

ACCESSING AND PROCESSING BRAZILIAN EARTH OBSERVATION DATA CUBES WITH THE OPEN DATA CUBE PLATFORM[†]

Vitor C. F. Gomes^{1,2,*}, Felipe M. Carlos², Gilberto R. Queiroz², Karine R. Ferreira² and Rafael Santos²

¹C4ISR Division, Institute for Advanced Studies (IEAv), São José dos Campos, SP 12228-001, Brazil

²Image Processing Division, National Institute for Space Research (INPE), São José dos Campos, SP 12227-010, Brazil

[†] This paper is an extended version of our paper "Integração dos ambientes Brazil Data Cube e Open Data Cube" published in XXI Brazilian Symposium on Geoinformatics (2020)

KEY WORDS: Spatial Data Infrastructure, Big Earth Observation Data, Data Cube, STAC

ABSTRACT:

Recently, several technologies have emerged to address the need to process and analyze large volumes of Earth Observations (EO) data. The concept of Earth Observations Data Cubes (EODC) appears, in this context, as the paradigm of technologies that aim to structure and facilitate the way users handle this type of data. Some projects have adopted this concept in developing their technologies, such as the Open Data Cube (ODC) framework and the Brazil Data Cube (BDC) platform, which provide open-source tools capable of managing, processing, analyzing, and disseminating EO data. This work presents an approach to integrate these technologies through the access and processing of data products from the BDC platform in the ODC framework. For this, we developed a tool to automate the process of searching, converting, and indexing data between these two systems. Besides, four ODC functional modules have been customized to work with BDC data. The tool developed and the changes made to the ODC modules expand the potential for other initiatives to take advantage of the features available in the ODC.

1. INTRODUCTION

In recent years, the amount of Earth Observation data freely available has grown, motivated by technological advances in acquisition and storage equipment and space agencies' policies that make their data repositories available. The estimated volume of EO data produced in 2019 by Landsat (7 and 8), MODIS (Terra and Aqua units), and the Sentinel missions (1, 2, and 3) exceeded 5 PB (Soille et al., 2018). The technological challenges to store, process, and analyze these large data sets impose significant restrictions for EO community scientists to take advantage of all potential of these resources (Camara et al., 2016, Stromann et al., 2020).

A Spatial Data Infrastructure (SDI) provides an environment that allows people and systems to interact with technologies to foster activities for using, managing, and producing geographic data (Rajabifard and Williamson, 2001). In the last years, SDIs have been built using technological components that implement standards proposed by Open Geospatial Consortium (ODC) and Organization for Standardization (ISO) to represent, store and disseminate spatial data. Even with these standards, most current SDIs are focused on EO data sharing and dissemination in the form of individual files through web portals and HTTP, FTP, and SSH protocols (Müller, 2016).

In the scenario of big EO data, the proper management, processing and dissemination of this vast volume of data poses a challenge for EO system infrastructures. The needs in this scenario demand more precise and structured research services, automated acquisition, calibration and availability processes, and the possibility of data being processed without having to be moved through the network (Woodcock et al., 2016).

To address these challenges, the scientific community recently adopted the Earth Observation Data Cube paradigm, which,

through specialized technologies, seeks to change the way researchers deal with these large volumes of EO data (Giuliani et al., 2019b). Even though there is no consensus on the definition of the term EODC, these systems are software infrastructures that manage large time series of EO data using multidimensional array concepts, facilitating the access and the use of Analysis Ready Data (ARD) by users (Nativi et al., 2017, Giuliani et al., 2019b).

Many institutions have been developing technologies that use these concepts and can be considered EODC (Giuliani et al., 2019b), such as Open Data Cube (ODC) (Open Data Cube, 2021a), Google Earth Engine (GEE) (Gorelick et al., 2017), JRC Earth Observation Data and Processing Platform (JEODPP) (Soille et al., 2018), Sentinel Hub (SH) (Sinergise, 2021) or Brazil Data Cube (BDC) Platform (Ferreira et al., 2020). These platforms adopt different data abstractions, standards, or technological solutions to provide their functionalities despite similar functionalities (Gomes et al., 2020). ODC and BDC, for example, are both open source solutions that use different data abstractions and methods to process and prepare ARD. BDC Platform uses an R package, called *sits* (Camara et al., 2018) for land use and land cover mapping, while ODC focus all data processing and analysis using Python language and *xarray* package. GEE and Sentinel Hub are commercial solutions that allow users to access and process large EO catalogs using, each one, specialized APIs and different web services. For data abstraction, GEE uses *Image/ImageCollection* and *Feature/FeatureCollection*, while Sentinel Hub uses *Data Source*, *Instances*, and *Layer* concepts.

As a consequence of the difference in the way each platform manages data and provides access and processing to users, there is a lack of interoperability between these solutions, making it a challenge to discover, access, and share processes between EODCs (Nativi et al., 2017, Giuliani et al., 2019b).

One of the leading platforms in the EODC scenario (Gomes et

*Corresponding Author

al., 2020) is the Open Data Cube, which is composed of a set of tools and services for the management, processing, and access to EO data and which has been used by several initiatives and institutions around the world. In July 2019, there were 56 initiatives for ODC, 9 of which are operational, 14 under development, and 33 under review (Open Data Cube, 2019). The expectation is that there will be 22 operational instances of ODC in 2022 (Killough, 2018).

In the Brazilian context, the National Institute for Space Research (INPE) leads the development of the Brazil Data Cube Platform for management and analyzing massive EO data. The main objectives of the BDC project are to produce cubes with Analysis Ready Data in medium resolution for the Brazilian territory that allows the analysis of time series and the use and development of technologies for processing and storage of those data cubes, including cloud computing and distributed processing (Ferreira et al., 2020).

This work presents the integration process between the data products of the BDC project and the ODC framework. This integration aims to expand the services and tools that can be used to access, view, and analyze the EODC produced by the BDC project. Besides, the integration intends to allow algorithms previously developed for ODC technology to be more easily adapted for use with the BDC data.

The remainder of this article is organized as follows. In Section 2, we present an overview of the Open Data Cube and Brazil Data Cube platforms. Section 3 the process used to integrate the systems and the developed tools is presented. Section 4 presents the results of the integration of the ODC components with the data and services of the BDC. Finally, in Section 5, we make the final considerations regarding the challenges encountered and the next steps that will be developed.

2. EARTH OBSERVATIONS DATA CUBES

Earth Observations Data Cube is commonly used to refer to multidimensional arrays with space, time dimensions, and spectral derived properties created from remote sensing images (Appel and Pebesma, 2019). This term is often used to refer to analytical technology solutions that make use of these data structures. The term can also be found to refer to analytical, technological solutions that allow the management, processing, and analysis of these data structures (Giuliani et al., 2019a).

Recently, many institutions have been creating EO data cubes from remote sensing images to specific regions, such as Australian Data Cube (Lewis et al., 2017), Swiss Data Cube (Giuliani et al., 2017), Armenian Data Cube (Asmaryan et al., 2019), Africa Regional Data Cube (Killough, 2018), Colombian Data Cube (CDCol) (Ariza-Porras et al., 2017), and Brazil Data Cube (Ferreira et al., 2020). Except for the last one, the other initiatives use the framework Open Data Cube as the core technology to index and handle the EO data. These initiatives use ODC as a start point and build custom applications to accomplish specific demands. The CDCol initiative, for example, implements tools to handle a bank of algorithms and its live cycle (in development, published, obsolete, and deleted) and user roles to distinguish the responsibilities of users in the platform (Ariza-Porras et al., 2017). On the other hand, the Brazil Data Cube platform uses its own tools to produce and manipulate data cubes, but it has been developing tools to integrate its solutions with the ODC framework, such as the one presented

in this work. The following subsections present details about the Open Data Cube framework and the Brazil Data Cube platform.

2.1 Open Data Cube

The Open Data Cube is a framework that allows the cataloging and analysis of EO data. It consists of a set of data structures and Python libraries that allow the manipulation, visualization, and analysis of that data. The source code of ODC is available under Apache 2.0 license and is distributed as modules on github.

The module `datacube-core` is responsible for indexing, searching, and retrieving cataloged data. It consists of a Python package and command-line tools that use a PostgreSQL database to store metadata for managed data.

In indexing the data, the description and metadata of a *Product* is initially registered. *Product* is the abstraction used by ODC for data collections that share the same measures (bands) and metadata. Then, information about the *Datasets* that together represent a *Product* are recorded. *Datasets* represent the smallest aggregation of managed data, they usually are scenes of a given *Product* stored in files (Open Data Cube, 2021b).

After indexing the EO data in an ODC instance, it is possible to use an ODC Python API to access and process it. ODC also provides other modules for: data visualization (`datacube-explorer`); performing temporal statistical analysis (`datacube-stats`); disseminating data through OGC services (`datacube-ows`); illustrates the usage of ODC Python interface (`datacube-notebooks`); etc.

Also, ODC has a catalog with the source code of applications that use the ODC Python API to access data cubes and perform specific analyzes. The Committee on Earth Observation Satellites (CEOS) also provides a repository (CEOS, 2021) with algorithms for processing data accessed in an ODC catalog.

2.2 Brazil Data Cube

The Brazil Data Cube project has objectives to produce Analysis Ready Data structured in data cubes in medium resolution for the entire Brazilian territory. For this, the BDC project use and develop technologies necessary for the processing, analysis, and dissemination of these data products and to produce information on Land Use and Land Cover from the cubes using machine learning methods and image processing techniques (Ferreira et al., 2020).

Figure 1 illustrates the services and products of the Brazil Data Cube platform. The first layer, below, represents the generating process of ARD data sets used to create the data cubes. The layer above represents the cataloging of the metadata of the generated data products.

The data and metadata of the data products produced by the BDC are accessed and processed through web services presented in the Services layer. Services using standard protocols are available, such as Tile Map Service (TMS), Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS). In addition to these, there are also special-purpose services developed by the project team, such as Web

<http://github.com/opendatacube>
<https://www.opendatacube.org/dca1>

Time Series Service (WTSS) (Vinhas et al., 2016) and Web Land Trajectory Service (WLTS).

The Brazil Data Cube platform also makes its data products publicly available through a STAC service .

The applications, available to end-users, use these web services to access the data cubes produced by the BDC. In this project's structure, ODC behaves like an application that uses BDC services to consume and index metadata and EO data.

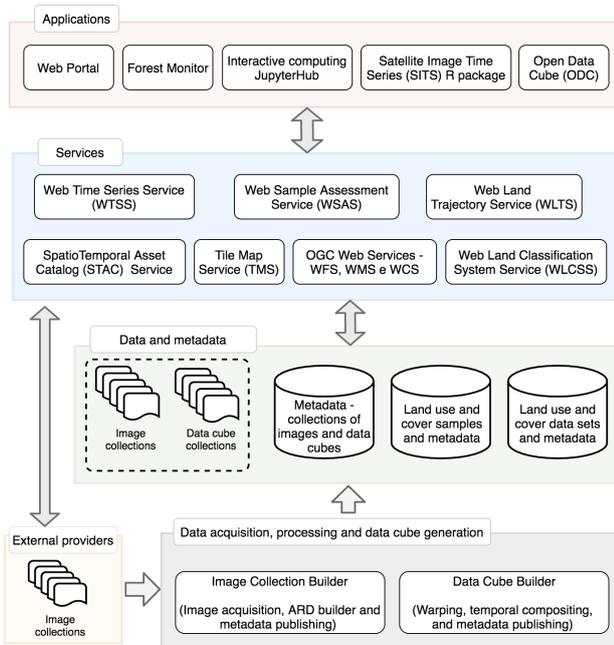


Figure 1. Brazil Data Cube Project: data and software products. Source: (Ferreira et al., 2020)

All software products developed by BDC are available under MIT license in the project's source code repository .

3. METHODOLOGY

Four steps were necessary to accomplish the integration between BDC's data products in the ODC platform and the availability of a computational environment to use the new integrated functionalities. Figure 2 illustrates these four steps and the associated data flow from BDC Platform to the BDC's ODC instance.

The first step was to prepare an ODC instance of the BDC project's computing infrastructure. For this, an instance of the PostgreSQL DBMS was implemented to store the products and datasets catalog metadata (ODC-db). Also, a Docker container (ODC-core) with the datacube-core module of the ODC framework was deployed, which provides the necessary tools to index and manage the metadata catalog. This instance and all the others created in this integration process have direct access to a file system that contains the BDC data repository. Thus, no data replication is required for shared use between the BDC tools and the ODC instance. To facilitate readability, we will use the acronym BDC-ODC to refer to the ODC instance running on the BDC project infrastructure.

<http://brazildatacube.dpi.inpe.br/stac>
<https://github.com/brazil-data-cube>

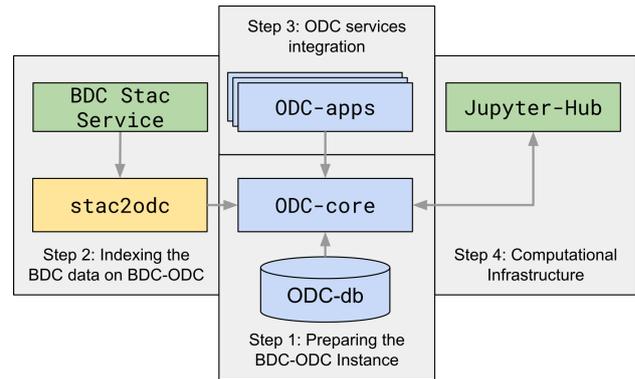


Figure 2. Overview of data flow from BDC data to the BDC-ODC instance.

The other three integration steps are the indexing of BDC data products in the BDC-ODC (Step 2), the adaptation and configuration of ODC framework services in the BDC-ODC (Step 3), and the provision of a multi-user computational infrastructure to access and processing of data indexed in the BDC-ODC (Step 4). The following subsections detail these tasks.

3.1 Data indexing

The BDC STAC service was chosen as the source for access the BDC project products' data and metadata. This choice over direct access to the BDC database was motivated by leveraging other future available catalogs' indexing through this specification.

From this choice and motivated by the need to manipulate large volumes of data from the BDC, an application was developed, called *stac2odc*, to automate reading the STAC catalog and indexing the metadata in the BDC-ODC.

Figure 2 presents an overview of the data flow used in this integration. It can be seen that the *stac2odc* tool is responsible for collecting the data in the BDC catalog and, after converting it, storing it in the BDC-ODC catalog through the *datacube-core* module. From that moment on, the data will be available for use in other applications.

The *stac2odc* tool maps the BDC STAC catalog metadata to the format accepted by the ODC. The mapping specification is done through a configuration file. This option allows changes to the BDC or ODC metadata structures to be easily incorporated into the tool.

Listing 1 shows an example of the configuration file in JSON format used by *stac2odc* tool. In this example, it is possible to view the mapping definition structure. The JSON keys represent the values' source, and the respective values represent the destination in the ODC metadata file to be generated. This file can also define external scripts (*custom_mapping.py*) for the translation of metadata or constant values.

This configuration file also allows the user to inform the location where the data products consulted at STAC are stored. With this information, *stac2odc* does not need to duplicate the data in the BDC project infrastructure. On the other hand, if it is of interest to the user, it is possible to inform a target folder to download the data being indexed. This feature is handy for researchers who want to populate a particular instance of the ODC with data from the BDC project.

Listing 1. Example of an stac2odc configuration file

```

1 {
2   "product": {
3     "fromSTAC": {
4       "name": {
5         "from": "id",
6         "customMapFunction": {
7           "functionName": "transform_id",
8           "functionFile": "custom_mapping.py"
9         }
10      },
11     "storage.crs": "bdc:crs",
12     "fromConstant": {
13       "metadata_type": "eo",
14       "measurements.units": "meters"
15     }
16   }
17 }
```

3.2 ODC services integration

With BDC data products indexed in BDC-ODC, they are now ready for use through the command line tools and the ODC Python API, available on the datacube-core module.

The ODC provides a wide range of tools and services in its ecosystem in a modular way. For this first phase of integrating BDC data products into the ODC, the datacube-ows, data-explorer, datacube-stats, and datacube-ui modules have been chosen. The following paragraphs present the functionalities and considerations made for each module's configuration during this integration.

The datacube-ows module allows the dissemination of data indexed in the ODC catalog through web services in the OGC WMS, WMTS, and WCS standards. Its configuration is done through a command-line tool, which helps to create indexes used by the service in the database. This module also requires creating a configuration file with the description of the products to be published and their associated styles.

Changes to datacube-ows module source code were necessary to use this module with the BDC data products. These changes were made to adapt how a Coordinate Reference System (CRS) is handled in the module internal operations. The BDC project uses a CRS generated specifically for its data products to reduce geometric distortions in the data generated for South American territories (Ferreira et al., 2020). The CRS used by the BDC project does not have a standard EPSG code, which is required by the original code of the datacube-ows module.

The datacube-explorer module provides a simple web interface for searching data and metadata indexed in the ODC catalog. This application also provides a STAC API for advanced searches. Similar to datacube-ows, this module is also configured using a command-line tool. This tool creates new tables in the database used by the ODC and populates them with the information used by the module. During the configuration of the datacube-explorer, it was also identified the lack of support for the use of CRS that does not have a standard EPSG code. For this reason, it was necessary to modify the source code of the tool in order to make possible the use of custom CRS. This capability was implemented by adding a new command line parameter, `--custom-crs-definition-file`, which allows users to inform the location of an additional configuration file. This file allows linking a CRS, defined through a PROJ String, with an arbitrary EPSG code. These settings are made in JSON

format. Listing 2 presents the configuration used for the use of the BDC data products in the datacube-explorer module.

Listing 2. BDC custom CRS definition file used in datacube-explorer module.

```

1 { "epsg:100001": "+proj=aea +lat_0=-12 +
   lon_0=-54 +lat_1=-2 +lat_2=-22 +x_0=5
   000000 +y_0=10000000 +ellps=GRS80 +
   units=m +no_defs" }
```

The need to process a large volume of EO data often requires the researchers to use advanced techniques such as parallel processing and efficient data management in memory. These requirements can represent a barrier for users who want to process and analyze data but do not have enough technical expertise for this complex task.

To address this needs, the ODC datacube-stats module provides a simple interface for processing the data indexed in the ODC catalog and automatically managing the computational resources used in this process. In the integration of this module in the BDC-ODC, the configuration of the Dask tool was also done to allow the parallel and distributed analyzes. Dask is a Python library that provides data structures and tools for scheduling tasks across multiple processing nodes.

Finally, the datacube-ui module was configured. Unlike the other modules used, the datacube-ui is available in the CEOS github repository. This module provides a full-stack Python web application for performing analysis on data indexed in an ODC catalog. This application's objective is to provide a high-level interface for users to access the indexed data, perform analyses previously configured in the tool, and easily access the metadata of the analyses performed (CEOS, 2021). This module currently comes with analysis application code such as Cloud coverage detection, Coastal change, Water detection, and spectral indices calculation. It is also possible to include new applications through a template included in the documentation of this module.

Currently, the applications previously available in datacube-ui are not in use in the BDC project due to the incompatibility between the data and metadata required by the applications present in this module and those available in BDC-ODC. The usage of this module is currently more focused on the access and visualization of indexed data and metadata on the BDC project.

The implementation of datacube-stats and datacube-ui modules in the BDC-ODC did not require changes in their source codes.

3.3 Computational infrastructure

The processing and analysis of the large volume of EO data indexed in the BDC-ODC may require many computational resources. To allow BDC researchers to consume this data, a computational infrastructure was prepared. This infrastructure, illustrated in Figure 3, provides a multi-user web interface for processing the data indexed in the BDC-ODC.

The infrastructure presented was done using JupyterHub technology, which provides for each user a ready-to-use isolated environment. The available environments are created by the

<http://dask.org>
<https://github.com/ceos-seo>

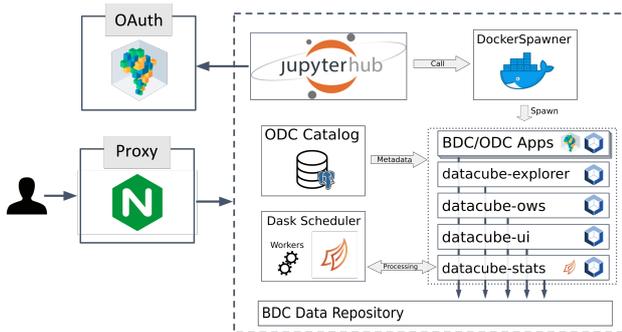


Figure 3. Computational infrastructure BDC-ODC.

DockerSpawner module of JupyterHub, through Docker images. In these images are stored a set of instructions for creating the environment. Different images can be created and used in the DockerSpawner, which makes the definition of environments flexible to users' needs. User authentication is done with the BDC-OAuth service, previously used in other services of the BDC project.

This infrastructure was configured in the data processing servers of the BDC project. For this, Docker images were prepared for the use of datacube-core and datacube-stats tools, being added in the images all the software dependencies of each one of these tools. An instance of Dask Scheduler and Dask Workers was also configured to enable distributed processing in datacube-stats. When an environment is generated by DockerSpawner, it is configured to have direct access to the file system where the BDC data repository is stored. The environment also has access to the database server used by BDC-ODC.

Besides the Docker images for the use of the BDC-ODC, there are also available in the JupyterLab Docker images with other tools used by the BDC project team, such as SITS (Camara et al., 2018) and GDALCubes (Appel and Pebesma, 2019).

4. RESULTS

This section presents the use of the modules and tools made available after the ODC integration in the BDC project platform. Other examples of using the tools presented and the documentation of the changes made can be found in the code repository of BDC project.

The configuration of the datacube-ows module made possible the consumption of the data indexed in the BDC-ODC through the WMS, WMTS, and WCS services. These services bring benefits to several applications by allowing quick access to data, making it possible to perform data processing and visualization. An example of the use of the WMS service can be seen in Figure 4. In the example, through a web page, Landsat-8/OLI data from all Brazilian territory is presented. The creation of the presented mosaic is done on-the-fly by the datacube-ows module.

The modifications made to the datacube-explorer made it possible to use it to perform space-time searches of the data indexed in the BDC-ODC. All operations can be performed through a web interface, which allows, besides the search, to inspect the metadata of each Dataset identified. Figure 5 presents the result for a search made for January 2020 in all Brazilian territory.

<https://jupyter.org/hub>

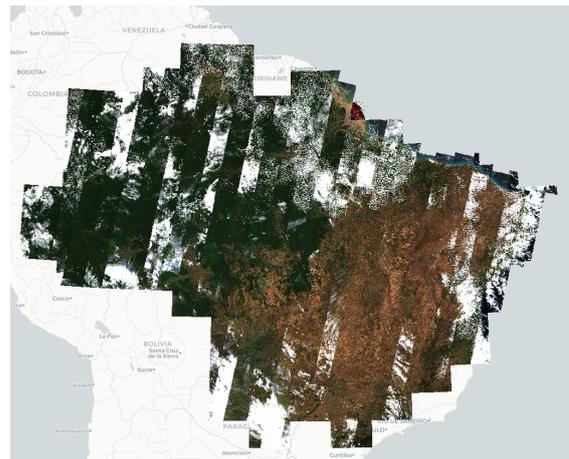


Figure 4. Consumption of data products through WMS via datacube-ows



Figure 5. Explorer presenting data from the CBERS-4 collection indexed in BDC-ODC

For data processing, the datacube-stats tool facilitated the extraction of time statistics from indexed data. With this module, the processing of large spatial extensions to extract metrics from the data can be done by users without technical knowledge. This tool hides from the user the complexity of the processing distribution. To illustrate the use of datacube-stats, the extraction of the temporal average of NDVI data CBERS-4/WFI was realized in the whole Amazon biome between January 2018 to July 2020. The results are presented in Figure 6.

The extraction of time metrics from the data, presented in Figure 6, was performed in the prepared computational infrastructure based on JupyterHub. The possibility of adding ready-to-use environments, easily customized with the users' needs, made it simple to apply the indexed data in different contexts. To show another customization of this infrastructure and the use of existing applications for the ODC, Figure 7 presents the result of the execution of a code developed by CEOS researchers to cluster pixels of a Sentinel-2/MSI scene (CEOS, 2021). The scene chosen for this analysis is located in the state of Roraima, in Brazil.

4.1 Code and data availability

All the tools necessary to reproduce the activities presented in this work are available in the repository <https://github.com/brazil-data-cube/bdc-odc>.



Figure 6. Temporal NDVI Mean from Landsat-8/OLI collection in 01/2018 to 07/2020

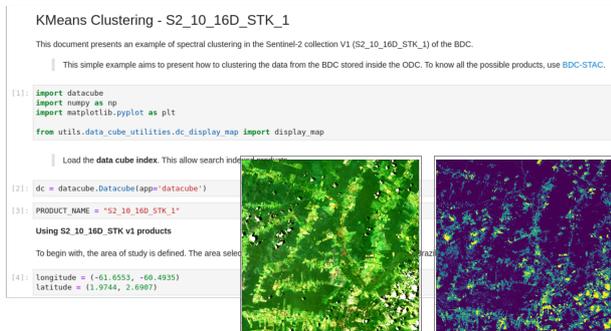


Figure 7. Computational infrastructure used to clustering Sentinel-2/MSI pixels in Roraima state, Brazil

In this repository, the following are available: i) the source code of the `stac2odc` tool and usage documentation; ii) the modified ODC modules; iii) documentation of changes made to the ODC modules; iv) the configuration files used in the deployment of ODC modules; and v) Dockerfiles to generate the images used in this integration. We believe that the content of this repository allows an interested researcher to prepare an ODC instance in his local infrastructure with the data available in the BDC's STAC catalog, including the modules presented in our integration.

The added usage documentation was created following (Kilgough, 2018) recommendations. Thus, details have been passed on so that others can understand and consume the lessons we have learned during the integration process.

5. DISCUSSION AND FINAL REMARKS

In this work, we presented the integration process between the Open Data Cube framework with the Brazil Data Cube project's data products. The results obtained indicate that the integration approach, done through the conversion of metadata formats and linking with the existing data repository, can be used as an initial form of interoperability between the technologies. The tool created for such a process is independent of the source of the data. This tool can be used to consume other data services that implement the STAC specification.

The addition of the ODC ecosystem modules to the BDC-ODC complemented the BDC's technology base to disseminate and

process the large volumes of data generated by the project. With the use of `datacube-ows`, the BDC data cubes became available through standardized and interoperable interfaces, such as OGC WMS, WMTS, and WCS. These services allow that researchers use well-known geographic information systems (GIS) to consume these data.

In this study, an example of using the `datacube-stats` for the processing of NDVI time averages for the entire Amazon biome was presented. This tool, the functionalities of the `datacube-core` module and the computational infrastructure defined in this work, can be used as a reference setup for the application of *time-first*, *space-later* approaches that analyses the temporal variation of EO datasets.

Even with the benefits presented and the ODC tools' maturity, the integration required adaptations to some modules' source code for their adequacy and use with the BDC data. Such modifications were made to the `datacube-ows` and `datacube-explorer` modules, which out-of-the-box expected data that had a coordinate reference system (CRS) with a standard EPSG code. The modifications made improve the ODC ecosystem tools, making their use possible in more general scopes of data cubes. However, further testing is needed to verify that any defined coordinate reference system is compatible.

Although the process of this work still represents the beginning of the integration process between ODC and BDC technologies, we believe that there are two main contributions presented: the provision of data from the BDC project to the EO community already familiar with ODC tools; and the possibility for users of the BDC platform to take advantage of the large catalog of algorithms available for the ODC framework.

For the next steps, we hope that more tests will be done on the modifications made to the modules to be sent to the official ODC repositories. It can help the ODC community reach out to more projects and initiatives that seek to use the tools available. It is also essential to consider that during integration, modules such as `datacube-ui`, due to attribute incompatibilities, could not be readily used to apply the algorithms to the indexed data. It is expected that the adaptation or even the development of new algorithms will be performed for this tool.

Although there is still a long way to go for optimal interoperability between EODCs, this work of integrating ODC tools with the BDC platform paves the way to increase integration between these technologies. Currently, few efforts are known in the domain of EODCs integration. These initiatives are important to prevent these solutions from becoming silos of information (Giuliani et al., 2019b).

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and 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 (ARS, AS, MP).

REFERENCES

- Appel, M., Pebesma, E., 2019. On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library. *Data*, 4(3), 16.
- Ariza-Porras, C., Bravo, G., Villamizar, M., Moreno, A., Castro, H., Galindo, G., Cabera, E., Valbuena, S., Lozano, P., 2017. CDCol: A Geoscience Data Cube that Meets Colombian Needs. *Advances in Computing*, Springer International Publishing, Cham, 87–99.
- Asmaryan, S., Muradyan, V., Tepanosyan, G., Hovsepyan, A., Saghatelyan, A., Astatsryan, H., Grigoryan, H., Abrahamyan, R., Guigoz, Y., Giuliani, G., 2019. Paving the Way towards an Armenian Data Cube. *Data*, 4(3), 11.
- Camara, G., Assis, L. F., Ribeiro, G., Ferreira, K. R., Llapa, E., Vinhas, L., 2016. Big earth observation data analytics: Matching requirements to system architectures. *Proceedings of the 5th ACM SIGSPATIAL international workshop on analytics for big geospatial data*, ACM, 1–6.
- Camara, G., Simoes, R., Andrade, P. R., Maus, V., Sánchez, A., de Assis, L. F. F. G., Santos, L. A., Ywata, A. C., Maciel, A. M., Vinhas, L., Queiroz, G., 2018. e-sensing/sits: version 1.12.5; Zenodo: Geneva, Switzerland.
- CEOS, 2021. CEOS repository. Available online: <https://github.com/ceos-seo/> (accessed on 25 jan 2021).
- Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F., Simoes, R. E., Picoli, M. C., Camara, G., Cartaxo, R., Gomes, V. C., Santos, L. A., Sanchez, A. H., Arcanjo, J. S., Fronza, J. G., Noronha, C. A., Costa, R. W., Zaglia, M. C., Zioti, F., Korting, T. S., Soares, A. R., Chaves, M. E., Fonseca, L. M., 2020. Earth observation data cubes for Brazil: Requirements, methodology and products. *Remote Sensing*, 12(24), 1–19.
- Giuliani, G., Camara, G., Killough, B., Minchin, S., 2019a. Earth observation open science: enhancing reproducible science using data cubes. *Data*, 4(4), 4–9.
- Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J.-P., Allenbach, K., Dao, H., Peduzzi, P., 2017. Building an Earth Observations Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data*, 1(1-2), 100–117.
- Giuliani, G., Masó, J., Mazzetti, P., Nativi, S., Zabala, A., 2019b. Paving the Way to Increased Interoperability of Earth Observations Data Cubes. *Data*, 4(3), 23.
- Gomes, V. C. F., Queiroz, G. R., Ferreira, K. R., 2020. An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sensing*, 12(8), 25.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202(2016), 18–27.
- Killough, B., 2018. Overview of the open data cube initiative. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 8629–8632.
- Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., Dekker, A., Dhu, T., Hicks, A., Ip, A., Purss, M., Richards, C., Sagar, S., Trenham, C., Wang, P., Wang, L. W., 2017. The Australian Geoscience Data Cube - Foundations and lessons learned. *Remote Sensing of Environment*, 202, 276–292.
- Müller, M., 2016. Service-oriented Geoprocessing in Spatial Data Infrastructures. PhD thesis, Technische Universität Dresden.
- Nativi, S., Mazzetti, P., Craglia, M., 2017. A view-based model of data-cube to support big earth data systems interoperability. *Big Earth Data*, 1(1-2), 75–99.
- Open Data Cube, 2019. The "Road to 20" International Data Cube Deployments. Technical report, Open Data Cube.
- Open Data Cube, 2021a. Open data cube. Available online: <https://www.opendatacube.org/> (accessed on 25 jan 2021).
- Open Data Cube, 2021b. Open data cube manual. Available online: <https://datacube-core.readthedocs.io/en/latest/> (accessed on 25 jan 2021).
- Rajabifard, A., Williamson, I. P., 2001. Spatial data infrastructures: concept, SDI hierarchy and future directions. *Proceedings, of GEOMATICS'80 Conference*, In: Proceedings Geomatics, Australia, 2001, 11.
- Sinergise, 2021. Sentinel Hub by Sinergise. Available online: <https://www.sentinel-hub.com/> (accessed on 15 apr 2021).
- 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 Generation Computer Systems*, 81, 30–40.
- Stromann, O., Nascetti, A., Yousif, O., Ban, Y., 2020. Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine. *Remote Sensing*, 12(1), 24.
- Vinhas, L., Ribeiro De Queiroz, G., Ferreira, K. R., Camara, G., 2016. Web Services for Big Earth Observation Data. *Geoinfo 2016*, 166–177.
- Woodcock, R., Cecere, T., Mitchell, A., Killough, B., Dyke, G., Ross, J., Albani, M., Ward, S., Labahn, S., 2016. CEOS Future Data Access and Analysis Architectures Study. 33.