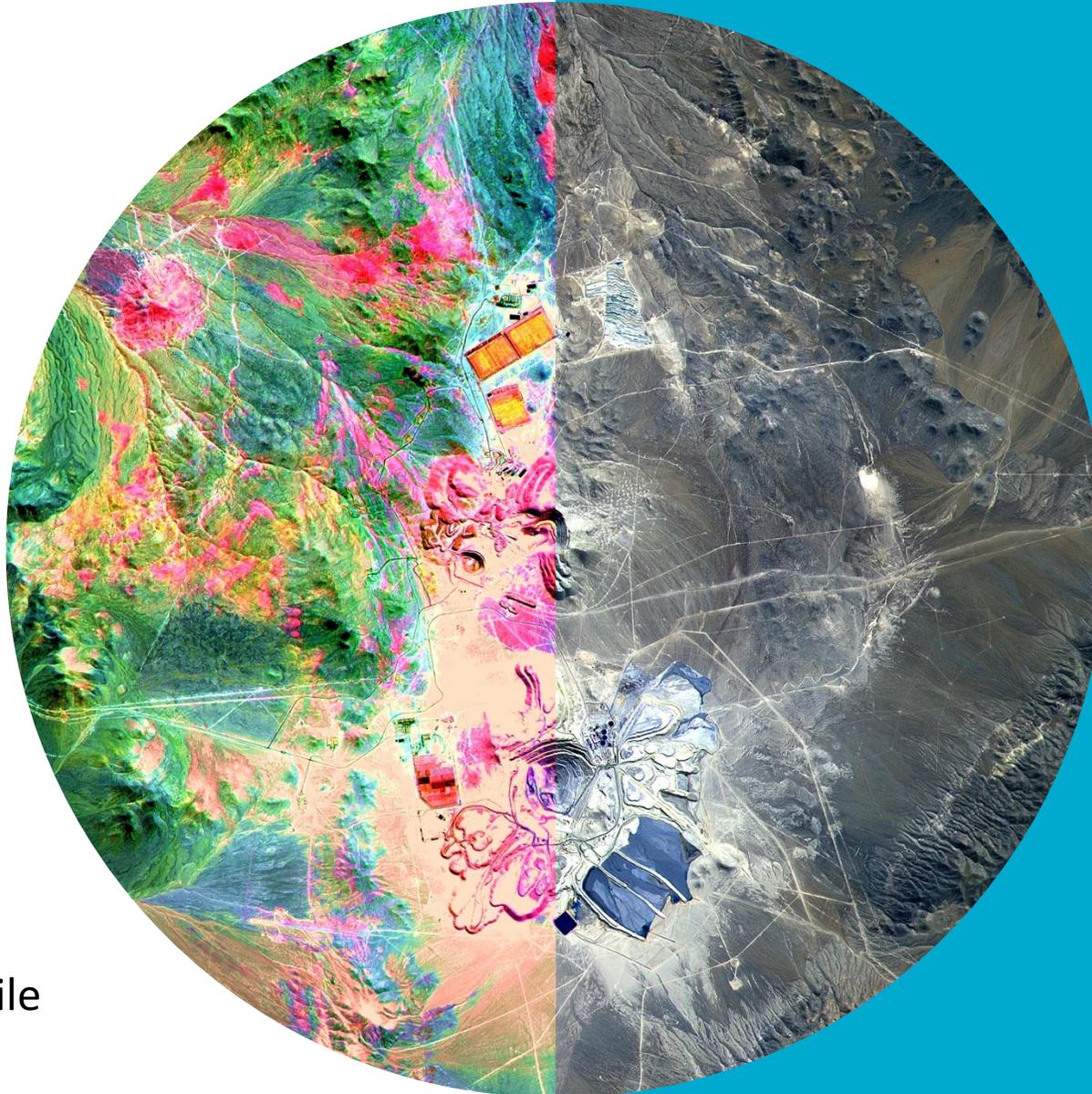




CSIRO

Chile



Earth Analytics Interoperability Lab (EAIL)

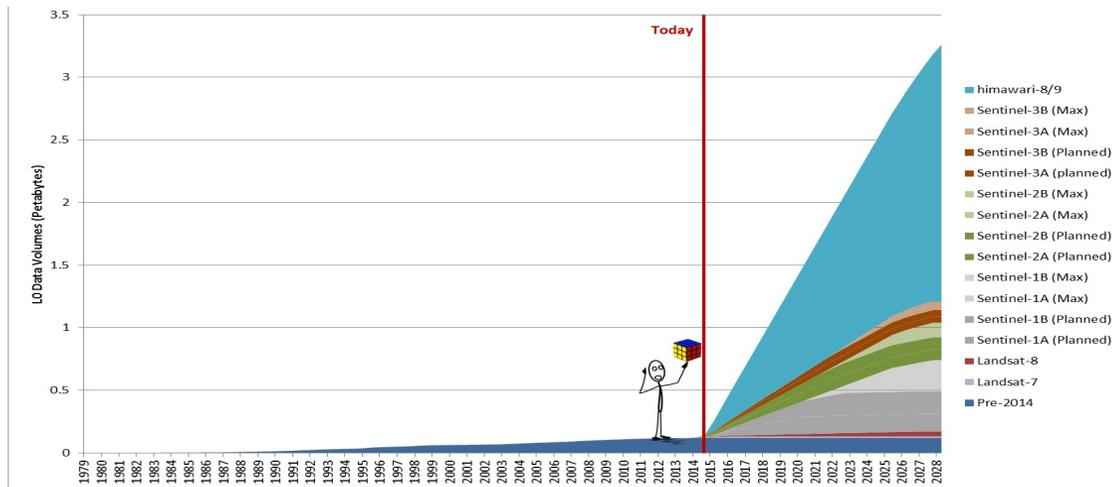
Jonathan Hodge
Program Director, CSIRO Chile

Challenge and Opportunity



- Over 400 EO sensors next decade
- New affordable satellites
- Sensor to user

Goal



Everyone has access to the best:

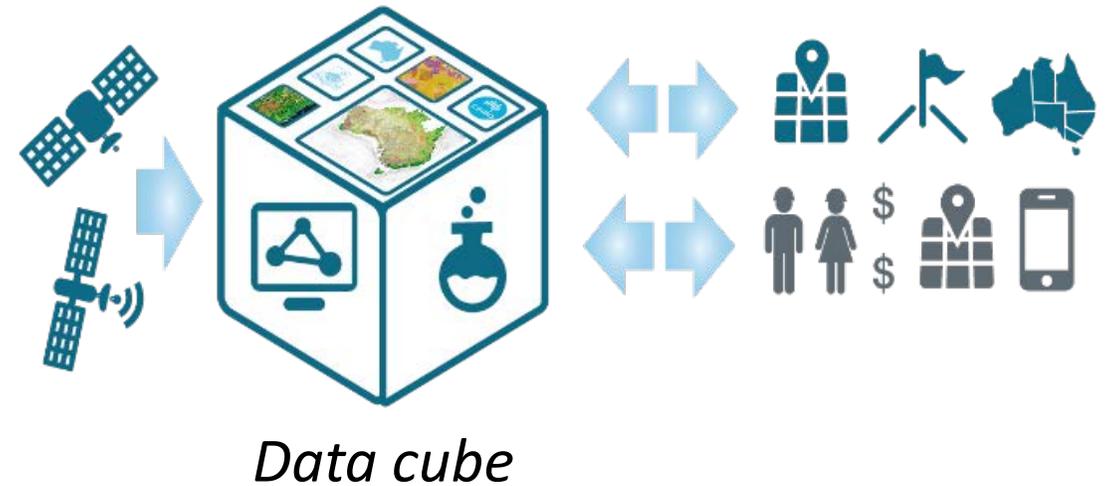
1. Data volumes and access
2. Processing and correcting
 - Analysis Ready Data
3. Products/algorithms and validation
 - Science literacy and use

Common and consistent data for everyone

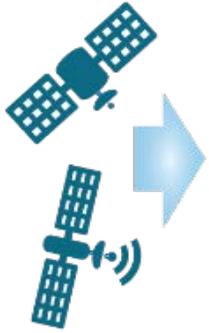
Before: Individuals invest in entire value chain



After: Data supply and big data analysis platform is shared. Individuals focus on delivery of specific value



A scalable platform for Data and Science



Data preparation and access

- National archives
- Instrument experts
- Atmospheric correction
- Data custodians

Output:

- **Analysis Ready Data**

Data cube

- Prepared data sets (ARD)
- Spatially-aligned pixels
- Programming interface
- Desktop, Cloud and HPC

Output:

- Search and read data
- Large data array (cube)

Products and users

- **Do science**
- Combine data
- Scale task

Output:

- Data products
- Share applications

Spatial data infrastructures



Open Data Cube



Rasdaman

Google Earth Engine Google Earth Engine



Thematic Exploitation
Platforms & Copernicus Hubs



SciDB

Goals

- Increase the impact and value of satellite data
- Provide an EO data analysis architecture
- Foster a community to grow the technology and applications

An Overview of Platforms for Big Earth Observation
Data Management and Analysis -

<https://www.mdpi.com/2072-4292/12/8/1253>

- Abstraction (data, processing, infrastructure)
- Scalability (storage, processing)
- Interoperability
- Extensibility
- Infrastructure requirements
- Reproducibility of science
- Governance

Where are the Data Cubes?

OpenDataCube.org



10 operational cubes (Australia, Colombia, Switzerland, Taiwan, Vietnam, Kenya, Tanzania, Ghana, Sierra Leone, Senegal)
67 data cubes in development – includes 49 additional countries in the Digital Earth Africa initiative
28 countries expressing interest or reviewing data cubes

Partners and platforms

Partners and Associates	Platforms and Initiatives
NASA (CEOS SEO)	CEOS Earth Analytics Interoperability Lab
Geoscience Australia	Digital Earth Australia Digital Earth Africa
CSIRO	Earth Analytics Science Innovation Hub
USGS	Landsat cloud archives
UK Catapult	Common Sensing
Switzerland	Swiss Data Cube
VNSC	Vietnam Data Cube
IDEAM	Colombia Data Cube

Open Earth Alliance

<https://www.openearthalliance.org>

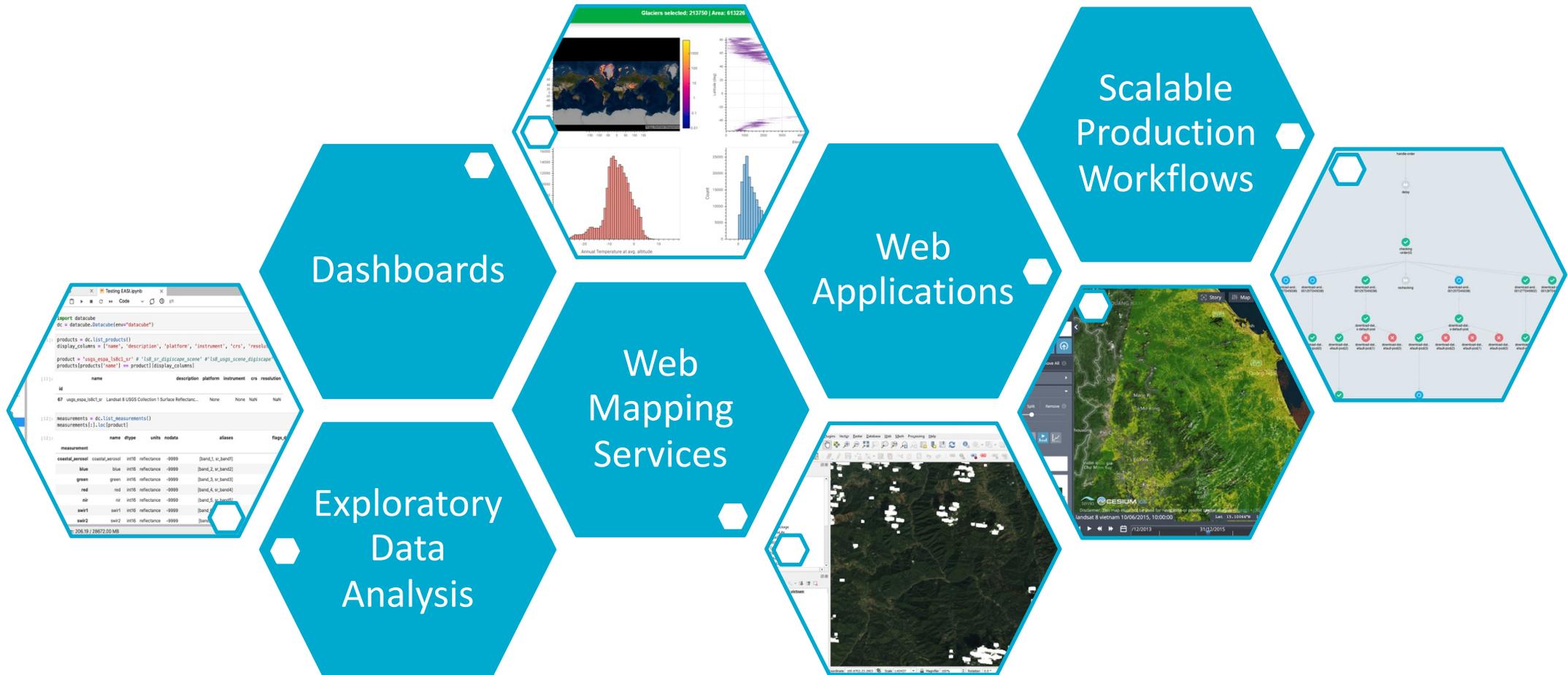
Open Data Cube

<https://www.opendatacube.org>

What is the Open Data Cube?

<https://medium.com/opendatacube>

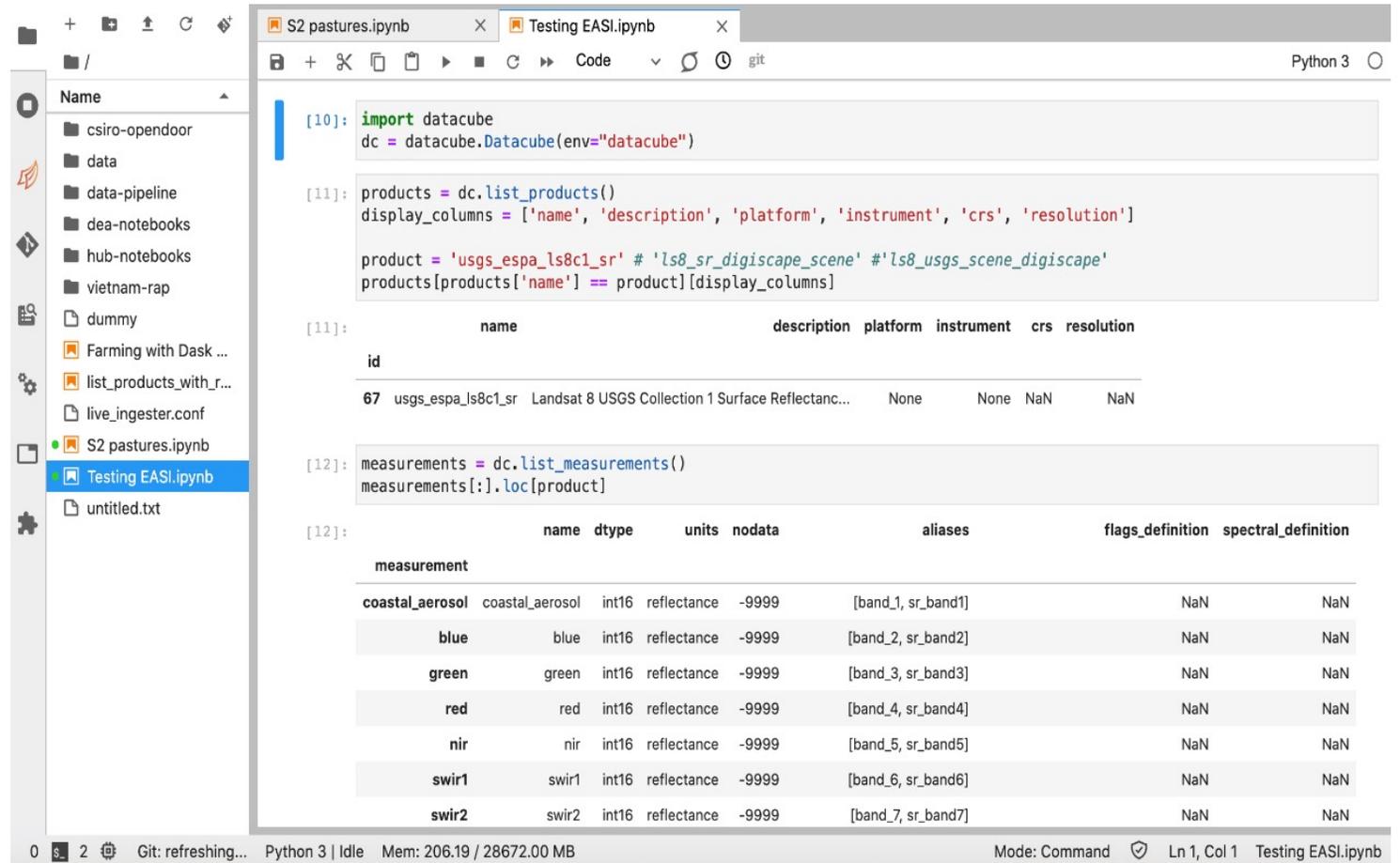
EASI is an ecosystem with a range of interfaces



Powered by Open Data Cube and the Python data science ecosystem

Exploratory data analysis with Jupyter Lab

- Interactive analysis with python notebooks
- User's have own project space (backed-up)
- Pre-configured for use with system datacube
- Copy and save notebooks via Git
- Per User Computing Clusters
- Interactive User Interfaces in a few lines of code



The screenshot displays a Jupyter Lab environment with two open notebooks: 'S2 pastures.ipynb' and 'Testing EASI.ipynb'. The 'Testing EASI.ipynb' notebook is active, showing the following code and outputs:

```
[10]: import datacube
dc = datacube.Datacube(env="datacube")

[11]: products = dc.list_products()
display_columns = ['name', 'description', 'platform', 'instrument', 'crs', 'resolution']

product = 'usgs_espa_ls8c1_sr' # 'ls8_sr_digiscape_scene' #'ls8_usgs_scene_digiscape'
products[products['name'] == product][display_columns]
```

	name	description	platform	instrument	crs	resolution
id						
67	usgs_espa_ls8c1_sr	Landsat 8 USGS Collection 1 Surface Reflectanc...	None	None	NaN	NaN

```
[12]: measurements = dc.list_measurements()
measurements[:].loc[product]
```

	name	dtype	units	nodata	aliases	flags_definition	spectral_definition
measurement							
coastal_aerosol	coastal_aerosol	int16	reflectance	-9999	[band_1, sr_band1]	NaN	NaN
blue	blue	int16	reflectance	-9999	[band_2, sr_band2]	NaN	NaN
green	green	int16	reflectance	-9999	[band_3, sr_band3]	NaN	NaN
red	red	int16	reflectance	-9999	[band_4, sr_band4]	NaN	NaN
nir	nir	int16	reflectance	-9999	[band_5, sr_band5]	NaN	NaN
swir1	swir1	int16	reflectance	-9999	[band_6, sr_band6]	NaN	NaN
swir2	swir2	int16	reflectance	-9999	[band_7, sr_band7]	NaN	NaN

The interface also shows a file browser on the left with a project space containing folders like 'csiro-opendoor', 'data', and 'data-pipeline', and files like 'dummy', 'live_ingester.conf', 'S2 pastures.ipynb', 'Testing EASI.ipynb', and 'untitled.txt'. The bottom status bar indicates 'Python 3 | Idle', 'Mem: 206.19 / 28672.00 MB', and 'Mode: Command'.

<https://jupyterlab.readthedocs.io/en/latest/user/interface.html>



EDA with Jupyter Lab – Scales with Cluster

The screenshot displays the Jupyter Lab interface with several open tabs and panels. The main window shows a satellite map of a region with a color scale on the right ranging from -0.4 (red) to 0.8 (green). The map axes are labeled 'x' and 'y' with values from 1.150e+6 to 1.300e+6 and -4.750e+6 to -4.550e+6 respectively.

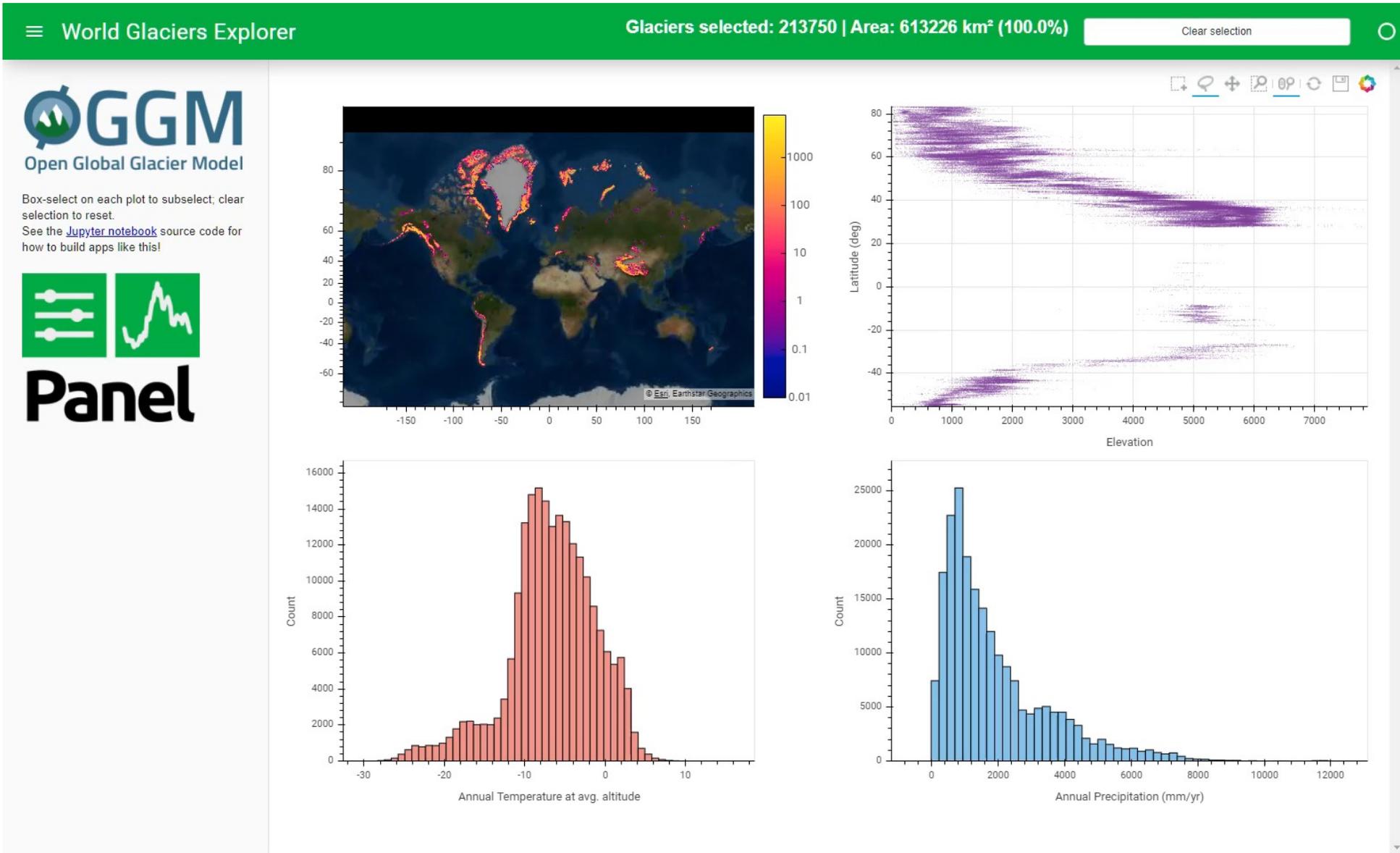
On the left, a sidebar contains various monitoring and configuration options such as BANDWIDTH TYPES, MEMORY BY KEY, and CLUSTER MAP. The bottom of this sidebar shows cluster information for KubeCluster 4, including scheduler address, dashboard URL, and resource limits.

On the right, a 'Task Stream' panel shows a Gantt chart of task execution. Below it, a 'Progress' panel displays a table of task progress:

Progress -- total: 25400, in-memory: 30, processing: 0, waiting: 0, erred: 0		
getitem	5850 / 5850	mean_combine660 / 660
dc_load_fm3600 / 3600	concatenate 630 / 630	array 1 / 1
dc_load_nba3600 / 3600	rechunk-merg_540 / 540	
dc_load_nba3600 / 3600	mean_agg-agg_30 / 30	
mean_chunk1950 / 1950	empty_2048x2048 6 / 6	
any-aggregate50 / 1950	empty_2048x1494 6 / 6	
any-any-agg1950 / 1950	empty_1487x1494 6 / 6	
dataset 1015 / 1015	empty_1487x2048 6 / 6	

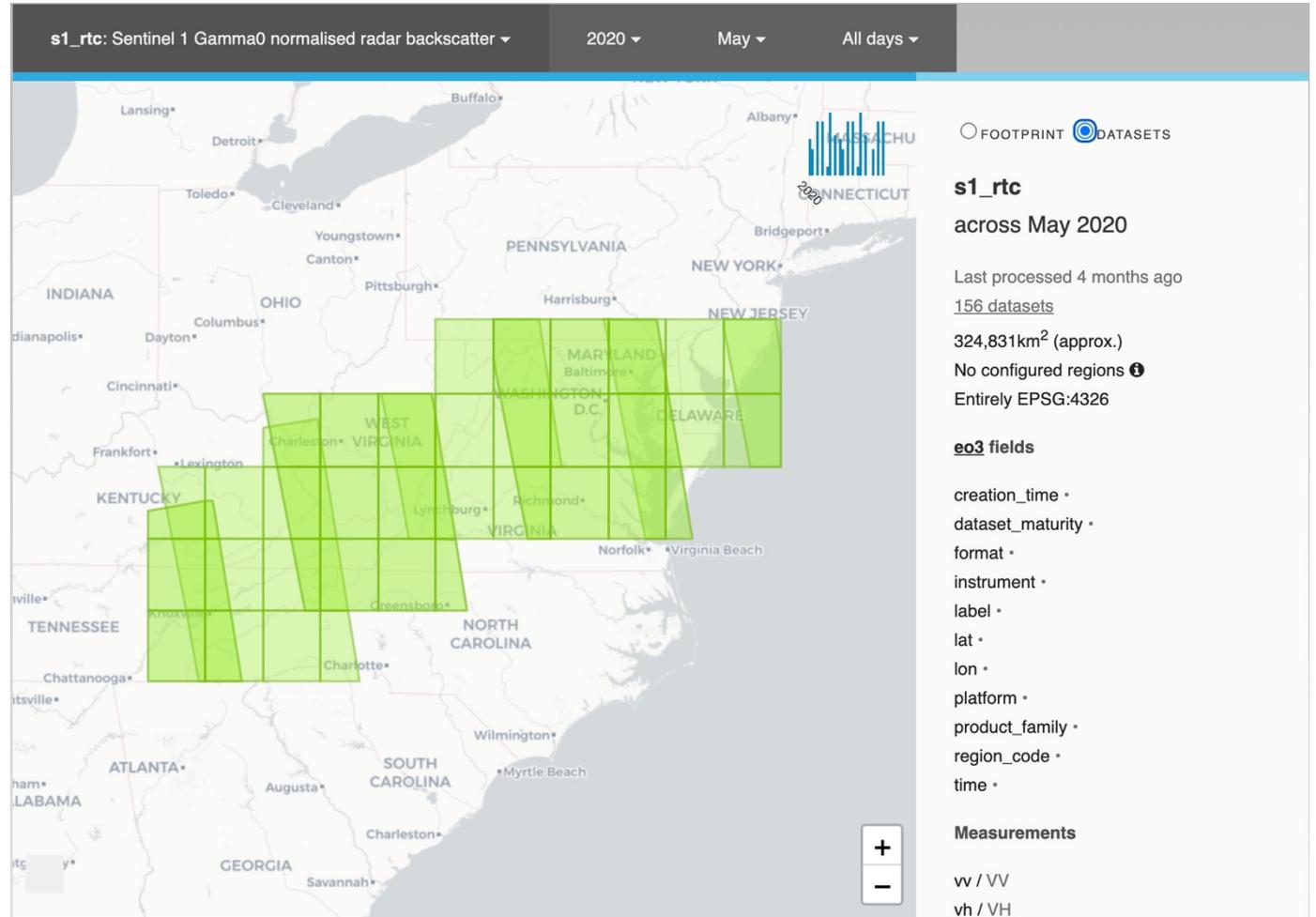
At the bottom of the interface, the status bar shows 'Git: refreshing...', 'Python 3 | Idle', 'Mem: 3.64 / 28.00 GB', and 'Saving completed'.

Evolution to Dashboards



Explorer

- Browse available datasets
- View specific scene coverage
- Discover product names and measurement names
- important for requesting data in `datacube.load()`



Web Map Services for GIS Applications

- Provides an OGC interface to the ODC API
- All you need is the URL
- Can perform some on-the-fly product generation e.g. NDMI, Cloud masking

This web-service serves georectified raster data from our very own special Open Datacube instance.

This URL is an end-point and is not intended for direct viewing. For more information:



OPEN DATA CUBE

[Open Web Services \(datacube-ows\) Repository](#)

This instance supports the following protocols:

WMS (OGC Web Map Service)

WMTS (OGC Web Map Tiling Service)

TerriaJS Web Map Application (Demonstrator)

- Uses the Web Map Services
- Can be extended for visualisation, analysis, etc



Search for locations

Add data

DATA SETS [2] Remove All

Total cover (vegetation and bare ground, two-band monthly) (copy)

Zoom To Extent About This Data Split Remove
Opacity: 100 %

Left Both Right

Time: 05/10/2009, 11:00:00

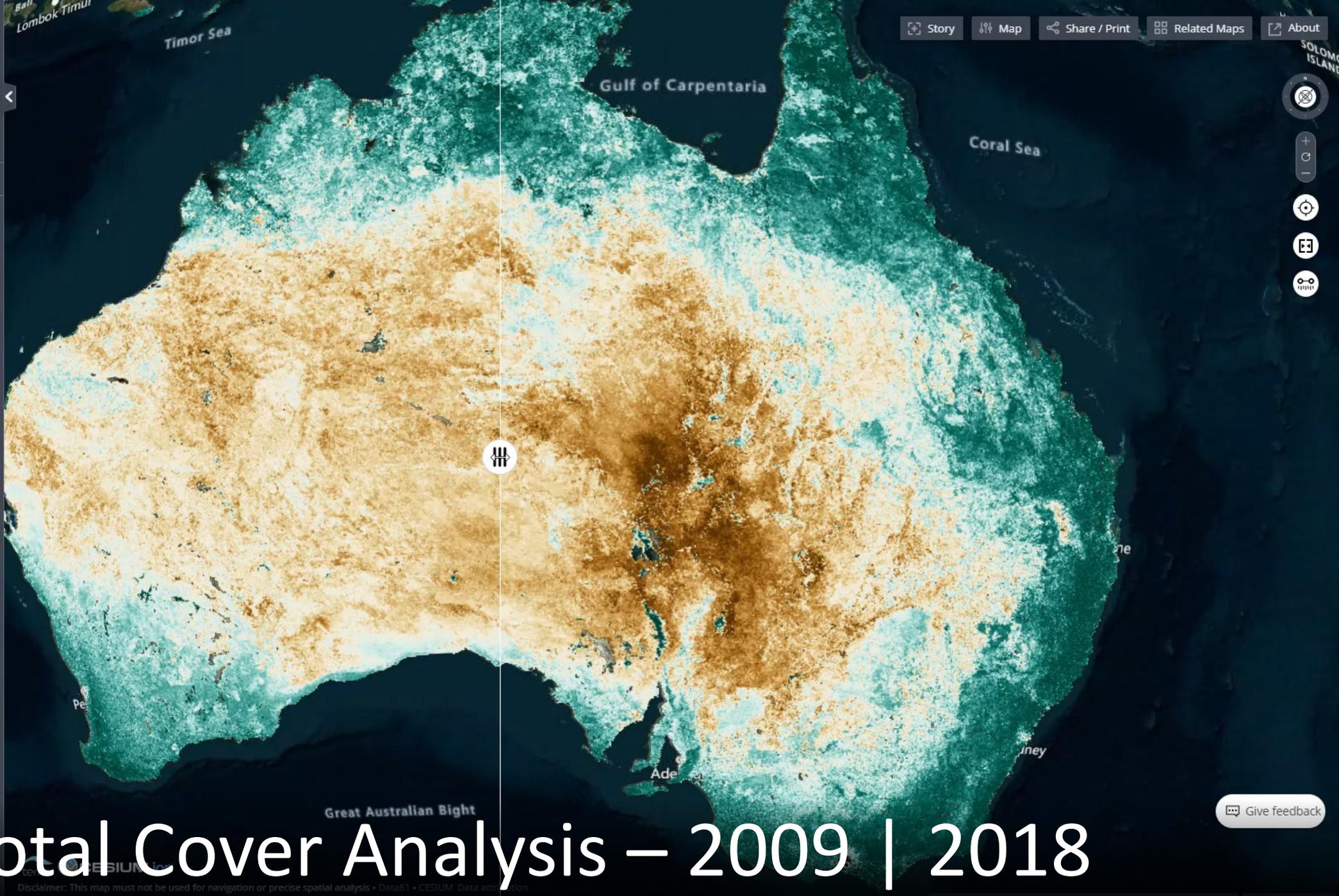


Total cover (vegetation and bare ground, two-band monthly)

Zoom To Extent About This Data Split Remove
Opacity: 100 %

Left Both Right

Time: 08/10/2018, 11:00:00



Map navigation controls: Home, Full Screen, Previous View, Next View, Layers, and a search icon.

Give feedback

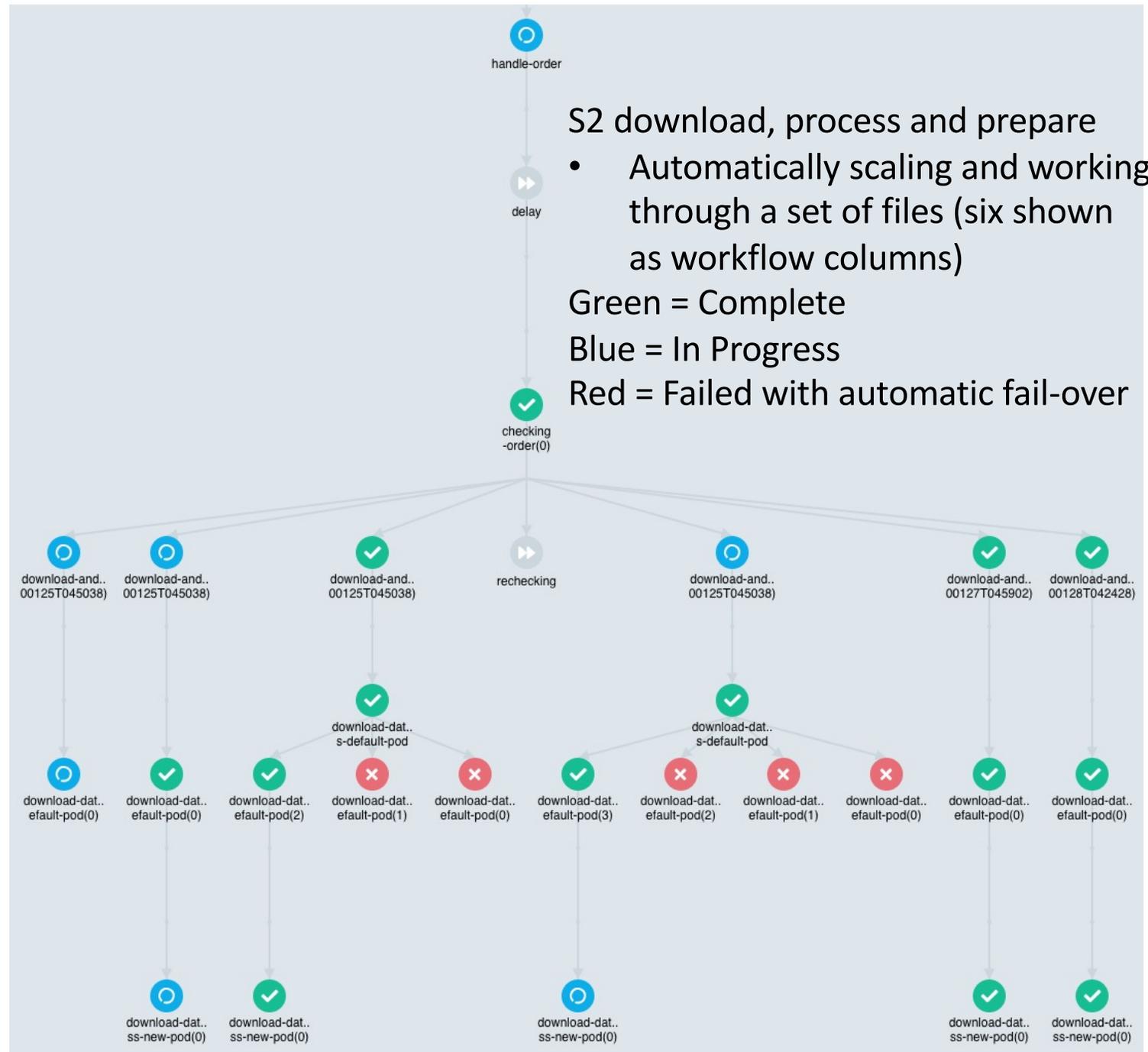
Total Cover Analysis – 2009 | 2018

Argo Workflows

- EASI's data-pipelines
- Custom production workflows

Features

- Scaling, Archive, Logging
- Web API to trigger workflows and monitor
- Single Sign On
- Resource tailoring for stages
- Resource Failover





Committee on Earth Observation Satellites

CEOS Earth Analytics Interoperability Lab Update

Dr. Robert Woodcock, CSIRO, WGISS Chair

Dr. Brian Killough, NASA, CEOS Systems Engineering Office

Mr. Jonathan Hodge, CSIRO Chile, Program Director

WGISS 51 Virtual Meeting

20-22 April 2021





- **Validating interoperability** between multiple CEOS organizations and working groups is complex
- **ARD and FDA are here** and there are WGISS technologies and Best Practices that need to change

The CEOS EAIL is intended to help CEOS WG & VCs explore these challenges *together*

EAIL is jointed operated by WGISS, CEOS SEO, CSIRO (Aus) and CSIRO Chile



- **15 registered users – more to come**
 - Large numbers is *not* the goal
- **CEOS engagement:**
 - WG Disasters Flood Pilot
 - CEOS COAST
 - DEMIX
 - Asia-RiCE
 - *In all cases EAIL is working with the groups on preparing data and regions of interest and assisting new users*
- **Open Earth Alliance community forum for EAIL**
 - <https://forum.symbios.space/c/ceos-eail/8>
 - *Very empty now but ready to use*

- **Fully deployed and operational**
- **For training purposes US Southeast states and coastline have been indexed** (<https://explorer.eail.easi-eo.solutions>):
 - USGS Collection 2 – full landsat archive, used-in-place, requester pays
 - Element 84 Sentinel 2 COGS, used-in-place, public
 - Sinergise S1 ARD service – ordered-and-stored, fee for service (but quick)
 - All of these are ARD (confirmed or under assessment for CARD4L)
- **CSIRO EASI Training notebooks:** customised for direct use in EAIL
- **Dask Gateway:**
 - The EAIL supports scalable computing
 - Customised based on project demand with larger nodes, GPU for ML, etc.