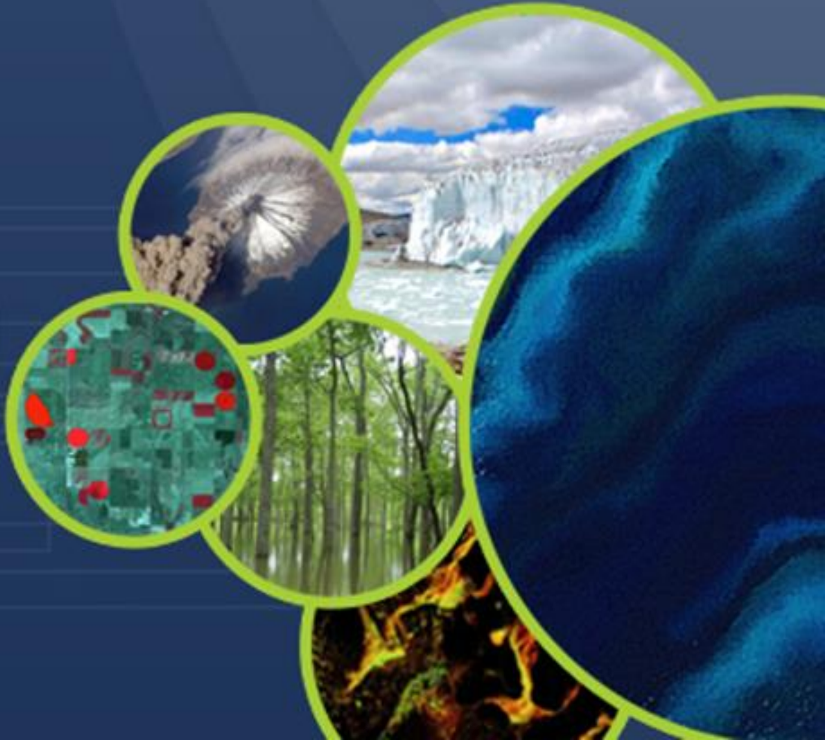
Committee on Earth Observation Satellites

Cloud Use Activities and Analysis

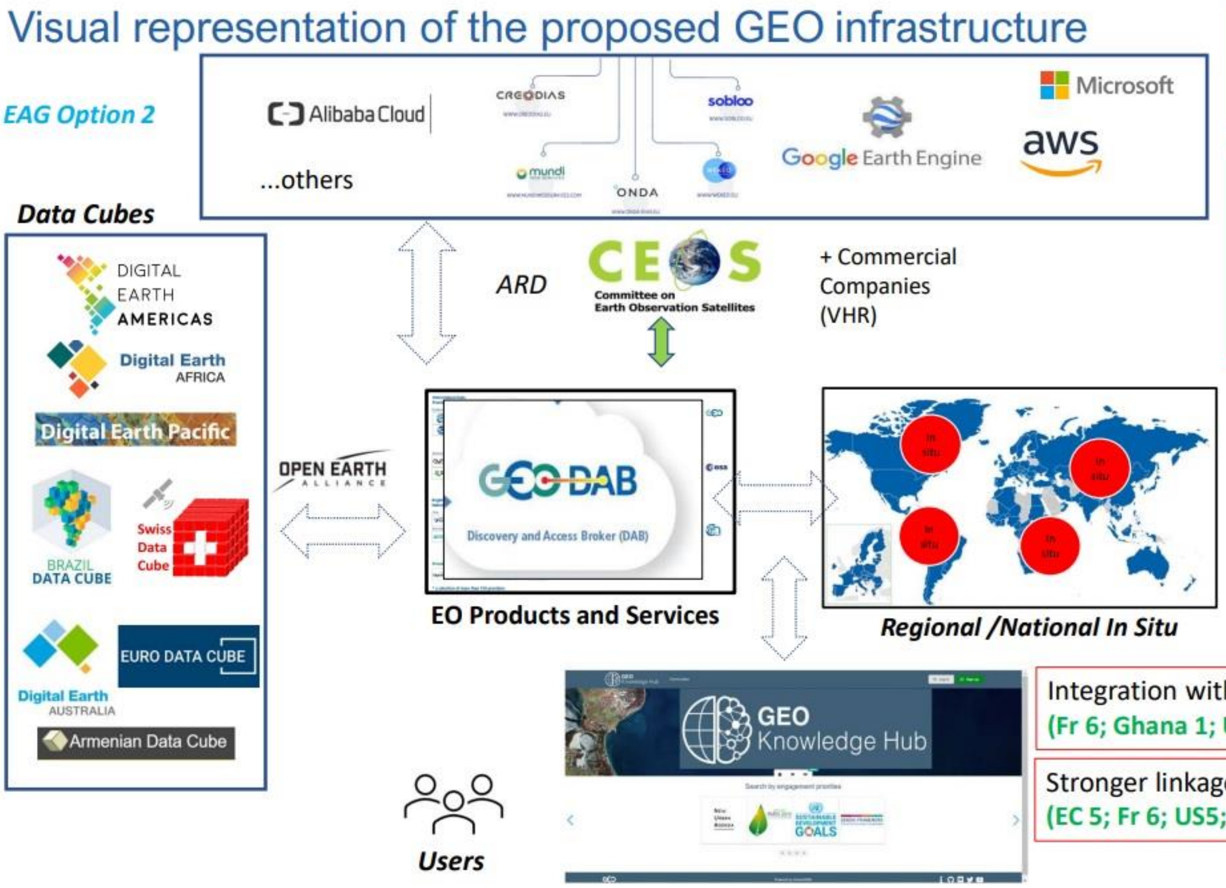


David Borges
CEOS Systems Engineering Office (SEO)
WGISS-55, Cordoba Argentina
18 April 2023

Motivation for a Cross-Cloud Study

- **Pecora 22 Workshop** - Session 2A - Implementing ARD – Common Approaches, Predictability, and Improving EO Data Interoperability
  - Key Finding #1 - NASA should capture lessons learned from its recent experiences with the three big cloud service providers (Google, Amazon, and Microsoft) – what are the advantages and disadvantages for using each one - and share them with other public EO agencies
- Request from GEO to CEOS SEO to inform the evolution of the **GEO Common Infrastructure (GCI).**
  - GEOSS Infrastructure and Data Task Team (GIDTT)
  - Existing Data Cubes are connected to the GCI via Open Earth Alliance (GEO Community Activity to promote Data Cubes)
- Improving the **User Experience**
  - Understand qualitative and quantitative differences in cloud providers
  - Leverage knowledge to improve community experiences
  - Develop plans for future benchmarking and optimizations
- Determine how SEO can support **CEOS Interoperability Roadmap**

GEO Common Infrastructure (GCI)



Global Governance with Representative of all components

Improve access to Datasets; Models; Computing Capabilities
ESA1,2; EC2,3;US 4; CH 3; China 1

Improve access to In situ (US 3)

Integration with GEO knowledge Hub
(Fr 6; Ghana 1; US 6; Costa Rica 6)

Stronger linkage with GWP
(EC 5; Fr 6; US5; China 2)

Cloud Computing Prototypes

The SEO has been testing several cloud computing frameworks to understand CEOS data access and technology capabilities.

<u>Supported Environments and Services:</u>

- **Google** – we use the Google Cloud (paid), Colab (free notebook platform, but limited) and Earth Engine (free satellite datasets).
- **Amazon** – we use the AWS Cloud (paid), SageMaker (free notebook platform, but limited), and the AWS Open Data Catalog (free satellite datasets).
- **Sentinel Hub** – we are working with Sinergise to test ODC integration with the Sentinel Hub via CreoDIAS (European cloud provider).
- **Microsoft** – we use the Microsoft Azure Cloud (paid), Azure Labs (free notebook platform, but limited) and the Planetary Computer Data Catalog (free satellite datasets).

Software Instances

Platform	vCPU	RAM	Storage	Operating System
AWS (EC2)	4 cores	16 GB	20 GB	Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-1022-aws x86_64)
Google (Compute Engine)	4 cores	16 GB	30 GB	Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-1023-gcp x86_64)
Azure (Virtual Machine)	4 cores	16 GB	30 GB	Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-1020-azure x86_64)

Software installations … We are using virtual private server machines. Docker Engine and docker-compose were installed in the instances. Docker container with JupyterHub and Postgres with PostGIS extensions was used for the analysis.

Data Source: Sentinel-2

Platform	Data Source
AWS EC2	S3 Bucket Digital Earth Africa Sentinel-2 Level-2A (arn:aws:s3:::deafrica-sentinel-2)
GCP Compute Engine	Google Earth Engine Data Catalog Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (ee.ImageCollection("COPERNICUS/S2_SR"))
Azure Virtual Machine	Microsoft Planetary Computers Sentinel-2 Level-2A (sentinel2l2a01.blob.core.windows.net/sentinel2-l2)

AWS = Amazon Web Services

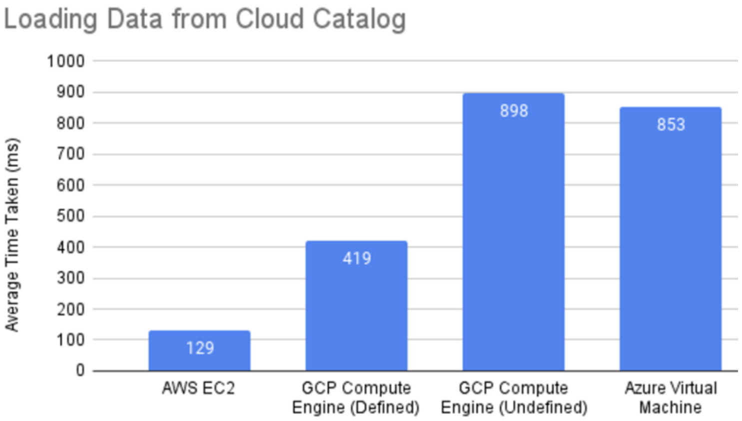
GCP = Google Cloud Platform

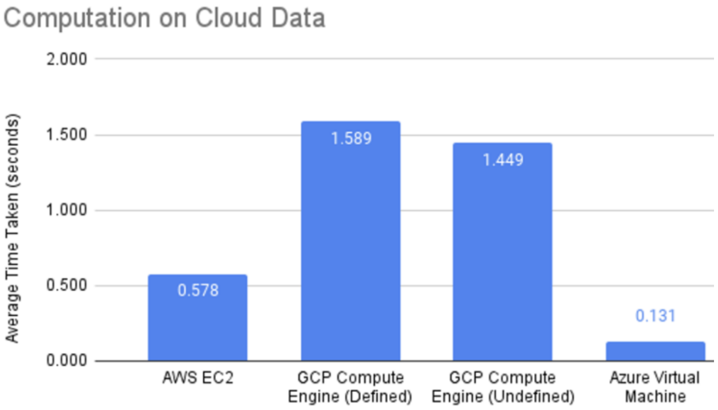
Early findings …

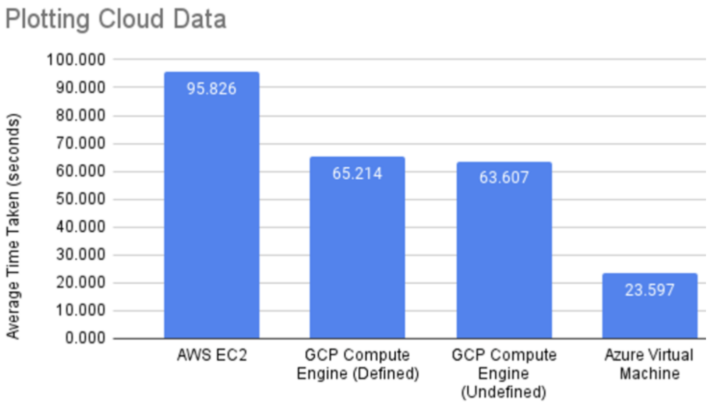
- Definitive differences in performance based on testing of standard options.
- This is a simple benchmark intended to fortify future refined benchmark methodologies.
- Requires larger statistical experiments over time to truly determine accuracy of benchmarks (daily service volume can vary, new systems can come online, etc.)

Benchmarking with Jupyter Notebooks

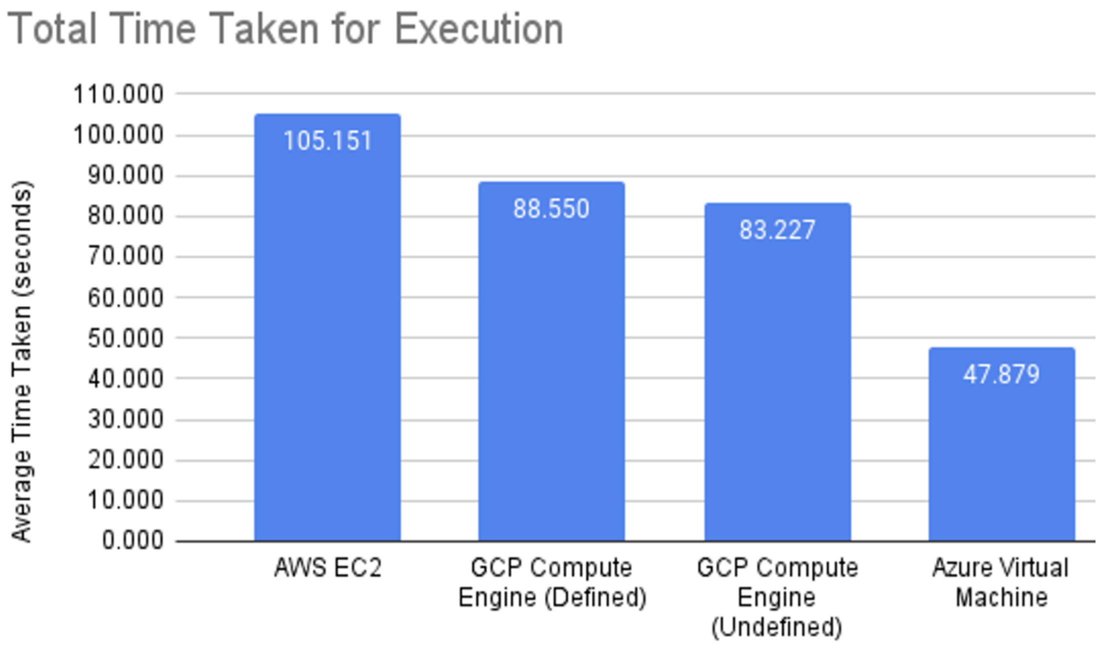
The following operations were tested for each cloud provider: data loading in an xarray, calculation of spectral indices (NDVI, NDWI, MNDWI), and plotting.







Benchmarking with Jupyter Notebooks



Disclaimer … this is the first known test of all 3 major U.S. cloud services!

Conclusions

- AWS 6x faster data loading than GCP and Azure
- Azure computation 4x faster than AWS and 10x faster than GCP
- Azure plotting 3x faster than GCP and 4x faster than AWS
- Total execution … Azure wins. 2x faster than GCP and AWS.

Satellite Dataset Summary

Dataset	Google Earth Engine (GEE) Datasets	Amazon (Open Data on AWS)	Microsoft Planetary Computer (PC) Datasets
MODIS	Many Level-2 and Level-3 products	Only 5 common land/vegetation products	Many Level-2 and Level-3 products
Landsat	Mission 1-9 (multiple collections and levels)	Mission 1-9 (multiple collections and levels)	Mission 1-9 (multiple collections and levels)
Sentinel-1	GRD (no RTC corrections)	GRD archive (no processing), DE-Africa with RTC	GRD with full RTC corrections (CEOS CARD4L)
Sentinel-2	Level-1 and Level-2 products	Level-1 and Level-2 products	Level-2A only
Sentinel-3	OLCI - 21 bands	OLCI - 21 bands	Multiple sensors and bands
ALOS	Global PALSAR annual mosaics, PALSAR-2 ScanSAR Level 2.2	DE-Africa (PALSAR and JERS), PALSAR-2 ScanSAR Level 2.2	Global PALSAR annual mosaic only
HLS			Only US, Europe and few other locations
Nightlights		VIIRS DNB (2012-2020)	
DEM	Copernicus and NASA 30m	Global 30m	Copernicus and NASA 30m, Copernicus 90m
Mangroves		DE-Africa GMW only	
JRC Water	Global Surface Water (1984-2022)		Global Surface Water (1984-2022)

Summary

- Major CEOS datasets available on all major cloud platforms.
- **Sentinel-1** radar with RTC … only on Microsoft PC
- **ALOS** mosaics and ScanSAR only on GEE
- Partial **HLS** only on Microsoft PC
- Global **JRC water** dataset not on AWS
- **Nightlights** (VIIRS DNB) only on AWS
- **DE-Africa** holds many datasets (S1 RTC, Fractional Cover, WOFS Water, GEOMAD, Chirps Rainfall, GMW, Coastlines)

Cloud Providers: Market Share and Datasets

Rank	Cloud Provider	Market Share	Well-Known Satellite Data Hosted
1	Amazon Web Services (AWS)	32%	Landsat, Sentinel, MODIS, ASTER, etc.
2	Microsoft Azure	20%	Landsat, Sentinel, MODIS, NAIP, CBERS, DigitalGlobe, etc.
3	Google Cloud Platform (GCP)	10%	Landsat, Sentinel, MODIS, ASTER, etc.
4	Alibaba Cloud	6%	
5	IBM Cloud	4%	
6	Tencent Cloud	3%	
7	Oracle Cloud	3%	
8	Salesforce	3%	
9	Baidu Cloud	2%	
NA	CREODIAS Cloud	<1%	Sentinel, Envisat, ERS, Radarsat-2, SMOS, CryoSat-2, Swarm
NA	DigitalGlobe (Maxar Tech)	<1%	WorldView, GeoEye, QuickBird, IKONOS
NA	Airbus OneAtlas	<1%	Pleiades, SPOT, TerraSAR-X
NA	Planet	<1%	PlanetScope, RapidEye, SkySat
NA	Orbital Insight	<1%	Sentinel, PlanetScope, SkySat
NA	Spire	<1%	ADS-B, AIS, GNSS-RO
NA	Descartes Labs	<1%	Landsat, Sentinel, MODIS
NA	SpaceKnow	<1%	Landsat, Sentinel, PlanetScope, SkySat
NA	BlackSky	<1%	SkySat, PlanetScope

- NA: Not Available
- Reference: Satellite Data Services. www.grandviewresearch.com/industry-analysis/satellite-data-services-market

Cloud Project with Sinergise

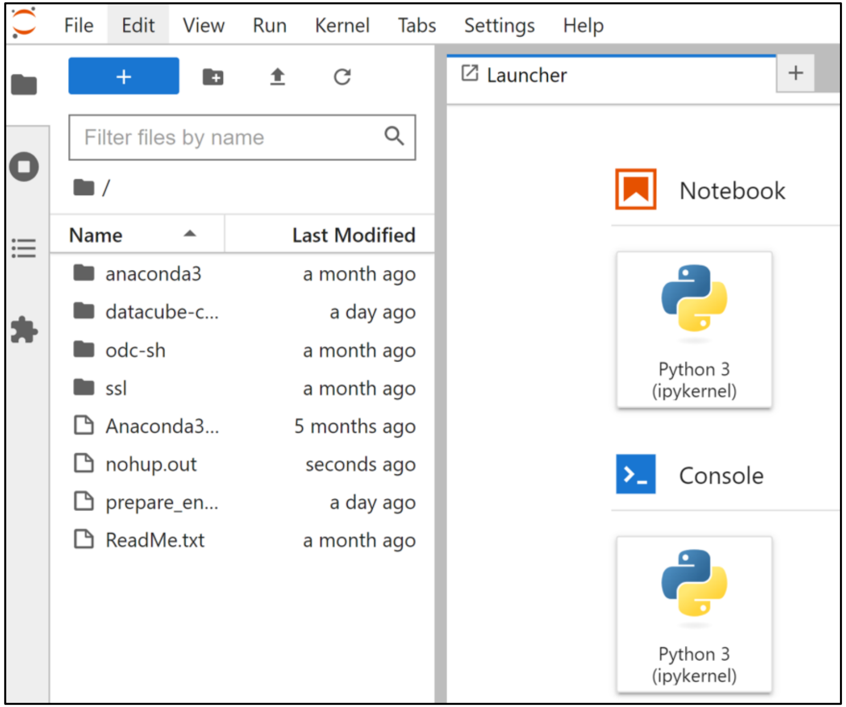
Integration of Sentinel Hub (SH) and the Open Data Cube (ODC)

- We tested the SH-ODC environment on **CreoDIAS**, a cloud-based platform funded by the European Commission that provides access to EO data from the Copernicus program.
- We tested a demo that confirmed our ability to use ODC on **CreoDIAS** and utilize the Sentinel Hub-Open Data Cube (SH-ODC) environment for importing Sentinel Hub datasets to be used in ODC applications.

Sentinel-2 image over Mombasa, Kenya
Produced by an SH-ODC demo notebook using SENTINEL2_L1C data from Sentinel Hub

How SH-ODC Works ...

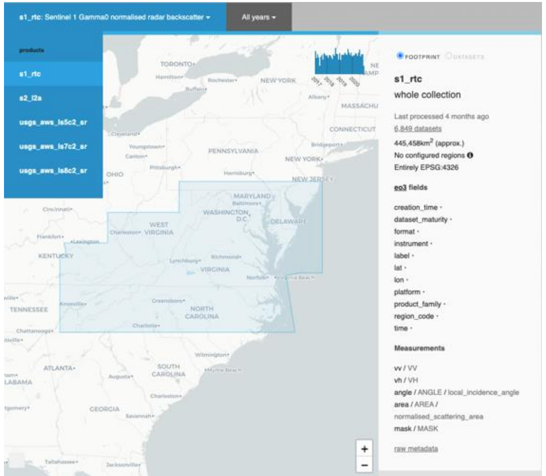
- CreoDIAS made an ODC "image" that spawns some ODC Jupyter notebooks and uses data from Sentinel Hub.
- An account was created in the CreoDIAS cloud environment and testing credits were added to the account.
- The cloud service with the account was configured to use the ODC VM image.
- JupyterLab was installed and configured to access the SH-ODC.



JupiterLab to access the SH-ODC environment

Earth Analytics Interoperability Lab (EAIL)

- Initiated in April 2020 as a CEOS WGISS initiative, **EAIL** is a data and analytics platform that uses **AWS Cloud and Open Data Cube.** Its advantages are Jupyter Hub, Dask scaling, customized ARD pipelines and GPU processing. There are 59 registered users!
- **Jonathan Hodge** (CSIRO-Chile) is the primary EAIL lead and architect. The SEO is working with CSIRO in 2023 to become trained on EAIL operations to support users.
- **EAIL** currently supports one active CEOS project > **COAST** (Chesapeake Bay study). Other projects interested in using EAIL include: **WGCV** (DEMIX Cal-Val campaigns), **DE-Americas** (Caribbean Pilot project), and CEOS **Ecosystem Extent** Pilot Project.
- **Datasets** include: Landsat, Sentinel-2, MODIS, Sentinel-3, Sentinel-1 (CARD4L with RTC), Copernicus DEM, and NASA DEM.



CSIRO Chile

Proposed CEOS Interoperability Roadmap

2023 ⟷ 2025

Framework
Lead: WGISS
- ★ SIT-38, Framework Endorsement
- ★ SIT-TW, Component Work Plans
- ★ Plenary, Framework Roadmap Endorsement
- Stakeholder Engagement
- Maturity Model Development
- Use Cases (Jupyter Notebooks, ODC)
- Capacity Building

Syntactic
Lead: WGISS
- CEOS Common Data Dictionary Updates
- TBD Activity
- TBD Activity
- TBD Activity

Semantic
Lead: WGCV
- TBD Activity
- TBD Activity
- TBD Activity

Data Architecture
Lead: CEOS-ARD OG
- TBD Activity
- TBD Activity
- TBD Activity

Data Accessibility
Lead: WGISS
- CEOS-ARD STAC Extension Defined
- TBD Activity
- TBD Activity

Common References
Lead: WGCV
- TBD Activity
- TBD Activity

Outreach Activities

- ARD23, May, California
  - CEOS-ARD Session
- GEO Open Data Open Knowledge Workshop, June, Geneva
  - Open Satellite Data Session
- IGARSS 2023, July, California
  - CEOS Exhibition Booth
  - Cloud-based Platform Environments for Earth Observation
- GEO Week 2023, November, South Africa
  - CEOS Exhibition Booth