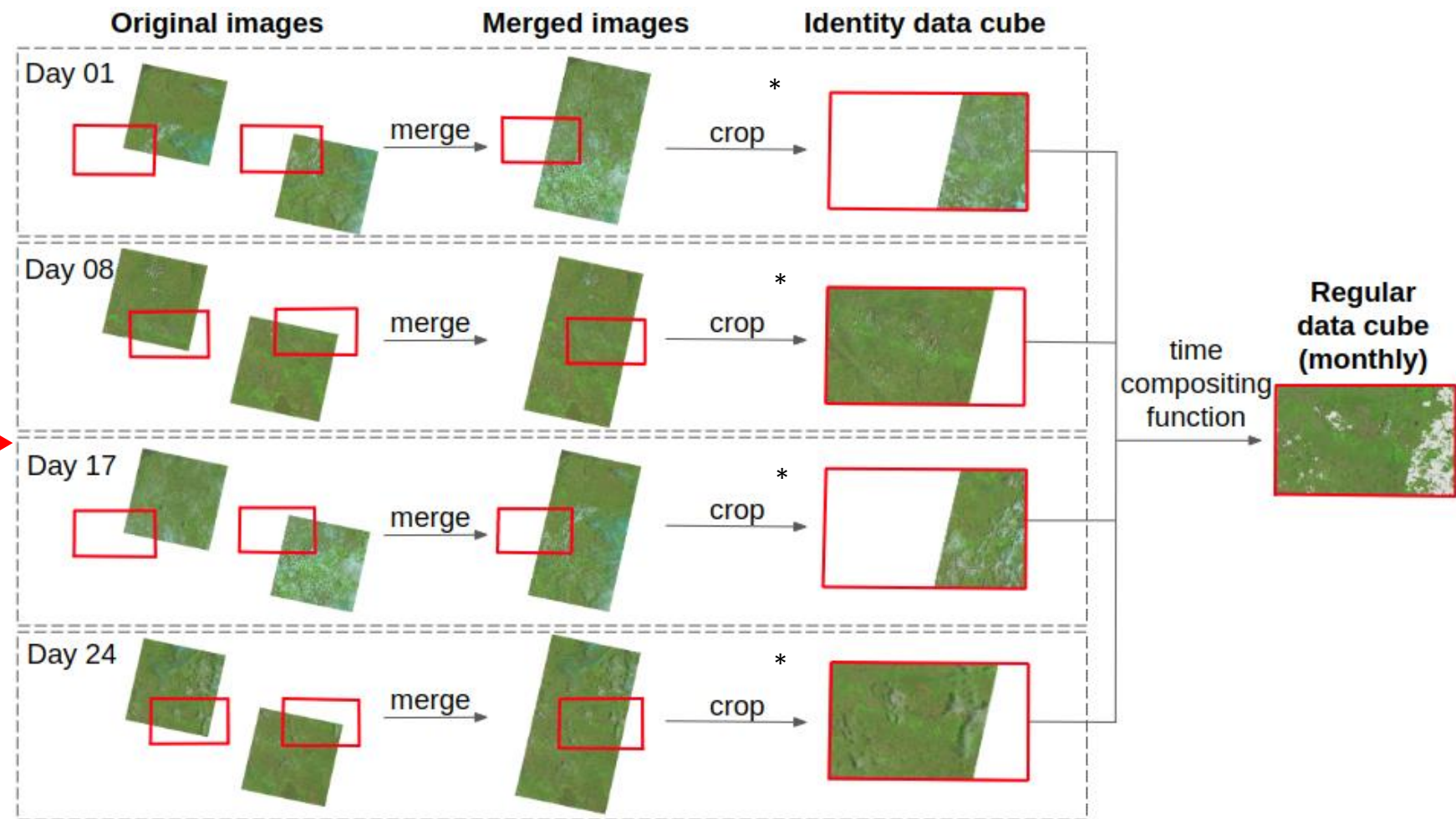
Collections of Earth observation satellite images – ARD

Data cubes – four-dimensional array.

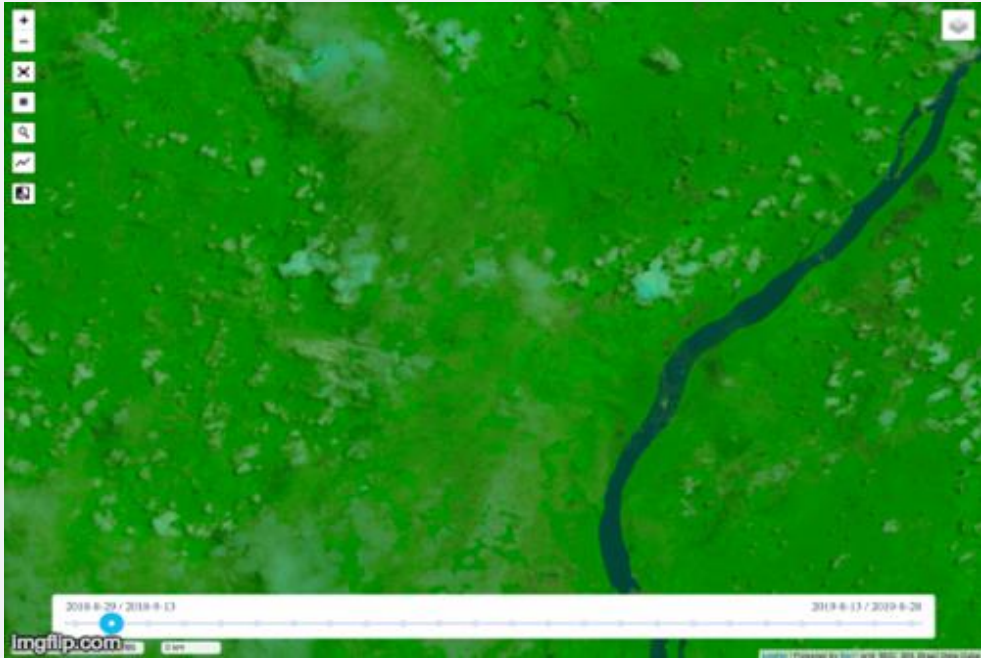
Building data cubes



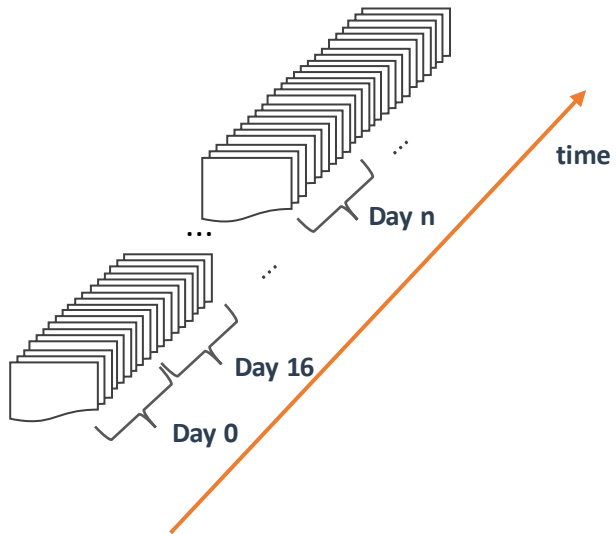
BDC Hierarchical Tiling System



*bilinear resampling for better spatial resolution band



brazildatacube.dpi.inpe.br/portal/explore



For each *tile* and *time step*, there are a set of COG (Cloud Optimized GeoTIFF) files:

- (1) Spectral bands from original images;
- (2) Vegetation indices (EVI and NDVI);
- (3) Cloud mask; (4) number of valid observations (excluding cloud, cloud shadow..); (5) data provenance; ...

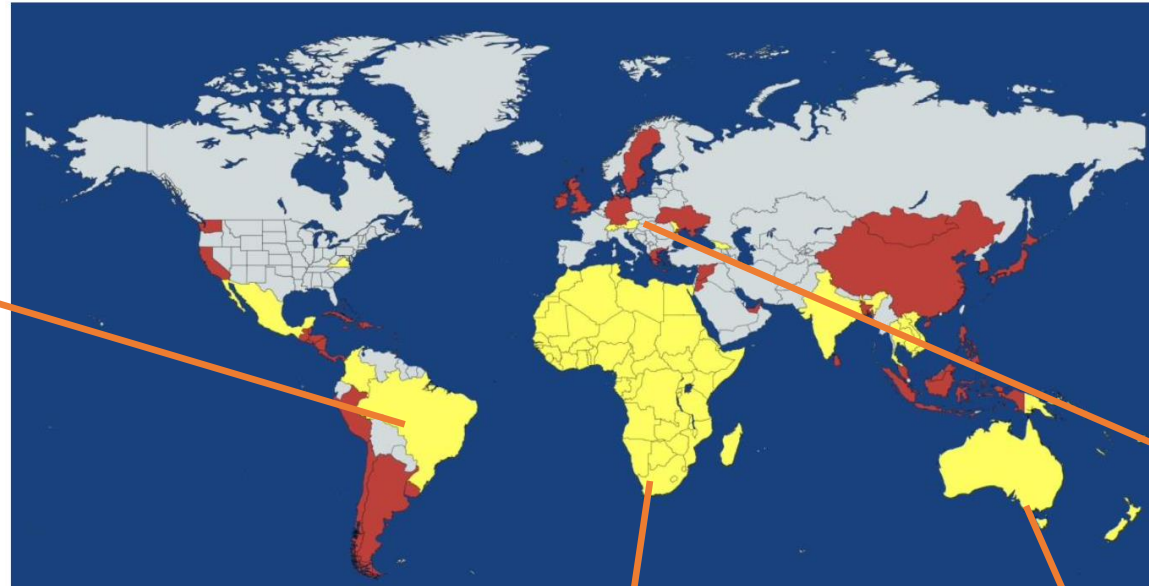
BDC – Small
 Each tile: 105600m x 105600m
 Sentinel-2/MSI – 10 meters
 Each file (band/tile): 400 MB
 Each tile: ~ 5.4 GB

Research and technological innovation

Partnership with international and similar initiatives

Source: [Sudmanns et al, 2022]

Yellow: operational
Red: under development



**BRAZIL
DATA CUBE**

<http://brazildatacube.org>

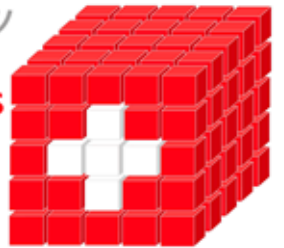


Article

Paving the Way towards an Armenian Data Cube



**Swiss
Data
Cube**



<https://www.swissdatacube.org/>

Digital Earth Americas
A vision for a solution that provides relevant Earth observation data for the benefit of the Americas



**Digital Earth
AFRICA**

<https://digitalearthafrika.org/>



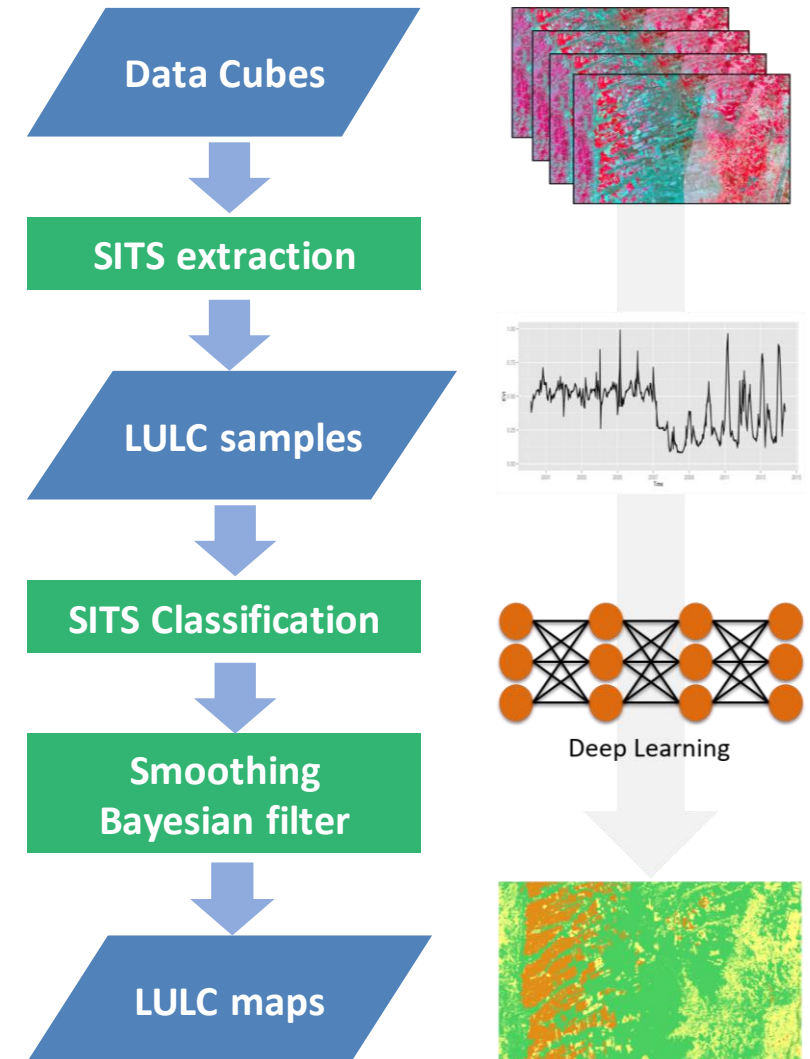
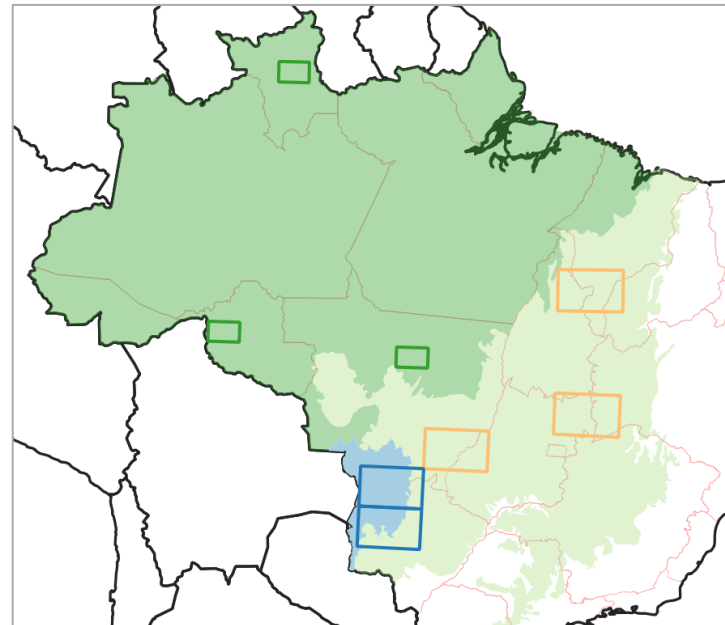
**Digital Earth
AUSTRALIA**

<https://www.dea.ga.gov.au/>

Objective

(3) Big data technologies, image time series analysis and machine learning methods .

(4) Land use and cover classification.



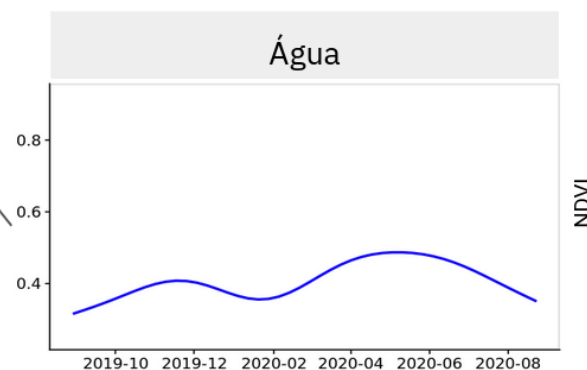
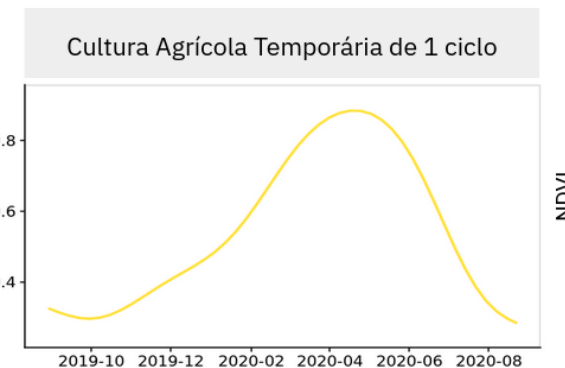
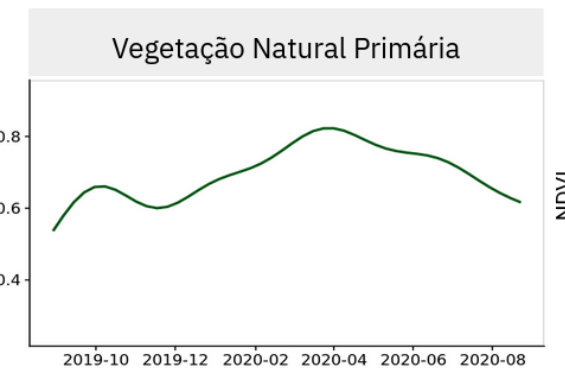
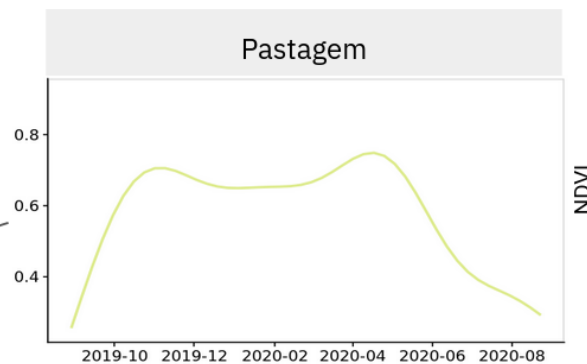
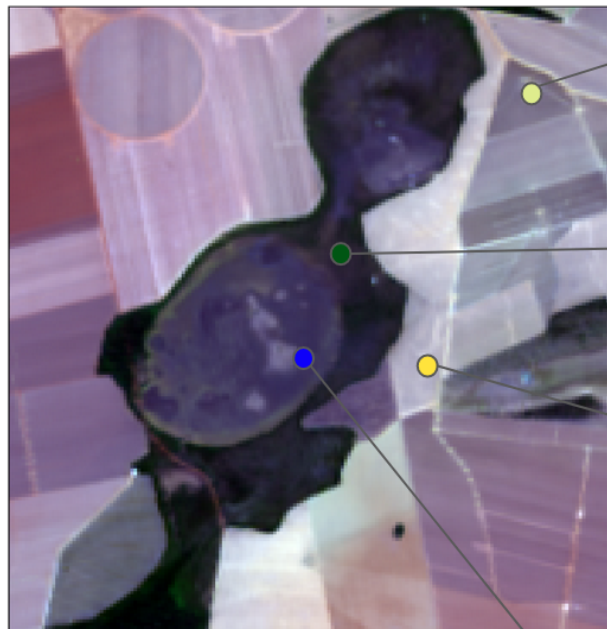
Source: [Ferreira et al, 2020]

Examples of NDVI time series

Time series extracted from Landsat8 data cubes

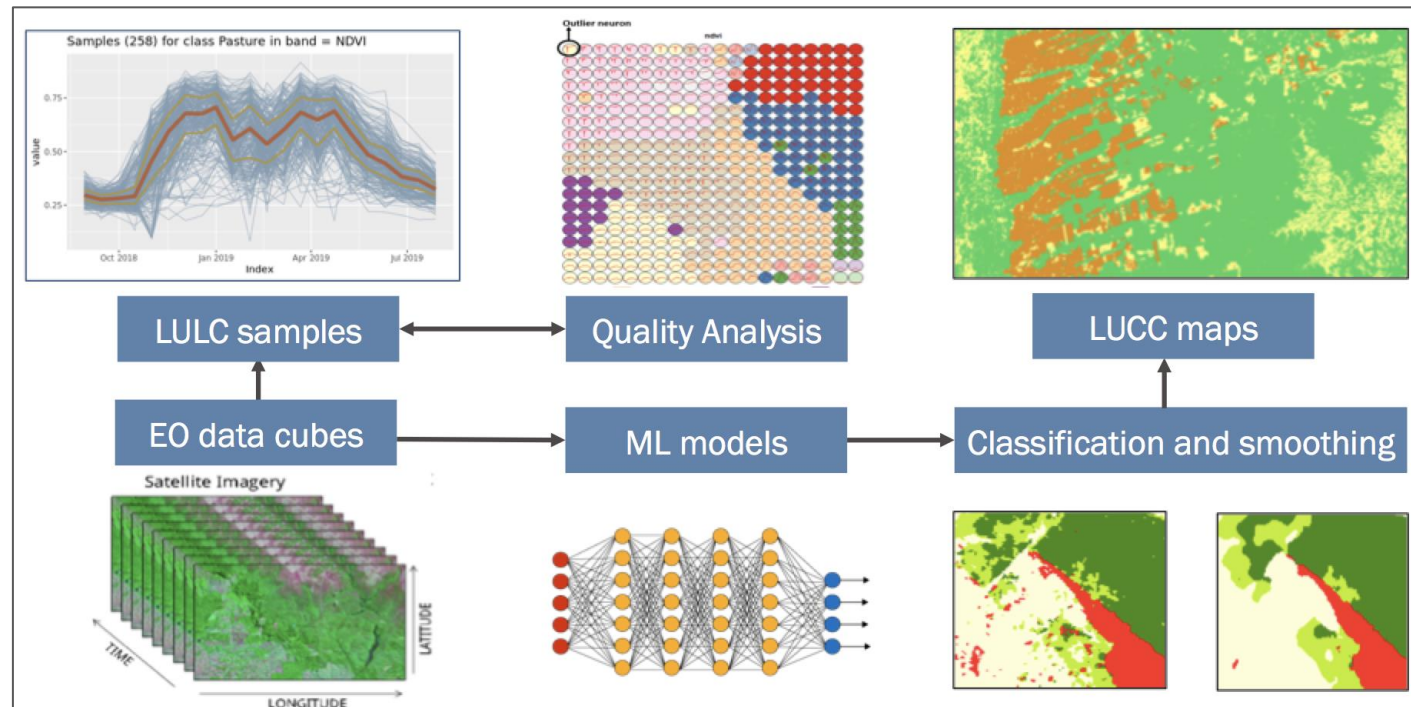
NDVI time series - September of 2019 to September of 2020

Landsat 8 image



SITS (Satellite Image Time Series) R package

<https://github.com/e-sensing>

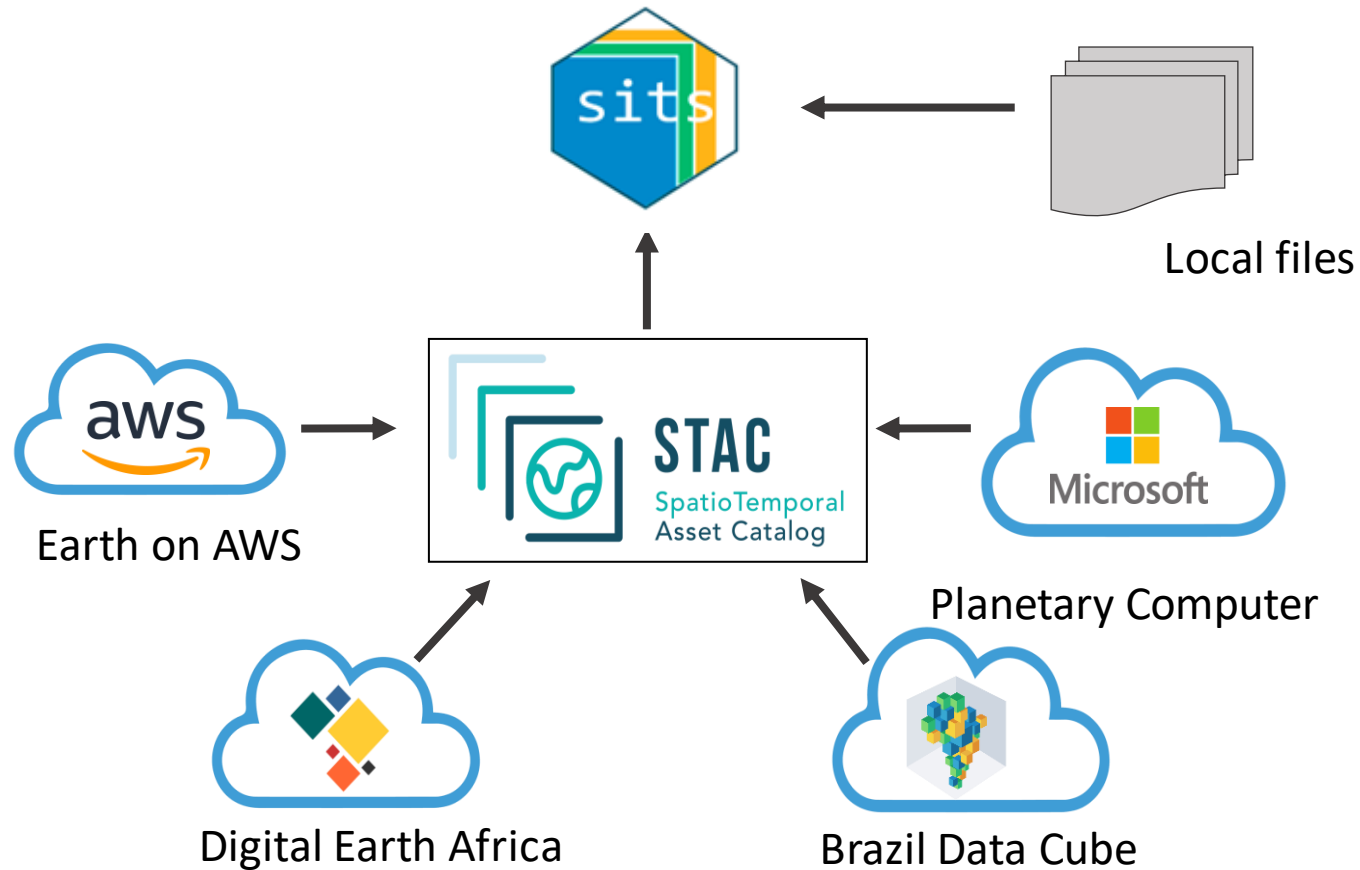


```
cube <- sits_cube(source = "BDC",  
                 collection = "S2_10_16D",  
                 name = "s2_cube",  
                 bands = c("NDVI", "EVI"),  
                 tiles = "022024", start_date = "2018-09-01", end_date = "2019-08-28")  
  
samples <- readRDS(url("https://bdc.dpi.inpe.br/rds/S2_10_16D.rds"))  
  
mlp_model <- sits_deeplearning(layers = c(512, 512, 512, 512, 512), activation = "relu")  
  
dl_model <- sits_train(samples, mlp_model)  
  
result <- sits_classify(data = cube, ml_model = dl_model, output_dir = tempdir())
```

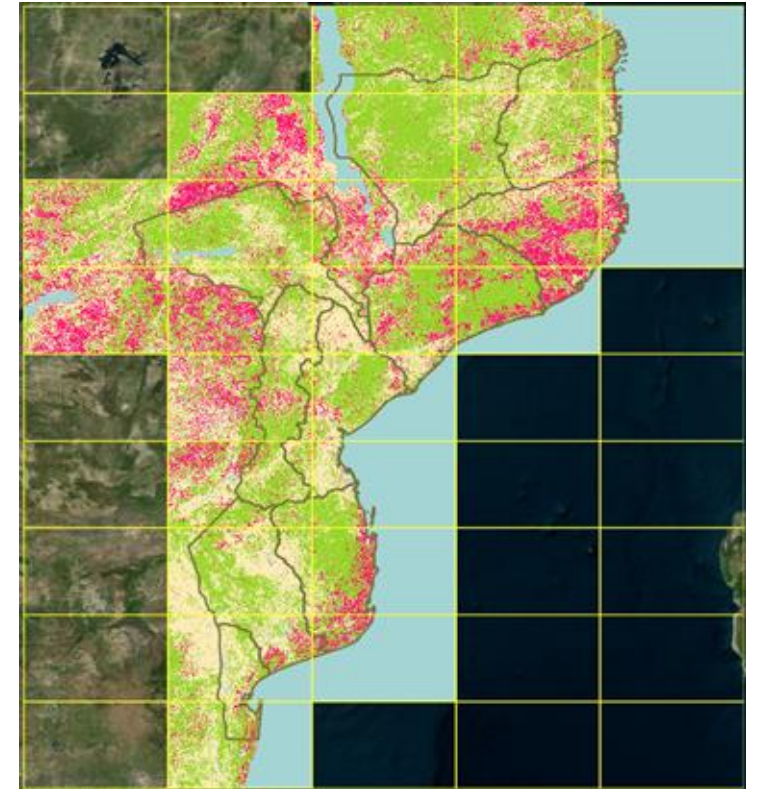


SITS (Satellite Image Time Series) R package

<https://github.com/e-sensing>



Land use and cover map for Mozambique - 2016



- 1 Area Alagada
- 2 Cultivo
- 3 Floresta
- 4 Outras terras
- 5 Pradarias

Land use and land cover maps

TerraClass Cerrado 2020
(Launched in December 2022)

Geospatial Data Science Computational Environment

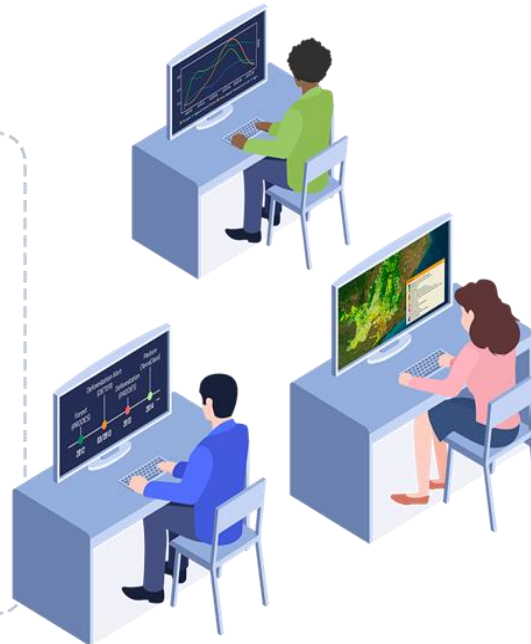
Retrieving the Time Series

```

In order to retrieve the time series for attributes 'red' and 'nir', in the location of latitude -12 and longitude -54 from January 1st, 2001 to December 31st, 2001, use the time_series function:

ts = Rstos::time_series(wts_inpa,
  attributes = c("red_reflectance", "NIR_reflectance"),
  attributes = c("red_reflectance", "NIR_reflectance"),
  latitude = -12.0,
  longitude = -54.0,
  name = "R200101-01",
  start_date = "2001-01-01",
  end_date = "2001-12-31")
    
```

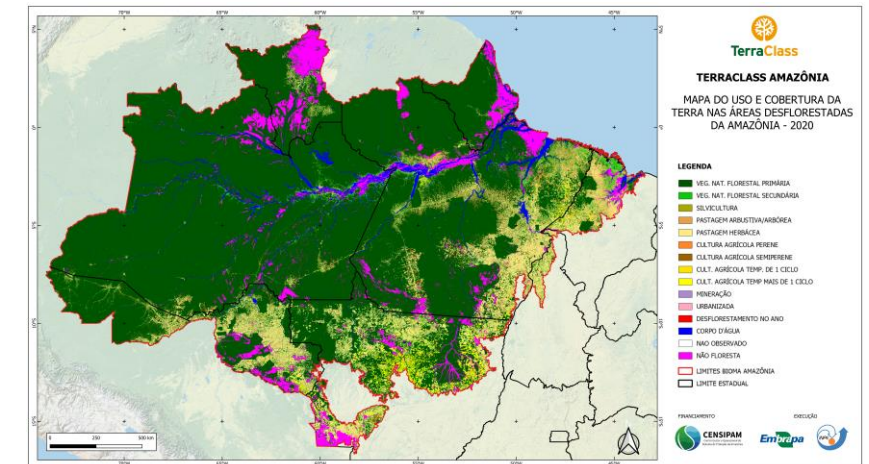
WTSS WLTS BDC STAC SITS



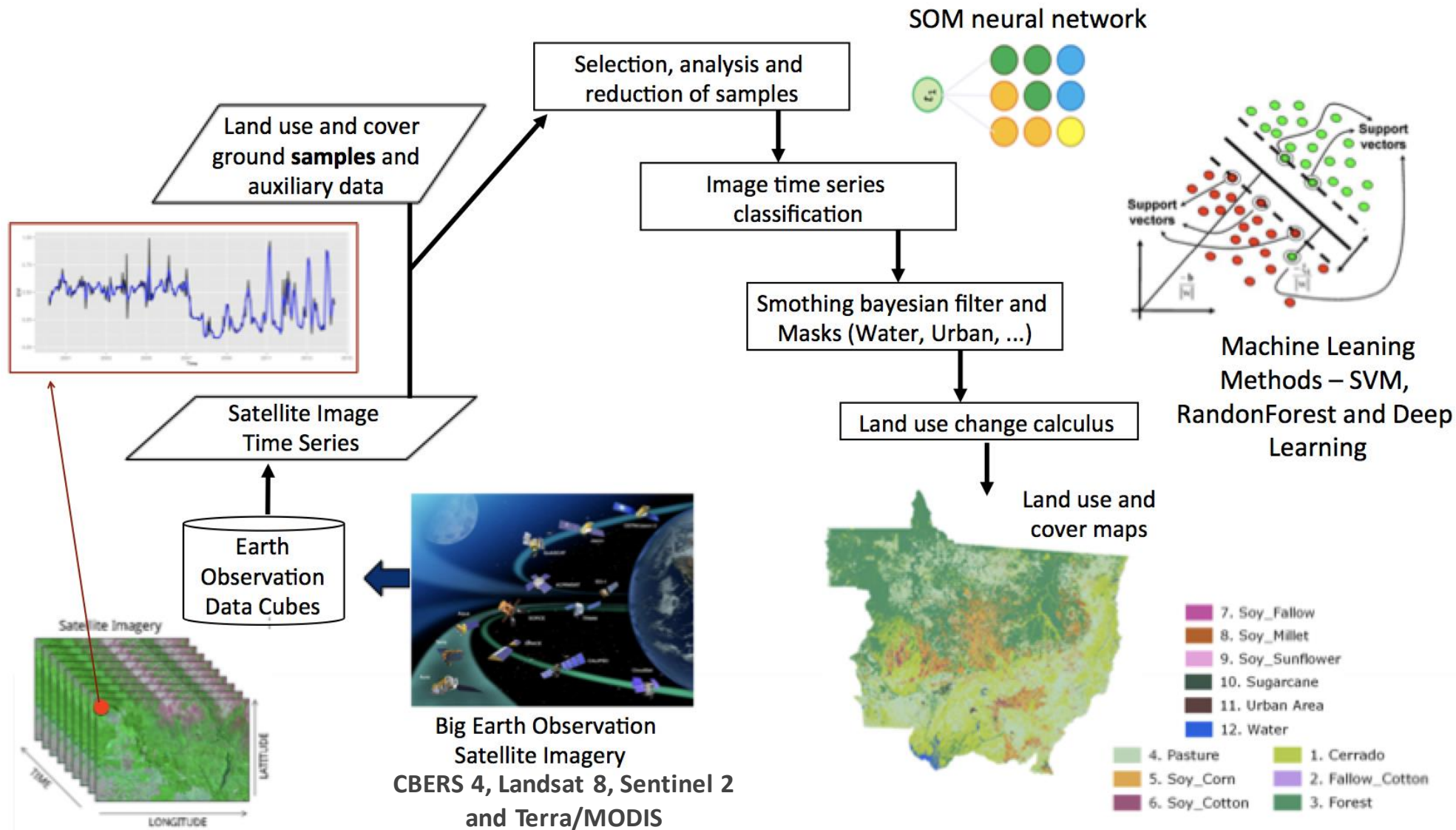
Server-side processing
infrastructure at INPE



TerraClass Amazônia 2020



~ 33 Terabytes - Sentinel-2 data cubes (16-days) - 292 tiles
~ 25,000 samples - Random Forest classifier



ARD and Data cubes available at:

<http://brazildatacube.dpi.inpe.br/portal/explore>

Land use and cover change maps:

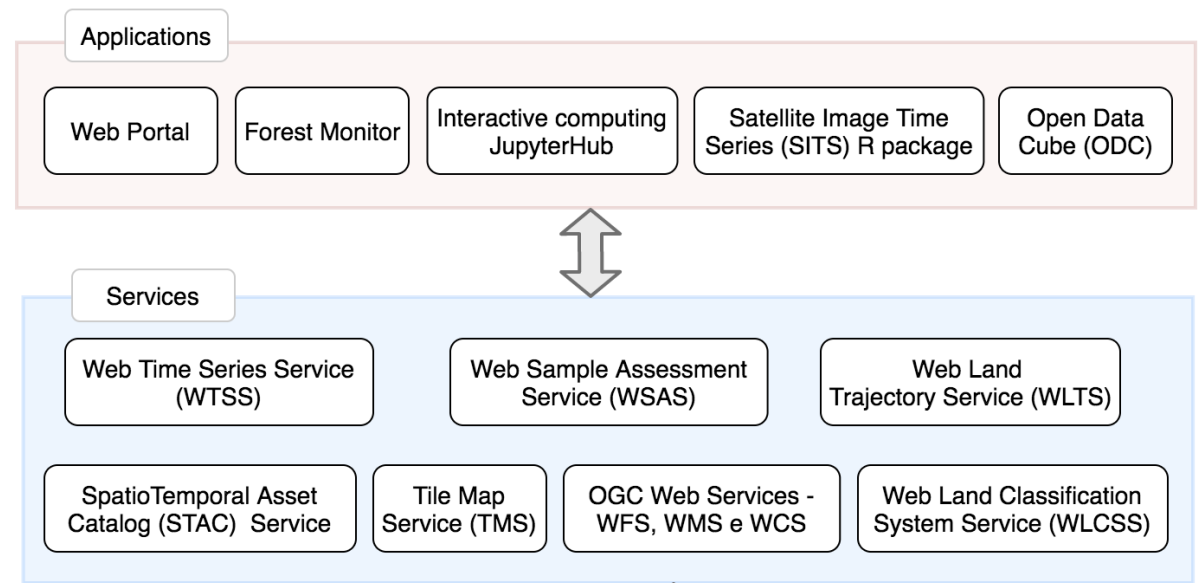
<http://brazildatacube.dpi.inpe.br/portal/explore>



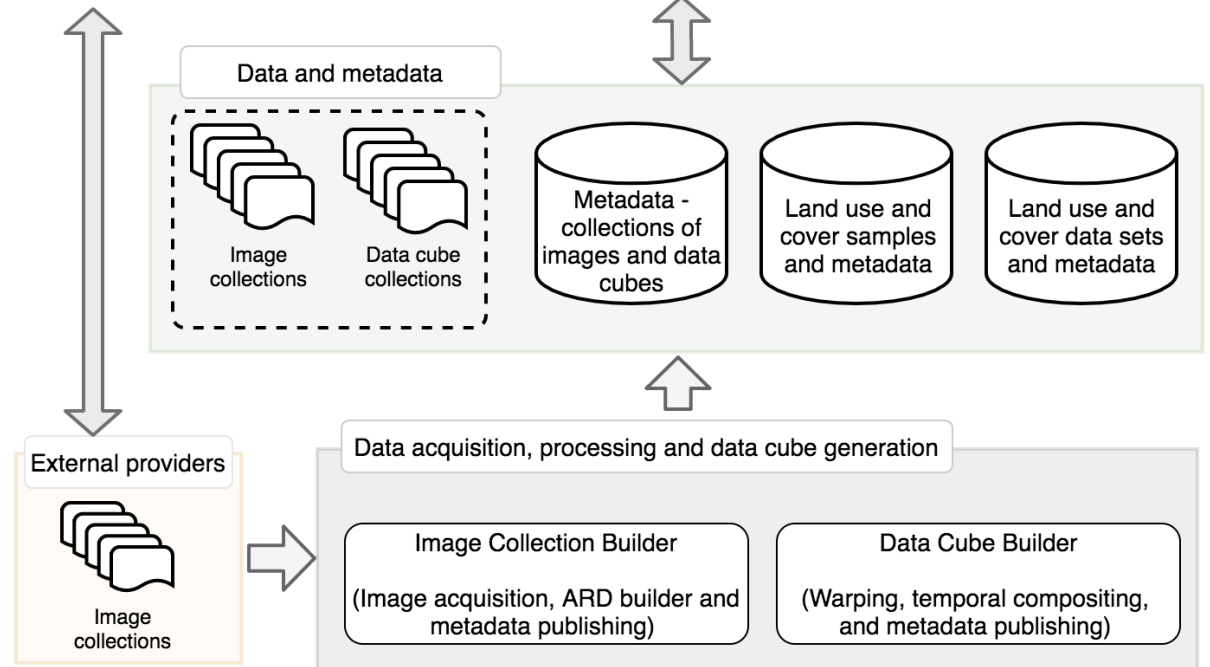
BRAZIL
DATA CUBE

Open
Data and
Software
Products

Software



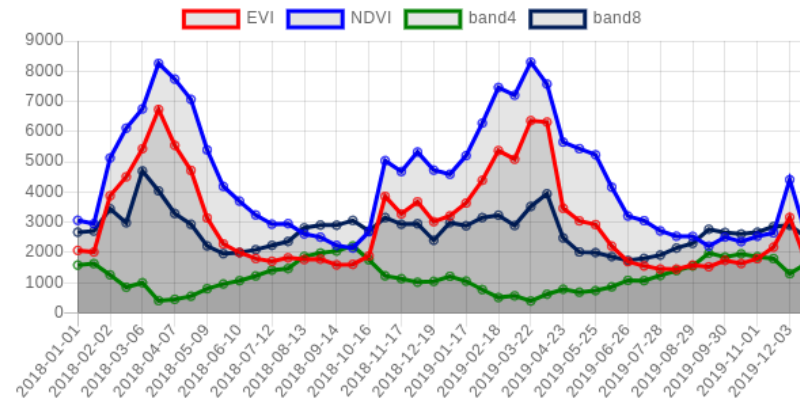
Data and metadata



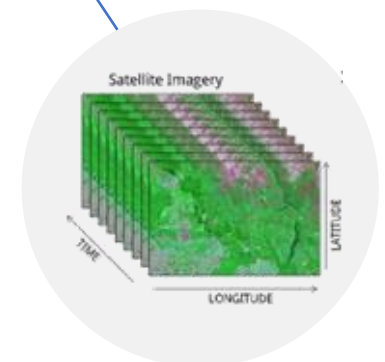
Source: [Ferreira et al., 2020]

Web Time Series Service (WTSS)

Image time series



Metadata

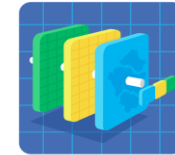


Data cubes

Web Land Trajectory Service (WLTS)

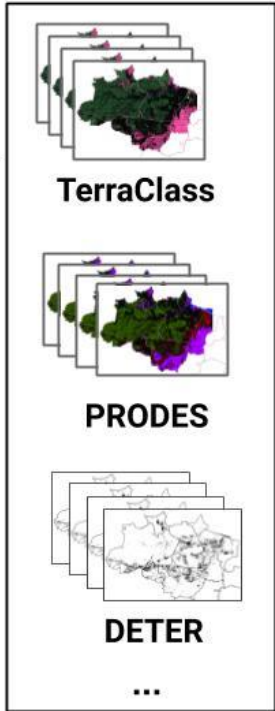


What is the trajectory of land use and cover at this location in PA?

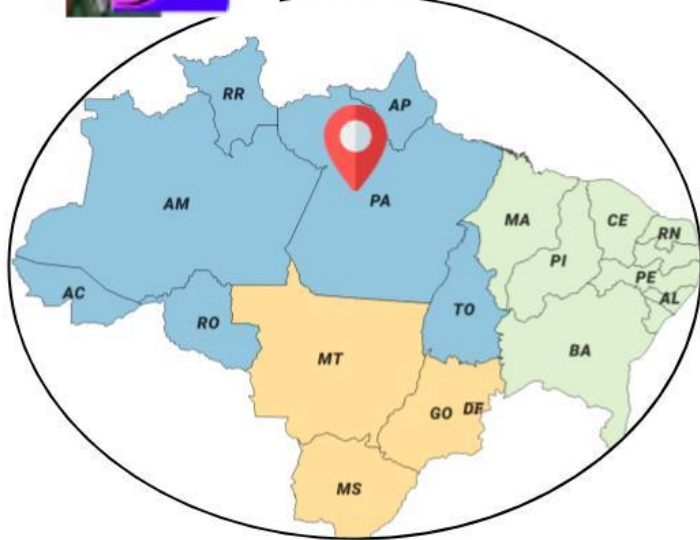


WLTS

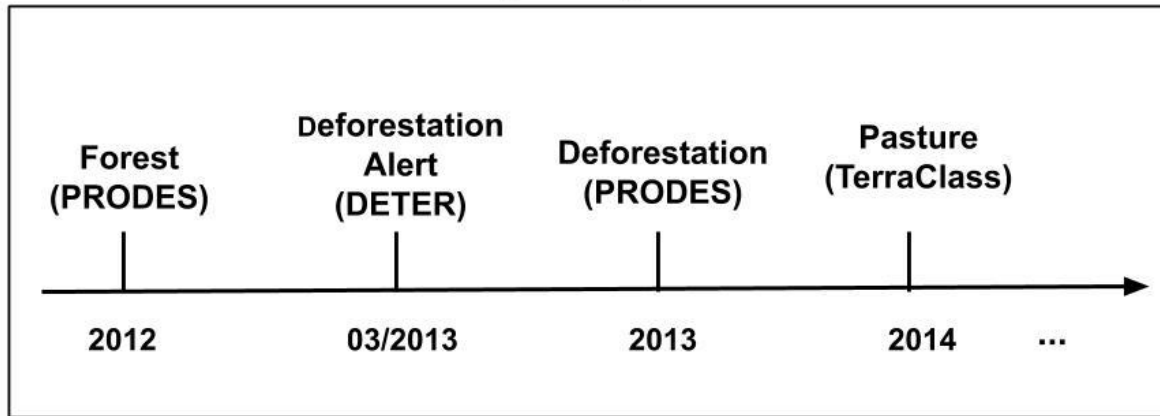
Web Land Trajectory Service



Land use and cover data sets

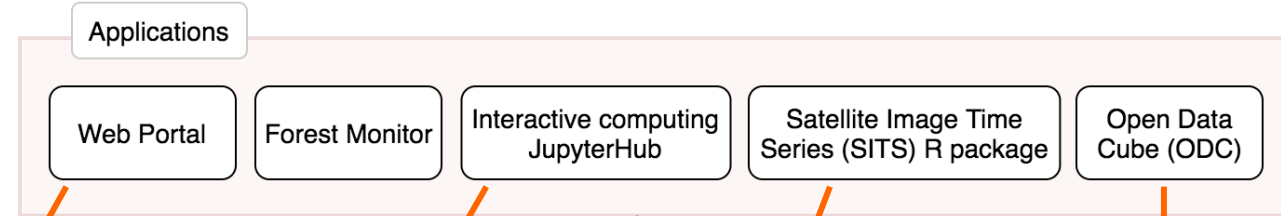
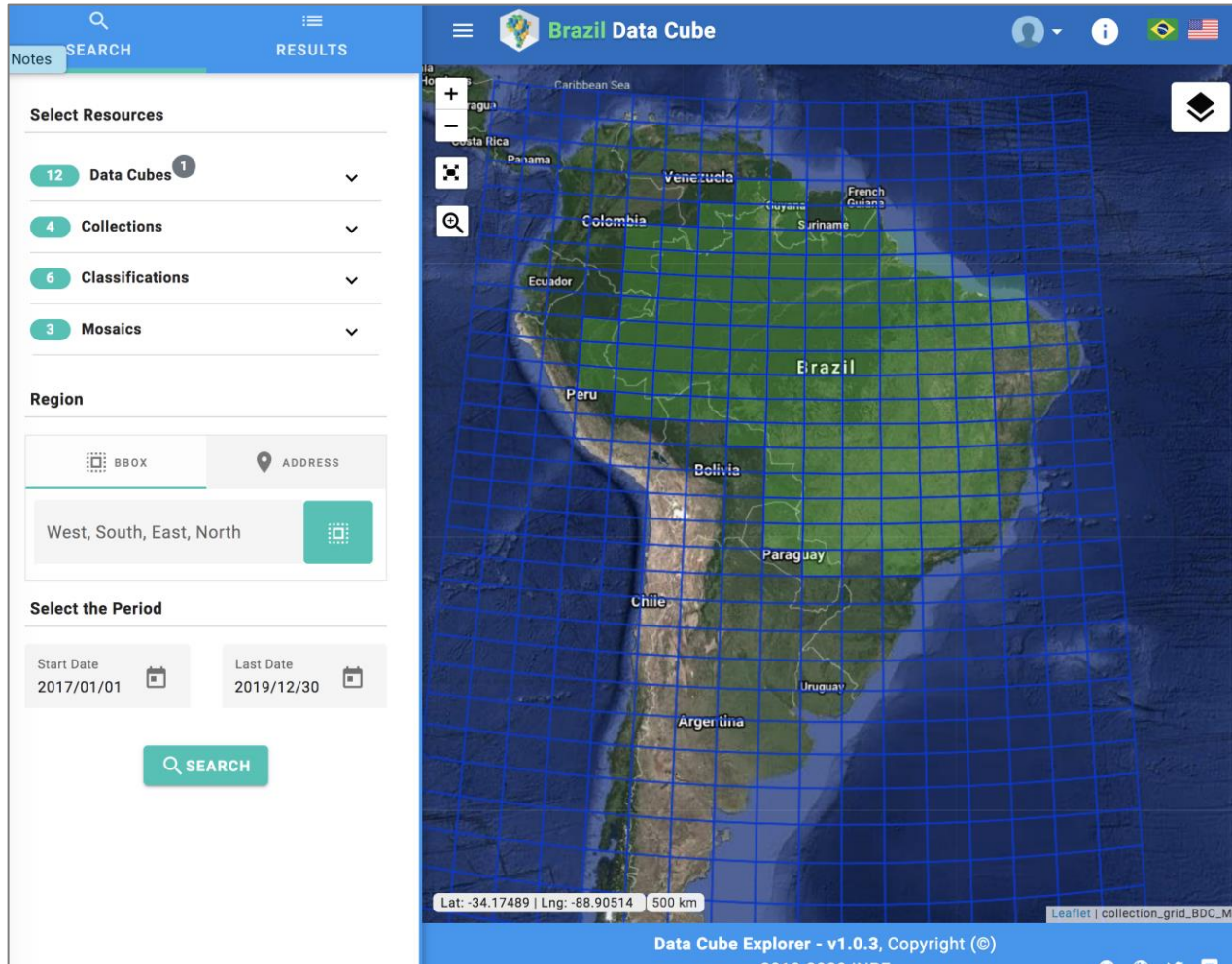


Land use and cover trajectory at this location in PA



BDC Explorer 3.0

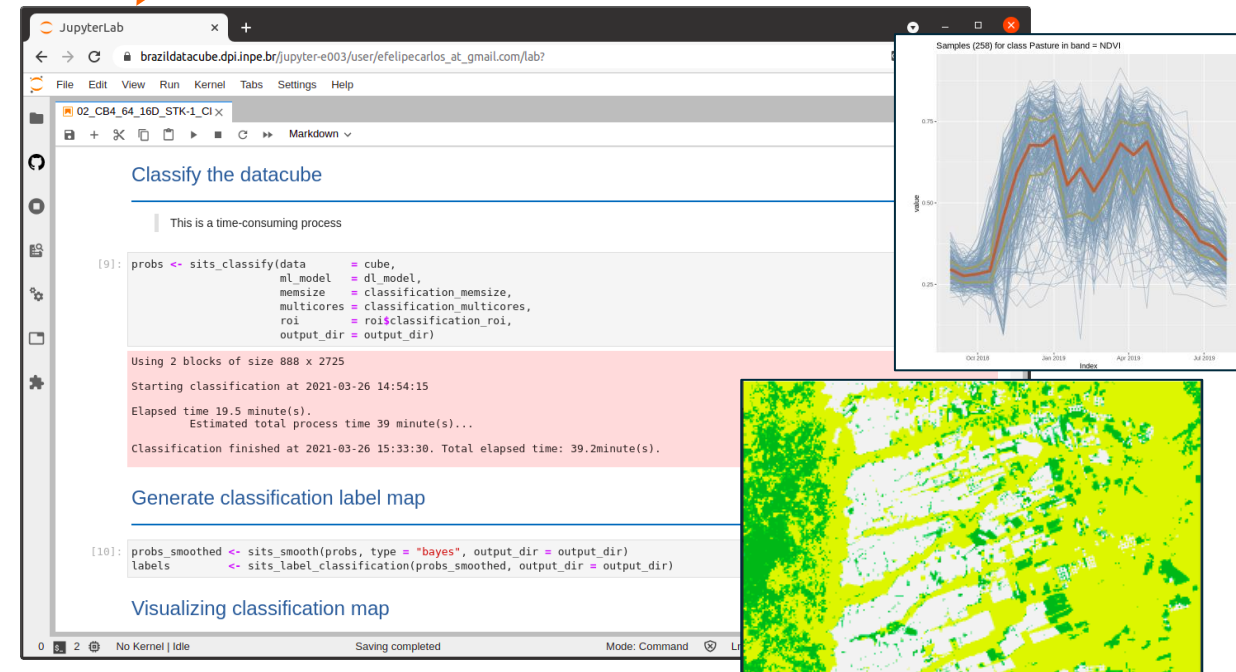
<http://brazildatacube.dpi.inpe.br/portal/explore>



SITS (Satellite Image Time Series) R package:
<https://github.com/e-sensing>



BDC JupyterHub : Interactive computing



```
[9]: probs <- sits_classify(data = cube,
ml_model = dl_model,
memsize = classification_memsize,
multicores = classification_multicores,
roi = roi,
output_dir = output_dir)

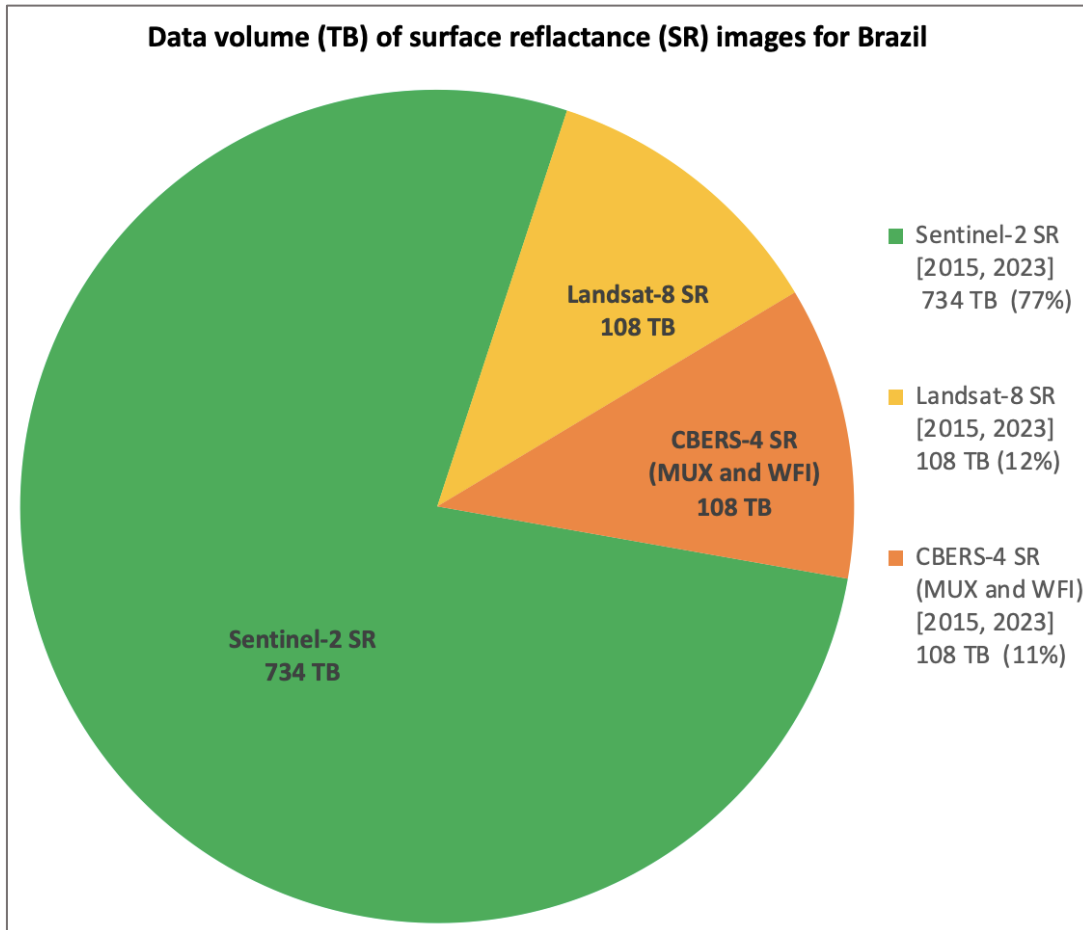
Using 2 blocks of size 888 x 2725
Starting classification at 2021-03-26 14:54:15
Elapsed time 19.5 minute(s).
Estimated total process time 39 minute(s)...
Classification finished at 2021-03-26 15:33:30. Total elapsed time: 39.2minute(s).
```

```
[10]: probs_smoothed <- sits_smooth(probs, type = "bayes", output_dir = output_dir)
labels <- sits_label_classification(probs_smoothed, output_dir = output_dir)
```

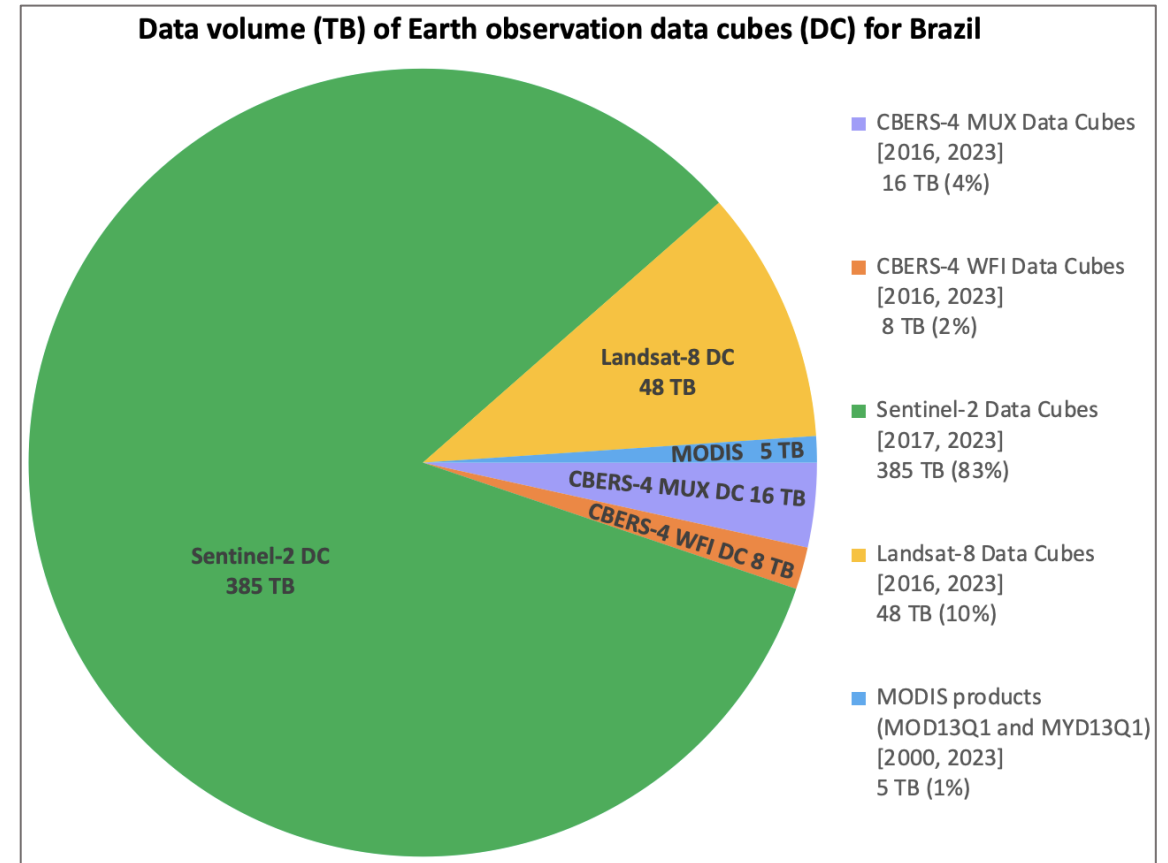

Big data

~ 2 Petabytes (PB)

Data volume (TB) of surface reflectance (SR) images for Brazil



Data volume (TB) of Earth observation data cubes (DC) for Brazil

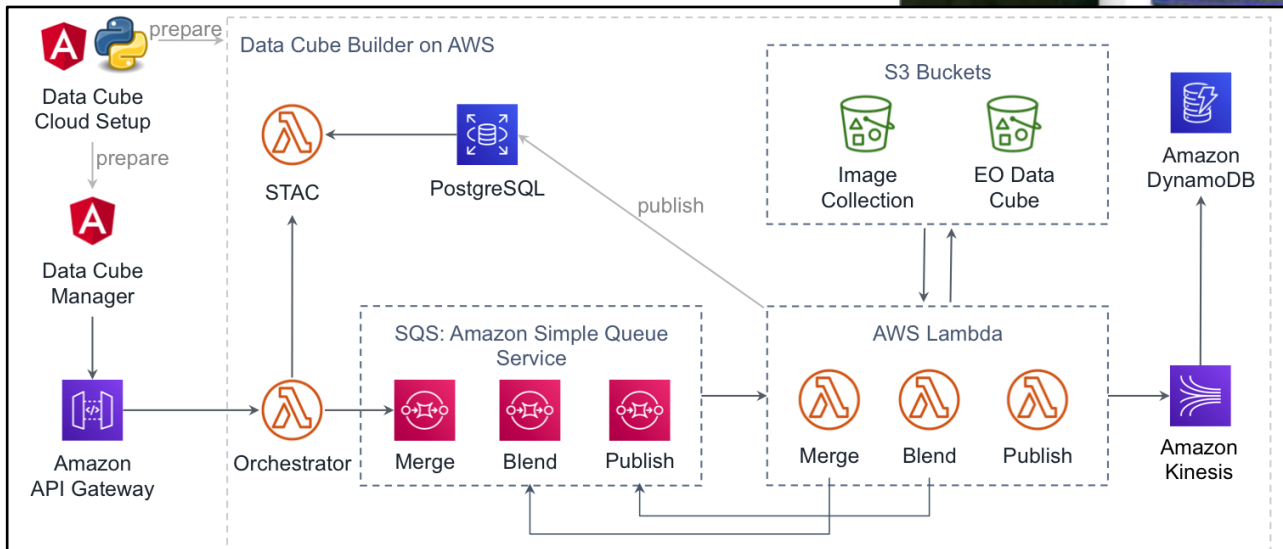
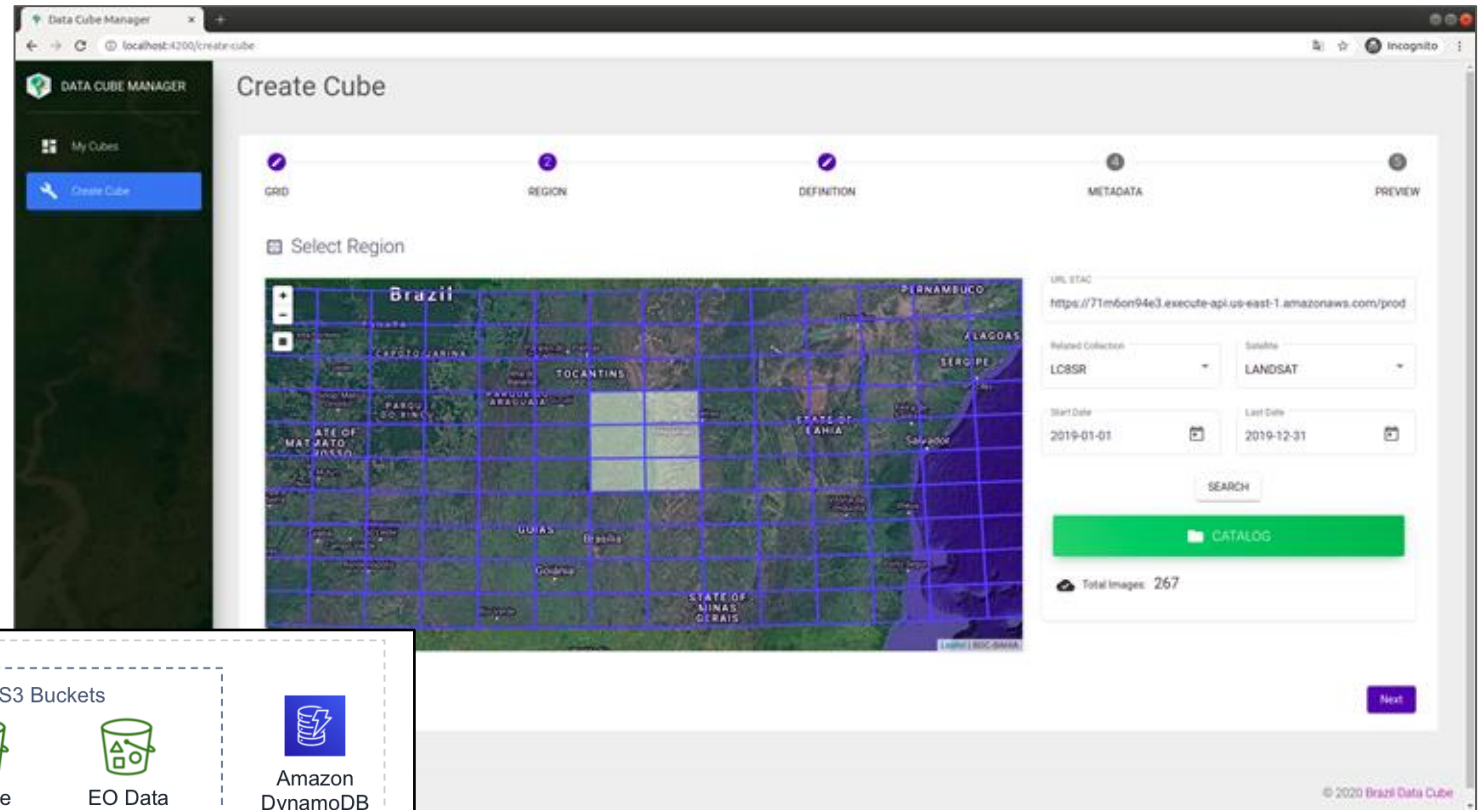


Source: [Ferreira et al., 2022]

Amazon Web Services (AWS)

Data Cube Builder On AWS

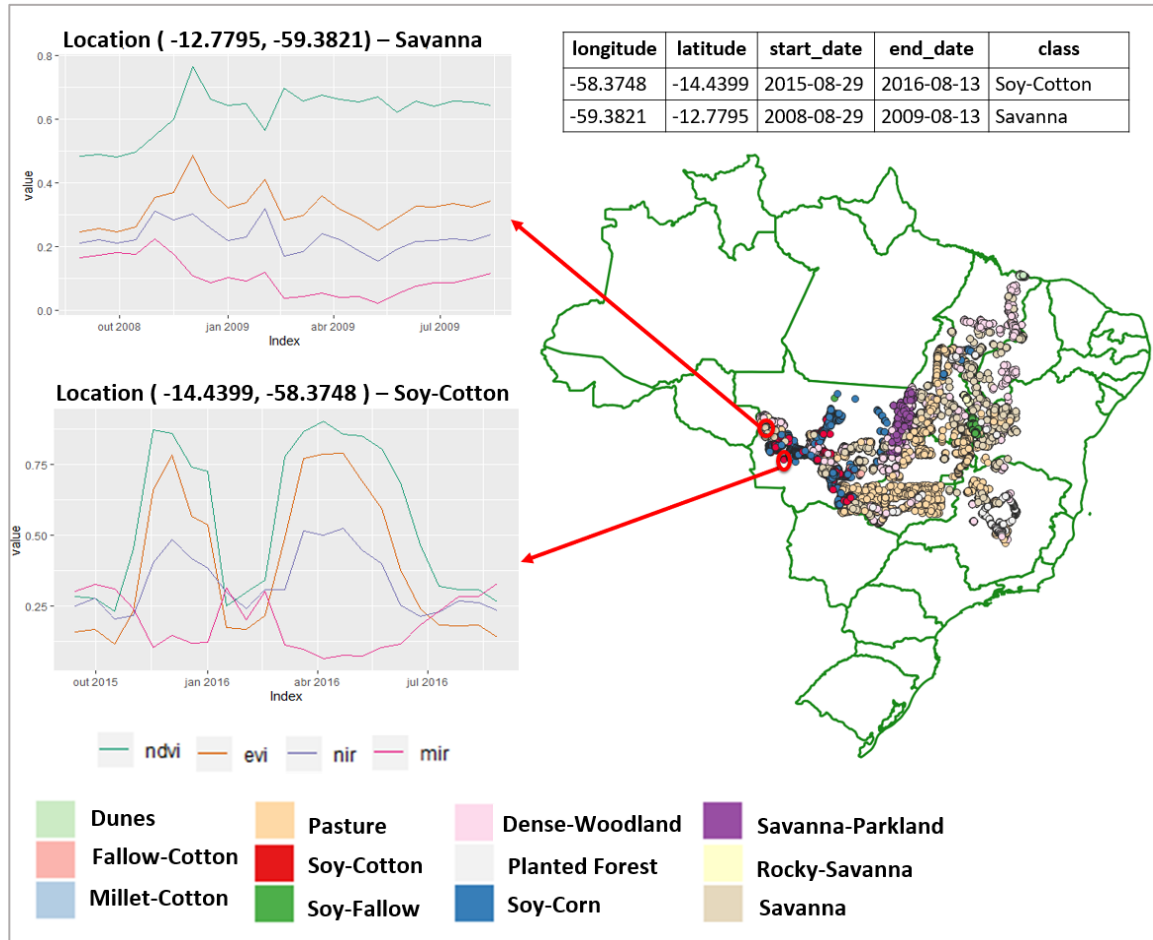
Build Sentinel-2 data cubes on AWS for 2021
AWS Open Data



GEO AWS Cloud Credit Program
U\$ 60,000

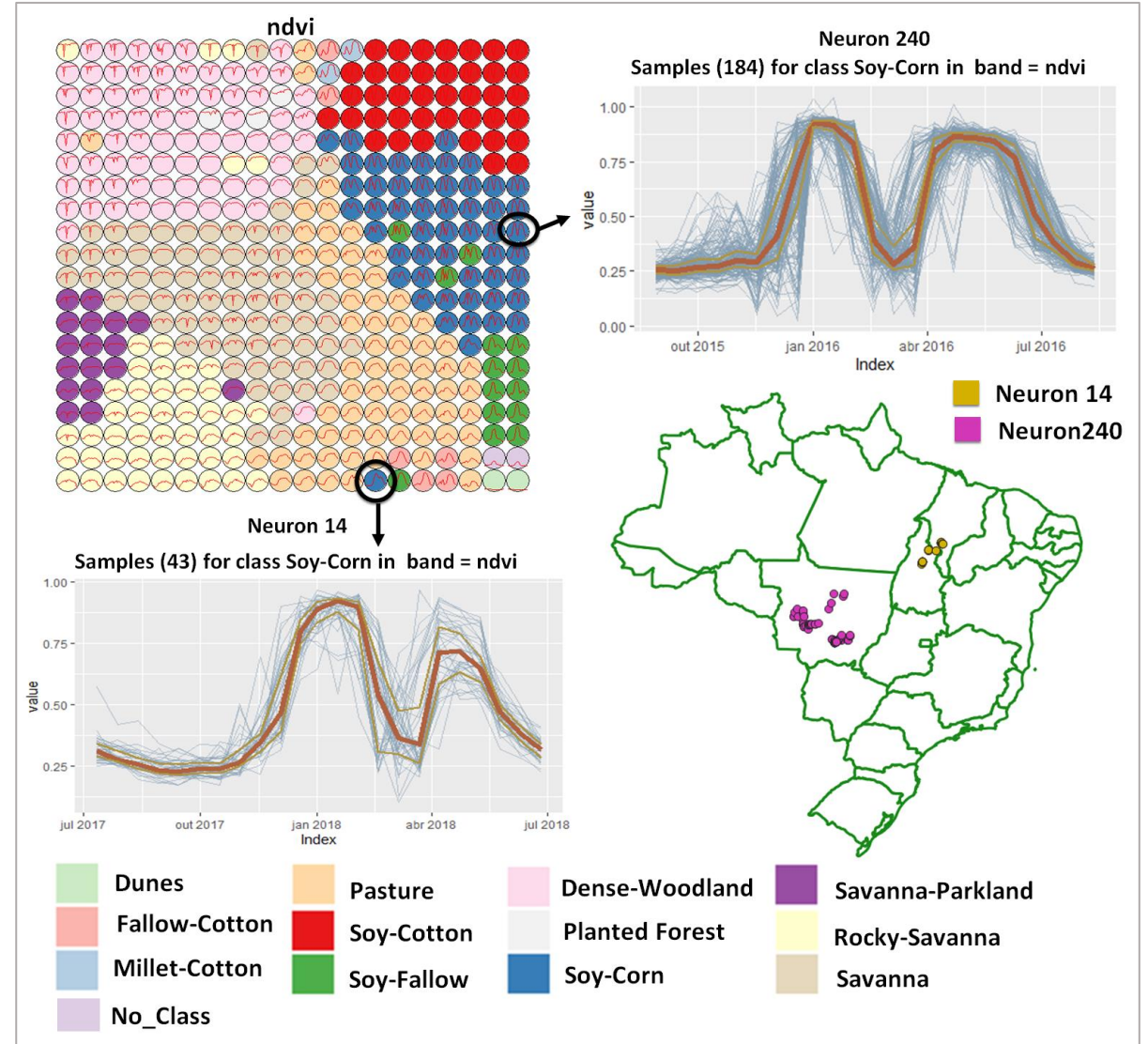


Challenge – Samples



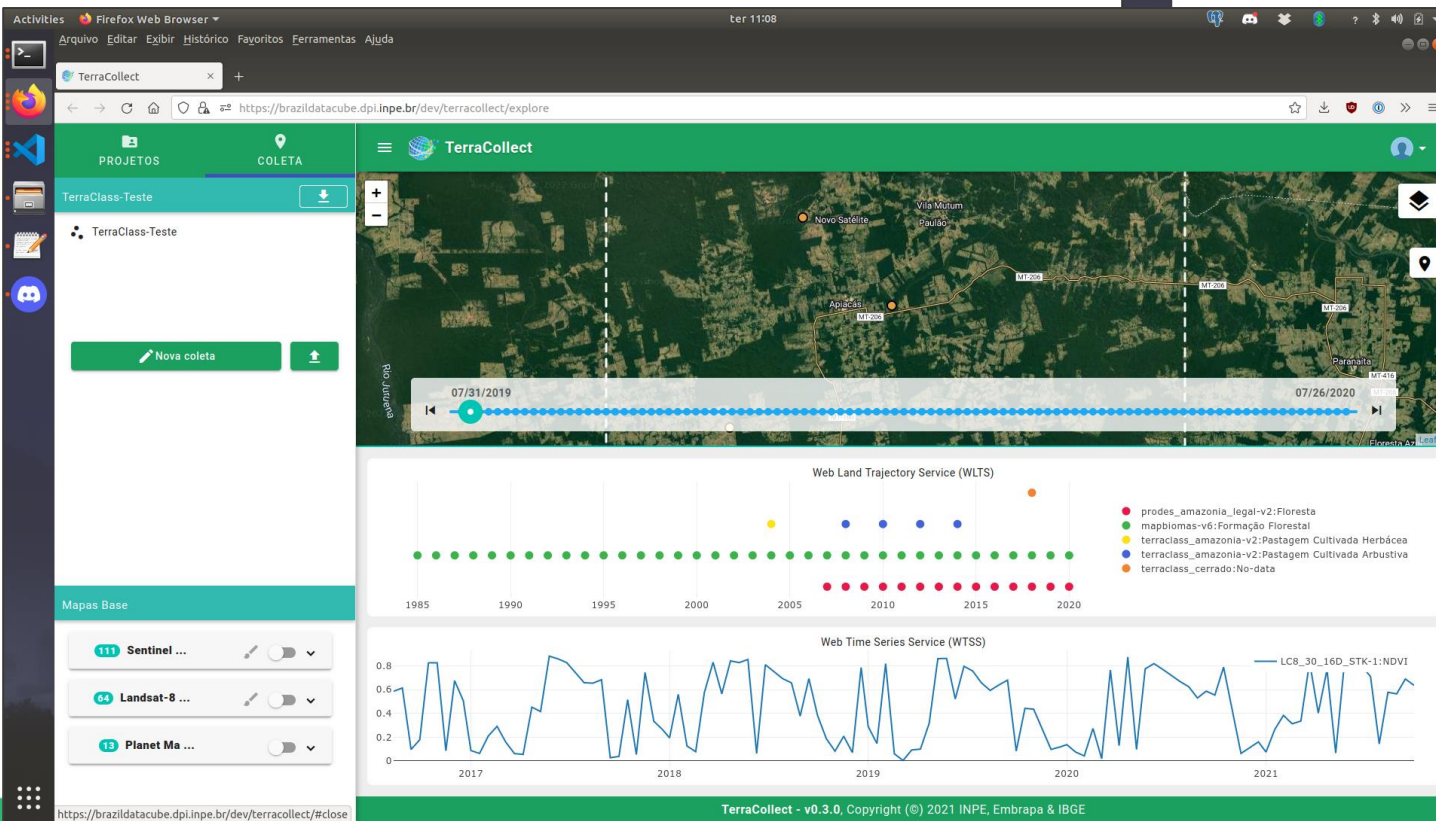
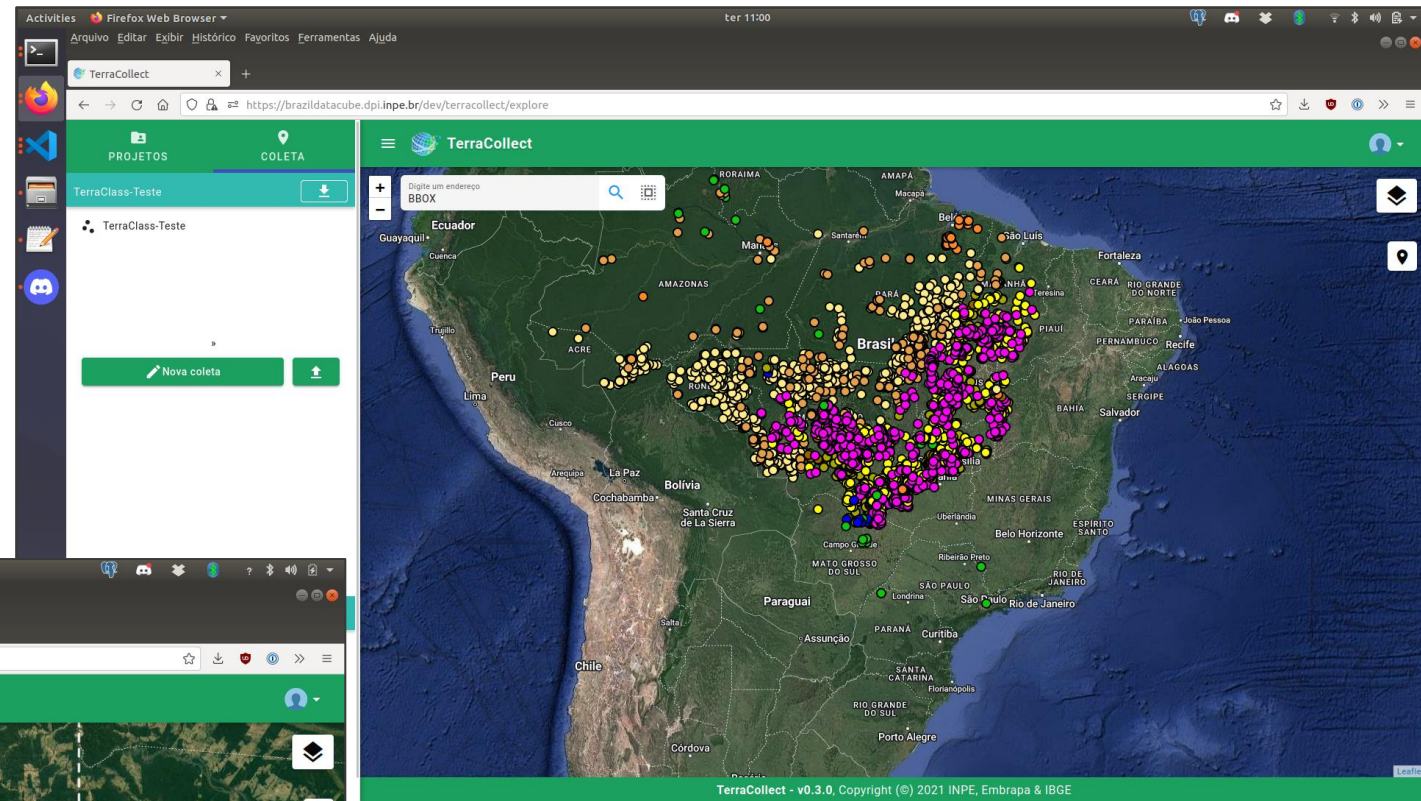
Source: [Lorena Santos et al., 2021]

Methods to assess and improve the quality of land use and cover samples

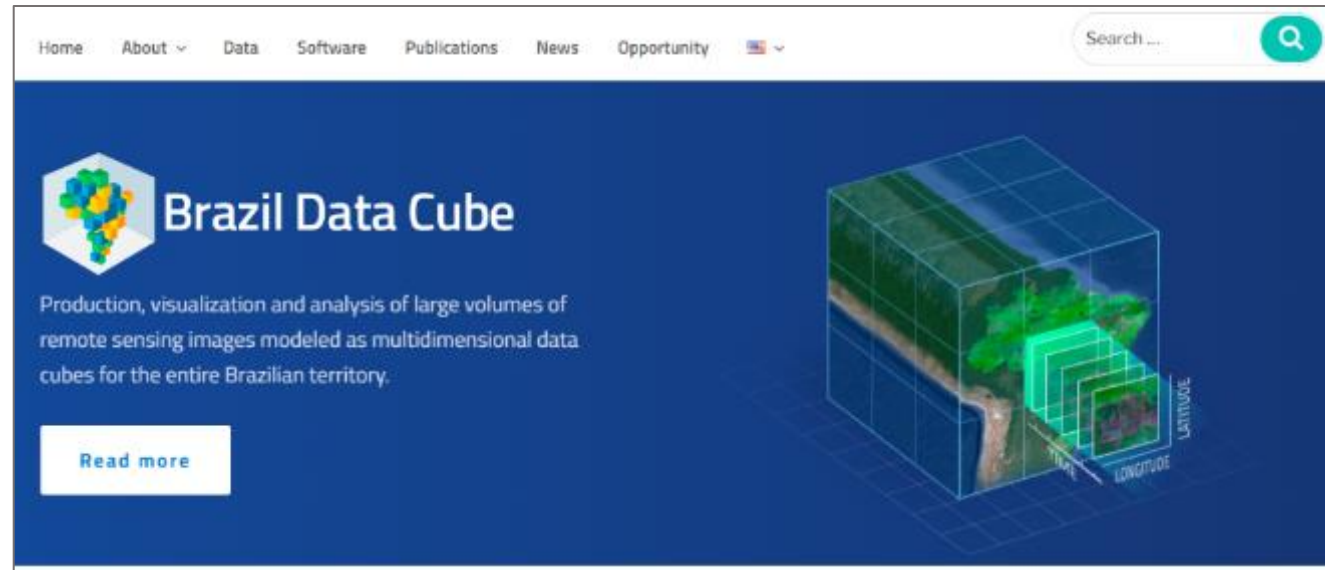


TerraCollect

web platform to collect and analyze land use and cover samples.



<http://brazildatacube.org>



Papers in
journals and
conferences:

2019: 9
2020: 14
2021: 12
2022: 10

<https://doi.org/10.3390/rs12244033>

remote sensing MDPI

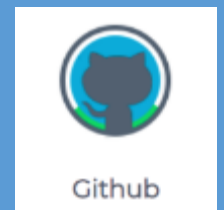
Article

Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products

Karine R. Ferreira ^{*}, Gilberto R. Queiroz, Lúbia Vinhas, Rennan F. B. Marujo, Rolf E. O. Simoes, Michelle C. A. Picoli, Gilberto Camara, Ricardo Cartaxo, Vitor C. F. Gomes, Lorena A. Santos, Alber H. Sanchez, Jeferson S. Arcanjo, José Guilherme Fronza, Carlos Alberto Noronha, Raphael W. Costa, Matheus C. Zaglia, Fabiana Zioti, Thales S. Korting, Anderson R. Soares, Michel E. D. Chaves and Leila M. G. Fonseca

<https://github.com/brazil-data-cube>

Software systems
and services: 57





BRAZIL
DATA CUBE

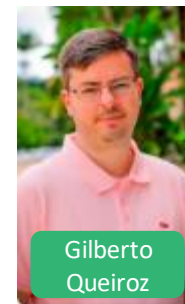
Software developers,
Associate researchers,
Master and PhD
students.



INPE Researchers



Karine
Ferreira



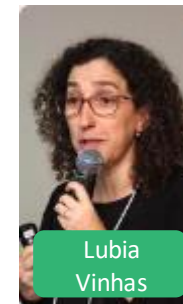
Gilberto
Queiroz



Claudio
Almeida



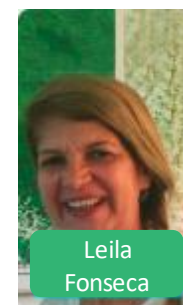
Gilberto
Camara



Lubia
Vinhas



Ieda
Del'Arco



Leila
Fonseca



Luis
Maurano



Ricardo
Cartaxo



Thales
Korting



BRAZIL DATA CUBE
brazildatacube.org



brazildatacube.org



brazildatacube.dpi.inpe.br/porta/explore

Karine R. Ferreira
Gilberto R. Queiroz

brazildatacube@inpe.br

OBRIGADO!

FUNPO
AMAZONIA

@brazildatacube



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÃO

