CEOS Interoperability Initiatives

WGDDisasters-14
September 2020

Dr. Robert Woodcock
CEOS WGISS Chair
CSIRO
WGISS Interoperability Initiatives

WIP

• CEOS Interoperability Terminology Report
• CEOS Earth Analytics Interoperability Lab
• Best Practices in:
  • Jupyter Notebooks for EO Analytics
  • Cloud data formats
  • CEOS Data Discovery and Access and Cloud
What is Interoperability?

- CEOS Interoperability Terminology Report (Draft):
  - ...the terms Analysis Ready Data (ARD), interoperability, and harmonization are often used and, to a large extent, used inconsistently
  - Interoperability represents a continuum of compatibility for products, services, algorithms.
  - “Making an unmanageable situation a manageable one” – or keeping it that way.
CEOS Interoperability Terminology Report

Five ARD terms are defined in this document:

1. Analysis Ready Data (ARD)
2. CEOS ARD for Land (CARD4L) Products
3. Interoperable Products
4. Harmonised Products
5. Fused Products
Three Cloud data format terms for ARD:
1. Cloud-friendly
2. Cloud-native
3. Cloud Data Access API

Three types of Analysis Interoperability:
1. Executable code (black box package)
2. Source code
3. Algorithm
Why do we need an Interoperability Lab?

- A significant number of CEOS activities are now engaged in the CEOS ARD and FDA strategies and in Integrated Earth observation data analysis.

- All recognize the broader implications – ARD relates to Discovery and Access, Analytics & Cloud Data Formats.

- Many are looking for advice or implementations. It can be difficult for a single project to cover all viewpoints and technologies.

- Validating interoperability between multiple CEOS organizations and working groups is complex.
The Opportunity

• The first Lab node is hosted by CSIRO and jointly operated by WGISS and SEO to provide:
  • Examples of FDA components in active use
  • A shared candidate ARD storage and access capability
  • Connection to existing and emergent services for Discoverability and Access
  • Collaboration on Analytics tools for integrated analysis using Jupyter Notebooks
  • Jointly develop CEOS Best Practices via interop experiments
• In doing this, validate interoperability approaches
Additional nodes?

CEOS SEO Node

Open Data Cube Infrastructure
- Application Library
- User Interface

Applications and Tools

Infrastructure/resources

ESA Node (TBC)

Web based GUI

Jupyter Notebook

CLI / REST API

DAS

OGC WCS Server

ESA Third Party Missions

ESA Heritage Missions

ESA Earth Explorer

Storage

Compute
Spawner Options

- **EASI Open Data Cube (CSIRO variant) environment**
  The EASI Open Data Cube with CSIRO variants. ODC 1.7
  EASI 1.0.5

- **Scipy environment**
  To avoid too much bells and whistles: Python, R, Julia, and Scipy.

- **Datascience environment**
  Includes libraries for data analytics such as Julia, Python, and R communities.

- **Python deep learning environment**
  Includes Keras, TensorFlow, and Scipy environment plus tensorflow and keras learning libraries.

**Experimental - EASI Open Data Cube (CSIRO variant) environment**
The EASI Open Data Cube with CSIRO variants. ODC 1.7
EASI Latest

Customise per CEOS activity as necessary

Interactive visualisation
Dashboards (from notebook)
Cloud data formats – Best Practices

Direct data access
- Compatibility (ARD, Geotiff)

Analytics APIs
- Services like OGC WPS
- Direct compute access:
  - python Dask

Cloud Compute Ready Data
- COGS – Cloud optimised geotiffs
- NetCDF in the Cloud
- Zarr – Cloud native array storage
CEOS Data in Cloud – Developing Best Practices

**Landsat 8**

**Description**
An ongoing collection of imagery by the Landsat 8 satellite.

**Update Frequency**
New Landsat 8 scenes are added regularly.

**License**
There are no restrictions on use. Requires a license from USGS.

**Documentation**
[https://docs.opendata.arcgis.com/](https://docs.opendata.arcgis.com/)

**Managed By**
[planet.](https://planet.com/)

---

**Sentinel-2**

**Description**
The Sentinel-2 mission is a series of satellites that provide high spatial resolution images with continuity for the current SPOT series. The mission provides a global coverage of the Earth.

**Update Frequency**
New Sentinel data are added regularly, and they are available on Copernicus.

**License**
Access to Sentinel data is free. For more information, visit [Copernicus](https://copernicus.eo.esa.int/).

**Documentation**
[Documentation](https://sentinelookep-copernicus.eo.esa.int/#/)

**Managed By**
[SINERGISE](https://www.sinergise.com/)

---

**Resources on AWS**

**Amazon Resource Name (ARN)** for [Landsat-8](https://aws.amazon.com/s3/), [Sentinel-2](https://aws.amazon.com/s3/), and [Sentinel-2 Imagery](https://aws.amazon.com/s3/)

**AWS Region**
- US West 2
- EU Central 1

---

**Resources on AWS**

**Amazon Resource Name (ARN)** for [Requester Pays S3](https://aws.amazon.com/s3/)

**AWS Region**
- US West 2
- EU Central 1

---

**Resources on AWS**

**Amazon Resource Name (ARN)** for [Sentinel-2 Inventory](https://aws.amazon.com/s3/)

**AWS Region**
- EU Central 1
Roadmap – Next steps

• Roadmap:
  • WGISS+SEO deploy the lab
  • CEOS COAST and others: inventory of data, analytics, etc
  • Define joint interop experiments for capabilities needed by these projects
  • Demonstrate and validate jointly in the Lab(s)
    • Cookbook of Jupyter notebook examples
  • Plenary endorse first edition of CEOS Interoperability Terminology and promote uptake
    • ARD terms already updated in WGISS Data Preservation Glossary