The CEOS Open Data Cube Initiative

A white paper to summarize the purpose, vision and implementation steps required to achieve an Open Data Cube (ODC) architecture and community that provides value to global Earth observation (EO) satellite data users and increases the impact of EO satellite data.

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) has started the Open Data Cube (ODC) initiative to provide a data architecture solution that has value to its global users and increases the impact of EO satellite data. Similar data architectures, such as Google Earth Engine (GEE), have transformed the EO satellite data user community. Their response to user demand has created a technological solution that removes the burden of data preparation, yields rapid results, and maintains an active and engaged global community of contributors. The ODC needs to duplicate many of these outcomes, but not the infrastructure. Hence, CEOS is committed to sustain and grow the ODC to meet the targeted needs of users in ways that differ from solutions such as GEE.

As the world is developing, so is its knowledge 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 scientific 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 purpose of the ODC is to provide a solution that has value and increases the impact of satellite data. Our goal is not to sell a product or give out another tool. CEOS believes the ODC is a solution that supports its key objectives, which includes the desire to build the capability of users to apply their 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 addition, CEOS Agencies desire to participate in the ODC through the provision of Analysis Ready Data (ARD) and by contributing to the development of the ODC architecture and its uptake.

In order to ensure success, the ODC must be a known "brand" that users can trust and it must promote a positive user experience. This is only 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 CEOS is making progress toward "acceptance" of the ODC as a trusted solution with user value.

How does this solution scale to meet the demand of many global users given limited resources? This scalability challenge is difficult and one of the most criticized elements of the ODC. Though our initial efforts have only resulted in 3 national data cubes (e.g. Australia, Colombia, Switzerland), there are many more countries with high interest. It is our goal to reach 20 country-level operational data cubes by 2020. This goal will not be achieved by CEOS alone, but will engage a number of other stakeholder organizations and current users to achieve the goal.

Through partnerships with global stakeholders, it will be possible to utilize their funds and existing projects to deploy data cubes in countries and demonstrate their potential to meet the needs of their projects, both efficiently and effectively. For example, World Bank has a high interest in the ODC to support a number of global water management projects. These projects use satellite data, but there is a lack of consistency and significant challenges with data preparation. Similarly, SERVIR is investigating the use of data cubes at its global regional data hubs. Should they determine the ODC is an acceptable and beneficial option for their local users, this would result in many more country-level data cubes, as their

hubs serve many countries. Through our current national data cube users, we can utilise their skills and lessons to support deployments in other countries. For example, Switzerland is considering a future capacity building project that may result in two new country-level data cubes where they will provide the user support for deployment and training. Finally, CEOS, through its Agencies and supporting groups (e.g. CEOS Systems Engineering Office) will support deployment and training for new country data cubes. Most recently, CSIRO has invested in a Python programmer to support Australian industry use of their data cube and to support data cube deployment to other countries in the Asia-Australia region.

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 our goals and build and sustain an ODC architecture and community? These steps are listed below.

- Develop instructions and open source code to obtain and/or produce ARD
- Create and manage an open source repository (e.g. GitHub) to host ODC content
- Develop an ODC installation package with sample inputs and outputs
- Develop an ODC Learning Center with documentation and training tools
- Create a user forum to share code and algorithms and obtain support
- Host common application algorithms desired by users (e.g. mosaics, water detection, change detection)
- Engage stakeholder organizations (e.g. GEO, World Bank, SERVIR) to identify country needs and provide infrastructure and funds for implementation
- Conduct pilot studies to test approaches and provide user feedback
- Develop a group of globally distributed Python programmers who are able to support local deployment and customer support

Conclusion

CEOS is the ideal group to develop and sustain an Open Data Cube solution. CEOS has the data, they understand how to produce ARD, they have the technical expertise (e.g. Agencies, SEO, Working Groups), they have the connections with key international groups and stakeholders, and they have proven that it can work through the Australian Geoscience Data Cube (AGDC) and recent functioning ODC prototypes in Colombia and Switzerland. Though acceptance and scaleability remain as the primary challenges, the vision of the ODC can still be accomplished. Such success will increase the use of EO satellite data and provide unquestionable value to global users.

For more information, visit: www.opendatacube.org