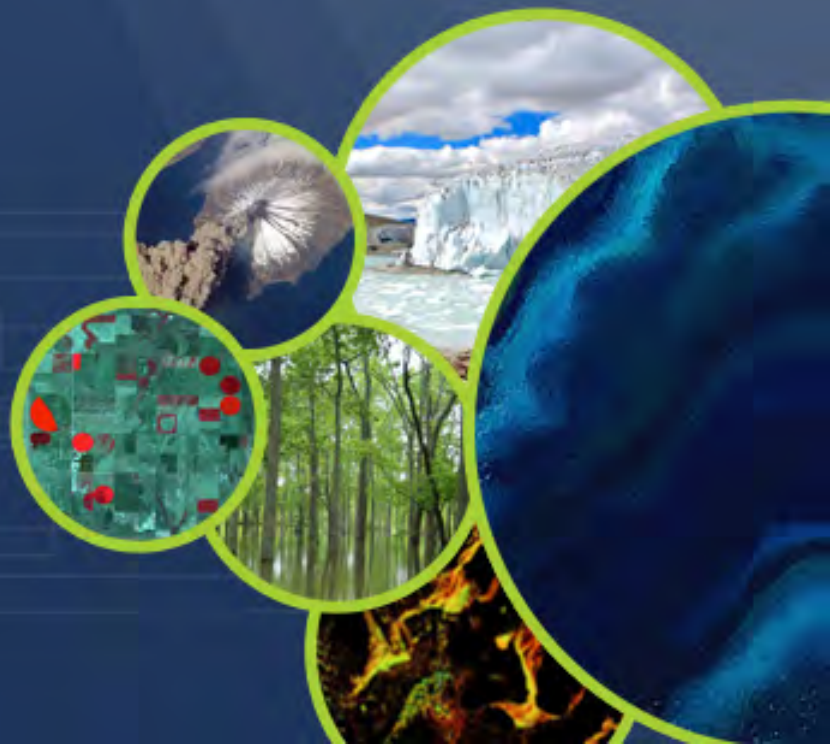
Status of the Open Data Cube

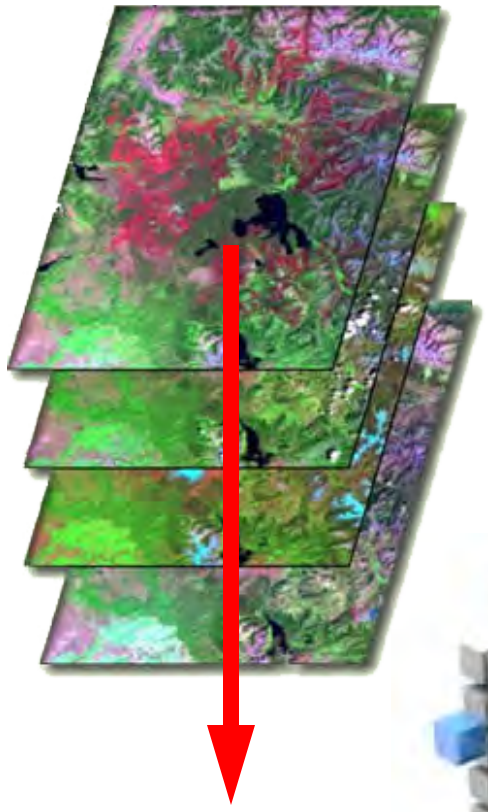
Brian Killough
CEOS Systems Engineering Office (SEO)

WGISS-44 Meeting
September 27, 2017



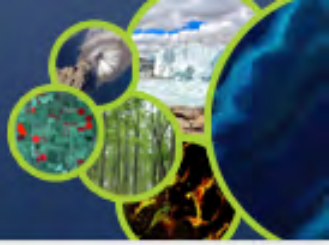
What are Data Cubes?

- **Data Cube** = Time-series multi-dimensional (space, time, data type) stack of spatially aligned pixels ready for analysis
- **Proven concept** by Australia with plans for global implementation
- **Analysis Ready Data (ARD) ...** Dependent on pre-processed products to reduce the burden on users
- **Open source** software approach allows free access, promotes expanded contributions, and increases data usage.
- **Unique features:** exploits time series, increases data interoperability, and supports many new applications.



Data Cubes *are a popular topic … "The Data Cube Manifesto" (Peter Baumann, EU) and "The Six Faces of the Data Cube" (Peter Strobl, EC)*

Benefits of Data Cubes



- Expanded use of CEOS satellite data … expanded user base
- Reduced processing burden .. dependency on ARD
- Enhanced interoperability … improved by MRI
- Efficient time series analyses
- Free and open access
- Flexible deployment (local or cloud)
- Use of a common architecture
- Community development and sharing … via GitHub

Our goal is **NOT** to sell a product or distribute a tool. Our goal is to provide a **SOLUTION** that has **VALUE** and increases the **IMPACT** of satellite data.

Open vs. CEOS Data Cubes

- The **ODC** initiative is larger than CEOS.
- The **Open Data Cube** (ODC) initiative was established by CEOS, with a goal to create and foster an open “community” of contributors.
- The **ODC** uses a common architecture among the various implementations so that all users can share tools and applications.
- The **CEOS Data Cube** (CDC) is one “implementation” of the ODC. Similarly, Digital Earth Australia (DEA) and USGS Land Change Monitoring, Assessment, and Projection (LCMAP) are implementations.
- The **CDC** goal is to focus on building global capacity to utilise satellite data and contribute to global initiatives (e.g. UN-SDG, GFOI, GEOGLAM) through the use of Data Cubes.

CEOS Data Cube Vision

A solution supporting CEOS objectives …

- Build capability of users to apply **CEOS satellite data**
- Supporting priority CEOS/GEO agendas and SDGs

CEOS Agencies wanting to participate …

- Through provision of **CEOS Analysis Ready Data (ARD) products**
- Contributing to development and uptake of solutions

Customer focused …

- Training materials and easy installation/maintenance
- A brand that people know and trust
- An active community of users

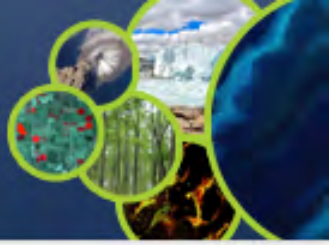
Scalable solution …

- Operational Data Cubes in **20 countries by 2022**
- Key partners (e.g. GEO, World Bank) supporting data cube projects

The "Road to 20" Operational Data Cubes by 2020

3 operational, 4 under development, 22 under review = **29 total**

The "Road to 20" Highlights



- **Colombia** has an operational Data Cube since Dec 2016 with over 25,000 historic Landsat images. They continue to expand the user base, applications and datasets. The Colombia Data Cube won the National Environmental Award of Colombian Society of Engineers in May 2017 and has been approved by the Colombia Government into 2018.
- **Switzerland** has an operational Data Cube since July 2017 with over 4,000 historic Landsat images. They have received Swiss government approval and developed a new website (swissdatacube.org). Their future plans include expanded datasets (Sentinel) and increased applications with both government and university involvement.
- **Vietnam** is slowly making progress by establishing pilot cubes in several regions using a new high performance computing system. Their focus is on forests, rice, and water applications. VNSC is hosting an internal Data Cube Workshop on Sept 17.
- **Taiwan** is making progress on a local HPC installation through support from CSIRO. Their focus is forests and water applications.
- **Uganda** has received support from the U.K. to install a demo cube for the Karamoja region on a cloud (AWS). They have made rapid progress with little CEOS support.
- See the "**Road to 20**" document on the ODC website for more!

Other Highlights

- Plans are in place to **leverage** the experience of several operational implementations to expand the presence of Data Cubes …
  - Switzerland >>> Georgia, Moldova
  - U.K. >>> Solomon Islands, Vanuatu, Nauru
  - Taiwan >>> Hondurus
- We are making progress with **World Bank** to support the deployment of a Data Cube in Uruguay to support an agriculture and water quality project with direct links to DINAMA (UN-SDG statistical agency).
- 4 Data Cube side events are planned for **GEO-17** on Oct 23-24. Each 1.5 hour segment will have a different Data Cube topic.
- Future **outreach opportunities** at Pecora-20 (USGS) and IGARSS-2018 (July in Valencia, Spain).

ODC Progress

- We have established an "**ODC Partners**" group which includes representatives from the NASA-SEO, GA, CSIRO, USGS, and UK-Catapult.
- We have established an "**ODC Steering**" group which include technical representatives from NASA-SEO, GA, CSIRO, and USGS.
- We have established a new website https://opendatacube.org
- We have developed "white papers" for the ODC and CDC that describe the goals of each initiative and an ODC governance document for code management.
- We conducted the first ODC Workshop at the recent IGARSS conference in Fort Worth, Texas, USA in July.
- We are planning the 2nd Annual ODC Technical Meeting in Canberra, Australia on Feb 14-16, 2018.

Technical Progress

- The CDC has established detailed content to support Data Cube deployments
  - **Installation** – system requirements, installation guide
  - **Data Preparation** – ARD guidance, data acquisition guidance
  - **Data Cube Creation** – ingestors for all popular datasets
  - **Applications** – AWS demo, Python notebooks, growing list of algorithms
  - **Forum** – discussion groups for user support
- Data Cube ingestion has demonstrated significant reduction in data storage requirements when comparing the ingested Data Cubes to the original data.
  - **Landsat** = 3x to 7x reduction (varies with data parameter selections). For example, a 1-deg x 1-deg x 1-year Landsat Data Cube is ~900 MB.
  - **Sentinel-1 GRD** = 6x reduction (based on 30m grid, VV and VH only)

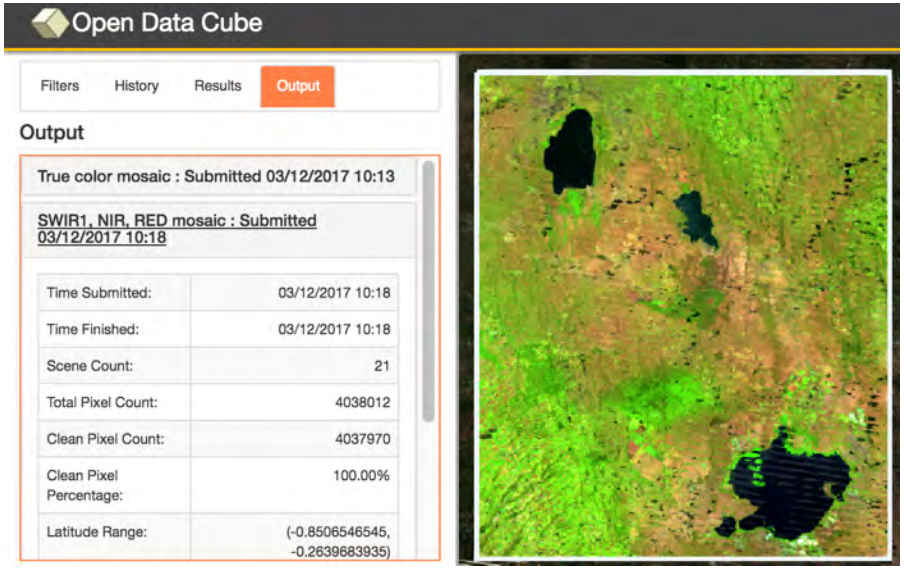
Amazon (AWS) Demo Portal

Data Cubes

- 16 cubes with 10+ years each.
- Kenya, Cameroon (Lake Chad), Togo (coastal Africa), Ghana, Colombia, Tonga (Pacific Island), Vietnam, Australia (Menindee Lakes), Bangladesh.

User Interface Features

- 9 applications: cloud coverage maps, custom cloud-free mosaics, fractional cover, NDVI anomaly, water detection, water quality, landslides, coastal change and urbanization.
- Outputs in GeoTIFF and GIF animation.
- **New features** added in Sept 2017: data visualization tools, ingestion "on demand" for new cubes or subsetting, indices, mosaics (medoid, geometric median)



This is the first "hands-on" global demo of the Data Cube to show its potential for rapid time series analysis and diverse applications

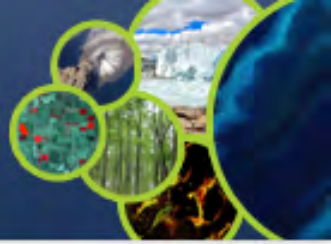
http://tinyurl.com/datacubeui

Free and Open!

Near-term Developments

- Develop a new **QGIS tool plugin** with a web-based (WCS) connection to a Data Cube hosted on AWS (cloud storage). This will be ready by Nov 2017.
- Develop and test sample **iPython Notebooks** on AWS to demonstrate interactive Data Cube applications and programming simplicity
- Test the **PyCCD** land change detection algorithm with radar datasets
- Develop and test a new **Water Quality algorithm** from Tony Vodacek (Landsat Science Team) based on a Look-Up-Table approach to infer Chlorophyll-A, CDOM and TSS concentrations.
- Test **Sentinel-1 GRD and SLC** cubes with the Random Forest land classification clustering algorithm

Lessons Learned



Through our initial country interactions, we have learned a number of **lessons** …

- Country users should have some Python programming skills
- It is important to clearly understand country needs and to guide them toward the needed satellite data and application tools
- It is important to maintain consistent customer communication (both face-to-face and remote) to sustain deployment progress and build trust
- It is important to utilise relationships with investment banks (e.g. World Bank) and GEO to increase access to country contacts and facilitate deployment
- The ODC community needs to continue to grow and expand to build confidence towards desired outcomes and to build the supply of open source tools and applications