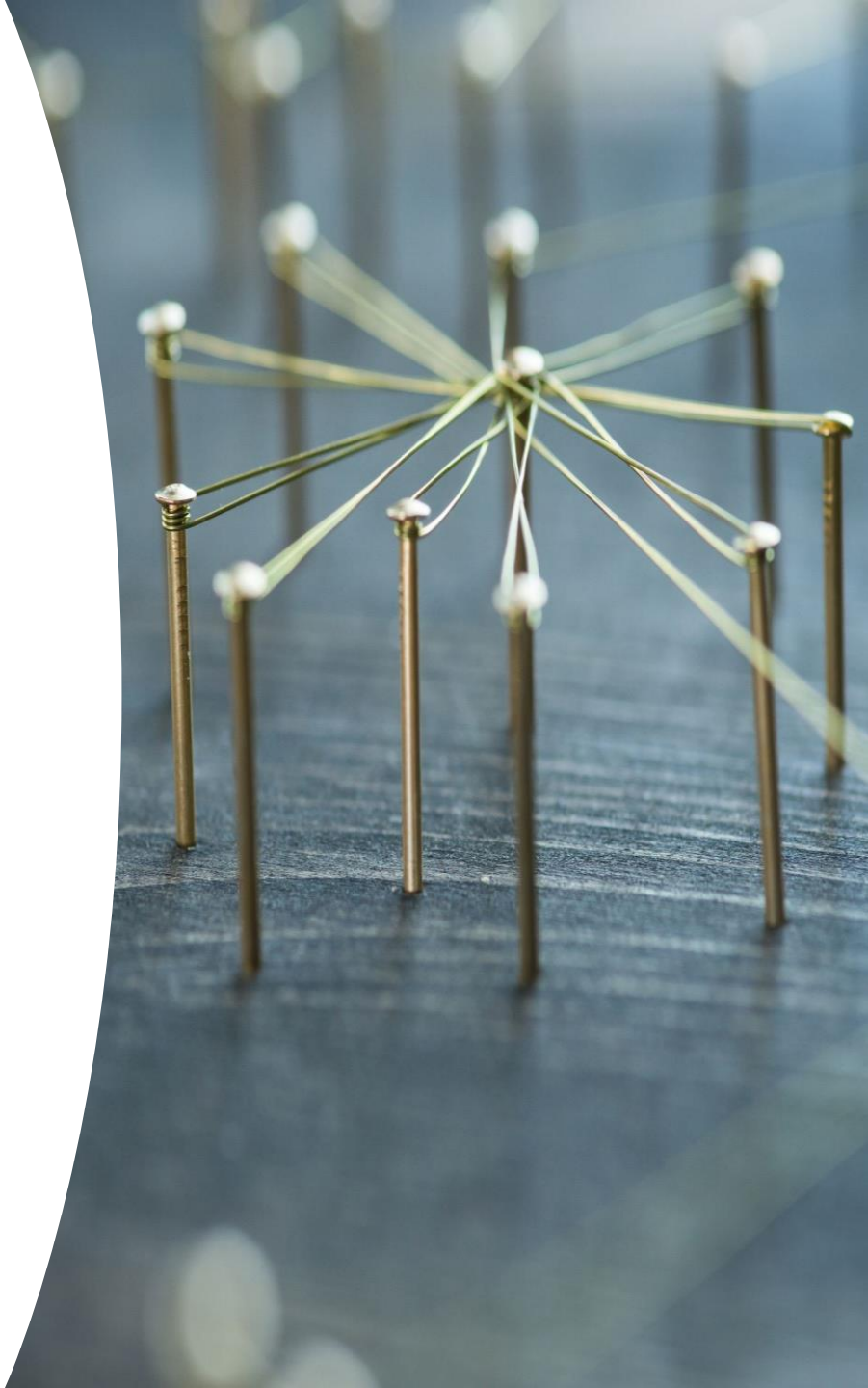
CEOS Interoperability Initiatives

WGDisasters-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



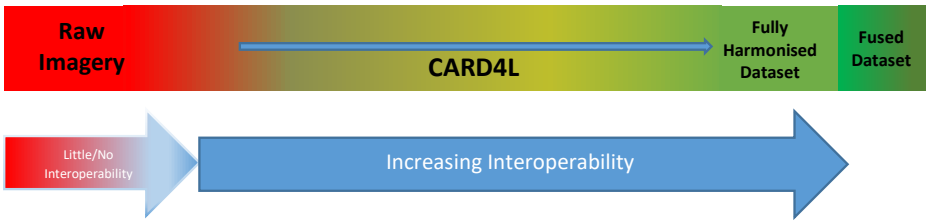
What is Interoperability?

- 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



CEOS Interoperability Terminology Report

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

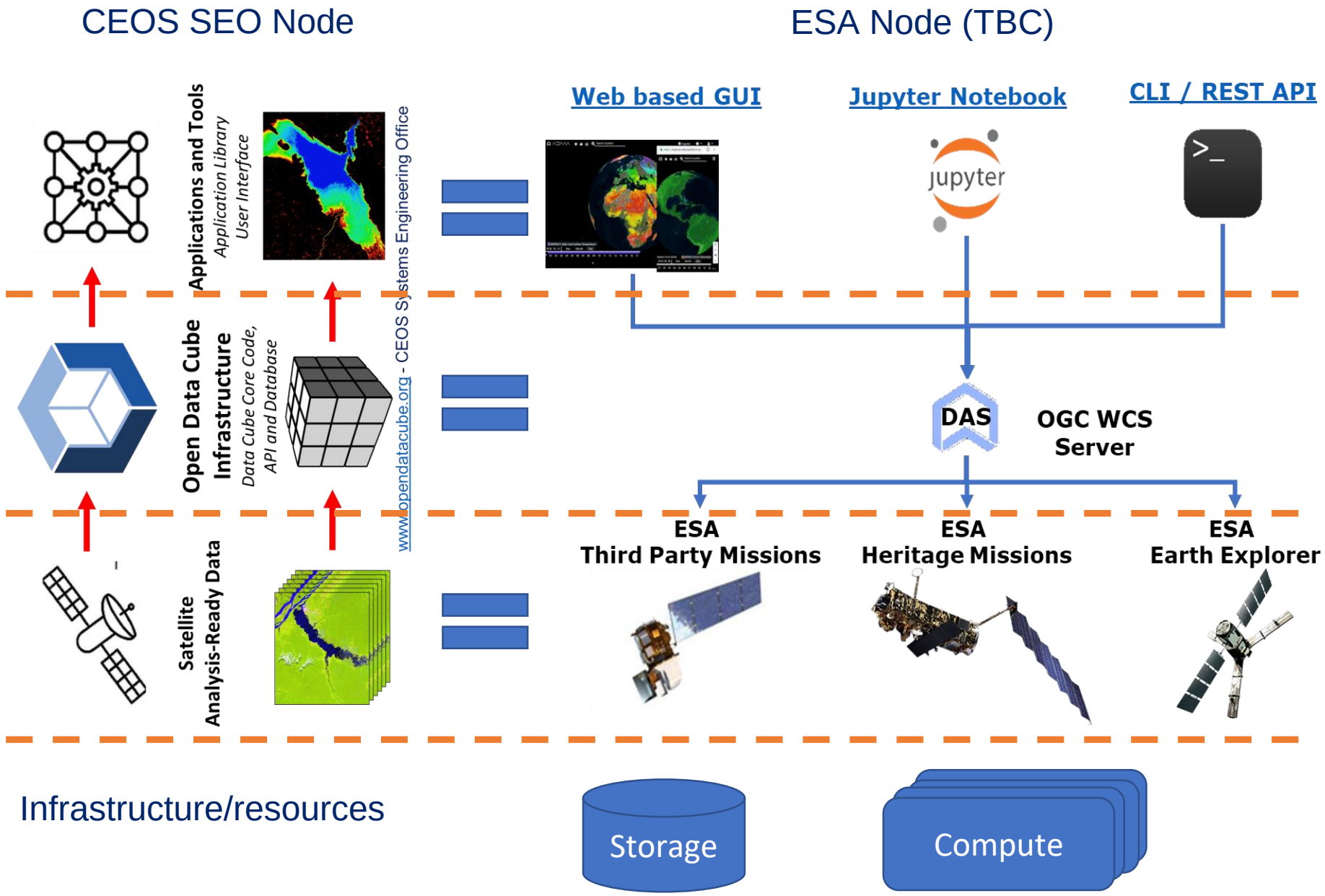
Why do we need an Interoperability Lab?

- 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**.

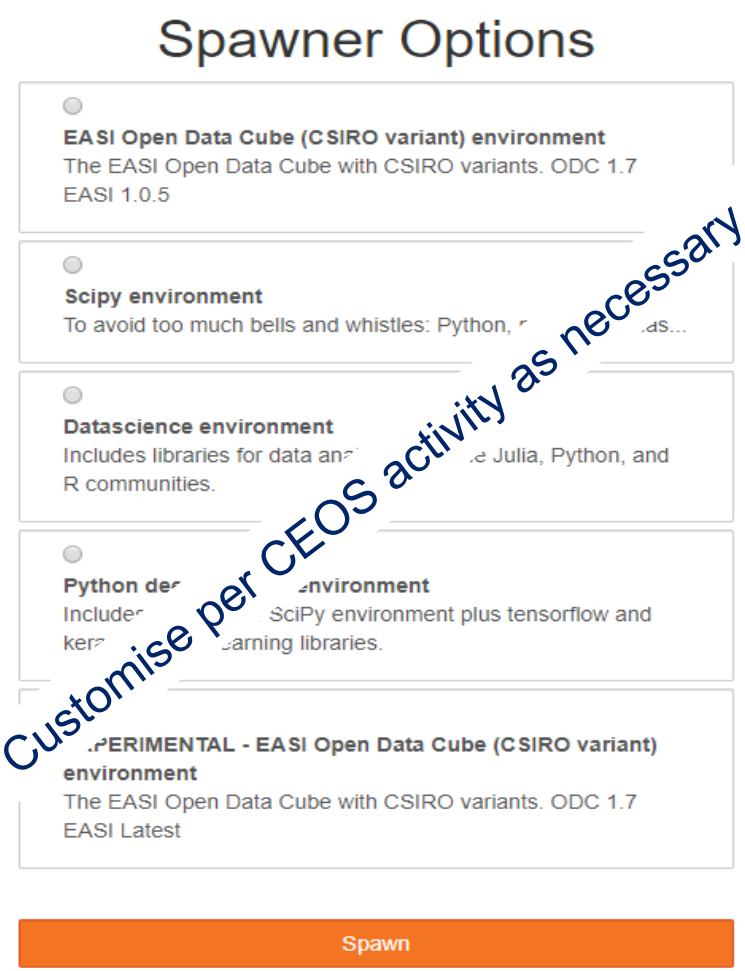
The Opportunity

- 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**

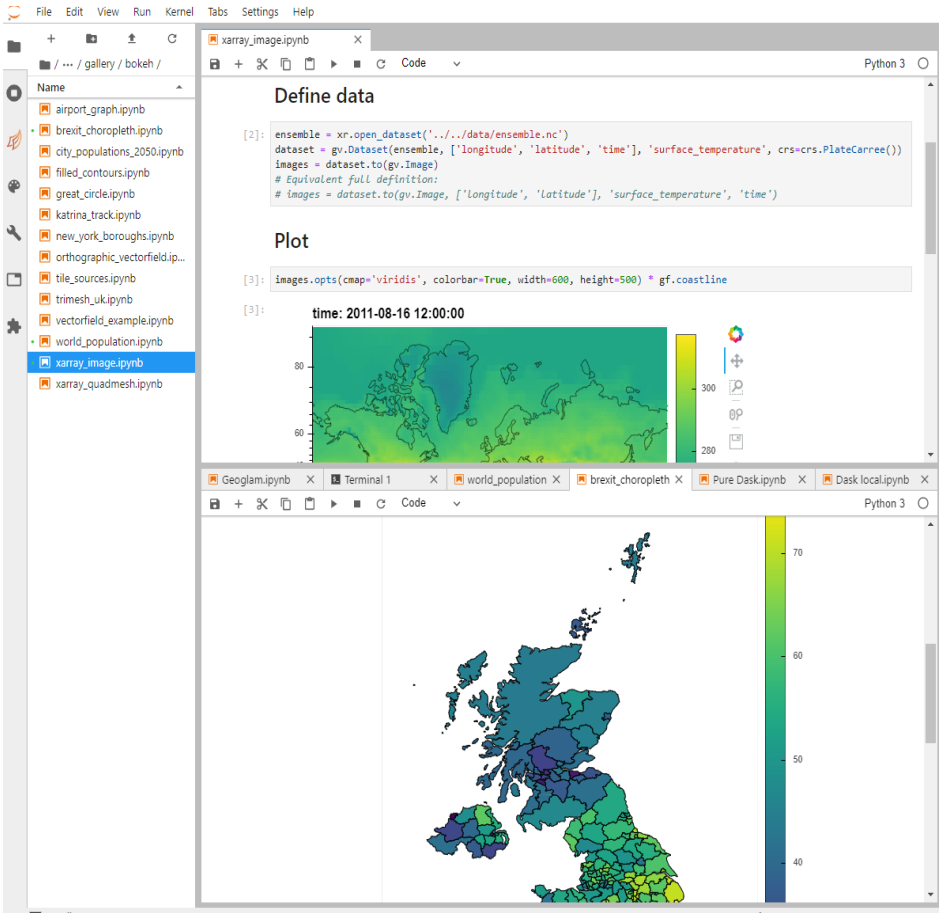
Additional nodes?



Jupyter Labs – exploratory data analytics
WGISS Survey on use of Jupyter in CEOS agencies

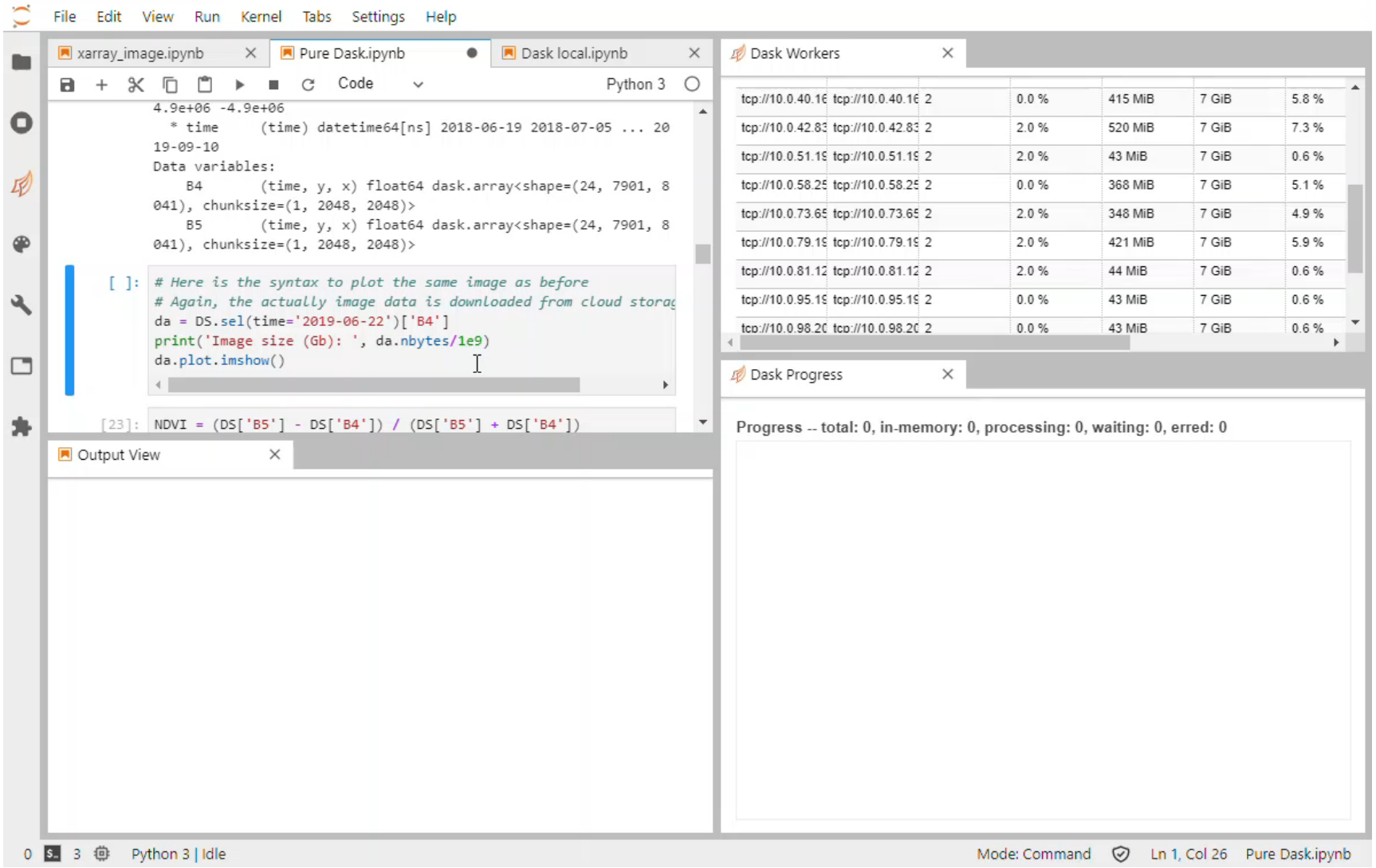


Customise per CEOS activity as necessary



Interactive visualisation

Dashboards (from notebook)



OPEN DATA CUBE

powered by easi

Cloud data formats – Best Practices

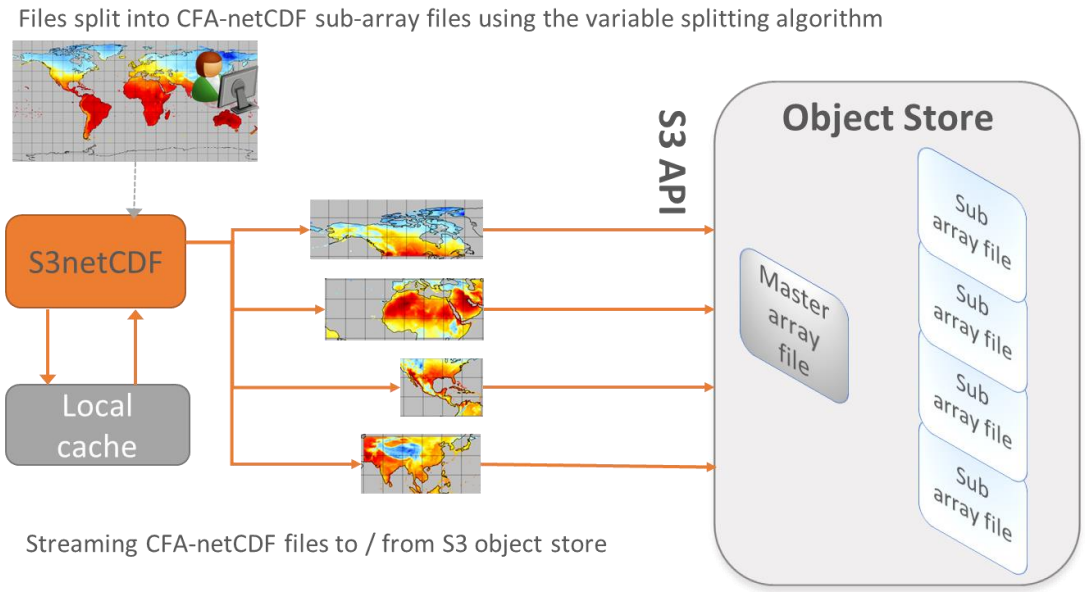
Direct data access

- Compatibility (ARD, Geotiff)

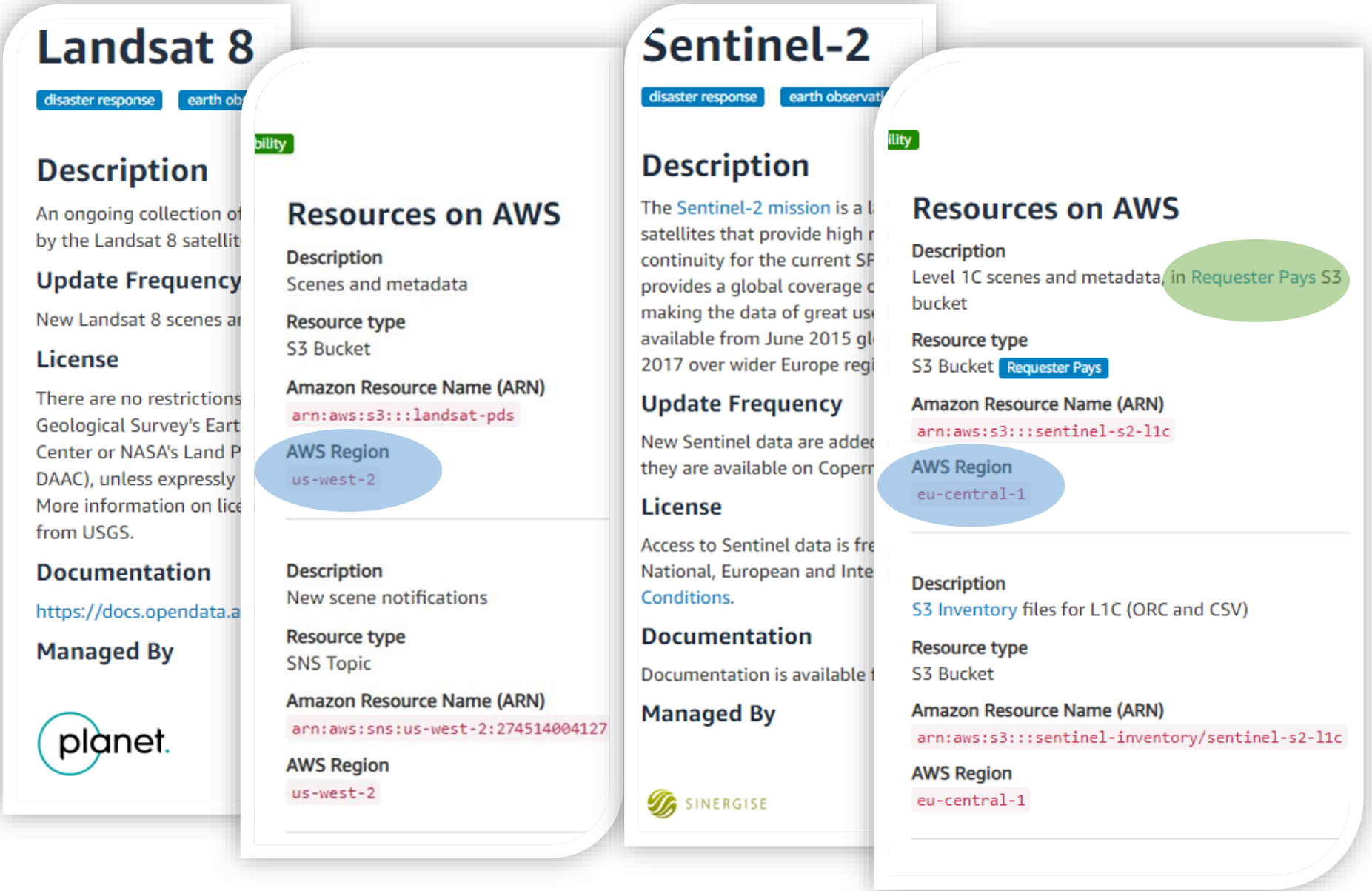
Analytics APIs

- 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



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