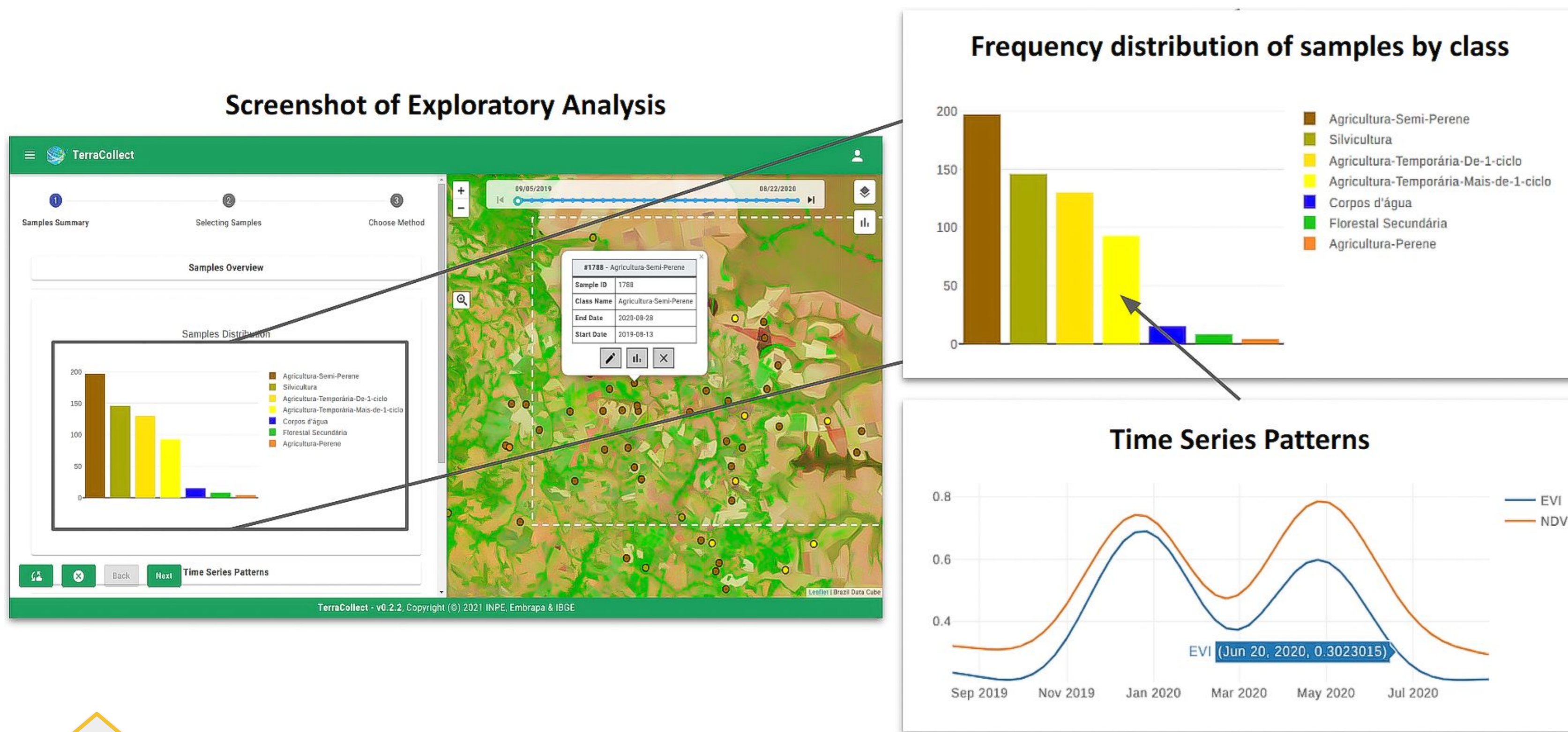
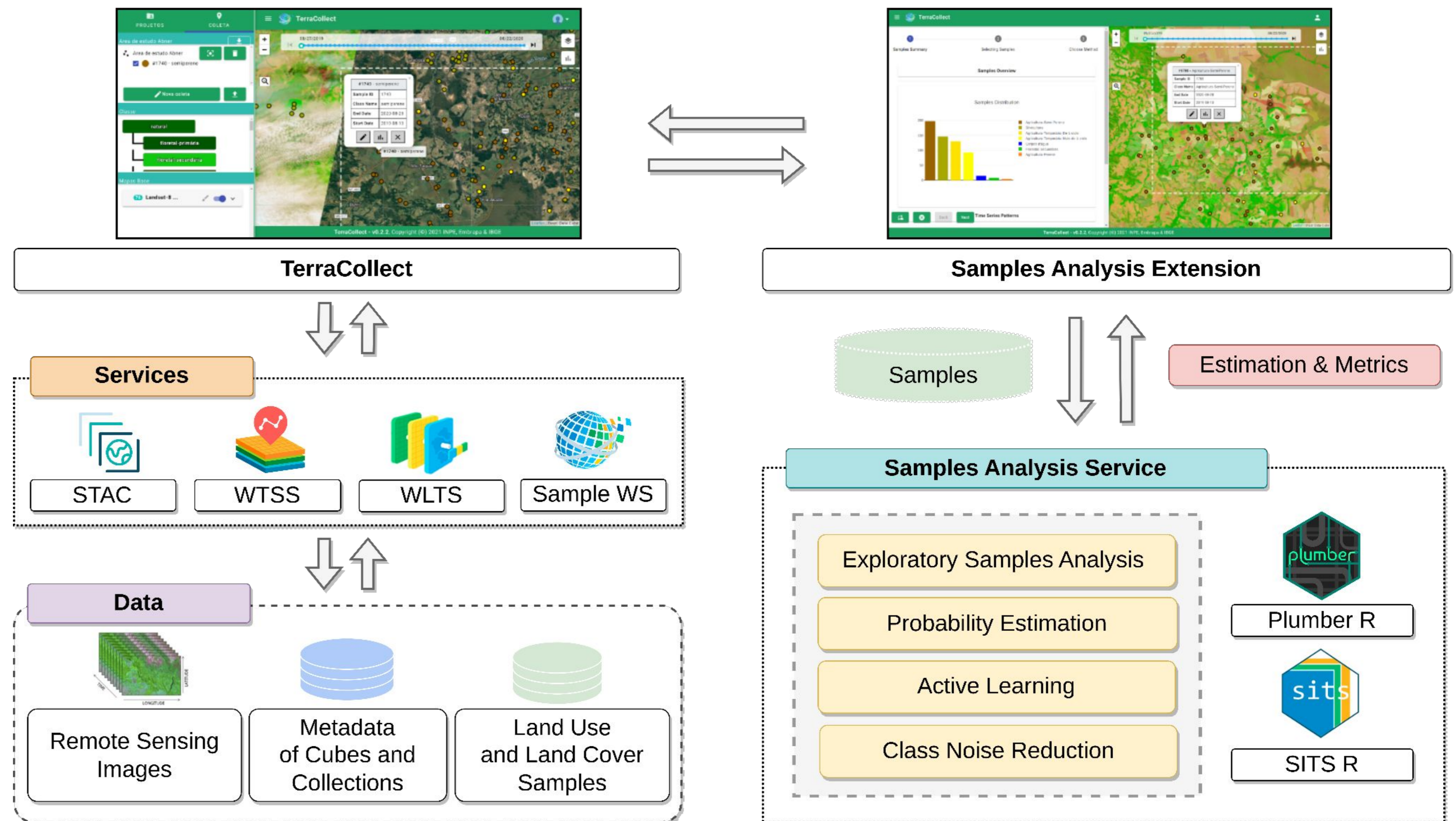


### Abstract

This project aims to provide auxiliary tools during sampling in the collecting system TerraCollect due to the lack of platforms for LULC data collection coupled with analytical tools. These tools are based on the integration of methods for **Exploratory Samples Analysis, Probability Estimation, Active Learning and Class Noise Reduction.**



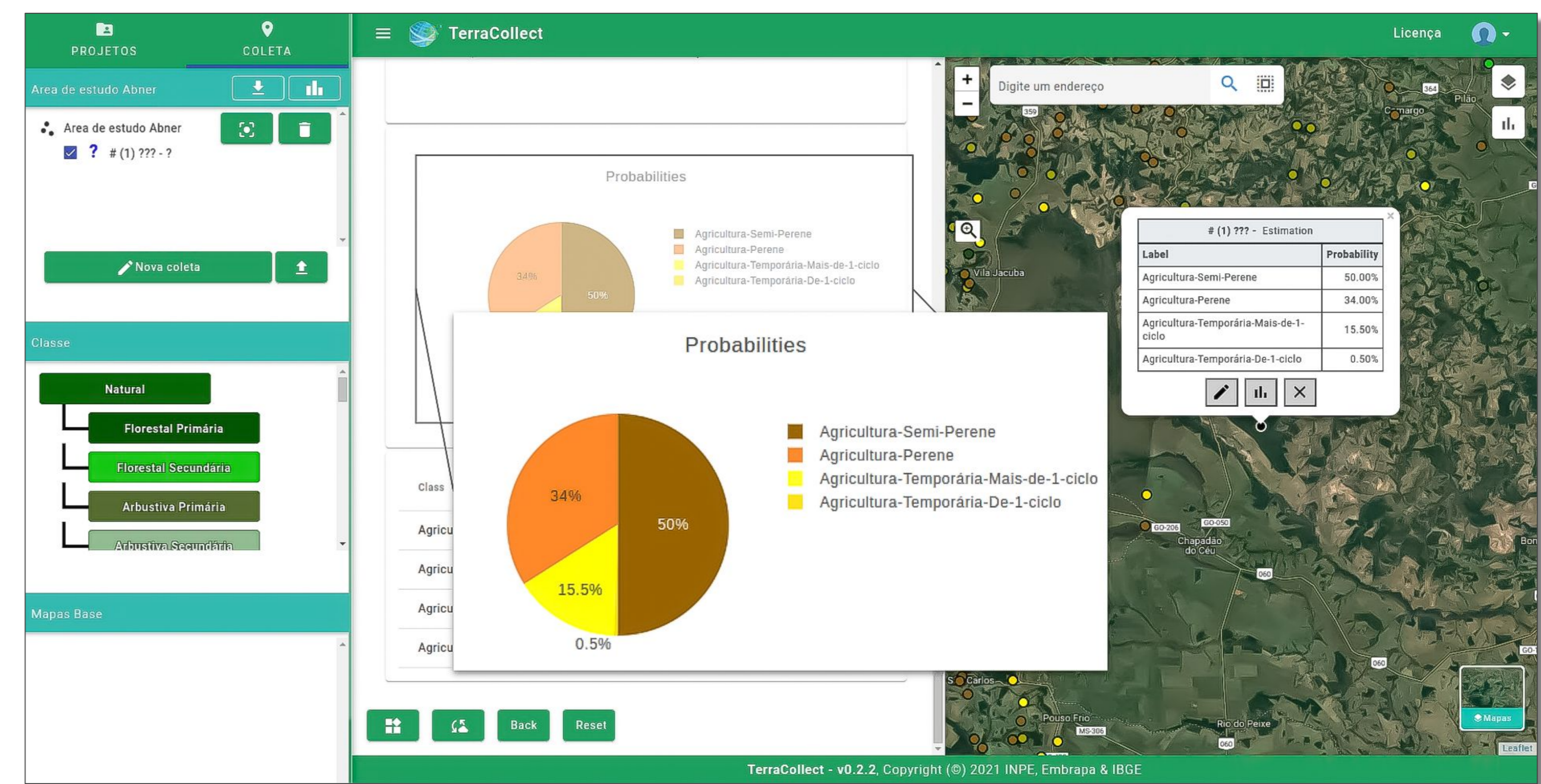
### Exploratory Samples Analysis

The component provides the visualization for histogram with the frequency distribution of samples by class and the visualization for time series patterns to summarize the dataset of available samples. Then, it is possible to identify anomalies and check out data, if this fits on the expectations or if it is in need of repairs.

### Probability Estimation

This method uses the samples in the TerraCollect database as training dataset for machine learning models to infer the probabilities of a new sample.

**Active Learning** component is derived from the same, however, the objective is to analyze divergences in model predictions.



### Class Noise Reduction

This component is based on method proposed by Santos et al. 2021 that aims to measure the probability of the label being correctly assigned using SOM Clustering. Then, it has the visualization of distribution of the samples tagged with clean, analysis and to remove.

