International Journal of Applied Earth Observations and Geoinformation 106 (2022) 102655

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

A platform for land use and land cover data integration and trajectory analysis

Fabiana Zioti, Karine R. Ferreira *, Gilberto R. Queiroz, Alana K. Neves, Felipe M. Carlos, Felipe C. Souza, Lorena A. Santos, Rolf E.O. Simoes

National Institute for Space Research (INPE), Av. dos Astronautas, 1625-1917 - Jardim da Granja, Sao Jose dos Campos 12227-010, SP, Brazil

ARTICLE INFO

Keywords:
Land use and land cover trajectory
Data integration and analysis

ABSTRACT

Information on land use and land cover (LULC) is essential to support governments in making decisions about the impact of human activities on the environment, planning the use of natural resources, conserving biodiversity, and monitoring climate change. Nowadays, different initiatives systematically produce information on LULC dynamics, on global, national, and regional scales. Examples of open and global LULC data products are Global Land-cover Classification with a Fine Classification System, Copernicus Global Land Service, and Global Land Cover by European Space Agency (ESA). At the national and regional level in Brazil, we can cite the data sets produced by PRODES, TerraClass, MapBiomas, and IBGE. Although these initiatives provide rich collections of open LULC maps, there is still a gap in tools that facilitate the integration of these data sets. The integrated analysis of these collections requires considerable effort by researchers who have to download, organize and harmonize them in their local computers, facing with different spatiotemporal resolutions and classification systems containing distinct class numbers, names and meanings. Besides that, these collections are distributed in different data formats through files or web services. To minimize these efforts, we propose a platform that allows users to access LULC collections from distinct sources, map their distinct classification systems, and retrieve LULC trajectories associated with spatial locations by integrating these collections. Besides the platform architecture description, this paper presents a case study that demonstrates its use in the integration and analysis.

1. Introduction

A consistent land use and land cover (LULC) mapping over time is essential to generate reliable environmental studies, such as the estimation of greenhouse gas emissions and the assessment of ecological services and food security (Brown, 2016). LULC maps are important data sources to measure global indicators of United Nations' Sustainable Development Goals (SDGs), including Indicator 15.3.1 on land degradation (Giuliani et al., 2020), Indicator 6.6.1. on freshwater ecosystems, or Indicators 11.3 and 11.7 on land use (Paganini et al., 2018).

LULC maps are mainly produced from remote sensing imagery (Saah et al., 2020). Using Earth observation (EO) satellite images, it is possible to monitor vast areas, locally and globally, especially in regions that are hard to reach (Gomez et al., 2016). Nowadays, different missions are launching EO satellites and producing big open data of images consistently over time and space (Soille et al., 2018). This scenario has promoted the production of many global, national and regional scale LULC maps based on EO satellite images.

On a global scale, initiatives that produce open LULC information include the Global Land-cover Classification with a Fine Classification System (GLC FCS30) (Zhang et al., 2021), Global Forest Change (Hansen et al., 2013), Global Land Cover Facility (GLCF) (Song et al., 2018), Copernicus Global Land Service (Buchhorn et al., 2020), Global surface water map (Pekel et al., 2016), Global Land Cover by ESA under the program Climate Change Initiative (CCI-LC) (Bontemps et al., 2013). GLC FCS30 is a global 30-meter LULC classification for 2015 with 30 classes, including grassland, wetland, water bodies as well as different types of cropland and forest. Global forest change provides annual 30-meter maps, from 2000 to 2019, of forest area, loss, and gain, based on Landsat data. Copernicus Global Land Service provides annual 100 m LULC maps, from 2015 to 2019, derived from PROBA-V satellite observations and ancillary data with 23 classes. Global surface water maps are annual classifications, from 1984 to 2020, of 30-meter based on Landsat satellite images. Challenges in producing high resolution

* Corresponding author.
E-mail address: karine.ferreira@inpe.br (K.R. Ferreira).

https://doi.org/10.1016/j.jag.2021.102655
Received 15 October 2021; Received in revised form 10 December 2021; Accepted 15 December 2021
Available online 18 December 2021

(around 30-meter) global LULC maps include unavailability of timely and accurate training and validation data and high-performance computing requirements (Giri et al., 2013).

In Brazil, different initiatives produce national and regional LULC maps from EO satellite images. These initiatives include PRODES (Brazilian Amazon Deforestation Monitoring Program), DETER (Near Real-Time Deforestation Detection System) (Diniz et al., 2015), TerraClass (ALMEIDA et al., 2016), MapBiomas (Souza et al., 2020) and LULC maps produced by the Brazilian Institute of Geography and Statistics (IBGE) (IBGE, 2020). These LULC data sets are described in Section 2.2.

Despite the effort of creating standardized legends based on global (e.g., LCCS - Land Cover Classification System by Food and Agriculture Organization) (Di Gregorio, 2005) or local (e.g., Technical Manual of the Brazilian Vegetation by IBGE) (IBGE, 2012) classification systems, each initiative follows its own methodology to produce LULC maps from EO satellite images. These maps have different spatial and temporal resolutions and classification systems with distinct class numbers, names, semantics, and meanings. Besides that, these maps are represented by distinct data types, raster or vector, and are served by different systems in several formats, such as files, database systems, or web services. According to (Di Gregorio et al., 2012), the semantic interoperability among LULC data is an important unsolved issue. The lack of harmonization among different classification systems hinders the integration and even the comparison of distinct LULC collections.

Some proposals to compare different LULC data products have been presented. (Congalton et al., 2014) compared and analyzed the uncertainty of four global LULC maps, IGBP DISCover, UMD Land Cover, Global Land Cover 2000, and GlobCover 2009. (Reinhart et al., 2021) provided the comparison of two LULC products, ESA CCI and CORINE for the Eastern Europe. For the Amazon biome in Brazil, (Neves et al., 2020) assessed the legend of two LULC products, TerraClass and MapBiomas, in terms of the LCCS system and compared them for five years. From global to regional scales, all initiatives reported the need for harmonization between two or more legends to perform the comparison and revealed inconsistencies regarding the classification schemes, especially related to mixed classes. No definition of land cover percentages is provided for them, increasing the spatial distribution uncertainty.

Another relevant aspect in comparing or integrating different LULC maps is the need to download and process all data locally, which requires considerable effort, computing power, and storage capacity. As pointed out by Chen (Chen et al., 2017), many spatial data web portals provide catalogs and Application Programming Interfaces (APIs) that allow users to access the LULC data. However, none of these alternatives provides a comprehensive solution to facilitate the integration and harmonization of LULC information with different semantic, spatial, and temporal components. To tackle this issue and facilitate the online access of LULC and other EO data products, some platforms have been released, such as Geo-Wiki (Fritz et al., 2012), EU-LULC WebGIS (Brovelli et al., 2016) and GEOCAB Portal (Desconnets et al., 2017). In the Geo-Wiki platform (Fritz et al., 2012), for instance, it is possible to visualize several global LULC products and other hybrid ones created to analyze the disagreements between them. But, this platform focuses on crowdsourcing and does not allow database manipulation to generate new reclassification or legend harmonization and to retrieve trajectory.

In short, although there are rich collections of open LULC data for global, national, or regional scales, the comparison and integrated use of these collections requires a great effort by researchers who have to download, organize and harmonize them in their local computers. In this paper, we propose a platform to minimize these efforts.

The main contribution of this work is a generic and simple software platform for LULC data integration and trajectory retrieval. It allows users to access LULC collections from distinct sources, map their distinct classification systems, and retrieve LULC trajectories associated with spatial locations by integrating these collections. The proposed platform is composed of three layers, (1) Data and metadata, (2) Web services, and (3) Clients. Besides the platform architecture description, this paper presents a case study that demonstrates its use in the integration and analysis of five different LULC collections in Brazil.

2. Methods

2.1. Study area

The Legal Amazon covers an area of approximately 5 million km^2, which encompasses almost 60% of the Brazilian territory. Over the last decades, this region has been facing high rates of deforestation. It is estimated that, by 2020, about 20.35% of its natural vegetation have already been converted to other land uses, such as agriculture and pasture (Almeida et al., 2021).

The selected study area is located in one of the five leading municipalities in deforestation in the Legal Amazon: São Félix do Xingu municipality, in the Pará state - Brazil. Until 2020, 23.60% of its primary forest had been deforested (Almeida et al., 2021). We created a regularly spaced grid of 1x1km, resulting in 4,472 points (Fig. 1). These points were used to retrieve the LULC trajectories in this region from 2000 to 2018, by integrating different LULC data sets, and to evaluate the agreement between these data sets.

2.2. Data sets

To validate the proposed platform, we used five different LULC data sets for Brazil: PRODES, DETER, TerraClass, MapBiomas, and IBGE. PRODES project has released annual deforestation data sets and rates for the Legal Amazon since 1988, assisting the government in decision making in public policies (Almeida et al., 2021). PRODES data sets have a spatial resolution of 30 m and are available in vector format from 2007 to 2020. The data for 2007 presents an accumulated amount of deforestation of previous years. Once a region is identified as Deforestation, it enters into the PRODES mask in the next year and is not analyzed again. DETER is an alert system that detects monthly disturbances in the Amazon forest through deforestation, degradation, or selective logging. PRODES and DETER data sets are available at TerraBrasilis platform (Assis et al., 2019).

TerraClass identifies the LULC classes that the deforested areas detected by PRODES have become (Almeida et al., 2016). It produces thematic maps with spatial resolution between 20 and 30 m. For the Amazon Biome, the LULC data products are available in the raster and vector formats, in biennial temporal resolution, with mappings for the years: 2004, 2008, 2010, 2012, and 2014. For Cerrado Biome, the products are available in raster formats for the 2013 and 2018 years. These data sets are available through the Terra Class portal.[1] In this work, we use the second version of TerraClass for Amazon Biome in the raster format.

MapBiomas produces annual LULC maps, using the concept of data collections, for the entire Brazilian territory from 1985 to 2019 (Souza et al., 2020). As of the date of this article, data collection 5 is available through its portal,[2] separately for each biome in Brazil in raster format. MapBiomas data sets have a maximum spatial resolution of 30 m.

The LULC maps provided by IBGE present information for the years 2000, 2010, 2012, 2014, 2016, and 2018. These data sets show the dynamics of occupation of the national territory, having cells of 1 km^2 as the basic spatial unit. LULC maps are available for all Brazilian states through the portal,[3] through files in vector format. Table 1 presents a description of the LULC data products used in this work, showing their features and variability.

[1] https://www.terraclass.gov.br/
[2] https://mapbiomas.org
[3] https://www.ibge.gov.br

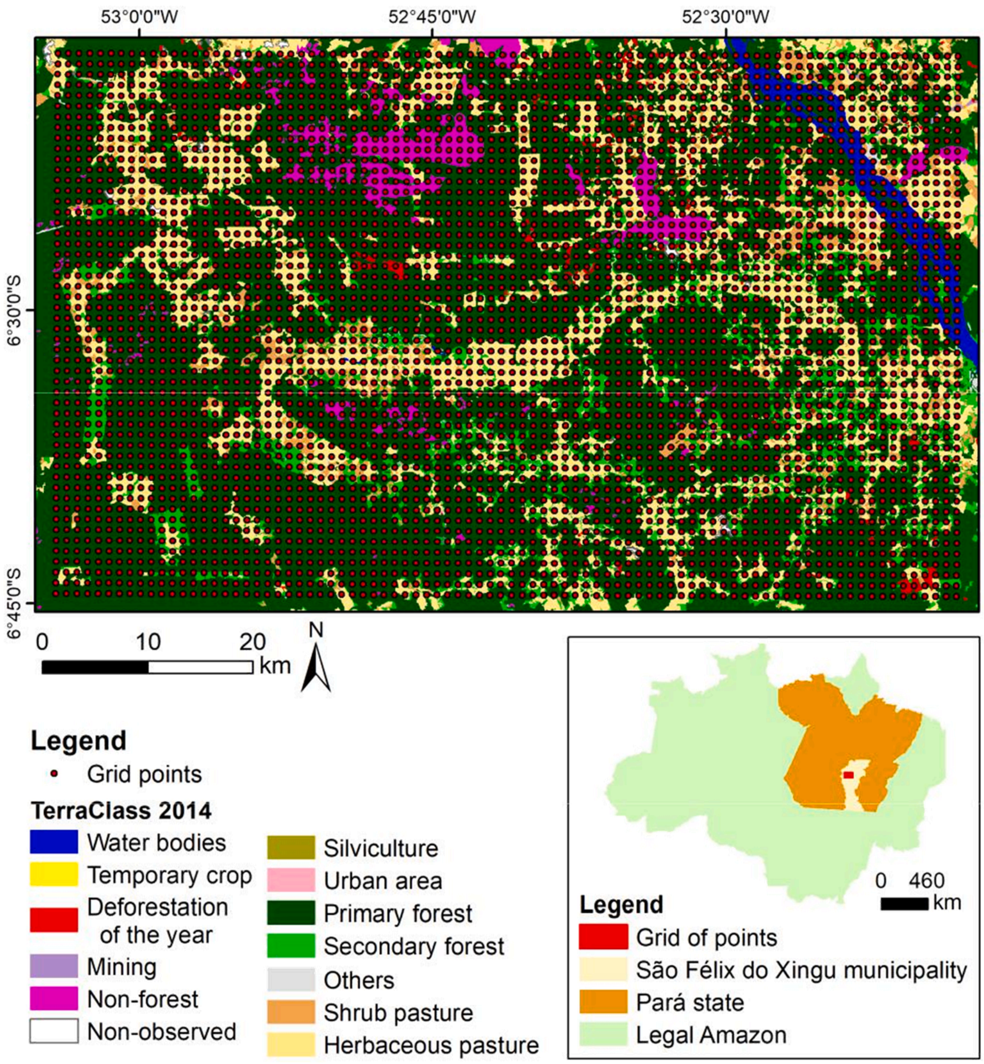

Fig. 1. Study area location and regularly spaced grid.

Table 1
LULC data sets information.

LULC Data set	Collection Name	Period	Temporal Resolution	Format
PRODES	prodes_amazonia_legal	2007–2020	Yearly	Vector
DETER	deter_amazonia_legal	2016–2020	Daily	Vector
TerraClass Amazonia	terraclass_amazonia-v2	2004, 2008, 2010, 2012, 2014	Biennial	Raster
MapBiomas	mapbiomas_amazonia-v5	1985–2019	Yearly	Raster
IBGE	ibge_cobertura_uso_terra	2000, 2010, 2012, 2014, 2016, 2018	Biennial	Vector

2.3. Platform architecture

The architecture of the proposed platform, shown in Fig. 2, is composed of three layers: (1) Data and Metadata, (2) Services and (3) Clients. It allows users to access LULC data sets from distinct sources, to map their distinct classification systems, and to retrieve LULC trajectories associated to spatial locations, by integrating their data sets.

The platform supports LULC data sets stored as Shapefile or GeoTif files; stored as PostGIS databases or provided as Web Feature Service (WFS) and Web Coverage Service (WCS). Therefore, if a LULC data set is available through WFS or WCS by its provider, the platform can access it without downloading it. All metadata about how LULC data products are stored or provided and about their properties are described as JSON files. Fig. 3 shows an example of a JSON file with metadata about LULC data sources. The types of data sources are webservice_source, files_sources, dbms_sources. Each type is described as a specific set of attributes needed to access the LULC data sets.

If users are interested in mapping different classification systems of LULC data sets, they can inform the metadata about hierarchical or non-hierarchical classification systems and their respective classes. The metadata about the classification systems and their mapping is stored as a PostGIS database, following the database model presented in Fig. 4. This database model represents the mapping with similarity measures between classes of distinct classification systems in order to simplify joint data analysis. Besides that, this model allows users to associate a proper visual style for each class that is considered in the rendering task.

The platform is also composed of two web services, Web Land Trajectory Service (WLTS) and the Web Land Classification System Service (WLCSS), Python client libraries, R client packages, and a QGIS plugin. All of them are described in the next sections.

2.4. LULC trajectory and web land trajectory service

WLTS is a web service designed to access, integrate and retrieve LULC trajectories from different data sources associated with specific

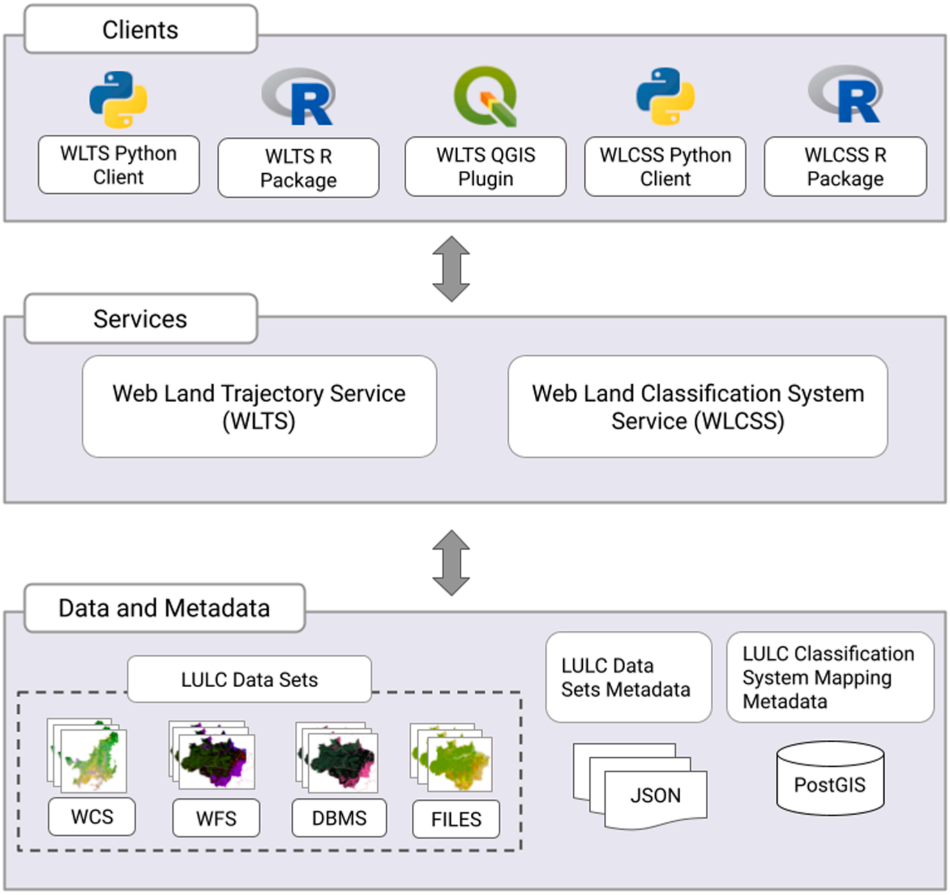

Fig. 2. Architecture of the platform for LULC data integration and trajectory analysis.

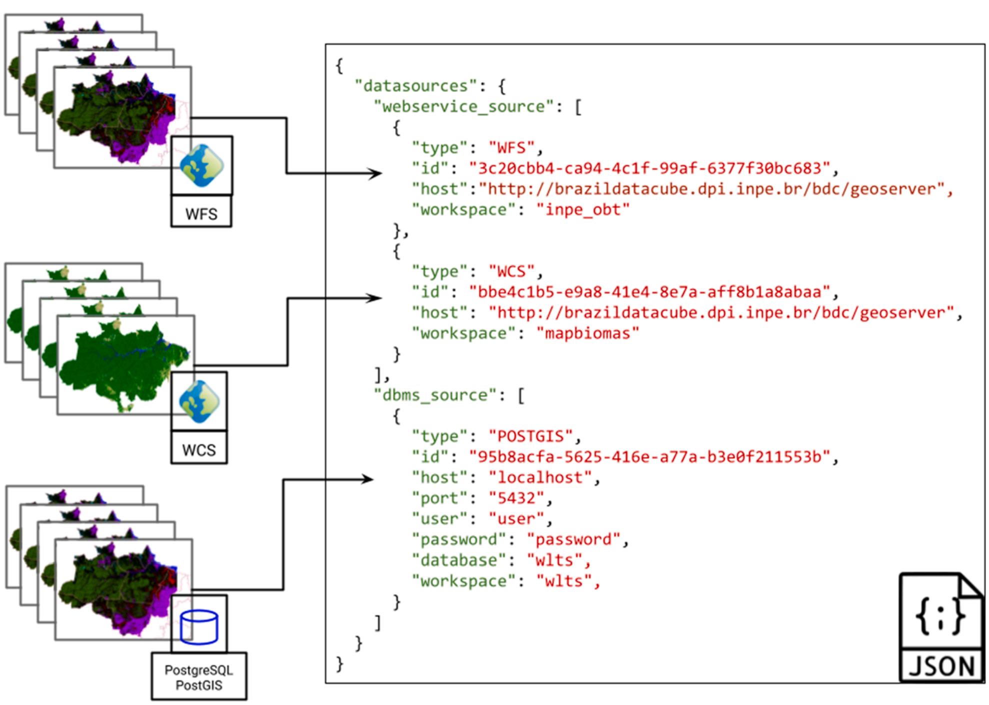

Fig. 3. Metadata about LULC data sources as JSON file.

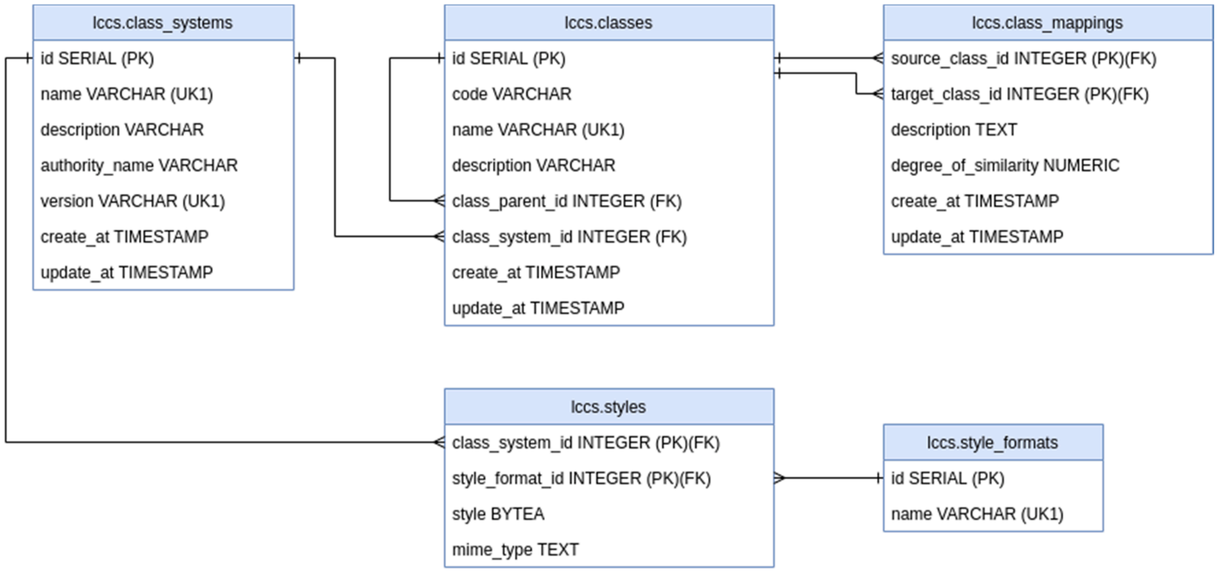

Fig. 4. Database model for metadata about classification systems of distinct LULC data sets and their mapping.

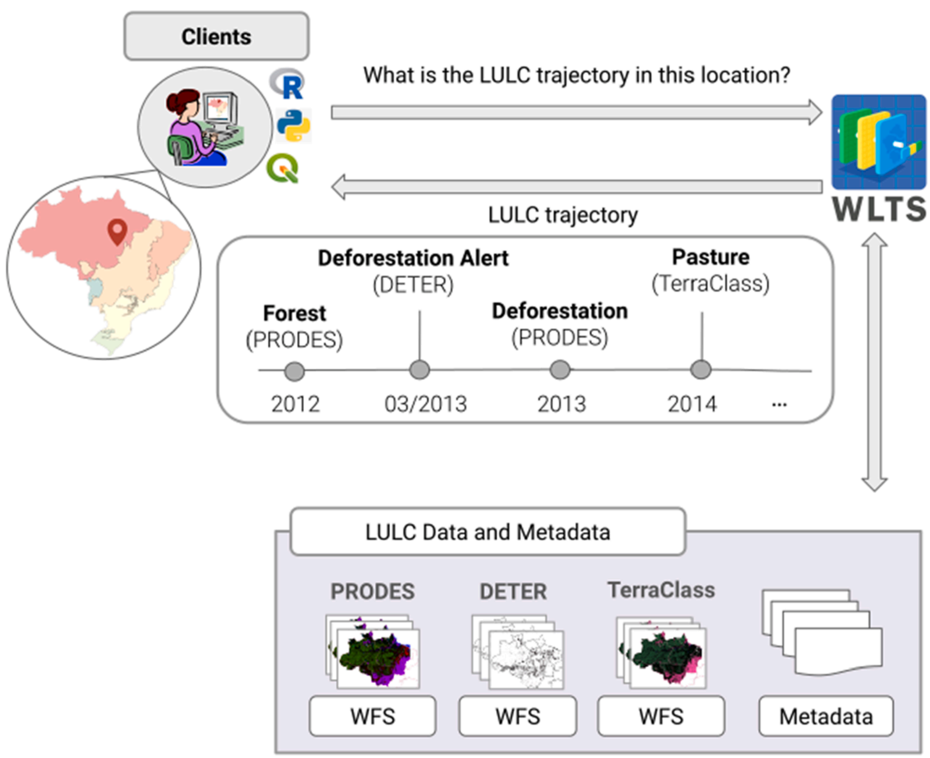

Fig. 5. Web Land Trajectory Service (WLTS) retrieving a LULC trajectory associated to a spatial location in the Amazonia biome, Brazil, integrating three LULC collections TerraClass, PRODES and DETER.

spatial locations, as illustrated in Fig. 5. Through a simple API, users indicate a spatial location and WLTS is able to stick different LULC data sets in that location and retrieve how it was classified over time by these data sets. The term LULC trajectory refers to a sequence of LULC classes, ordered in time, associated with a spatial location. Fig. 5 shows an example of a LULC trajectory associated with a spatial location in the Brazilian Amazonia biome, according to three different LULC data collections TerraClass, PRODES, and DETER. This location was classified as *Forest* in 2012 by PRODES; then as *Deforestation Alert* in 2013 by DETER; then by *Deforestation* in 2013 by PRODES; and by *Pasture* in 2014 by TerraClass.

Definition. LULC Trajectory (*Trj*): Given a set of n LULC Collections (*Cl*) represented by $Cl = \{Cl_{i=1}, \ldots, Cl_{i=n}\}$ where each collection Cl_i has a set of m LULC Classes (*C*) represented by $C = \{C_{i,j=1}, \ldots, C_{i,j=m}\}$, a LULC Trajectory (*Trj*) is represented by a set of observations $Trj = \{(t_{k=tmin}, Cl_i, C_{ij}), \ldots, (t_{k=tmax}, Cl_i, C_{ij})\}$ associated to a spatial location (x, y). Each observation (t_k, Cl_i, C_{ij}) contains the LULC class *Cij* of a LULC collection (Cl_i) in the location x, y at time t_k. Temporally, a trajectory is delimited by $(tmin, tmax)$ which are the minimum and maximum time instants associated to all LULC collections (*Cl*) and all observations of *Trj* satisfy $tmin < k < tmax$.

To access the LULC data collections from distinct sources, WLTS uses the JSON metadata file that describes how they are stored and provided, as illustrated in Fig. 3. It allows WLTS to be extensible. New LULC data sources can be added to this JSON metadata file and thus the WLTS can access them, without needing major source code changes. WLTS has a specification[4] that presents the complete description of its available operations. The WLTS API is composed of three operations:

- **list_collections**: returns a list with the LULC data collections names available by an instance of the service. The names returned by this operation are used in subsequent operations to identify the collections.
- **describe_collection**: retrieves metadata of a LULC data collection identified by its name. The metadata includes a description of the collection, its spatial and temporal extent, and its classification system.
- **trajectory**: returns a LULC trajectory of a given spatial location from the LULC data collections. The location is indicated by the latitude and longitude parameters in decimal degrees adopting the Datum WGS84 (EPSG 4326).

2.5. Web land classification system service

WLCSS is designed to represent and map classification systems of LULC data collections. Furthermore, it facilitates the access of applications, sym bology, or graphic styles associated with classification systems, such as Styled Layer Descriptor (SLD) or QGIS Style file. WLCSS provides a simple interface to use and query the metadata database, shown in Fig. 4, that represents different classification systems, hierarchical or non-hierarchical, and their respective classes. The WLCSS service has a specification[5] that presents a complete description of its available operations. Some operations supported by the WLCSS API are:

- **classification**_systems: returns the list of LULC classification systems available in the service. The classification systems returned by this operation can be used in subsequent operations.
- **classification_systems**/**sytem_id**: returns a document with full information about the classification system.
- **classification_systems/sytem_id**/classes: returns the list of classes from a classification system.
- **classification_systems/sytem_id/styles**: returns the list of styles formats available for the classification.
- **classification_systems/system_id_source/system_id_target**: returns the mapping between two classification systems.

WLCSS was designed considering the issues of data reusability. However, it does not aim to solve semantic interoperability problems. It provides an environment that allows researchers and specialists to describe classification systems associated with different LULC data collections and how they can be related. This helps users to integrate, harmonize and compare distinct LULC data collections.

2.6. Clients

WLTS and WLCSS are services that can be accessed through their API operations, described in Sections 2.4 and 2.5. To enable the use of these operations in analytical applications, we developed client libraries for R and Python languages. To use WLTS, the Python library wlts.py and the R package rwlts were developed. For WLCCS usage, the Python library lccs.py and the R package rlccs were implemented.

These libraries are integrated with other ones to facilitate analysis of the data sets returned by the services. For example, to allow quick manipulation of LULC trajectories, the wlts.py library retrieves data as Pandas DataFrame (McKinney et al., 2010), a high-level data structure for tabular data manipulation in Python. In rwlts, trajectories are returned as tibble DataFrame (Müller and Wickham, 2021) in R. Besides that, these libraries can be easily used in interactive computing environments such as Jupyter Notebooks for data visualization and interpretation.

These client libraries can be used as basis to develop other tools, such as the WLTS Plugin[6] for QGIS software. This plugin provides user-friendly graphical interfaces that allows researchers, students, and users in general to access the WLTS service in the geographical information system QGIS. Fig. 6 shows the WLTS Plugin for QGIS. Using its graphical interfaces, users can select the LULC data collections and the period to be considered in the trajectory retrieving. Then, when users select a spatial location of an active layer in QGIS, the LULC trajectory is returned. The resulting trajectory is displayed to the user in a new window as a table. The user can also enable the Get geometries option to return the geometries of the LULC data collections that intersect the given spatial location. Using this plugin, users can also export the returned trajectories in CSV or JSON files and save Python scripts.

3. Results

This section presents the results of using the proposed platform to integrate the LULC data sets in the study area described in Sections 2.1 and 2.2.

3.1. Retrieving LULC trajectories

The code in Listing. 9 shows the use of the wlts.py library and its method t_j, that uses the WLTS operation trajectory. The code retrieves a trajectory in a specific spatial location (latitude = −6.2829, longitude = −52.3106) by integrating the LULC data collections, PRODES, DETER, TerraClass, IBGE and MapBiomas. The resulting trajectory is presented in Table 2. It shows the variation of LULC classes over the time in that spatial location. It did not retrieve any class for the DETER data set, which means that this point had no alert (e.g., deforestation or degradation) in the analyzed period. We can note that before 2008 there was still forest in this location: Primary Forest Natural Vegetation in TerraClass, Forest formation in MapBiomas, and Forest vegetation in IBGE.

The resulting trajectory is presented in Table 2. It shows the variation of LULC classes over the time in that spatial location. It did not retrieve any class for the DETER data set, which means that this point had no alert (e.g., deforestation or degradation) in the analyzed period. We can note that before 2008 there was still forest in this location: Primary Forest Natural Vegetation in TerraClass, Forest formation in MapBiomas, and Forest vegetation in IBGE.

In 2008, the PRODES project detected that this point had been deforested and, in this same year, the TerraClass project classified it as Deforestation of the year. This class represents areas that have undergone removal of primary forest that year but have not yet received a new use (e.g., pasture or agriculture). This was the only year in which TerraClass disagreed with MapBiomas, which classified this point as Pasture. From 2010 onwards, the three projects agreed on the class, since it was classified as Shrub and Herbaceous Pasture by TerraClass, as Pasture by MapBiomas, and as Mosaic of uses in forest area by IBGE. This IBGE class represents a mixture of agriculture, pasture, and/or silviculture associated or not with forest remnants where, due to the spatial resolution (1 km), the mapping does not individualize the land uses.

We use the R package rwlts to retrieve all LULC trajectories of the spatial locations presented in the study area in Section 2.1 from the TerraClassAmazonia data. To improve the visualization of these trajectories, we created an Alluvial diagram using the plot function provided

[4] https://github.com/brazil-data-cube/wlts-spec
[5] https://github.com/brazil-data-cube/lccs-ws-spec
[6] https://github.com/brazil-data-cube/wlts-qgis

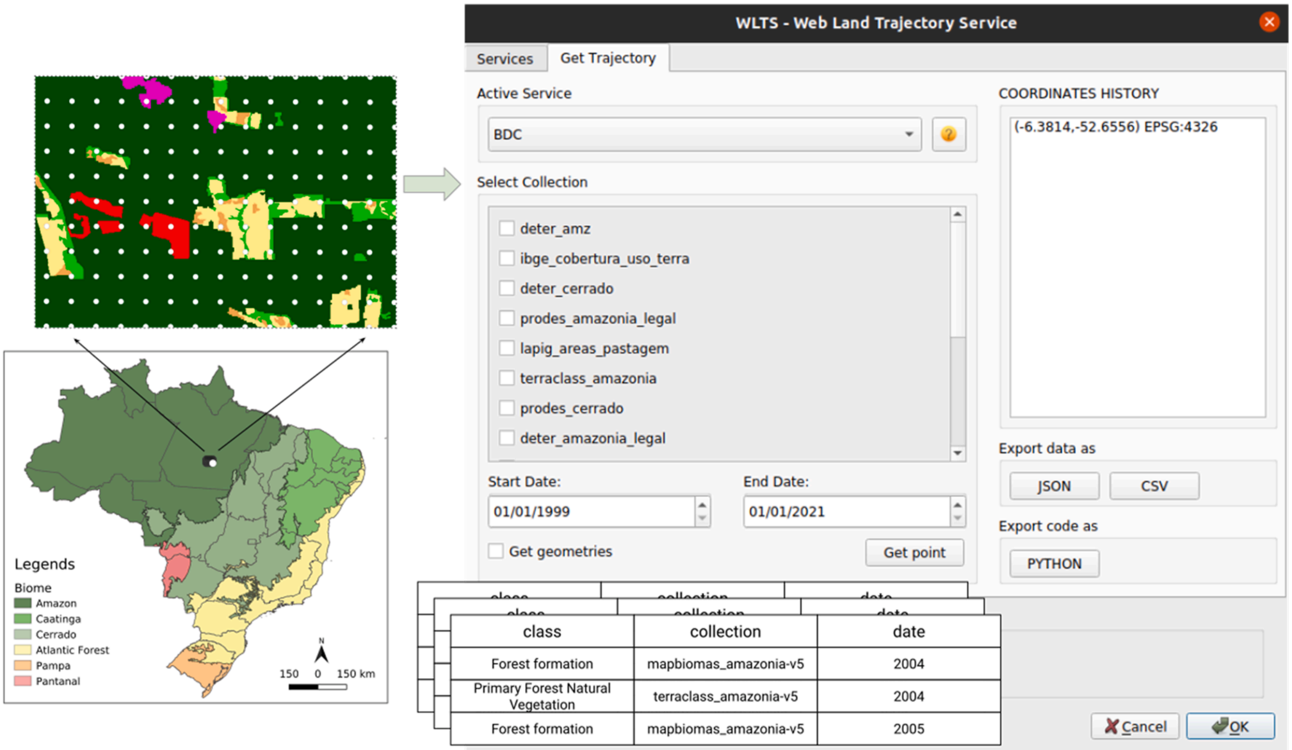

Fig. 6. WLTS Plugin for QGIS.

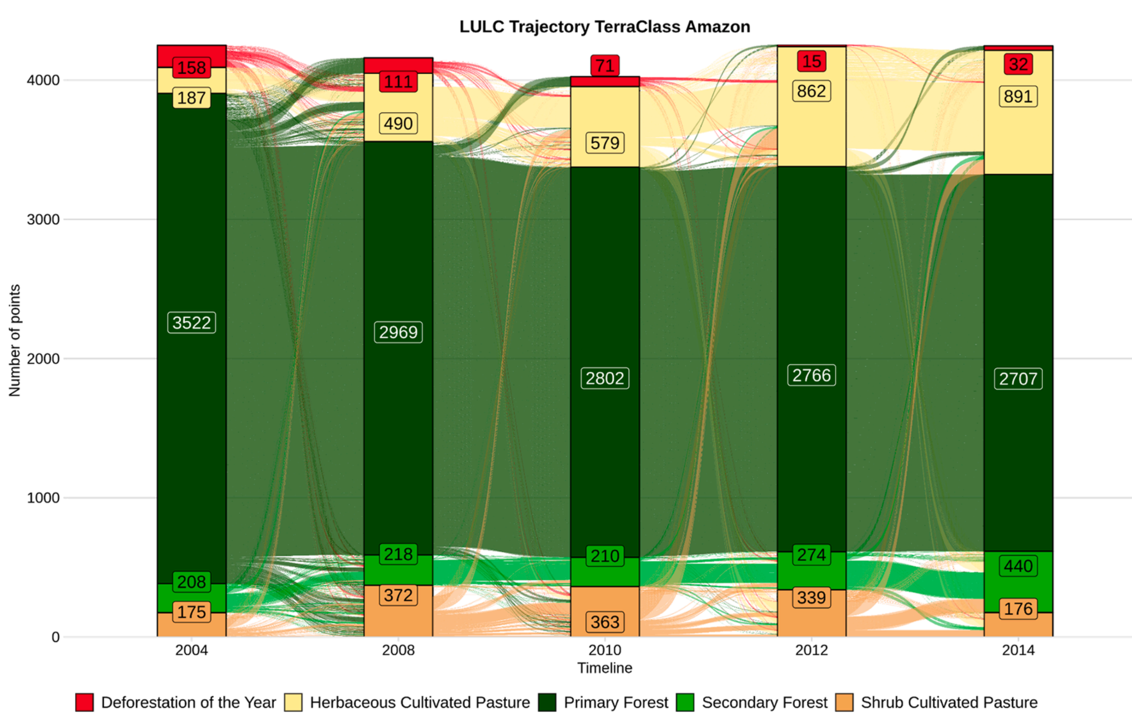

Fig. 7. Alluvial plot of LULC trajectories for TerraClass collection using rwlts.

by rwlts. The Alluvial diagram produced from all LULC trajectories of the 4,472 spatial locations is presented in Fig. 7. It shows LULC trajectories of five classes (Deforestation of the year, Shrub Pasture, Herbaceous pasture, Primary forest, and Secondary forest) of the TerraClass Amazonia collection through the analyzed years.

The Alluvial diagram enables clear visualization of the Primary forest reduction while Herbaceous and Shrub pasture considerably increased in the period. Regarding all of these points (4,472), 3,522 were Forest in 2004, representing 78.76%. In 2014, this number dropped to 2,707 (60.53%). In this last studied year, 1,081 (24.17%) points represented anthropic classes: Deforestation of the year, Mining, Urban area, Herbaceous pasture, Shrub pasture, or Others. From these points with anthropic classes in 2014, 589 were still Forest in 2004, which means that more than half (54.49%) of the anthropic uses in 2014 derived from

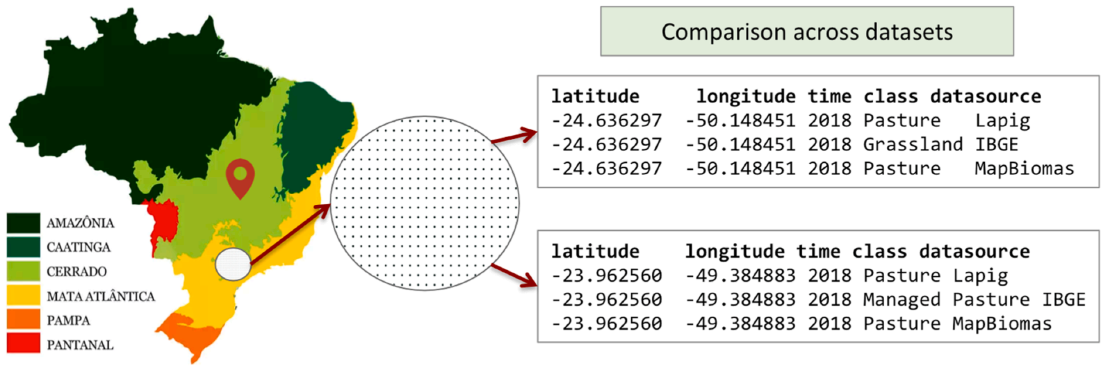

Fig. 8. Extracting training samples based on agreements among different LULC data collections.

```python
import wlts

# Defines the URL of an instance of the WLTS
wlts_url = 'https://brazildatacube.dpi.inpe.br/wlts'
service = wlts.WLTS(wlts_url, access_token='change-me')

# Defines the collections to be consulted in the service
c = "prodes_amazonia_legal,deter_amazonia_legal," \
        "terraclass_amazonia,mapbiomas_amazonia-v5," \
        "ibge_cobertura_uso_terra"

# Retrieving the trajectory
point_tj = service.tj(latitude=-6.2829, longitude=-52.3106, collections=c)

#Visualizing the trajectory with Pandas
    point_tj.df()
```

Listing 9. Retrieving LULC trajectory of one point from multiple collections (TerraClass Amazon, PRODES Amazon, DETER, IBGE and MapBiomas) using wlts Python package.

Table 2
LULC Trajectory in the location of latitude −6.2829 and longitude −52.3106 (EPSG:4326) retrieved by tj method.

Date	Class	Collection
2000	Forest formation	mapbiomas_amazonia-v5
2000	Forest vegetation	ibge_cobertura_uso_terra
2004	Primary Forest Natural Vegetation	terraclass_amazonia
2004	Forest formation	mapbiomas_amazonia-v5
2008	Deforestation	prodes_amazonia_legal
2008	Desforestation of the year	terraclass_amazonia
2008	Pasture	mapbiomas_amazonia-v5
2010	Shrub pasture	terraclass_amazonia
2010	Pasture	mapbiomas_amazonia-v5
2010	Mosaic of uses in forest area	ibge_cobertura_uso_terra
2012	Herbaceous pasture	terraclass_amazonia
2012	Pasture	mapbiomas_amazonia-v5
2012	Mosaic of uses in forest area	ibge_cobertura_uso_terra
2014	Herbaceous pasture	terraclass_amazonia
2014	Pasture	mapbiomas_amazonia-v5
2014	Mosaic of uses in forest area	ibge_cobertura_uso_terra
2016	Pasture	mapbiomas_amazonia-v5
2016	Mosaic of uses in forest area	ibge_cobertura_uso_terra
2018	Pasture	mapbiomas_amazonia-v5
2018	Mosaic of uses in forest area	ibge_cobertura_uso_terra

Primary Forest Natural Vegetation.

3.2. Comparing LULC data sets

In most cases, to compare different LULC data sets, it is necessary to harmonize their distinct classification systems and LULC classes. WLCSS provides methods to register and map LULC classes from different classification systems.

Table 3
Classes harmonization - reclassification of TerraClass Amazon and MapBiomas legends for a simplified legend. NF corresponds to Non Forest.

Classes harmonization		
TerraClass Amazon	MapBiomas	Simplified legend
Primary forest	Forest formation	Forest
Secondary forest		
–	Mangrove	
Silviculture	Forest plantation	Silviculture
Shrub pasture	Pasture	Pasture
Herbaceous pasture		
Perennial crop	Perennial Crop	Agriculture
Semiperennial crop	Sugarcane	
	Other temporary crop	
Temporary crop	Annual crop	
	Soybean	
Mining	Mining	Mining
Urban area	Urban infrastructure	Urban area
Others	Beach and dune	Others
	Other non vegetated area	
	Salt flat	
Non observed	Non observed	Non observed
NF	Savanna formation	NF natural formation
	Wetland	
	Grassland formation	
	Other NF natural formation	
	Rocky outcrop	
Water bodies	Aquaculture	Water
	River, lake and ocean	
–	Mosaic of ag. and past.	Mosaic of ag. and past.
Defor. of the year	–	Defor. of the year

Table 3 shows the harmonization of the LULC classes of TerraClass and MapBiomas collections and their mapping to a simplified legend, considered here as our target classification system. The classes Deforestation of the year, from TerraClass, and Mosaic of Agriculture and Pasture, from MapBiomas, do not present a similar class in the other project legends. Thus, we kept the same names in the simplified legend.

The harmonization between the TerraClass and MapBiomas classes as well as their mapping to a simplified legend, showed in Table 3, were produced based on the previous paper by (Neves et al., 2020). We registered the new classes of the simplified legend in the WLCSS and then included the mapping between each TerraClass and MapBiomas class to its related class of the simplified legend. Each mapping in the WLCSS is done between a pair of classes.

To register a new classification system in WLCSS, with its classes and mappings, users can use the client libraries in Python or R. Users can describe the new classification system and the mapping among their classes in a JSON file and use the rwlccs or wlccs.py clients to read this file and include the information in the database whose model is shown in Fig. 4. Thus, users can inform it in a high-level way, without needing to know details about the database model.

WLTS and WLCCS work in an integrated way. Therefore, it is possible to return harmonized trajectories from LULC data collections, based on the mappings available in the WLCSS. To return LULC trajectory in a single classification system, it is necessary to inform the desired destination system. This is done in method tj using the target system as one of the parameters of this function, as shown in the code in Listing 10.

Table 4 presents the result of the harmonized LULC trajectory produced by the code in Listing 10 using the Python client wlts.py integrated with lccs.py. The classes shown in this trajectory are from the simplified legend whose mapping is presented in Table 3.

Based on the LULC trajectories of all 4,472 spatial locations shown in the study area in Fig. 1, we computed the agreement between TerraClass and Mapbiomas for 2014. Table 5 presents the agreement analysis, considering the simplified legend presented in Table 3. Considering the 4,472 points, the overall agreement was of 85.89%.

4. Discussion

This work addresses the need for mechanisms that help users to integrate existing LULC data sets and to analyze LULC trajectories and changes that occur over time. In the Amazon region, where is the study site, the characterization of LULC classes in deforested areas has been performed by national and international projects (Almeida et al., 2016; Mapbiomas, Mapbiomas General Handbook, 2021). However, there is still a lack of information regarding the trajectories followed by these classes. Some efforts have been made to tackle this issue (Gibbs et al., 2010; Rosan and Anderson, 2017; Assis et al., 2020) and all of them required the download of the evaluated LULC data and computational power to perform the analysis. In this paper, we demonstrate that the proposed platform and its services facilitate the access and manipulation of several years of LULC data collections in an easy way without having to download the data.

The results in Section 3 shows that the proposed platform is useful for the integration, harmonization, and comparison of different LULC data collections as well as for analysis of LULC trajectories. Besides that, it is very useful for extracting training samples based on agreements among different LULC data sets. (Simoes et al., 2021) used the platform to

Table 4
Trajectory in the location of latitude −6.282 and longitude −52.536 (EPSG:4326) retrieved by tj method considering the simplified legend presented in Table 3.

Trajectory harmonization		
Date	Class	Collection
2004	Forest	terraclass_amazonia-v2
2004	Forest	mapbiomas_amazonia-v5
2005	Forest	mapbiomas_amazonia-v5
2006	Forest	mapbiomas_amazonia-v5
2007	Forest	mapbiomas_amazonia-v5
2008	Deforestation of the year	terraclass_amazonia-v2
2008	Pasture	mapbiomas_amazonia-v5
2009	Pasture	mapbiomas_amazonia-v5
2010	Pasture	terraclass_amazonia-v5
2010	Pasture	mapbiomas_amazonia-v5
2011	Pasture	mapbiomas_amazonia-v5
2012	Pasture	terraclass_amazonia-v2
2012	Pasture	mapbiomas_amazonia-v5
2013	Pasture	mapbiomas_amazonia-v5
2014	Pasture	terraclass_amazonia-v5
2014	Pasture	mapbiomas_amazonia-v5

```
from wlts import WLTS
from lccs import LCCS

# Define Services
wlts_url = 'https://brazildatacube.dpi.inpe.br/wlts'
lccs_url = 'https://brazildatacube.dpi.inpe.br/lccs/'

service_wlts = WLTS(url=wlts_url, access_token='change-me')
service_lccs = LCCS(url=lccs_url, access_token='change-me')

# Find the classification system of collections
    print(service_wlts['terraclass_amazonia-v2'])
    print(service_wlts['mapbiomas_amazonia-v5'])

# The return from previous commands is used in the command below
# Show available classifications system mappings
    print(service_lccs.available_mappings('TerraClass-AMZ-2'))
    print(service_lccs.available_mappings('MapBiomas-5'))

# Get a trajectory with the Simplified-legend-TM
tj = service_wlts.tj(latitude=-6.282295, longitude=-52.53655,
    collections='terraclass_amazonia-v2,mapbiomas_amazonia-v5',
    target_system='Simplified-legend-TM-1')

    print(tj.df())
```

Listing 10. Retrieving the harmonize trajectory using wlts and lccs Python client packages.

Table 5
Agreement analysis between TerraClass (TC) and MapBiomas for 2014, considering 4472 points in the study area shown in Fig. 1. Overall agreement is 85.89% and the classes abbreviations are: Non observed (NO), Deforestation of the year (DY), Urban area (UA) and Non forest natural formation (NFNF)

		MapBiomas								
		Forest	Pasture	NO	Water	DY	Mining	Others	UA	NFNF
	Forest	2806	330	0	6	–	0	0	0	4
	Pasture	91	975	0	1	–	0	0	0	1
TC	NO	106	15	0	0	–	0	0	0	27
	Water	5	0	0	60	–	0	0	0	0
	DY	14	18	0	0	–	0	0	0	0
	Mining	0	3	0	1	–	0	0	0	0
	Others	2	5	0	1	–	0	0	0	0
	UA	0	1	0	0	–	0	0	0	0
	NFNF	0	0	0	0	–	0	0	0	0

extract LULC training samples for the Cerrado biome in Brazil for the year 2018. These samples were used to train a Convolutional Neural Network (CNN) model that classified Landsat data cubes and produced a LULC map for Cerrado (Simoes et al., 2021).

The Cerrado biome has a large latitude gradient occupying 22% of the Brazilian territory (2 million km^2). Since the Cerrado biome includes different soils and climate regimes, the training data acquisition becomes difficult for humans because of the necessity to obtain a large and high-quality data set of training data. As illustrated in Fig. 8, to obtain these training samples, (Simoes et al., 2021) performed a systematic sampling using a grid of 5 $\times$ 5 km for the entire Cerrado totaling 85,026 spatial locations. Using the platform, they extracted the LULC trajectories associated with these spatial locations considering the collections IBGE (IBGE, 2020), MapBiomas (Souza et al., 2020), and the pastureland maps provided by The Image Processing and Geoprocessing Laboratory (Lapig) (Parente et al., 2019). Thus, they selected only the most reliable spatial locations where there was no disagreement between the label classes provided by the three collections. For example, in the point ($-$24.636297, $-$50.148451), the labels provided by the three collections disagree. Therefore, the sample equivalent to this point is discarded, whereas the samples collected in point ($-$24.962560, $-$49.384883) were kept. In total 48,850 samples were in agreement, and they were used to classify the entire biome.

Besides the two web services WLTS and WLCCS, the platform provides clients for the most used language for data science, Python and R. It allows users to write high-level programming language scripts, taking advantage of the analytical power of R and Python environments. Through simple scripts, users are able to access LULC trajectories from different data collections, without needing to download and organize these collections on their own computers.

5. Conclusion

This paper describes the architecture of a platform for LULC data integration and trajectory analysis and its use in a region in the Amazon biome in Brazil. We have an instance of this platform running in the Brazil Data Cube (BDC) project at the National Institute for Space Research (INPE) with national and regional LULC data collections for Brazil. This platform instance is public and can be used by anyone. It has been used by the BDC project team mainly to extract LULC training samples based on the agreements among different collections and to validate LULC maps produced from Earth observation data cubes using machine learning and image time series analysis.

The platform is simple, generic and its code is open source. Thus, anyone from other institutions or countries can use it and create an instance with existing LULC collections of areas or regions of interest. In the BDC project instance, we have only national and regional LULC data collections for Brazil. Therefore, we intend to include global data sets soon.

To replicate the platform installation in other institution, it is necessary to describe the LULC data sets in JSON files according to the proposed metadata model. This metadata model includes information about how these data sets are stored or provided (PostGIS database or WFS and WCS services) and about their properties, as presented in Fig. 3. Besides that, metadata about the hierarchical or non-hierarchical classification systems of the LULC data sets and their mappings must be stored in a PostGIS database, according to the data model shown in Fig. 4. Based on the metadata stored in the JSON files and PostGIS database, WLTS and WLCSS services as well as their clients are able to run. Step-by-step instructions for building and running the WLTS and WLCSS services and their clients in Python and R, as well as examples of metadata descriptions, can be found in a set of documents in Markdown format available on the project's Github.

To include a new LULC data collection in the platform, the manager has to edit an internal JSON file with metadata about how to access it and its properties. The platform reads this file and automatically considers the newly added collection. If this new collection is disseminated by its provider as WFS or WCS service, the platform accesses it through this service, without needing to download and maintain a copy of it internally. It allows the platform to be extensible in an easy way, without major code and data changes.

For future works, we aim to include the legend harmonization in terms of LCCS (Di Gregorio, 2005), which is an internationally accepted classification system that could facilitate the use or comparison of different LULC data collections. Additionally, we also intend to enable the inclusion of new data sets by the users and not only by the platform manager. To do that, we need to extend the platform with a user control system to maintain the users' context and preferences.

6. Code and data availability

The platform components are available in the Brazil Data Cube project organization on GitHub (https://github.com/brazil-data-cube) in the following repositories: (i) WLCCS Clients for R and Python; (ii) WLTS Clients for R and Python; (iii) WLTS service and the WLCCS service.

All components are licensed under the MIT License All material needed to reproduce the analyses presented in Section 3 is available as a Research Compendium in the repository: https://github.com/brazil-data-cube/wlts-paper. In this repository, there are: (i) Scripts for generating the analyses; (ii) LULC trajectory data sets; (iii) Dockerfile with the description of the computing environment; and (iv) Description of the execution workflow.

Funding

This research was supported by the Environmental Monitoring of Brazilian Biomes project (*Brazil Data Cube*), funded by the Amazon Fund through the financial collaboration of the Brazilian Development Bank (BNDES) and the Foundation for Science, Technology and Space Applications (FUNCATE), Process 17.2.0536.1. We also acknowledge support from "*Coordenação de Aperfeiçoamento de Pessoal de Nível*

Superior" - Brasil (CAPES) - Finance Code 001 and from the "*Conselho Nacional de Desenvolvimento Científico e Tecnológico*" - Brasil (CNPq) under grant 133573/2018-4 (F.Z.) and 130727/2020-2 (F.C.).

CRediT authorship contribution statement

Fabiana Zioti: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **Karine R. Ferreira:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft. **Gilberto R Queiroz:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft. **Alana K. Neves:** Validation, Writing – original draft, Writing – review & editing. **Felipe Carlos:** Visualization. **Felipe Souza:** Visualization. **Lorena Santos:** Validation. **Rolf Simoes:** Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Almeida, Cláudio Aparecido de, Coutinho, Alexandre Camargo, ESQUERDO, Júlio César Dalla Mora, Adami, Marcos, Venturieri, Adriano, Diniz, Cesar Guerreiro, Dessay, Nadine, Durieux, Laurent, Gomes, Alessandra Rodrigues, 2016. using Landsat-5/TM and MODIS data. Acta Amazon 46 (3), 291–302.

Almeida, C.A., Maurano, L.E.P., Valeriano, D.M., Camara, G., Vinhas, L., G.A. R, Monteiro, A.M.V., Souza, A.A.A., Renno, C.D., Silva, D.E., Adami, M., Escada, M.I.S., Motta, M., Amaral, S., 2021. Methodology for forest monitoring used in PRODES and DETER projects, Available at: http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2021/01.25.19.14/doc/publicacao.pdf, acess: September 21, 2021.

Assis, Talita Oliveira, de Aguiar, Ana Paula Dutra, von Randow, Celso, Melo de Paula Gomes, Diego, Kury, Juliana Nunes, Ometto, Jean Pierre H B, Nobre, Carlos A, 2020. Co_2 emissions from forest degradation in Brazilian amazon. Environ. Res. Lett. 15 (10), 104035. https://doi.org/10.1088/1748-9326/ab9cfc.

Assis, L.F.F.G., Ferreira, K.R., Vinhas, L., Maurano, L., Almeida, C., Carvalho, A., Rodrigues, J., Maciel, A., Camargo, C., 2019. TerraBrasilis: A spatial data analytics infrastructure for large-scale thematic mapping. ISPRS Int. J. Geo-Inf. 8 (11), 513.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., et al.,2013. Consistent global land cover maps for climate modelling communities: current achievements of the esa's land cover cci. In: Proceedings of the ESA living planet symposium, Edimburgh, pp. 9–13.

Brovelli, M.A., Fahl, F.C., Minghini, M., Molinari, M.E., et al., 2016. Land user and land cover maps of Europe: a webgis platform. Int. Arch. Photogramm. XLI-B 7, 913–917.

Brown, Molly E., 2016. Remote sensing technology and land use analysis in food security assessment. J. Land Use Sci. 11 (6), 623–641.

Buchhorn, M., Smets, B., Bertels, L., Roo, B.D., Lesiv, M., Tsend bazar, N.-E., Herold, M., Fritz, S., 2020. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2018: Globe (Sep. 2020). doi: 10.5281/zenodo.3518038. URL https://doi.org/10.5281/zenodo.3518038.

Chen, Jun, Li, Songnian, Wu, Hao, Chen, Xianjun, 2017. Towards a collaborative global land cover information service. Int. J. Digit Earth 10 (4), 356–370.

Congalton, R.G., Gu, J., Yadav, K., Thenkabail, P., Ozdogan, M., 2014. Global land cover mapping: A review and uncertainty analysis. Remote Sens. Basel 6 (12), 12070–12093.

Desconnets, J.-C., Giuliani, G., Guigoz, Y., Lacroix, P., Mlisa, A., Noort, M., Ray, N., Searby, N.D., 2017. Geocab portal: A gateway for discovering and accessing capacity building resources in earth observation. Int. J. Appl. Earth Obs 54, 95–104.

Di Gregorio, A., 2005. Land cover classification system: classification concepts and user manual: LCCS. Food Agric. Org. 2.

Di Gregorio, A., O'Brien, D., 2012. Overview of land-cover classifications and their interoperability. In: Giri, C.P. (Ed.), Remote Sensing of Land Use and Land Cover: Principles and Applications. CRC Press, pp. 37–47.

Diniz, C.G., de Almeida Souza, A.A., Santos, D.C., Dias, M.C., da Luz, N.C., de Moraes, D.R.V., Maia, J.S.A., Gomes, A.R., da Silva Narvaes, I., Valeriano, D.M., Maurano, L.E.P., Adami, M., 2015. DETER-b: The new amazon near real-time deforestation detection system. IEEE J. Sel. Top Appl. 8 (7), 3619–3628.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., Van der Velde, M., Kraxner, F., Obersteiner, M., 2012. Geo-wiki: An online platform for improving global land cover. Environ. Modell. Softw. 31, 110–123.

Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proc. Natl. Acad. Sci. USA 107 (38) (2010) 16732–16737.

Giri, C., Pengra, B., Long, J., Loveland, T.R., 2013. Next generation of global land cover characterization, mapping, and monitoring. Int. J. Appl. Earth Obs. 25, 30–37.

Giuliani, Gregory, Mazzetti, Paolo, Santoro, Mattia, Nativi, Stefano, Van Bemmelen, Joost, Colangeli, Guido, Lehmann, Anthony, 2020. Knowledge generation using satellite earth observations to support sustainable development goals (sdg): A use case on land degradation. Int. J. Appl. Earth Obs. 88, 102068. https://doi.org/10.1016/j.jag.2020.102068.

Gomez, Cristina, White, Joanne, Wulder, Michael, 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing 116, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S., Goetz, S.J., Loveland, T.R., et al., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.

IBGE, 2012. Manual técnico da vegetação brasileira, Available in: https://www.terrabrasilis.org.br/ecotecadigital/pdf/manual-tecnico-da-vegetacao-brasileira.pdf, access on: 20 sept 2021.

IBGE, 2020. Monitoramento da cobertura e uso da terra do brasil: 201-2018, Available in: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101703.pdf, access on: 15 sept 2021 (2020).

MapBiomas, Mapbiomas general handbook. algorithm theoretical basis document (atbd). collection 4, Available in: "https://mapbiomas-br-site.s3.amazonaws.com/ATBD_Collection_4_v2_Dez2019.pdf", access on: 15 sept 2021.

McKinney, Wes, 2010. Data structures for statistical computing in python. In: van der Walt, Stéfan, Millman, Jarrod (Eds.), Proceedings of the 9th Python in Science Conference, pp. 56–61. doi:10.25080/Majora-92bf1922-00a.

Müller, K., Wickham, H., 2021. tibble: Simple Data Frames, r package version 3.1.4 (2021). URL https://CRAN.R-project.org/package=tibble.

NEVES, Alana Kasahara, KÖRTING, Thales Sehn, FONSECA, Leila Maria Garcia, ESCADA, Maria Isabel Sobral, 2020. Assessment of terraclass and mapbiomas data on legend and map agreement for the Brazilian amazon biome. Acta Amazon 50 (2), 170–182.

Paganini, M., Petiteville, I., Ward, S., Dyke, G., Steventon, M., Harry, J., Kerblat, F., 2018. Satellite earth observations in support of the sustainable development goals, The CEOS Earth Observation Handbook.

Parente, L., Mesquita, V., Miziara, F., Baumann, L., Ferreira, L., 2019. Assessing the pasturelands and livestock dynamics in Brazil, from 1985 to 2017: A novel approach based on high spatial resolution imagery and Google Earth Engine cloud computing. Remote Sens. Environ. 232, 111301.

Pekel, Jean-François, Cottam, Andrew, Gorelick, Noel, Belward, Alan S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540 (7633), 418–422.

Reinhart, V., Fonte, C.C., Hoffmann, P., Bechtel, B., Rechid, D., Böhner, J., 2021. Comparison of ESA climate change initiative land cover to CORINE land cover over Eastern Europe and the Baltic States from a regional climate modeling perspective. Int. J. Appl. Earth Obs. 94, 102221.

Rosan, T.M., Anderson, L.O., 2017. Land cover change trajectories in western Amazonia. In: Anais..., XVIII Simpósio Brasileiro de Sensoriamento Remoto—SBSR, Santos - SP, 2017, pp. 4290–4297.

Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., San Aung, K., Markert, K. N., Clinton, N., Anderson, E.R., Cutter, P., et al., 2020. Primitives as building blocks for constructing land cover maps. Int. J. Appl. Earth Obs. 85, 101979.

Simoes, Rolf, Camara, Gilberto, Queiroz, Gilberto, Souza, Felipe, Andrade, Pedro R., Santos, Lorena, Carvalho, Alexandre, Ferreira, Karine, 2021. Satellite image time series analysis for big earth observation data. Remote Sens.-Basel 13 (13), 2428. https://doi.org/10.3390/rs13132428.

Soille, P., Burger, A., De Marchi, D., Kempeneers, P., Rodriguez, D., Syrris, V., Vasilev, V., 2018. A versatile data-intensive computing platform for information retrieval from big geospatial data. Future Gener. Comp. Sy. 81, 30–40.

Song, Xiao-Peng, Hansen, Matthew C., Stehman, Stephen V., Potapov, Peter V., Tyukavina, Alexandra, Vermote, Eric F., Townshend, John R., 2018. Global land change from 1982 to 2016. Nature 560 (7720), 639–643.

Souza, C., Zanin Shimbo, J., Rosa, M., Parente, L., Alencar, A., Rudorff, B., Hasenack, H., Matsumoto, M., Ferreira, L., Souza-Filho, P., Oliveira, S., Rocha, W., Fonseca, A., Balzani, C., Diniz, C., Costa, D., Monteiro, D., Rosa, E., Vélez-Martin, E., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in Brazilian biomes with landsat archive and Earth engine. Remote Sens.-Basel 12.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021. Glc fcs30: global land- cover product with fine classification system at 30 m using time-series landsat imagery. Earth Syst. Sci. Data 13 (6), 2753–2776.