The Open Data Cube Initiative

The Open Data Cube (ODC) initiative seeks to increase the value and impact of global Earth observation satellite data by providing an open and freely accessible exploitation architecture and to foster a community to develop, sustain, and grow the technology and the breadth and depth of its applications for societal benefit.

Background

With each passing year, new generations of EO satellites are creating increasingly significant volumes of data with such comprehensive global coverage that for many applications, the lack of data is no longer a limiting factor. Extensive research and development activity has delivered new data applications that offer significant potential to deliver great impact to important environmental, economic and social challenges, including at the local, regional and global scales. Such applications highlight the value of EO, though the challenge is in providing the proper connections between data, applications and users. Even today, much archived EO satellite data is underutilised despite modern computing and analysis infrastructures.

Addressing this challenge is difficult for advanced economies and even more challenging for developing countries with an interest in using EO satellite data. It is simply not technically feasible or financially affordable to consider traditional local processing and data distribution methods (e.g. scene-based file download over the internet) to address this "scaling" challenge in many economies, as the size of the data and complexities in preparation, handling, storage, and analysis remain significant obstacles.

Fortunately, just as satellite Earth observation technology has advanced significantly, so too has information technology. The data management and analysis challenges arising from the huge increase in free and open data volumes can be overcome with new computing infrastructures, technologies and data architectures, such as the "Open Data Cube". Such a solution has great potential to streamline data distribution and management for providers while simultaneously lowering the technical barriers for users to exploit the data to its full potential.

Purpose

The Committee on Earth Observation Satellites (CEOS) is a founding partner in the Open Data Cube (ODC) initiative which seeks to provide a data architecture solution that has value to its global users and increases the impact of EO satellite data. Technologies such as the Australian Geoscience Data Cube (AGDC) and Google Earth Engine (GEE) have transformed the EO satellite data user community. In response to user demand, such technological solutions remove the burden of data preparation, yield rapid results, and foster an active and engaged global community of contributors. Hence, CEOS is committed to stewarding and contributing to the ODC architecture as part of the ODC community. They seek to encourage others to join the initiative with an ultimate goal to meet the targeted needs of users, similar to the objectives of the AGDC and GEE, but differing in implementation.

As the world develops, so does its knowledge of, and demand for EO satellite data. The primary problems for users are data access, data preparation, and efficient analyses to support user applications. CEOS, through its network of global connections, has determined that global users share many common needs that can be met through the ODC initiative. Some of those needs are listed below.

- Minimize time and specialised knowledge required to access and prepare satellite data
- Free and open EO satellite data and application algorithms
- Open source software solutions that are advanced through community contributions
- Consistent data architectures that allow sharing of code, tools and algorithms
- Efficient time series analyses to support land change applications
- Use of multiple datasets together (e.g., interoperability and complementarity)
- Use of common GIS tools (e.g. QGIS, ArcGIS)
- Local and regional solutions that avoid commercial and internet dependence
- Sustained customer service and user support

Vision

The objective of the ODC is to increase the impact of satellite data by providing an open and freely accessible exploitation tool, and to foster a community to develop, sustain, and grow the breadth and depth of applications. This solution intends to support key objectives, which include building the capacity of users to apply EO satellite data and to support global priority agendas, such as those found in the United Nations Sustainable Development Goals (UN-SDG) and the Paris and Sendai Agreements. In order to ensure success, the ODC must establish a "brand" that users can trust and it must promote a positive user experience. This should be made possible through the development of an open source ODC community that is actively engaged and contributes to the core code, shares algorithms and provides support to each other for the resolution of problems.

With this vision, comes challenges, such as acceptance and scalability. Will EO satellite data providers and key global stakeholders agree that the ODC vision is achievable and worth their contribution? CEOS has utilised prototype efforts to demonstrate the functionality and impact of the ODC and has reached out to global stakeholders (e.g. GEO, World Bank, SERVIR, AWS) to explore common objectives. Global interest in the ODC initiative has been promising and continues to make progress toward "acceptance" as a trusted solution which users value.

How does this solution scale to meet the demand of many global users given limited resources? Scalability is one of the key challenges faced by the ODC. Though initial efforts have only resulted in several national-scale Data Cube implementations (e.g. Australia, Colombia, Switzerland), there are many more countries and global organizations with high interest in the ODC. Through engagement with stakeholder organizations and current users, it is expected that the number of ODC implementations will increase and the ODC community will flourish with expanding contributions and impact. Such open source solutions will also take advantage of prior deployments for increased efficiency by reusing content and applying "lessons learned" for new deployments.

Implementation

We have now defined the purpose and vision for the ODC, but none of it is possible without an implementation plan. How do we reach these goals and build and sustain an ODC architecture and community? The initial steps include:

- Create and manage an open source repository (e.g. GitHub) to host ODC content
- Develop documentation and open source code for: installation, preparation of analysis-ready data, creation of Data Cubes (ingestion), and common applications (e.g. cloud-filtered mosaics, water detection, land change detection)
- Develop an ODC Learning Center with documentation and training tools
- Create an online forum where users can share Data Cube code, application algorithms, and access community support

For more information, visit: www.opendatacube.org