TERRABRASILIS: A SPATIAL DATA INFRASTRUCTURE FOR DISSEMINATING DEFORESTATION DATA FROM BRAZIL

Luiz Fernando Ferreira Gomes de Assis[1], *Karine Reis Ferreira.*[1], *Lúbia Vinhas*[1], *Luis Maurano*[1], *Cláudio Aparecido de Almeida*[1], *Jether Rodrigues Nascimento*[1], *André Fernandes Araújo de Carvalho*[1], *Claudinei Camargo*[1], *Adeline Marinho Maciel*[1]

[1] Image Processing Division, National Institute for Space Research (INPE),
Astronautas Avenue 1758, São José dos Campos, São Paulo, 12227-001, Brazil
{luizffga, jetherrodrigues, claudineicam}@gmail.com, {karine.ferreira, lubia.vinhas, luis.maurano, claudio.almeida, andre.carvalho, adeline.maciel}@inpe.br

ABSTRACT

DETER and PRODES projects have been pioneerers at monitoring large deforestation mapping areas in Brazil. They have proved with quality-assured data their commitment to bridge the gap between society and better management of resources. Such environmental programs complexity, however, requires the designing, implementation and deployment of a spatial data infrastructure able to easily disseminate those data. TerraBrasilis thus plays an important role by supporting storage and query web services using modularized and virtualized environments to enable high availability and performance. The key idea behind TerraBrasilis is to facilitate the access and analysis of deforestation data generated by PRODES and DETER, and perhaps further thematic mapping projects. In this paper, we demonstrated TerraBrasilis within GIS and analytics environments with deforestation data of the Cerrado biome.

Key words *– Spatial Data Infrastructure, Deforestation Data, PRODES, DETER, Cerrado Biome*

1. INTRODUCTION

Internationally acclaimed projects such as the projects of deforestation monitoring (*PRODES*) and near-real time deforestation detection (*DETER*), conducted by the Brazilian National Institute for Space Research (INPE) have provided quality-assured, up-to-date, and spatially extensive official deforestation data. They together have improved the socio-economic and environmental processes linked to the implementation of public policies for the sustainable management of natural resources in Brazil [1–3]. Based on GIScience techniques and satellite imagery analysis of different resolutions, interpreters monitor critical areas to conceive annual clear-cut deforestation rates and increments, as well as alerts in near-real time of potentially dangerous forest suppression circumstances. DETER and PRODES projects also try to mitigate the loss of information caused by seasonal cloud cover and canopy shadows, specific issues related to tropical areas. As a result of the complex nature of spatio-temporal data combined with a demand for processing and visualization capabilities of large mapping areas, new possibilities have been opened up to leverage emergent technologies that enable their easy and flexible manipulation.

That is exactly where Spatial Data Infrastructure (SDI) plays an important role. It aims to facilitate the discovery of spatial data, by using metadata base catalogue and standard web services (e.g., Web Mapping Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS)). SDIs take advantage of large data storage connected to spatial service plug-ins, for instance, object-relational database such as PostgreSQL combined with GIS extensions (e.g., PostGIS) can keep vectorial data following Open Geospatial Consortium (OGC) specifications. These tools are essential to manage the high variability of thematic mapping projects specially in large areas such as Cerrado biome in Brazil. Cerrado is the second largest biome in Brazil, covering a fourth of its territory. Over the last few years it has lost almost 24% of its original coverage due to the agriculture expansion (e.g., soybean, cotton, and corn production), supressed vegetation and pasture cattle. Its biodiversity richness has been highlighted for the maintenance of environmental resources and prevention against vanishing of thousand of annimals and plants [4]. Cerrado's degree of destruction has reached such alarming rates that if it continues it will be difficult to recover its biodiversity. With that in mind, much of the attention that has flowed towards Amazon Forest over the last few years while other biomes stayed in the background, has cloven to Cerrado now.

For example, *PRODES* Cerrado project have generated more than 2.5 million features from 2000 to 2017. In 2004, more than 500 thousand features were created with about 60000 km^2. Large features reach 213 km^2, while small ones reach meters scales. If one wants to analyze more regionally information, further geoprocessing and adjusts are required and that takes time. SDI thus must support both visualization and downloading of pre-filtered spatial data, which is one of the main goals proposed by TerraBrasilis [5]. TerraBrasilis helps to organize, access and use spatial data produced by INPE's environmental monitoring programs, but throughout a web portal, it makes possible based on customized views to aggregate other types of spatial data. Rather than just relying on geoservices, it uses ubiquitous clear and simple APIs accross a cluster of virtualized machines to make spatial data analysis easier. Furthermore, TerraBrasilis enables the management of dynamic environments such as those found in *DETER* project that produces daily data. That means, it becomes reasonable to trace forest degradation and fire scars areas every day even before they are deforestated.

Dealing with all these issues is not a trivial task, since account must be taken of the following: (1) the influence of regional governamental policies to increase the resilience of Cerrado biome and to preserve its biodiversity

(2) the concern for handling the integrated and adaptive management of historical and near-real time deforestation-related rates, increments and alerts; (3) the expensiveness to afford constantly the technology innovation transformations that often follow SDI evolution; and (4) the degree of SDI modularity with benefit of generic and flexible implementations to other biomes .

Thus, the main contributions made by this work are the following:

1. Engineering requirements, designing, implementing, and evaluating an open-source SDI to organize and disseminate deforestation data obtained from consolidate thematic mapping projects such as *DETER* and *PRODES*;

2. Learning lessons from the application of the proposed approach in a real-world deforestation scenario that has called attention for its fast natural anthropological conversion, complex formation and high correlation to soybean cultivation in Cerrado biome, Brazil.

2. TERRABRASILIS PLATFORM

TerraBrasilis aims to transform GIS society into large data analysts by means of Geoinformatics and visual analytics technologies. It leverages modularized processing components of traditional SDI architectures to increase data availability and interoperability with OGC standards, and flexible and rapid API designs, developments and delivery.

2.1. TerraBrasilis - Combining Web Services for Maps and Dashboards

Traditional SDIs such as TerraBrasilis require servers for sharing geospatial data (e.g., Geoserver[1]), and robust metadata search and catalogues (e.g., Geonetwork[2]) following the ISO standard 19115:2003 Geographic Metadata structure recommended by the Brazilian SDI (INDE – Infraestrutura Nacional de Dados Espaciais) initiative[3]. They also support heterogeneous databases for dealing with spatio-temporal, thematic and business data (e.g., PostgreSQL[4], MongoDB[5] and Redis[6]). They provide Web GIS libraries to build simple and responsive interactive maps (e.g., Leaflet Javascript[7]), to visualize high performance components (e.g., d3 Javascript[8]) and to filter large multivariate datasets (e.g., crossfilter Javascript[9]) based on modern web browsers' standards (see PRODES Cerrado deforestation in Figure 1 and 2).

The services of TerraBrasilis are modularized accross a cluster of virtualized machines to facilitate future deployments in different information technology (IT)

[1]http://geoserver.org/
[2]https://geonetwork-opensource.org/
[3]http://www.inde.gov.br/
[4]https://www.postgresql.org/
[5]https://www.mongodb.com/
[6]https://redis.io/
[7]https://leafletjs.com/
[8]https://d3js.org/
[9]http://square.github.io/crossfilter/

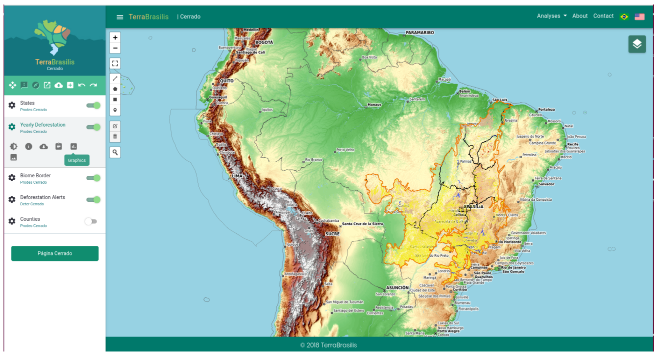

Figure 1: TerraBrasilis - Deforestation Layers.

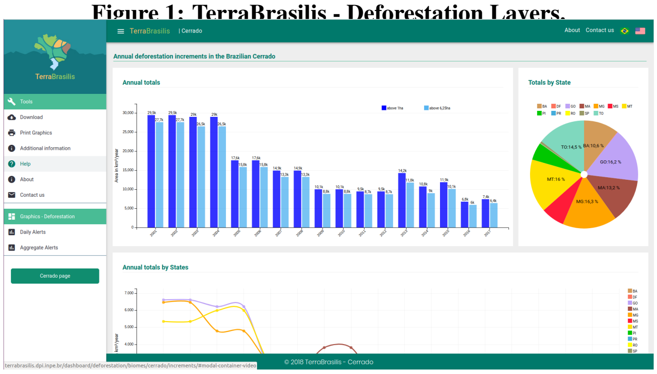

Figure 2: TerraBrasilis - Deforestation Rates Charts.

infrastructure. All of them are packaged up using a lightweight and standalone manner [10] so they can be wrapped with all their dependencies in order to improve performance and reliability of the computing environment. An http server and a reverse proxy[11] are also used to balance the web requests.

2.2. Spatial Data Infrastructure: Improving GIS Interoperability

A SDI aims to integrate a set of technologies, policies, and standards for a more efficient exploitation, access and dissemination of spatial data. It harmonizes and catalogues those data so that any client who has access to the Internet can manipulate them without specialized knowledge. In this sense, TerraBrasilis offers data based on WMS, WFS and WCS to monitor and coordinate environmental programs. It allows users to open layers such as deforestated features using any consolidate GIS platform (e.g., QGIS and ArcGIS). Users just need to add a new layer, for instance, from a WMS Server and TerraBrasilis data is there at their's desktop. A simple way using QGIS is demonstrated in Figure 3.

2.3. Transforming GIS Experts into Data Science Analysts

To increase the data availability, we consider not only users' GIS expertise, but also by means of the designing and development of an API, we leverage their empowerment with scripts' language such as *R* for further data analysis. This API follows a design pattern in which the write and read information are different, used specially in scenarios

[10]https://docs.docker.com/engine/reference/builder/
[11]https://www.nginx.com/

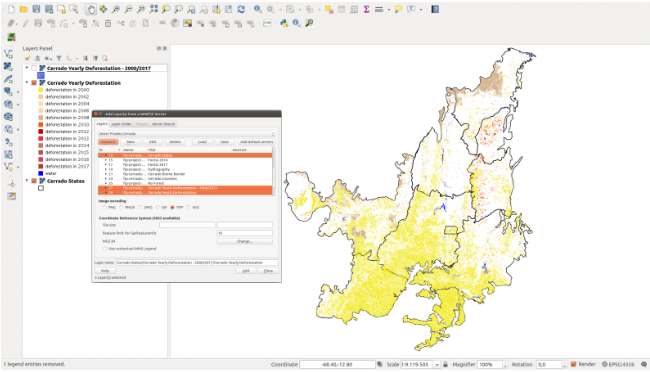

Figure 3: Adding new Layer from a TerraBrasilis WMS Server in QGIS.

where multiple representations are needed[12]. This approach mitigates the need for creating additional columns or use a multi-table approach when dealing with diverse data. Users are then able to check layers information for a plenty of local of interests by the seggregation of updates and queries responsabilities. This helps to improve the representation of a fraction or a portion of the overall data. The Listing 1 represents the resulting API paths operations to get all the available data, its additional information such as their local of interests (e.g., states, municipalities, conservation units, and indigeneous areas) as well as a queryable data, by defining classes (e.g., pasture), local of interest names (e.g., Mato Grosso), and periods (e.g., from 2000 to 2017).

Listing 1: TerraBrasilis API Calls

```
GET api_path/list_available_data
GET api_path/config/list_available_loiname
GET api_path/data/defined_class/defined_loiname
GET api_path/query?class=...&loiname=...&time=...
```

In order to ensure the API interoperability with analytics environments as well, an *R* script demonstration with deforestation data can be seen in Figure 4. At first, it is necessary to manipulate handlers with customized headers as inputs to curl requests to list all the available data. With those information, it is possible then to list the available local of interests associated with that data. Those server responses are casted to JSON format and finally a simple analysis such as a linear estimation method to check a regression for each state can be performed (see Figure 5). Without an automatized manner, it would be difficult to see that Piaui is the only state in which its deforestation rates has a positive coefficient although its absolute values only represents 5%.

3. RESULTS AND DISCUSSIONS

We deployed an alpha version of TerraBrasilis so users can access the Cerrado view of annual deforestation rates (PRODES) and deforestation alerts (DETER) layers. Throughout a nicely designed querying and visualization front-end Web GIS, we demonstrated how to take advantage and observe the deforestation dynamics in Cerrado. Furthermore, it is highlighted how spatial database operations were optimized and the fast web app loading performance.

[12]https://martinfowler.com/bliki/CQRS.html

```r
library(curl)
library(jsonlite)
library(tidyverse)

#### GET apps identifier
configHeader <- function(header) {
    curl::handle_setheaders(header,
                            "Content-Type" = "application/json",
                            "Access-Control-Allow-Origin" = "*")
    return(header)
}

dataHeader <- function(header) {
    curl::handle_setheaders(header,
                            "Content-Type" = "application/json",
                            "Access-Control-Allow-Origin" = "*",
                            "App-Identifier" = res1_json$identifier[1])
    return(header)
}

request <- function(URL, h){
    req <- curl::curl_fetch_memory(URL, handle = h)
    res_json <- jsonlite::fromJSON(rawToChar(req$content))
    return(res_json)
}

h <-  configHeader(curl::new_handle())

res1_json  <- request("api_path/list_available_data", h)
```

```r
res1_json
```

identifier	name	created
prodes_cerrado	Dashboard of the Prodes in the Cerrado	2018-08-31 11:22
prodes_amazon	Dashboard of the Prodes in the Amazon Forest	2018-08-31 11:22
deter_cerrado	Dashboard of the Deter in the Cerrado	2018-08-31 11:22
deter_amazon	Dashboard of the Deter in the Amazon Forest	2018-08-31 11:22

```r
#### GET loi ids
h2 <-  dataHeader(curl::new_handle())

res2_json  <- request("api_path/config/loinames", h2)
```

```r
head(res2_json)
```

gid	loiname
1	BAHIA
2	DISTRITO FEDERAL
3	GOIÃ·S
4	MARANHÃO
5	MATO GROSSO
6	MATO GROSSO DO SUL

```r
#### GET increase cerrado
res3_json  <- request("api_path/data/deforestation/25", h)

nro = nrow(res3_json$periods)

features = do.call(rbind.data.frame, res3_json$periods$features)

result = tibble(name = rep(res3_json$name, nro),
                class = rep(res3_json$clazz, nro),
                startDate = as.Date(paste(res3_json$periods$startDate$year, res3_json$periods$startDate$month, res3_
json$periods$startDate$day, sep ="-")),
                endDate = as.Date(paste(res3_json$periods$endDate$year, res3_json$periods$endDate$month, res3_json$p
eriods$endDate$day, sep ="-")),
                loiname = features$loiname,
                area_km2 = features$area)
```

```r
result
```

name	class	startDate	endDate	loiname	area_km2
PRODES CERRADO	deforestation	1988-08-01	2000-07-31	25	2304.95772
PRODES CERRADO	deforestation	2000-08-01	2002-07-31	25	400.74171
PRODES CERRADO	deforestation	2002-08-01	2004-07-31	25	164.84204
PRODES CERRADO	deforestation	2004-08-01	2006-07-31	25	86.21742
PRODES CERRADO	deforestation	2006-08-01	2008-07-31	25	77.85633
PRODES CERRADO	deforestation	2008-08-01	2010-07-31	25	65.87552
PRODES CERRADO	deforestation	2010-08-01	2012-07-31	25	32.70518
PRODES CERRADO	deforestation	2012-08-01	2013-07-31	25	43.52680
PRODES CERRADO	deforestation	2013-08-01	2014-07-31	25	27.01974
PRODES CERRADO	deforestation	2014-08-01	2015-07-31	25	42.31998
PRODES CERRADO	deforestation	2015-08-01	2016-07-31	25	30.43939
PRODES CERRADO	deforestation	2016-08-01	2017-07-31	25	13.90514

Figure 4: Using the TerraBrasilis API within an analytics environment (R Jupyter).

3.1. PRODES and DETER Cerrado Projects Data Handling using TerraBrasilis: Lessons Learned from Deforestation Data in Cerrado Biome

Although the complexity of deforestation dynamics hampers its controlling and handling, SDI architectures such as TerraBrasilis make easier the access and use of historical and near-real time data. Together, they can improve planning actions and mitigate deforestation by realeasing an up-to-date overview to any interested citizen. There, PRODES and DETER projects are fundamentally important since they represent a surveillance strategy more suitable to dynamic environments such as those found in Cerrado (see Figure 6). Without an adequate management, it would be hard to depict the dynamics of such large areas in a long-term period.

3.2. Performance Analysis of Web App Loading

Measuring performance of web pages is essential to determine how much of our audience is satisfied. In this experiment, we gathered the time metrics acquired over ten

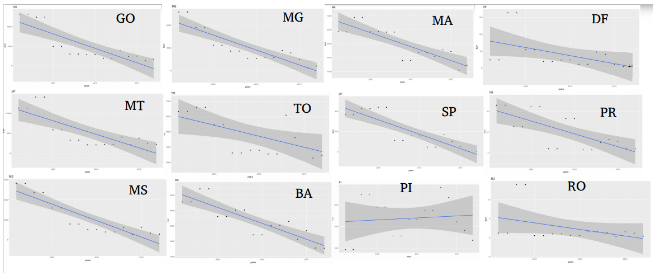

Figure 5: Linear smooth of deforestation data for all the Brazilian Cerrado States.

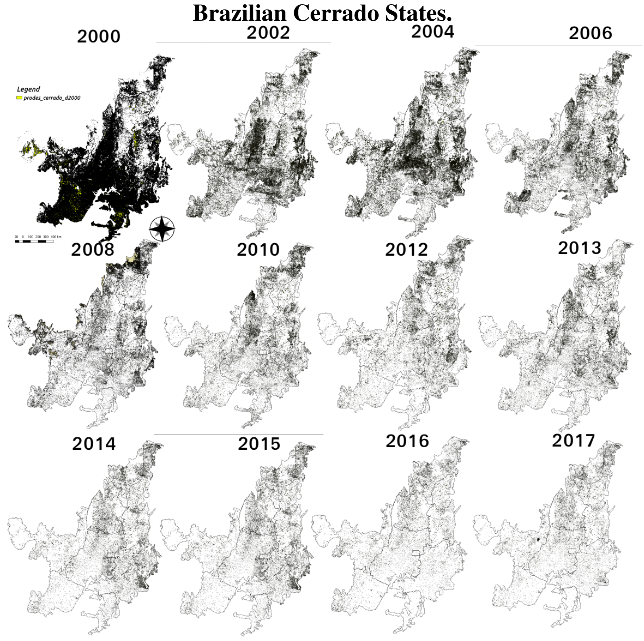

Figure 6: PRODES Cerrado Dynamics from 2000 to 2017 without 2001, 2003, 2005, 2007, 2009 and 2011.

requests repeatedly to emulate the behaviour of web browser to fetch all the web pages files (e.g., css, javascript, images). We highlighted that TerraBrasilis loads very quickly although the experiment has no statistical significance. We also emphasized the small amount of time users spent waiting the scripting, rendering and painting of those files (see Table 1).

Table 1: Time in ms Comparison.

	Mean	Standard Deviation
Loading	78.69	20.44
Scripting	2388.03	205.09
Rendering	298.56	28.87
Painting	65.52	12.59

3.3. "SubDivide" and Conquer: Tunning Spatial Database Operations for Query Performance Optimization

In computer science, a well-known technique (divide and conquer) to solve big problems aims to break them down into small parts so each of the remaining part can be simply answered. In order to process thousand of features faster and extract area values according to specific local of interests, we took advantage of the "subdivide" operation[13]. It converted recursively a single large feature into small fractions using as an input, the layer and a number of maximum vertex parameters to each fraction. It was recently optimized in PostGIS to tune intersect operations but has been neglected by the community [6]. This technique also explores spatial indexes, inevitably used for spatial data. An overview of how "subdivided" features and indexes are compared with regular ones is presented in Figure 7. The idea is to take advantage of small indexes to improve performance of database operations since they work faster than large indexes.

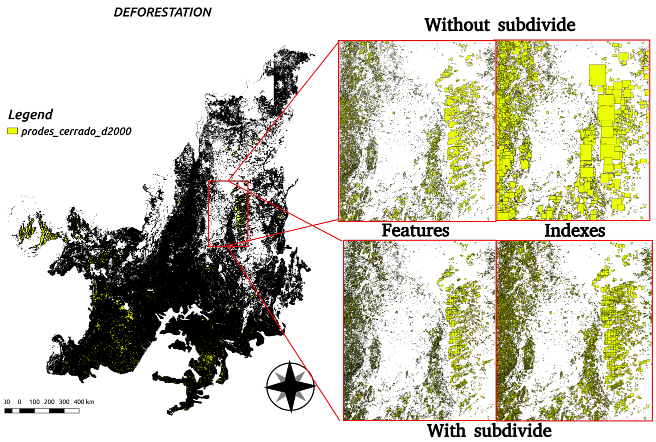

Figure 7: "Subdividing" PRODES Cerrado project deforestation features from 2000 for query performance optimization.

4. CONCLUSIONS

Remote sensing and geoprocessing techniques have been applied to identify critical areas of historical clear-cut deforestation rates and increments, as well as alerts in near-real time of potentially dangerous forest suppression circumstances. TerraBrasilis leverages emergent technologies to enable DETER and PRODES project easily and flexible manipulation. The results evidenced that TerraBrasilis is fast and lightweight to use with spatio-temporal deforestation data of large mapping areas. It provides APIs that interact with GIS and analytics environments, and is able to depict the dynamics of thematic mapping data such as high complex data such as deforestation.

5. REFERENCES

[1] ESCADA, M. I. S. et al. Avaliação de dados dos sistemas de alerta da amazônia: Deter e sad. *Simpósio Brasileiro de Sensoriamento Remoto*, v. 15, p. 2934–2943, 2011.

[2] DINIZ, C. G. et al. Deter-b: The new amazon near real-time deforestation detection system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 8, n. 7, p. 3619–3628, 2015.

[3] PRODES, P. Monitoramento da floresta amazônica brasileira por satélite. *Instituto Nacional de Pesquisas Espaciais Projeto Prodes. Available at http://www. obt. inpe. br/prodes/index. php Accessed*, v. 25, p. 2013, 2013.

[4] KLINK, C. A.; MACHADO, R. B. Conservation of the brazilian cerrado. *Conservation biology*, Wiley Online Library, v. 19, n. 3, p. 707–713, 2005.

[5] INPE. *Terrabrasilis*. 2018. <terrabrasilis.dpi.inpe.br/>. [Online; accessed 25-September-2018].

[6] RAMSEY, P. *Scaling PostgreSQL and PostGIS*. 2018. <https://vimeo.com/250498574>. [Online; accessed 05-October-2018].

[13] https://postgis.net/docs/ST_Subdivide.html