Aqua Monitor – Challenges and solutions from porting a multi-petabyte EO application from Google Earth Engine to openEO

Digital assets supporting SDG 13: Climate action

Arjen Haag (Deltares)

arjen.haag@deltares.nl

05.12.2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017529.

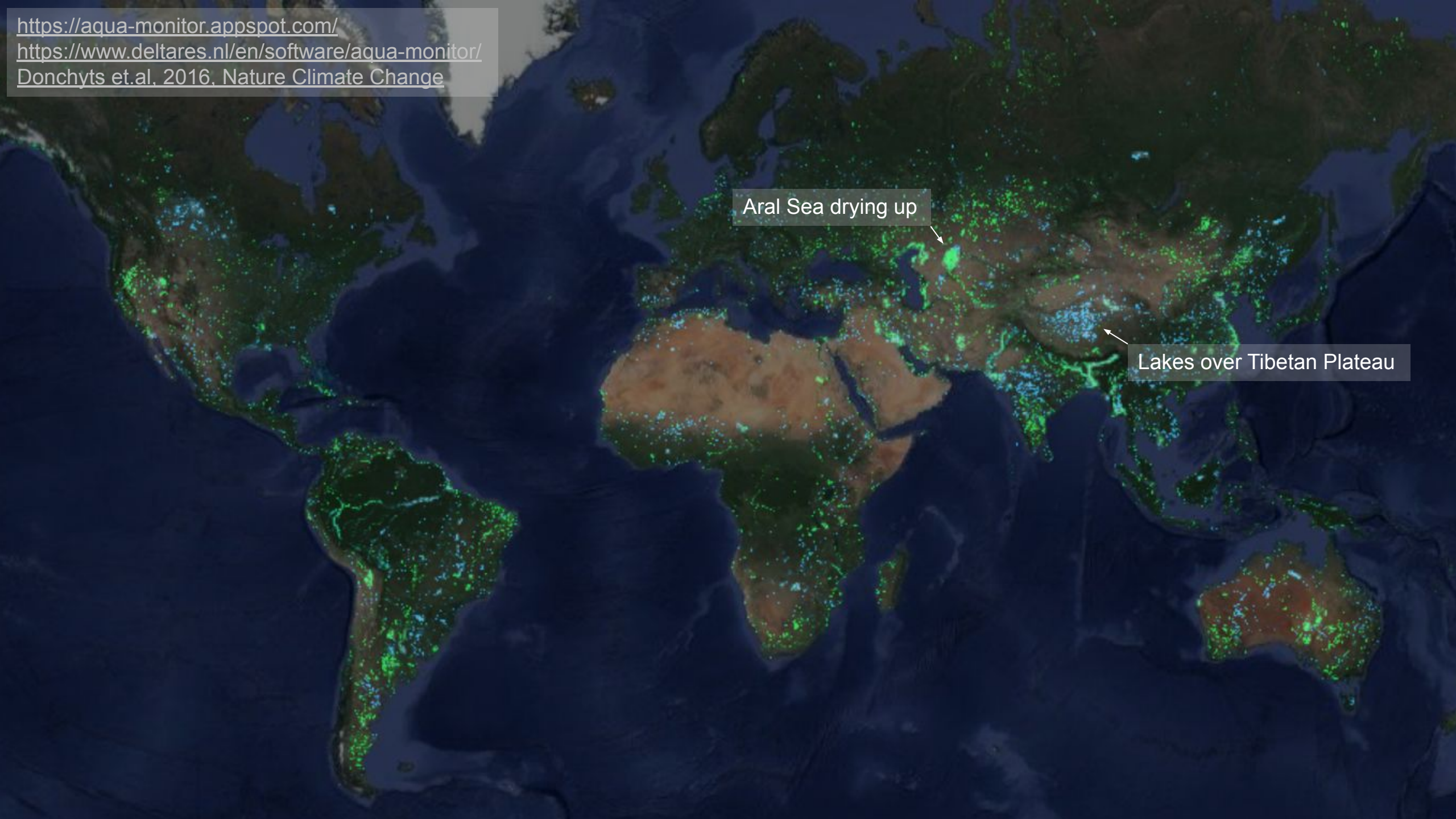
C-SCALE

Copernicus - eoSC AnaLytics Engine

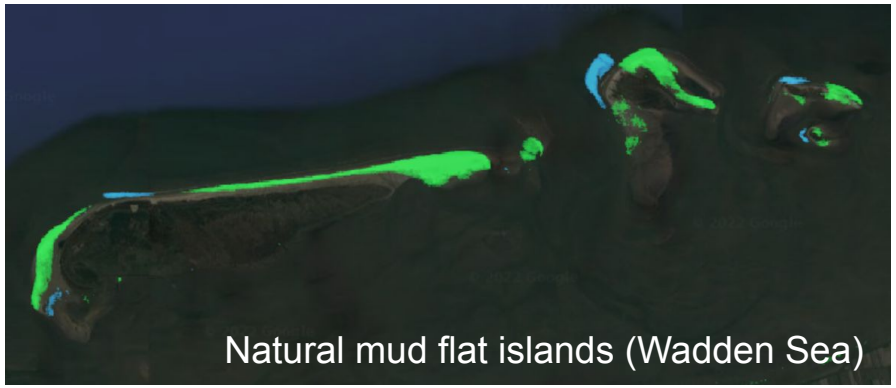
Arjen Haag (Deltares)
arjen.haag@deltares.nl

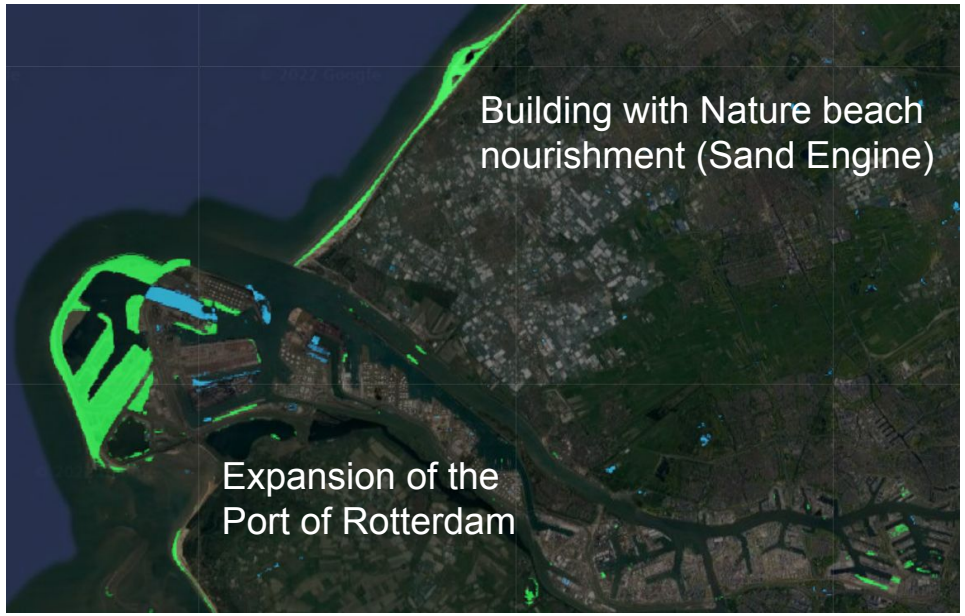
05.12.2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017529.


https://aqua-monitor.appspot.com/
https://www.deltares.nl/en/software/aqua-monitor/
Donchyts et.al, 2016, Nature Climate Change
Aral Sea drying up
Lakes over Tibetan Plateau

Aqua Monitor: the Netherlands





Google Earth Engine

Google Earth Engine

C-SCALE

The Earth Engine Public Data Catalog

Landsat and Sentinel
Raw, TOA, SR, ...

MODIS
Daily, NBAR, LST, ...

Terrain
SRTM, GTOPO, NED, ...

Land Cover
GlobCover, NLCD, ...

Atmospheric
NOAA NCEP, OMI, ...

... and many more, updating daily!

> 200 public datasets

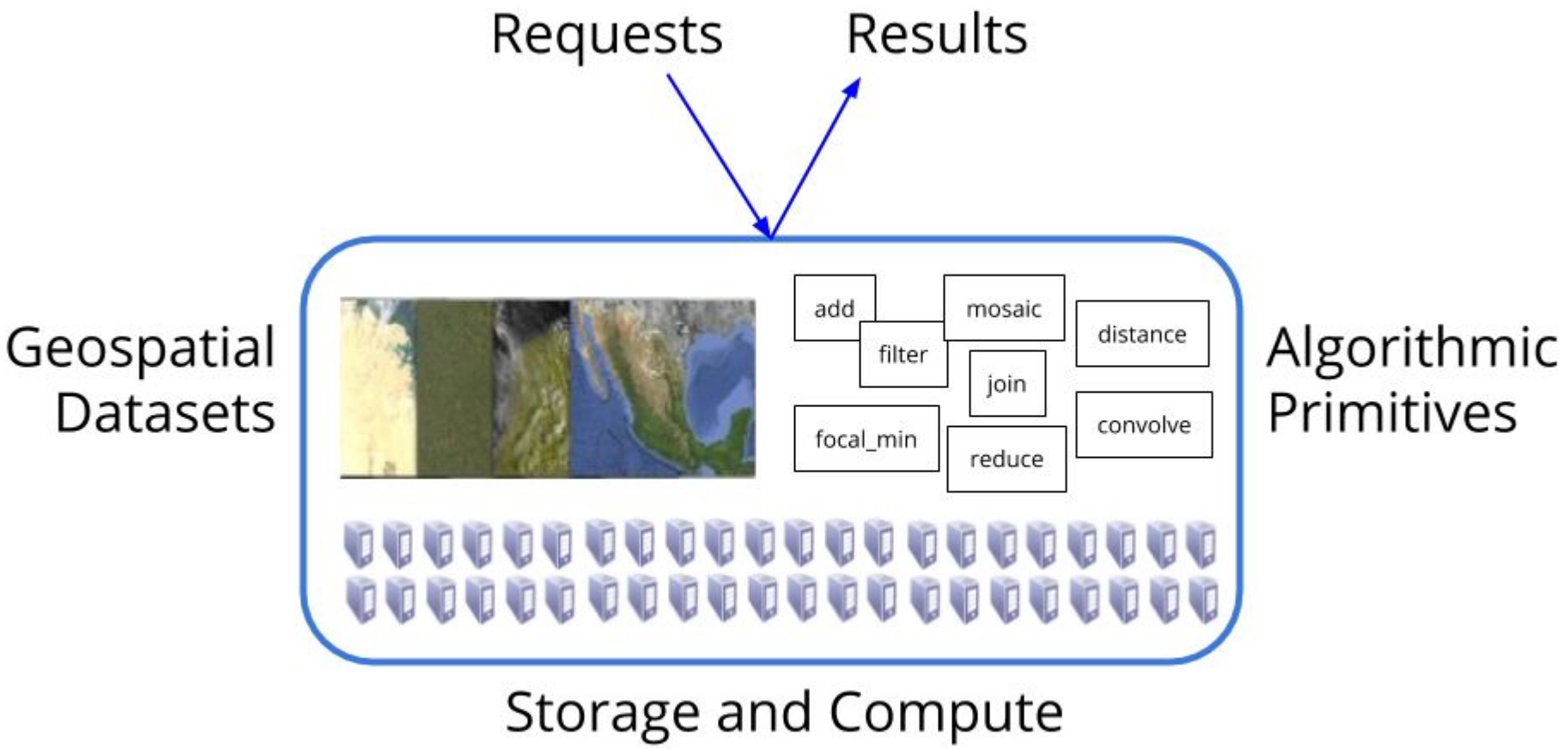
> 5 million images

> 4000 new images every day

> 5 petabytes of data

Google Earth Engine

Google Earth Engine



Copernicus

- The EU **Copernicus** programme has established itself globally as the **predominant global spatial data provider**.

- Copernicus provides **massive streams** of high resolution earth observation (EO) **data**.

Aims to significantly contribute to the vision of a digital copy of our Earth, i.e. a **digital twin of the earth**, to support current EC goals within the **Europe 2020 strategy**, the **Green Deal** and its related **Destination Earth** initiative.

The C-SCALE idea

Aim:

Federate European EO (DIASes, CollG nodes, etc.) and e-Infrastructure services to support Copernicus research and operations.

The C-SCALE consortium brings together expertise from:

EO sector: EODC, Deltares, VITO, CloudFerro, TUW, CESNET, GRNET

e-Infrastructure: EGI, CESNET, INFN, SURFsara, GRNET, INCD

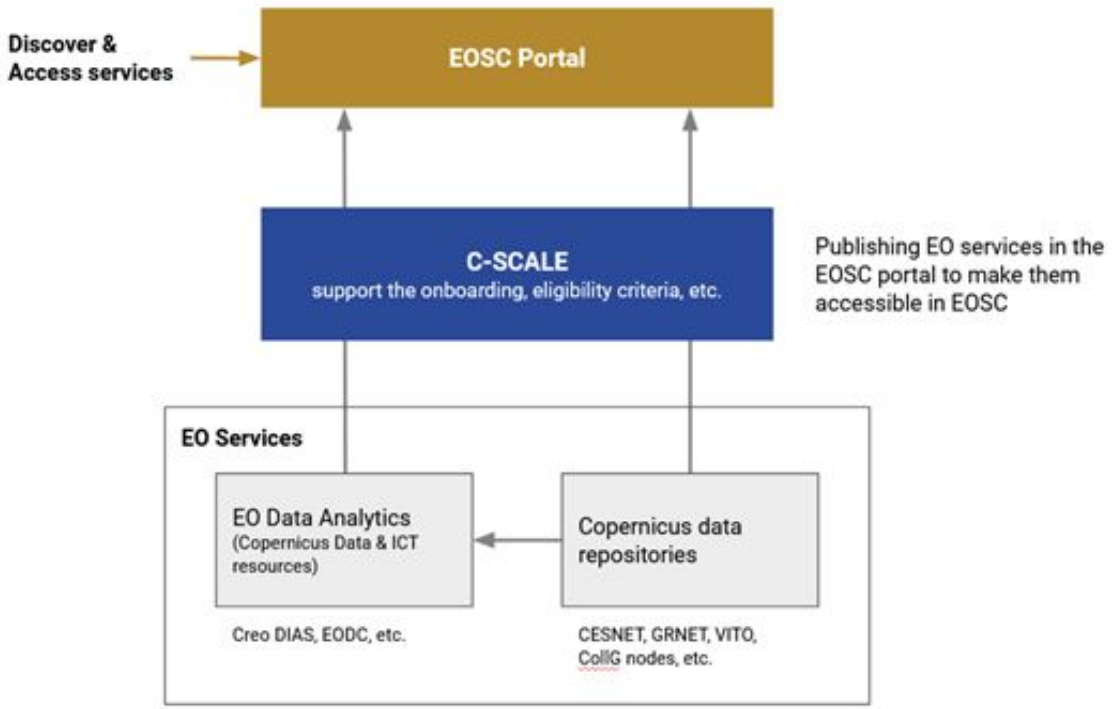
EUROPEAN OPEN SCIENCE CLOUD

C-SCALE will enrich EOSC and its Portal with scalable Big Copernicus Data Analytics federated services

C-SCALE Objectives – O1

O1: Scale-up the EOSC Portal integrating computing and data resources for Copernicus

- CREODIAS platform
- WEkEO DIAS
- EODC global Sentinel archive and processing infrastructure
- ESA SENTINEL Collaborative Ground Segment: CESNET, GRNET, VITO

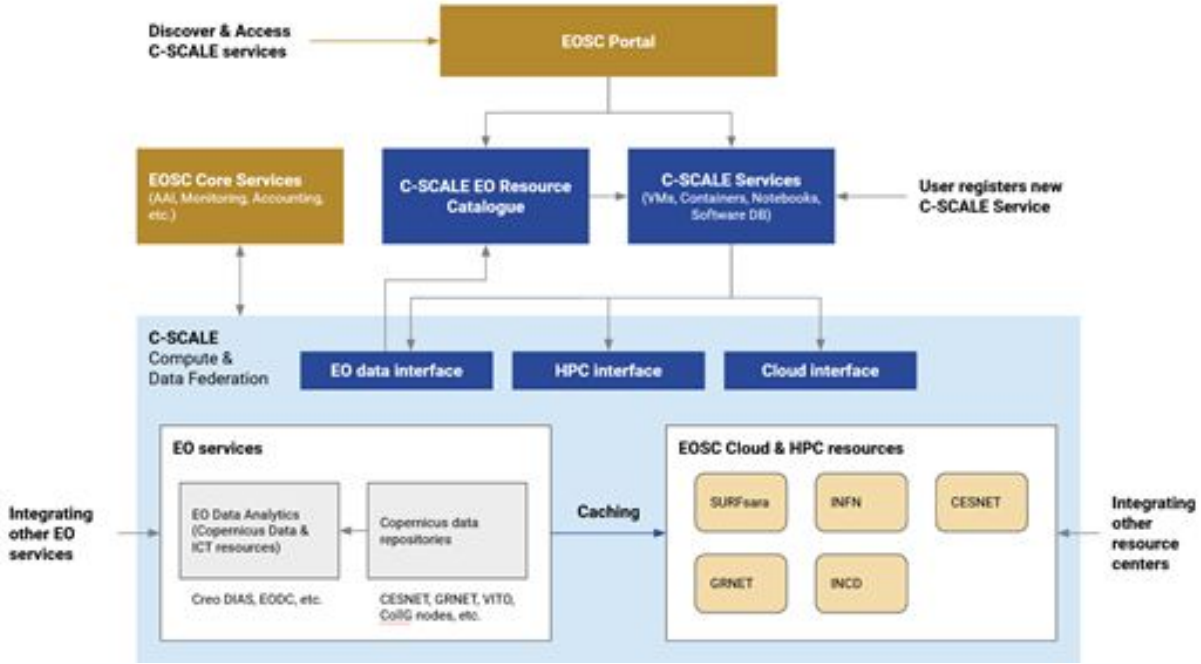


C-SCALE Objectives – O2

O2: Federated Copernicus resources with EOSC computing and storage providers to deliver a wide infrastructure optimised for very large FAIR Copernicus data handling

Federation principles:

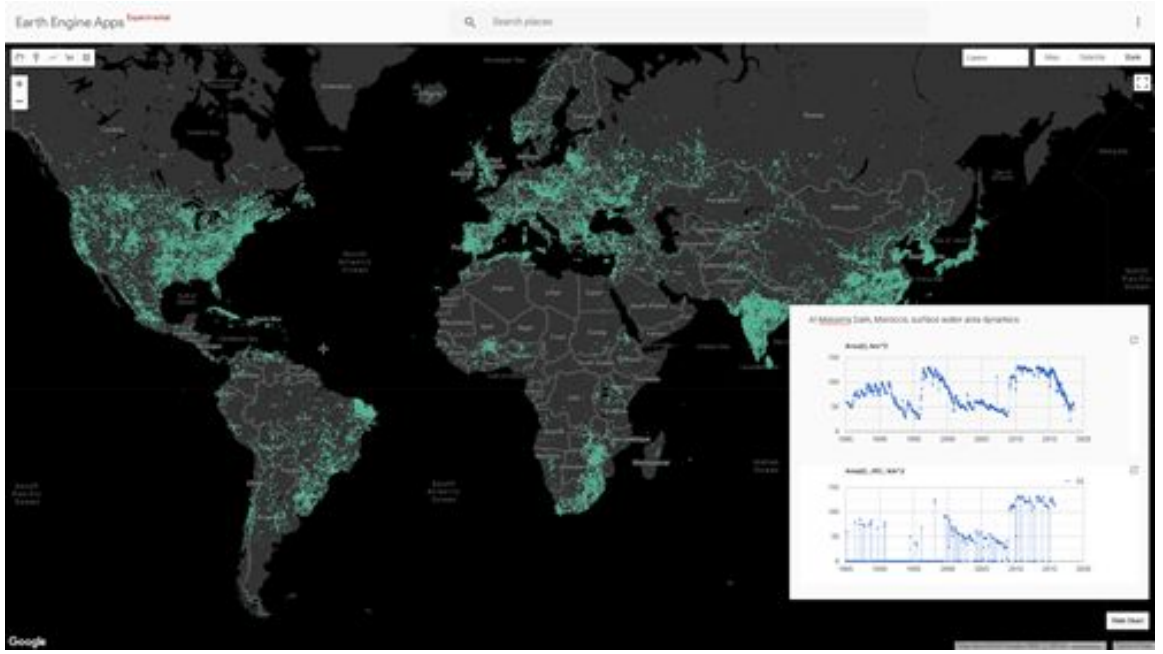
- Services accessible through homogeneous and standard interfaces
- EO data FAIR across the providers
- Burden to join the federation limited
- Adhere to EOSC policies and operational and technical requirements
- Basic and ops features (AAI, accounting, etc.) available through EOSC core services
- Maximise interoperability with other EOSC services



C-SCALE Objectives – O3

O3: Piloting the provision of a distributed online Sentinel long-term archive in EOSC

- No plans for Europe to maintain a full on-line archive of Copernicus data
- Cold storage not adequate for processing of long time series in arbitrary geographical area
- EOSC may fill this gap offering a Copernicus historical distributed data archive
- C-SCALE will develop a technical strategy to offer this archive in EOSC



Example of surface water area dynamics time series during nearly 40 years for one of 80 000 reservoirs quantified globally.

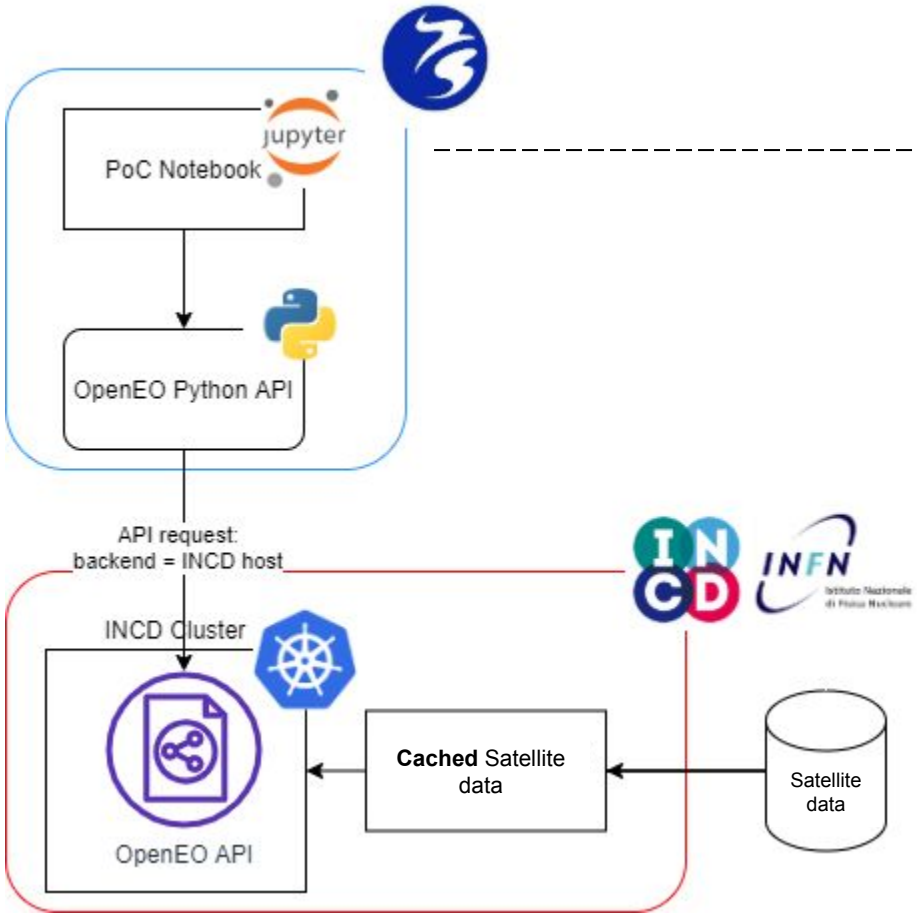
C-SCALE Objectives – O4

O4: Co-design of the federation with relevant scientific community across Europe

- 6 use cases
- Early Adopter programme to attract more use cases
- User forum to collect requirements for the co-design
- Automation to offer a seamless user experience
- Increased exploitation of Copernicus through the EOSC Portal

Use case	Description	EO Data
Aqua monitor (Deltares)	how the Earth's surface water has changed in the last 30 years	Landsat 4,5,7,8 Sentinel 2
Water watch (Deltares)	reanalysis of surface water changes in reservoir	Sentinel 1 Sentinel 2
HiSea (Deltares)	provide high resolution data of water quality at sea	Sentinel-3 ERA5
Land Surface Analysis (Deltares)	map with seasonal forecasts of drought indicators	Sentinel-1 ASCAT PROBA-V
RETURN (WUR)	monitoring tropical forest recovery capacity	Sentinel 1 Sentinel 2 Landsat 5,7,8
Virtual European Sentinel Data Cubes (TUW)	virtual Sentinel data cubes covering European countries and the whole of Europe	Sentinel 1 Sentinel 2

Aqua Monitor use case: architecture



https://github.com/c-scale-community/use-case-aquamonitor

Implements Aqua Monitor algorithm using openEO.

A notebook including course material was developed to help anyone get started with global EO data using openEO, and the algorithm is translated and broken down in detail for anyone to apply it to their use case.

The notebook:

- builds a set of instructions using the openEO Python client library
- sends it to the openEO Platform backend for processing
- results are downloaded for further analysis
- interactive plots are created for visualization

Aqua Monitor use case: lessons learned

The current algorithm uses a multi-petabyte dataset which requires parallel processing on large tiles

→ significant data challenges for the infrastructures in terms of storage, rapid access and compute, as openEO backend used keeps (most) data in memory to do calculations.

→ current implementation on the openEO backend is a proof-of-concept version; the algorithm is successfully ported and applied but limited in spatiotemporal scale.

→ Scaling this up requires considerable resources, which should be allocated in collaboration with backend providers.

In general:

- GEE algorithms can be successfully ported to openEO
- The openEO Platform has huge potential but, at the moment, falls short on several topics when compared to GEE
- However, openEO does give users more flexibility (e.g. control over parallelization and multiple programming languages/interfaces)

Thank you for your attention.

Arjen Haag (Deltares)
arjen.haag@deltares.nl

05.12.2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017529.