

THE CEOS DATA CUBE PORTAL: A USER-FRIENDLY, OPEN SOURCE SOFTWARE SOLUTION FOR THE DISTRIBUTION, EXPLORATION, ANALYSIS, AND VISUALIZATION OF ANALYSIS READY DATA

¹Syed R Rizvi, ²Brian Killough, ¹Andrew Cherry, ¹Sanjay Gowda

¹Analytical Mechanics Associates, Hampton, VA
²NASA Langley Research Center, Hampton, VA

ABSTRACT

There is an urgent need to increase the capacity of developing countries to take part in the study and monitoring of their environments through remote sensing and space-based Earth observation technologies. The Open Data Cube (ODC) provides a mechanism for efficient storage and a powerful framework for processing and analyzing satellite data. While this is ideal for scientific research, the expansive feature space can also be daunting for end-users and decision-makers who simply require a solution which provides easy exploration, analysis, and visualization of Analysis Ready Data (ARD). Utilizing innovative web-design and a modular architecture, the Committee on Earth Observation Satellites (CEOS) has created a web-based user interface (UI) which harnesses the power of the ODC yet provides a simple and familiar user experience: the CEOS Data Cube (CDC). This paper presents an overview of the CDC architecture and the salient features of the UI. In order to provide adaptability, flexibility, scalability, and robustness, we leverage widely-adopted and well-supported technologies such as the Django web framework and the AWS Cloud platform. The fully-customizable source code of the UI is available at our public repository. Interested parties can download the source and build their own UIs. The UI empowers users by providing features that assist with streamlining data preparation, data processing, data visualization, and sub-setting ARD products in order to achieve a wide variety of Earth imaging objectives through an easy to use web interface.

Index Terms— Open Data Cube, ODC, CEOS, Remote Sensing, Earth Observation, Satellite, User Interface

1. INTRODUCTION

The Committee on Earth Observation Satellites (CEOS) System Engineering Office (SEO) has acknowledged an urgent need to increase the capacity of developing countries to take part in studying and monitoring their environments

through remote sensing and observation of Earth from space. Forest preservation initiatives, carbon measurement initiatives, water management, agricultural monitoring, and urbanization studies are just a few examples of causes that can benefit greatly from remote sensing data. Currently, however, many developing nations lack the in-country expertise and computational infrastructure to utilize remote sensing data.

The CEOS SEO has played an important role in building software products and tools that leverage the group's extensive experience with space imaging systems and data products to provide significant and far-reaching capabilities to government stakeholders and scientific researchers. One such example is its support of the Open Data Cube (ODC) initiative to provide a data architecture solution that has value to its global users and increases the impact of Earth Observation (EO) satellite data [1-2]. The ODC platform provides a data ingestion framework that includes support for ingestion of a wide variety of remote sensing data products. The data products are ingested into an N-dimensional data array that abstracts away management of distinct acquisitions. The platform has a tiered API for data access and processing, and a data/application platform layer for higher-level access.

We have developed the CEOS Data Cube (CDC) web-based User Interface (UI) is one such well-known tool [3]. The UI has empowered users by providing features that assist with streamlining data retrieval, preparation, processing, and visualization, and facilitates exporting ingested data in order to achieve a wide variety of Earth imaging objectives (Figure 1). The main contributions in this paper are: (1) briefly describing the framework of the CDC UI, (2) enumerating its salient remote sensing applications, and (3) developing a concept - how exporting a subset of an ingested dataset can be leveraged in order to partially achieve Data Cube On-Demand. The web interface, available to the public at <http://ec2-52-201-154-0.compute-1.amazonaws.com/>, has been presented at multiple conferences, tutorials, training sessions, and international presentations [4-7].

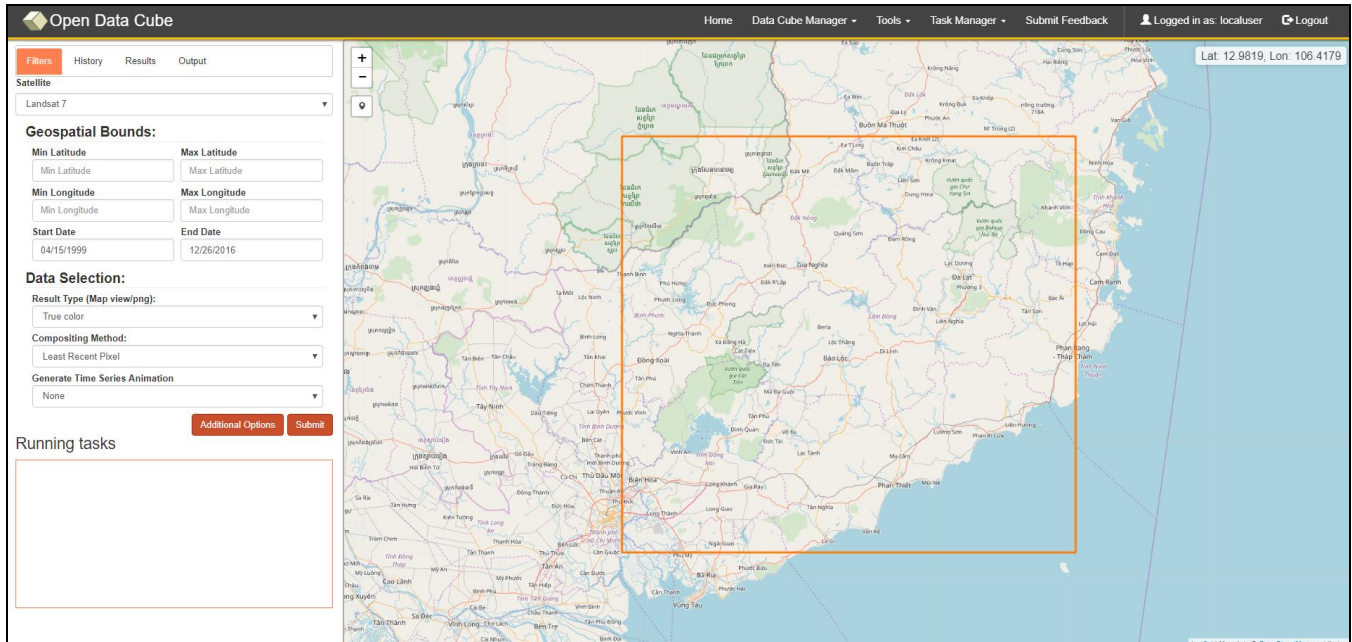


Figure 1. Web User Interface (UI) of CEOS Data Cube (CDC).

2. HIGH-LEVEL ARCHITECTURE

We have developed an interface that includes descriptions of our datasets, access to the datasets themselves, and instructions for the use of the data, as well as an administrative interface to manage the data and presentation of the UI itself. The fully-customizable source code of the UI is available at our public repository [3]. Interested parties can download the source and build their own UIs. The UI runs on the Ubuntu operating system and the source code is publicly available under the Apache License, Version 2.0. The CDC UI (along with the ODC core) utilizes a number of different software frameworks, including Python, JavaScript, PostgreSQL, and the Django web framework. AWS has been used as a one-stop solution for web hosting, parallel and distributed processing, and data storage, distribution, and analysis.

We have developed a UI based distributable asynchronous task processing system using a combination of Celery, Redis, and Django (Figure 2). Analysis cases are split into manageable chunks which are then processed by any number of Celery workers across any number of machines. Celery is a task queue with focus on real-time processing, while also supporting task scheduling. Redis is a message broker, handling task management between Django and Celery. Our AWS hosted UI currently consists of three worker nodes that can be spun up or shut down on demand, automatically adding or removing their workers from the pool. Using Celery/Redis, load balancing and job

scheduling are done automatically based on the number of workers subscribed to the cache.

The architecture of CDC web-based UI is modular and extensible, and the technology at the back-end, primarily the ODC, is agnostic to the original source of the data. As a result, new algorithms can be added to the system as they are developed, new bands can be accommodated when they are offered, and new satellite constellations can be incorporated when they come into operation. In addition, the system can process images from sources that go beyond optical satellites, such as Synthetic Aperture Radar (SAR) data, Digital Elevation Models (DEM), Global Precipitation Measurement (GPM), etc. This offers significant flexibility to the end users in expanding the usability of the tool, and utilizing other imaging systems that may be available to facilitate data analysis tasks.

3. INGESTED REGION VISUALIZER

One of the most frequently asked questions by the user community regards the regions (and their metadata) that are ingested in the system. The “Data Cube Visualization” feature displays each of the ingested data cubes that are available on the system. Filtering options allow the user to hide or show specific subsets of ingested data. Clicking a region on the map populates the lower panel with cards providing additional information about the selected regions. Selecting one of the cards creates a pop up with specific details about the area and a link where a sample cube can be created based on that area.

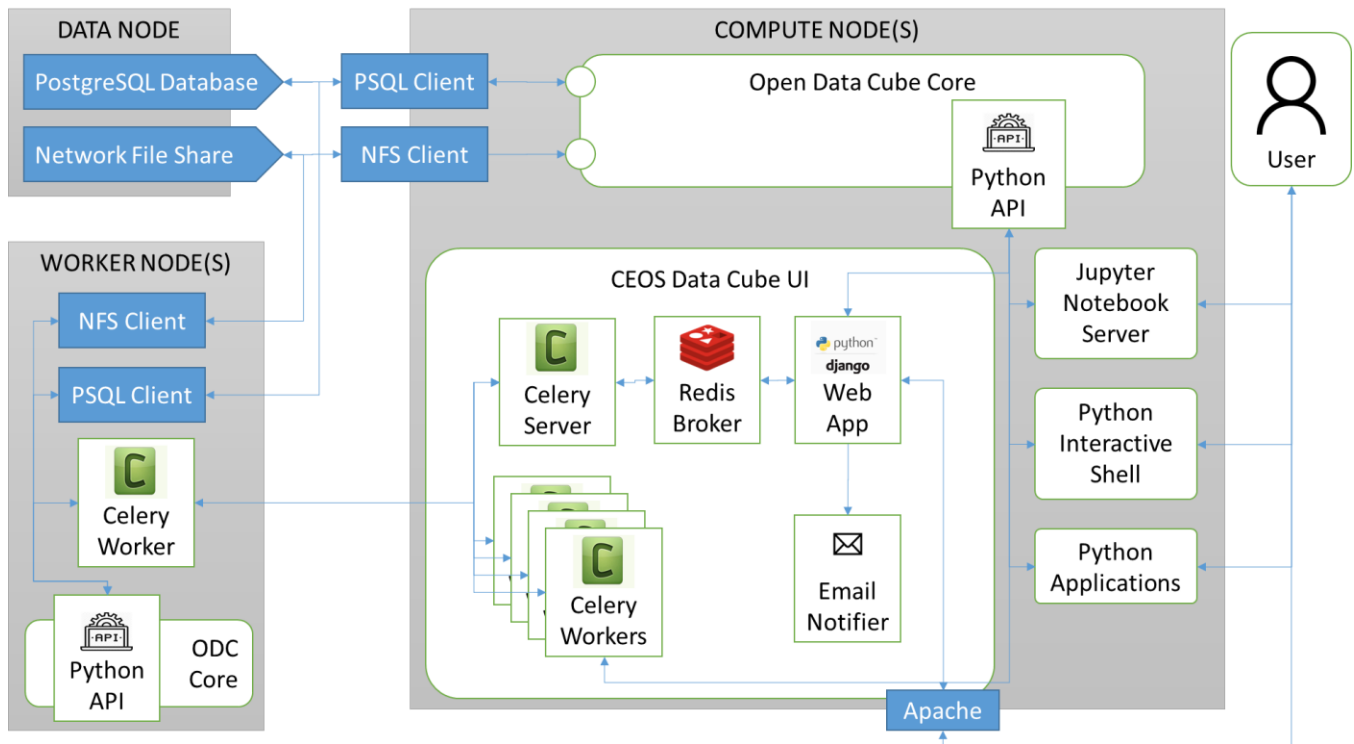


Figure 2. High-level architecture of the system.

4. REMOTE SENSING APPLICATIONS

Table 1 enumerates the applications available in CDC UI as of this writing, along with their contributing agencies.

TABLE 1: CDC applications and their contributors

Application	Purpose	Agency
WOFs (modified)	Water detection based on Australian WOFs algorithm.	GA, CEOS
Custom Mosaic	Cloud-free mosaic based on geomedian, median, max/min, recent pixel	CEOS
Fractional Cover	Land cover (bare soil, PV, NPV) fraction based on J. Gerschman algorithm	CSIRO, CEOS
PyCCD	Land change detection based on Boston University. and USGS PyCCD algorithm	USGS, CEOS
NDVI Anomaly	Change in NDVI	CEOS
SLIP	Landslide risk detection based on Dalia Kirschbaum algorithm	CEOS, NASA GSFC
Coastal Change	Change in coastline position based on Australian WOFs	CEOS, GA
TSM	Total Suspended Matter (water quality)	CSIRO, CEOS

Figure 3 illustrates a sample output analyzing urbanization around the Hồ Trĩ An Lake Area in Vietnam in 2015. The image is a false-color composite with the NDVI, NDWI, and NDBI indices mapped to red, green, and blue, respectively.

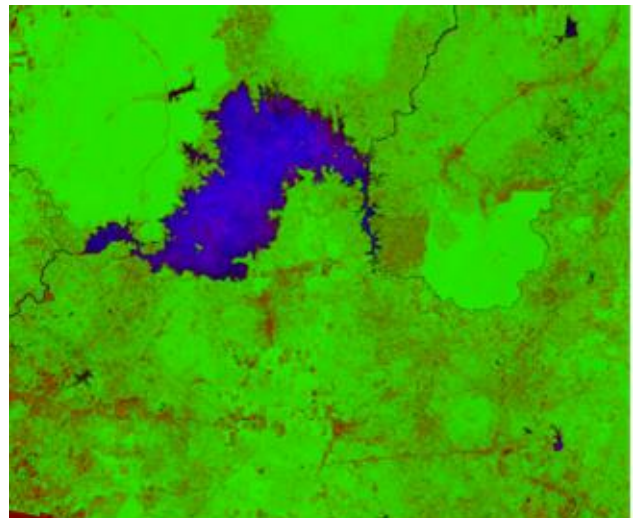


Figure 3. Visualization using CDC UI: Urbanization around the Hồ Trĩ An Lake Area in Vietnam in 2015.

5. TOWARD DATA CUBE ON-DEMAND

The world of computing is undergoing a rapid shift, from a product focus to a service orientation, as we increasingly access services through the cloud. The concept of Data Cube On-Demand follows the same philosophy. The Data Cube On-Demand mechanism would allow users to go to a web-based interface where they would specify a region of interest, time period, pixel spatial resolution and grid projection. The desired data will then be discovered automatically, returning a data cube which can then be downloaded to a local machine or moved to the cloud.

An innovative feature called "Create a Sample Data Cube" is available through the CDC UI. With it, a user can create and download a sample Data Cube for specific small-scale analysis cases. The user can enter a source product, bounding box, time range, and storage attributes to launch a custom ingestion-on-demand process to create their cube. The easiest way to create an ingestion request is to use the Dataset Viewer page in order to search for storage units in one of the *source only* datasets.

The screenshot shows the 'Open Data Cube' interface for creating a sample data cube. It features two main sections: 'Dataset Metadata' and 'Measurements'. The 'Dataset Metadata' section contains a dropdown for 'Source Dataset Type' (set to 'ls7_ledaps_vietnam - Ingested only'), a 'Create Sample Cube' button, and input fields for 'Start Date' (1999-09-08), 'End Date' (2016-12-29), 'Min Latitude' (9.1762906272858), 'Max Latitude' (13.9649399123443), 'Min Longitude' (102.404169465487), 'Max Longitude' (108.931058825317), and 'CRS' (EPSG:4326). The 'Measurements' section is a list of measurement types with checkboxes and delete icons: blue, green, red, nir, swir1, swir2, atmos_opacity, pixel_qa, radsat_qa, cloud_qa, solar_azimuth, solar_zenith, sensor_azimuth, and sensor_zenith.

Figure 4. Sub-setting a data cube via the CDC UI.

6. CONCLUSION

The AWS-hosted CDC UI plays a critical role in demonstrating how the Open Data Cube can take advantage of the AWS infrastructure and exploit open datasets in order to achieve several useful remote sensing applications such as water detection, custom mosaics, land cover change detection, etc. To date, our web UI based distributable asynchronous task processing system using Celery, Redis,

and Django has been instrumental in garnering participation from 39 additional countries with an interest in using the open source common architecture to manage their satellite data and take advantage of community algorithms and tools. The CDC has aided in developing the concept of how exporting a subset of an ingested dataset can be leveraged in order to partially achieve Data Cube On-Demand. The architecture of CDC UI is modular and extensible, and the technology at the back-end, primarily ODC, is agnostic to the sources of the data. This offers significant flexibility to the end users in expanding the usability of the tool, and utilizing other imaging systems that may be available. The web interface has been presented at multiple conferences, tutorials, training sessions, and international presentations.

7. ACKNOWLEDGMENT

The authors would like to acknowledge the efforts of the entire CEOS SEO development team on the development of the CEOS ODC web-based UI and associated applications.

8. DISCLAIMER

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

9. REFERENCES

- [1] Open Data Cube Website: <https://www.opendatacube.org>.
- [2] Open Data Cube GitHub Repository: <https://github.com/opendatacube>.
- [3] CEOS Data Cube User Interface GitHub Repository: https://github.com/ceos-seo/data_cube_ui.
- [4] The 1st CEOS Open Data Cube Workshop: <http://ceos.org/home-2/1st-ceos-open-data-cube-workshop/>.
- [5] A. Narain, "GEO on a strategic mission to deliver insights for a changing world," GeoSpatialWorld, Dec-2017. [Online]. Available: <https://www.geospatialworld.net/article/geo-week-2017>. [Accessed: 03-Jan-2018].
- [6] A. Singh, "New satellite data sharing system, Vietnam Data Cube, introduced," GeoSpatialWorld, Mar-2018. [Online]. Available: <https://www.geospatialworld.net/news/new-satellite-data-sharing-system-viet-nam-data-cube-introduced>. [Accessed: 15-Mar-2018].
- [7] Global Humanitarian Assistance Program (Development Initiatives), "New satellite technology tool transforms central Africa's ability to manage food security," ReliefWeb, Mar-2018. [Online]. Available: <https://reliefweb.int/report/kenya/new-satellite-technology-tool-transforms-central-africas-ability-manage-food-security>. [Accessed: 30-Mar-2018].