C-SCALE

Copernicus - eoSC AnaLytics Engine

Implementing a European Big Copernicus Data Analytics platform: C-SCALE services, Outcomes and Impact

Xavier Salazar (EODC), Zacarias Benta (INCD) and the C-SCALE Team

Xavier.Salazar@egi.eu , zacarias@lip.pt

Ibergrid 2023, Benasque, Spain & Online | 28 Sept 2023

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

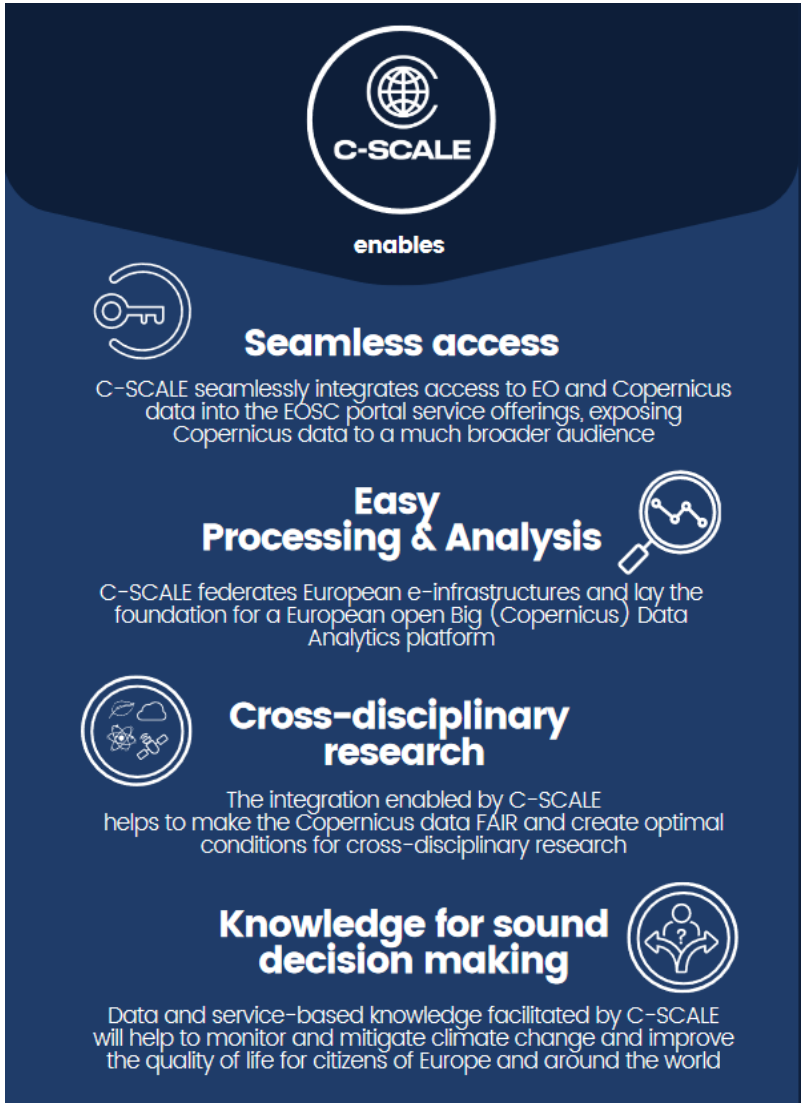
The C-SCALE Project

Europe lacks an integrated compute and storage infrastructure for the exploitation of **Copernicus** datasets in scientific and applied applications.

C-SCALE responds to that challenge by enhancing the **EOSC Portal** with pan-European federated data and computing infrastructure services for Copernicus.

C-SCALE: Copernicus - eoSC AnaLytics Engine

- Project duration: Jan 2021 – September 2023
- Budget: ~ 2 million Euros
- Consortium of 11 partners with pan-European coverage



Project Objectives and KERs

C-SCALE

Objectives

- O1: Scale-up the EOSC Portal integrating pan-European computing and data resources for Copernicus
- O2: Federate Copernicus resources with EOSC computing and storage providers
- O3: Piloting the provision of a distributed online Sentinel long-term archive in EOSC
- O4: Co-design of the federation with relevant scientific communities across Europe

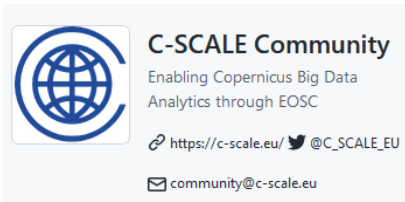
Key Exploitable Results:

FedEarthData

openEO Platform

Earth Observation Metadata Query Service

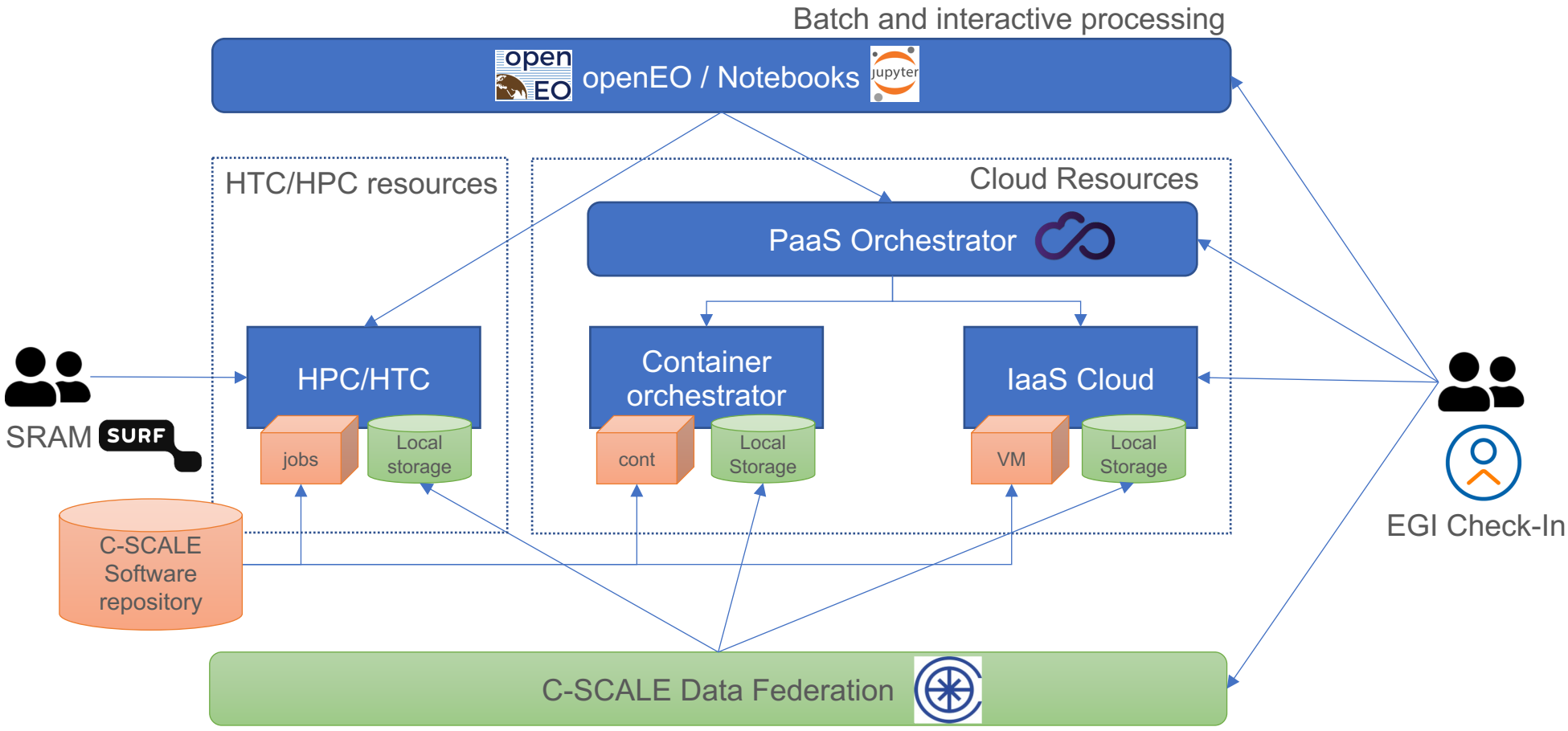
Workflow solutions



C-SCALE Support

FedEarthData

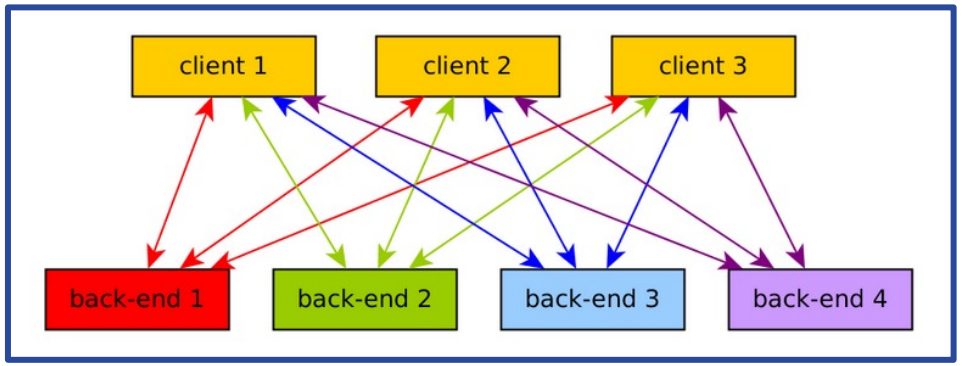
FedEarthData: federation of Earth observation data archives and computing resource providers, enabling execution of Earth observation processing workflows with seamless access to data



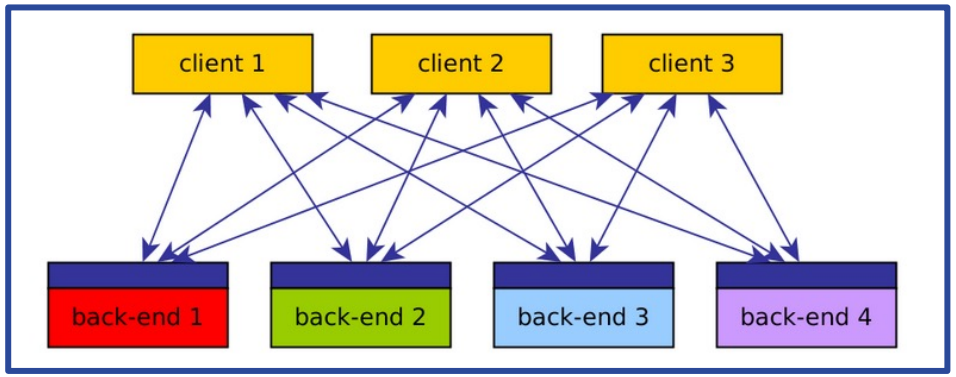
- Authentication and authorisation services
- Identity management infrastructure services
- Monitoring
- Accounting
- Virtual Machine or container image management and redistribution

openEO Platform

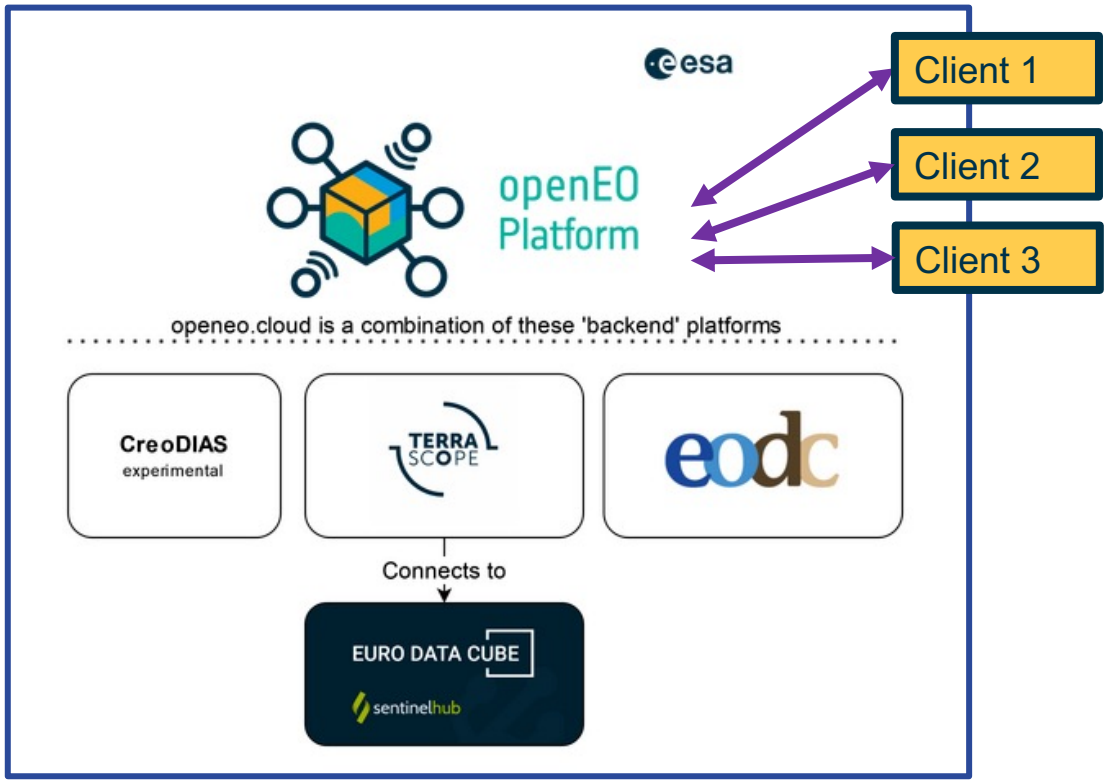
Situation before openEO:



openEO API:



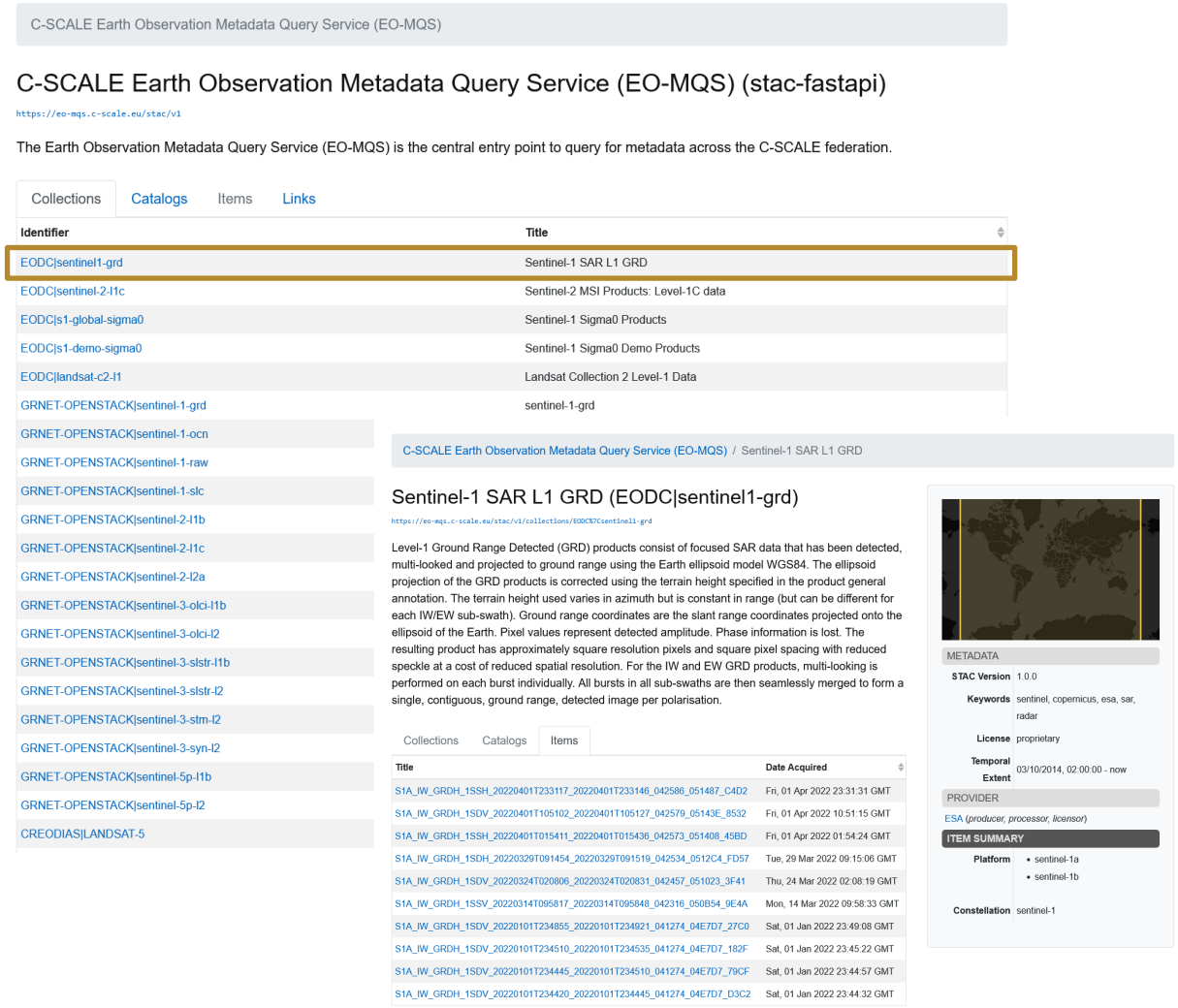
openEO Platform:



- Clients: Python, R, Javascript
- Web Editor
- JupyterLab

Earth Observation Metadata Query Service

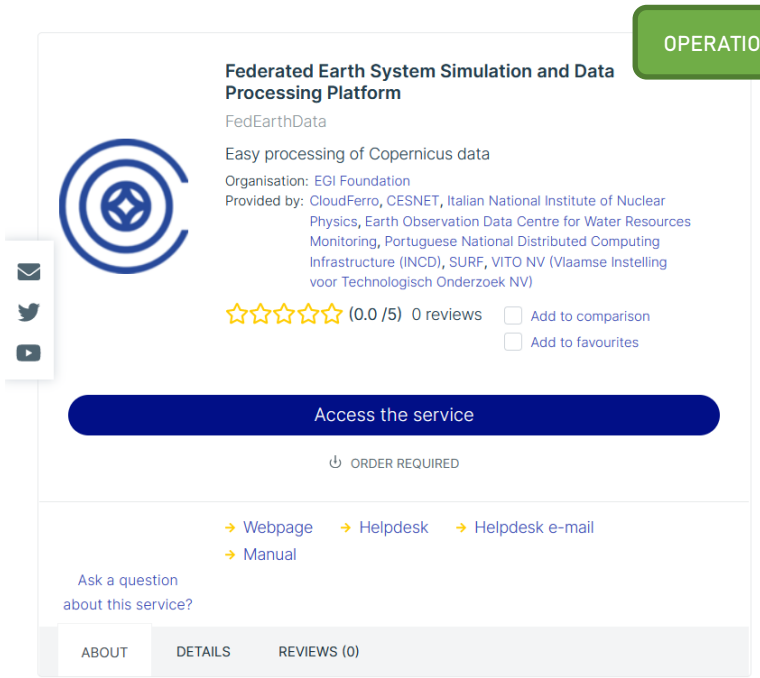
- Earth Observation data discovery service arching over FedEarthData member providers
- Data providers already know where their data are
  - Bring their discovery interfaces under a common one
    - single point
    - shared protocol
- Spatio-Temporal Asset Catalogue (STAC) interface to enable queries across the federation
- EO-MQS is a query broker and aggregator, it is **not yet another metadata database**.
- Focus and data retention policies at member sites - avoiding polling resources irrelevant to the given query



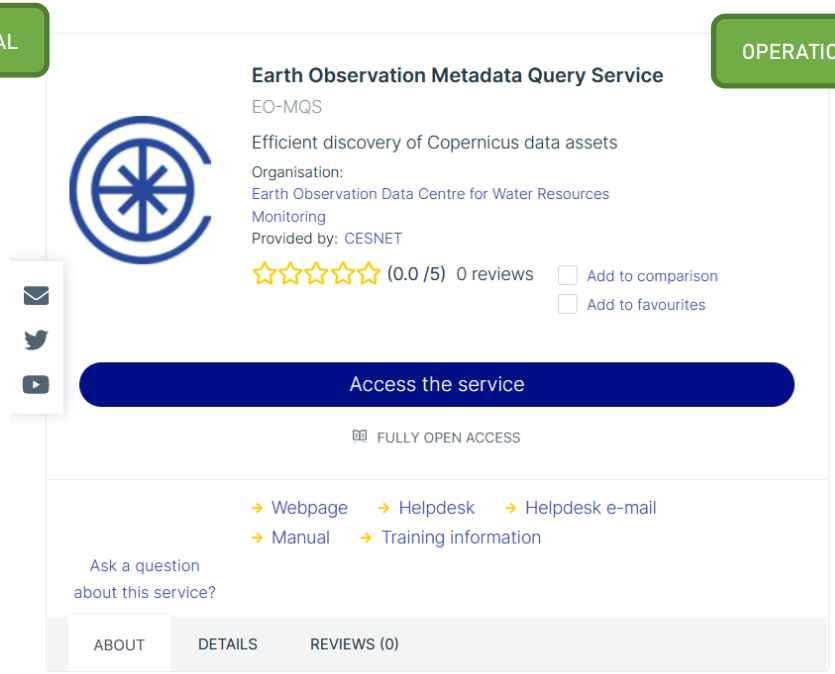
How to access the C-SCALE services

Main exploitation paths:

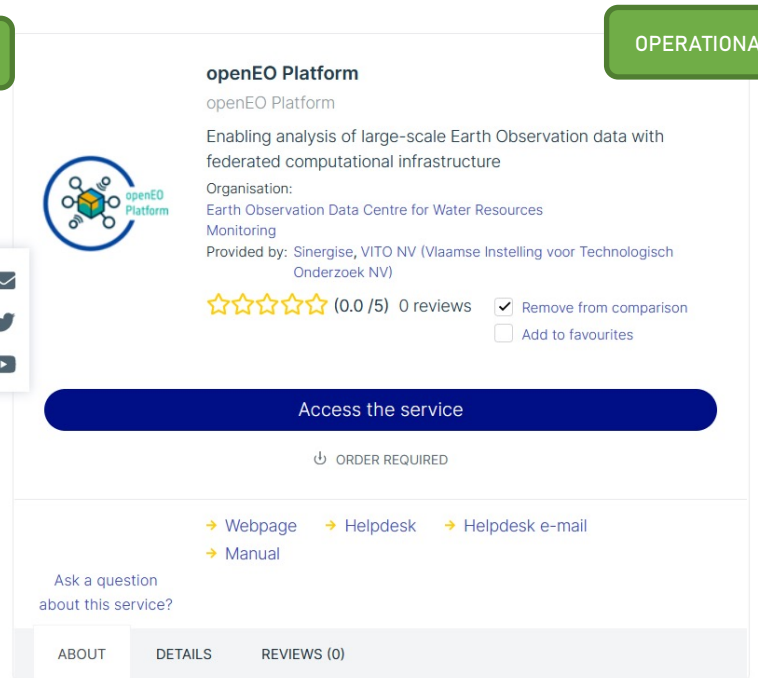
- EOSC Marketplace
- C-SCALE website
- openEO Platform website



https://marketplace.eosc-portal.eu/services/eosc.egi-fed.fedearthdata



https://marketplace.eosc-portal.eu/services/eosc.eodc.eo-mqs

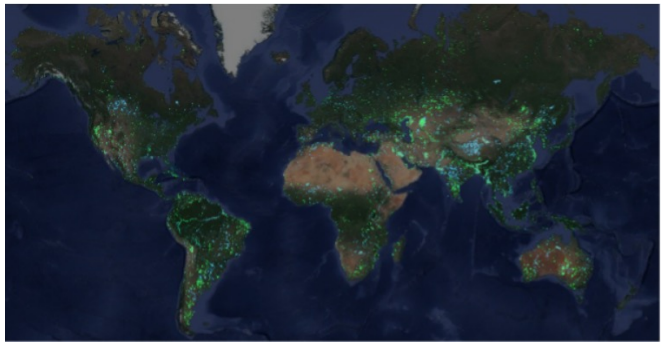


https://marketplace.eosc-portal.eu/services/openeo-platform/

C-SCALE workflow solutions

Workflows for Copernicus data processing: easy deployment of workflows supporting monitoring, modelling and forecasting of the Earth system

- Provided by C-SCALE Use Cases
- Templates and reusable components for users to build their own applications on FedEarthData

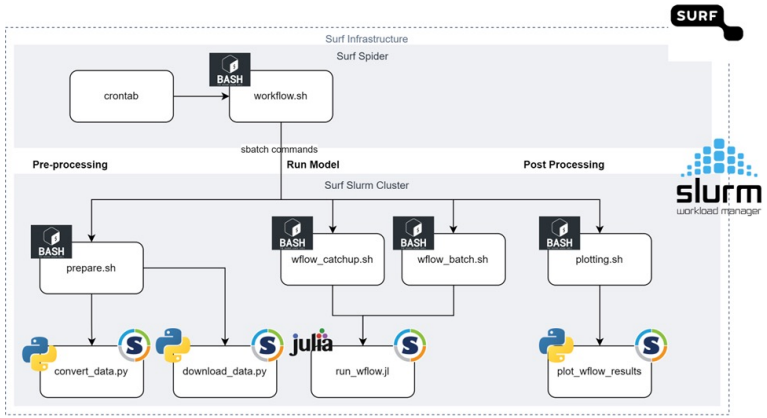


Aquamonitor using OpenEO on C-SCALE

RoHub Link zenodo Link

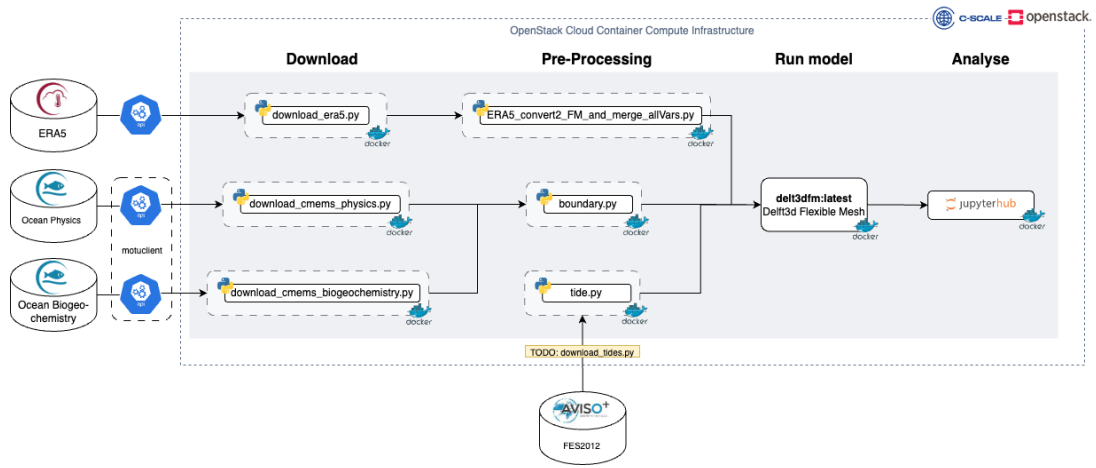
Global Water Watch using OpenEO on C-SCALE

RoHub Link



Automated monthly river forecasts using Wflow

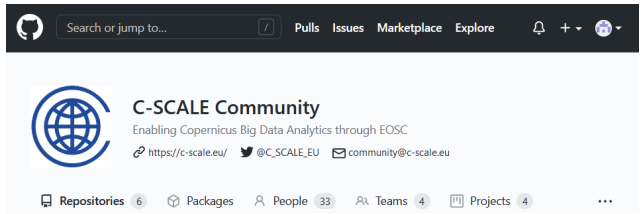
RoHub Link zenodo Link



Coastal hydrodynamic and water quality modelling using Delft3D FM

RoHub Link zenodo Link

User engagement

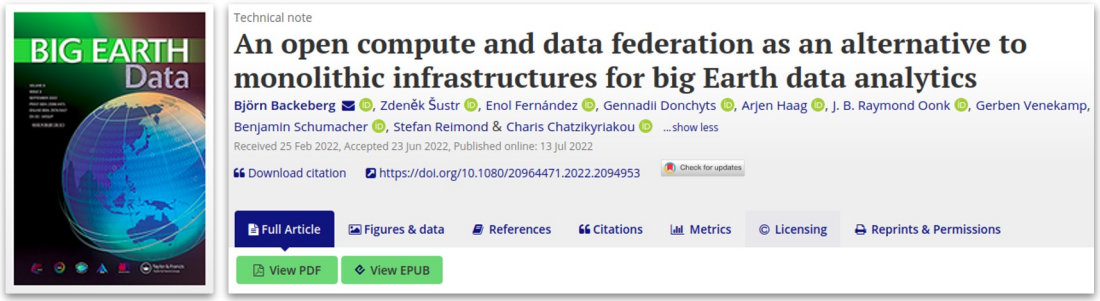


User forum and functional co-design

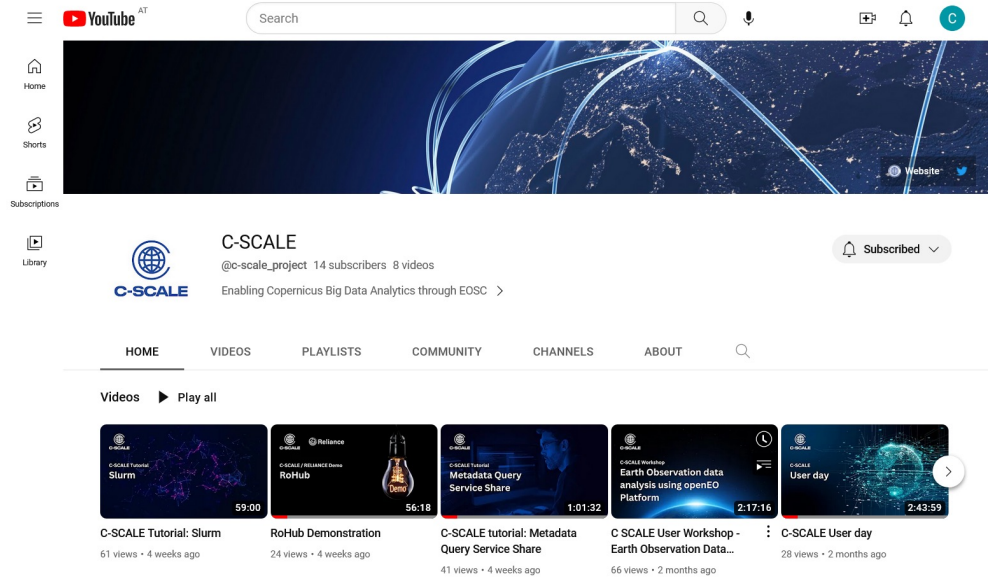
- C-SCALE community: https://github.com/c-scale-community/discussions
- encourages advanced users to become active participants in the development of the future C-SCALE services
- mechanism to engage with the national and international organisations invested in Copernicus services

C-SCALE documentation, training and support

- https://wiki.c-scale.eu/C-SCALE
- https://www.youtube.com/@c-scale_project



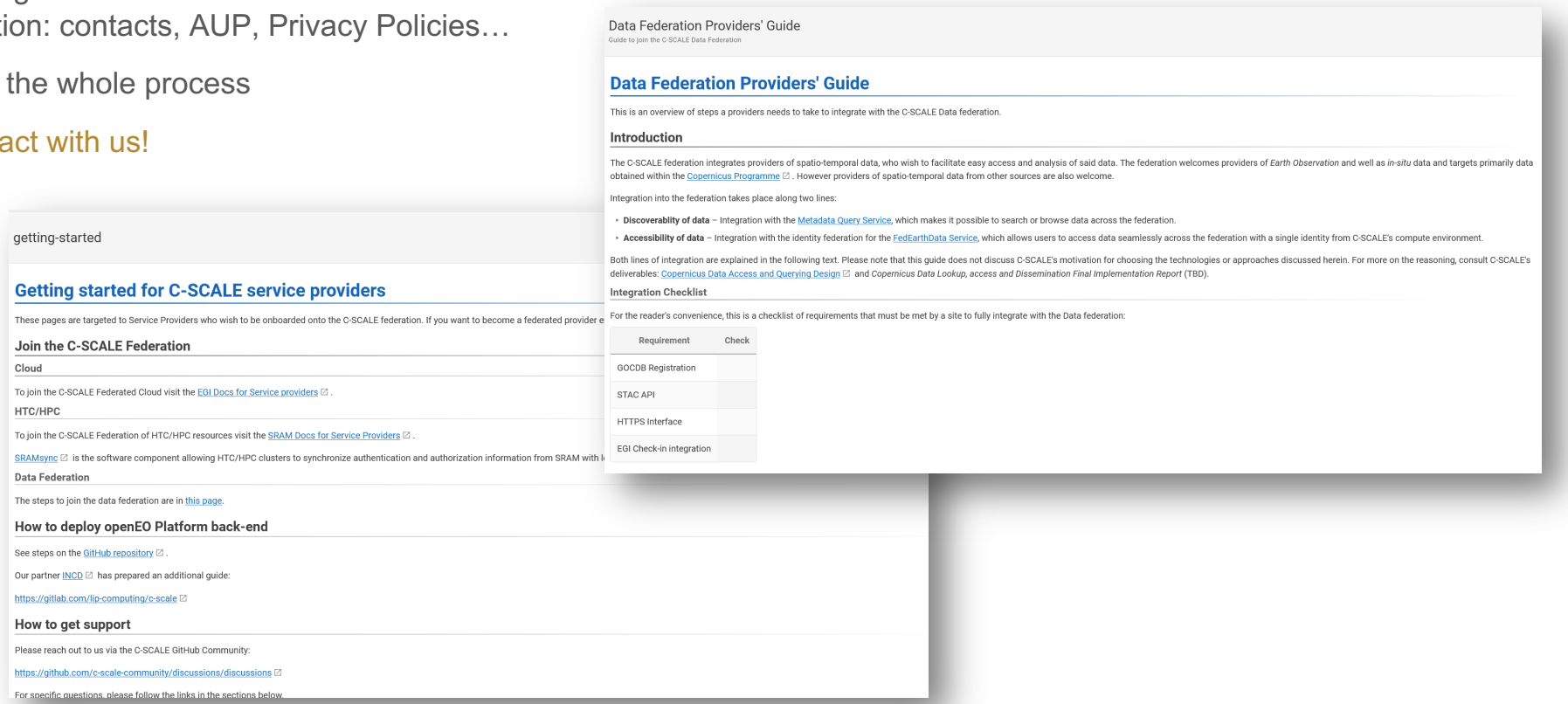
https://doi.org/10.1080/20964471.2022.2094953



Provider onboarding

C-SCALE aims to expand its Compute and Data federation with new service providers!

- Well-defined guidelines to join the federation
  - Technical integration
    - configure your system to allow federated identity
    - register in catalogues
  - Non-technical integration: contacts, AUP, Privacy Policies…
- Support is provided through the whole process
- If you interested, get in contact with us!



getting-started

Getting started for C-SCALE service providers

These pages are targeted to Service Providers who wish to be onboarded onto the C-SCALE federation. If you want to become a federated provider e

Join the C-SCALE Federation

Cloud

To join the C-SCALE Federated Cloud visit the EGI Docs for Service providers .

HTC/HPC

To join the C-SCALE Federation of HTC/HPC resources visit the SRAM Docs for Service Providers .

SRAMsync is the software component allowing HTC/HPC clusters to synchronize authentication and authorization information from SRAM with l

Data Federation

The steps to join the data federation are in this page.

How to deploy openEO Platform back-end

See steps on the GitHub repository .

Our partner INCD has prepared an additional guide:

https://gitlab.com/lip-computing/c-scale

How to get support

Please reach out to us via the C-SCALE GitHub Community:

https://github.com/c-scale-community/discussions/discussions

For specific questions, please follow the links in the sections below.

Data Federation Providers' Guide

Guide to join the C-SCALE Data Federation

Data Federation Providers' Guide

This is an overview of steps a providers needs to take to integrate with the C-SCALE Data federation.

Introduction

The C-SCALE federation integrates providers of spatio-temporal data, who wish to facilitate easy access and analysis of said data. The federation welcomes providers of *Earth Observation* and well as *in-situ* data and targets primarily data obtained within the Copernicus Programme . However providers of spatio-temporal data from other sources are also welcome.

Integration into the federation takes place along two lines:

- **Discoverablity of data** – Integration with the Metadata Query Service, which makes it possible to search or browse data across the federation.
- **Accessibility of data** – Integration with the identity federation for the FedEarthData Service, which allows users to access data seamlessly across the federation with a single identity from C-SCALE's compute environment.

Both lines of integration are explained in the following text. Please note that this guide does not discuss C-SCALE's motivation for choosing the technologies or approaches discussed herein. For more on the reasoning, consult C-SCALE's deliverables: Copernicus Data Access and Querying Design and *Copernicus Data Lookup, access and Dissemination Final Implementation Report* (TBD).

Integration Checklist

For the reader's convenience, this is a checklist of requirements that must be met by a site to fully integrate with the Data federation:

Requirement	Check
GOCDB Registration	
STAC API	
HTTPS Interface	
EGI Check-in integration	

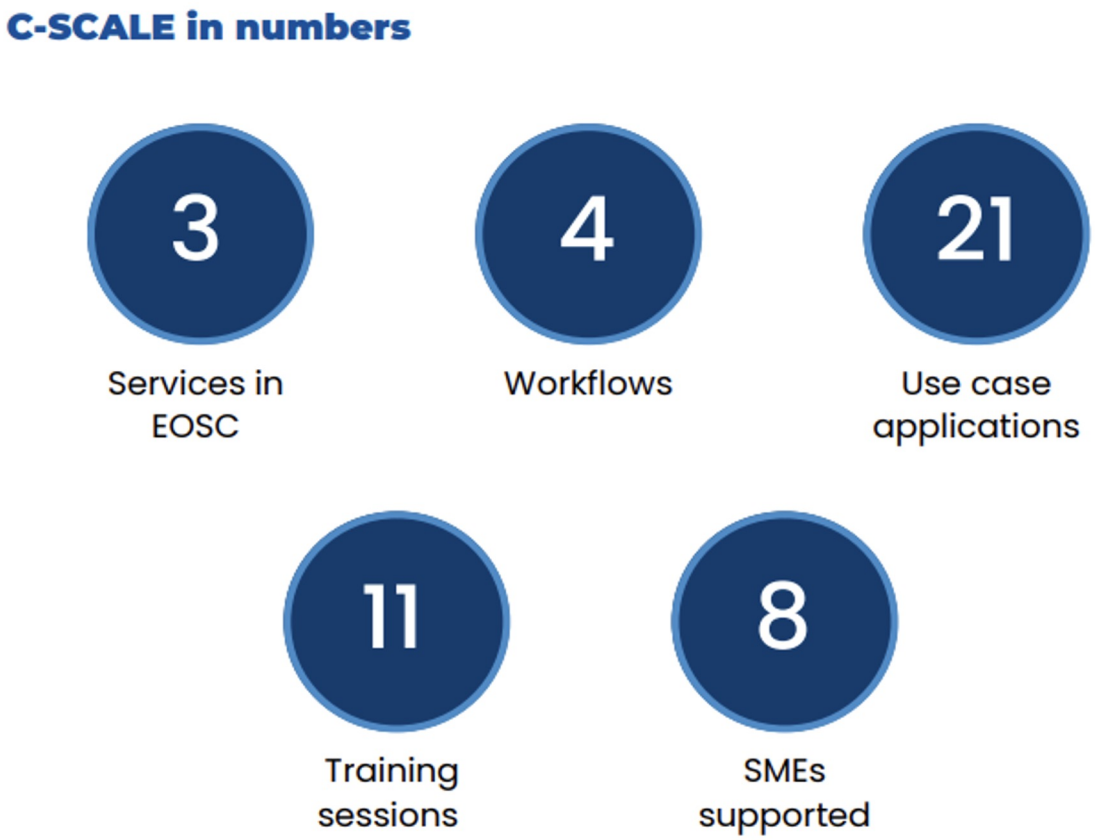
C-SCALE Outcomes & Sustainability

C-SCALE

	FedEarthData	EO-MQS	openEO Platform	Workflow solutions	C-SCALE Support
Key Exploitable Results					C-SCALE Community Enabling Copernicus Big Data Analytics through EOSC https://c-scale.eu/ @C_SCALE_EU community@c-scale.eu
Exploitation	Service Creation / Further Research Standardization	Product / Service Creation / Standardization	Service Creation / Further Research	Product Creation / Further Reaserch / Future Services	Product Creation / Part of Other services / Further Reaserch
Access Channel	EOSC Marketplace & C-SCALE	EOSC Marketplace Openly available on a public URL, C-SCALE	C-SCALE, EOSC Marketplace	EOSC Marketplace, C-SCALE and AppBD	C-SCALE website, GitHub, Twitter, etc
Sustainability	Depends on availability of providers LoC – > MoU & standard SLAs	Depends on availability of providers The STAC-APIs at the individual sites clear value for discoverability of data	Open Source Free - through C-SCALE & ESA Network Commercial License Free Training & documentation	Open Source with documentation (incl. training material). No warranties. SW containers hosted on AppDB.	Most of channels are free from operation point of view. Maintained by C-SCALE key partners

eodc egi Deltares cesnet INFN SURF vito grnet CloudFerro INCD TU WIEN

Impact

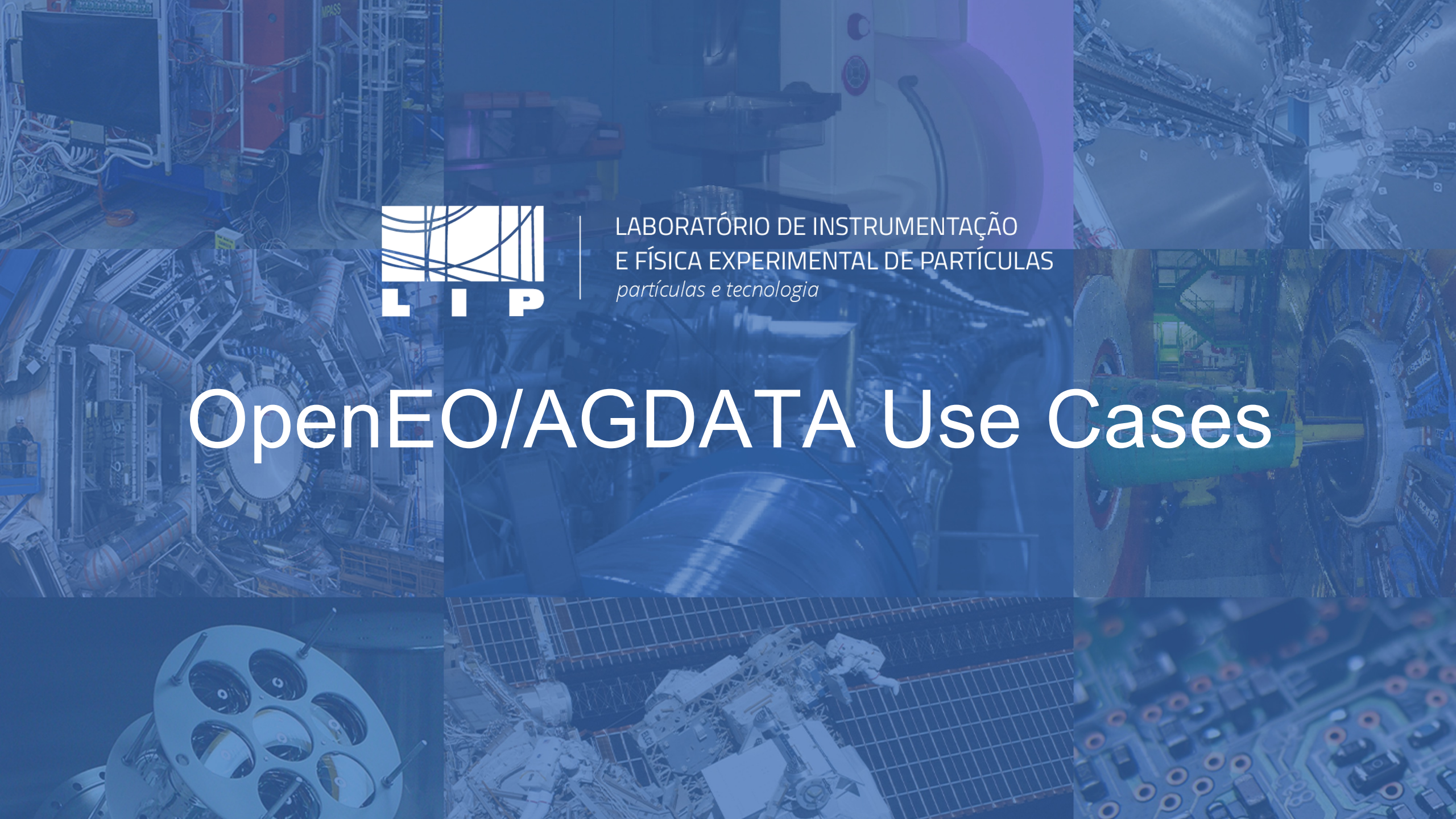


Scientific Impact

Societal Impact

Economic/ Technological Impact

https://c-scale.eu/impact-report/



LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

OpenEO/AGDATA Use Cases

What is OPENEO

The Openeo project, is a free and open source tool that provides an API to analyze and process satellite imagery;

In our CSCALE use case (Aqua Monitor) it uses Copernicus data to scan for land and water mass changes.

What is - Agriculture of data

Portugal aims to participate in the European Partnership “Agriculture of Data”

Currently starting a national pilot project under the calls for Advanced Computing Projects

Aiming at:

- sharing data from different entities and sources
  - including Copernicus data
  - to aid in the monitoring of soil status and of specific culture evolution and growth.
- Development and integration of applications and services
- Create added value while promoting sustainability of agricultural production

What is - AGDATA pilot 1

Monitoring of soil status:

- Develop water evaporation models;
- Check moisture levels on cultivated soil;
- Usage of several data providers
- Usage of several data sources
  - Satellites
  - ground stations
  - other equipment
- Better data = Better decisions.

What is - AGDATA pilot 2

Monitoring of cultivated areas:

- Monitor cultivated and abandoned fields;
- Identify different types of cultures;
- Monitor growth and evolution of crops;
- Use data to create and train predictive models.

Deploy - OPENEO & AGDATA

Deployment of a kubernetes cluster;

Creation of S3 buckets for storage;

Deployment of EODAG;

Deployment of OPENEO;

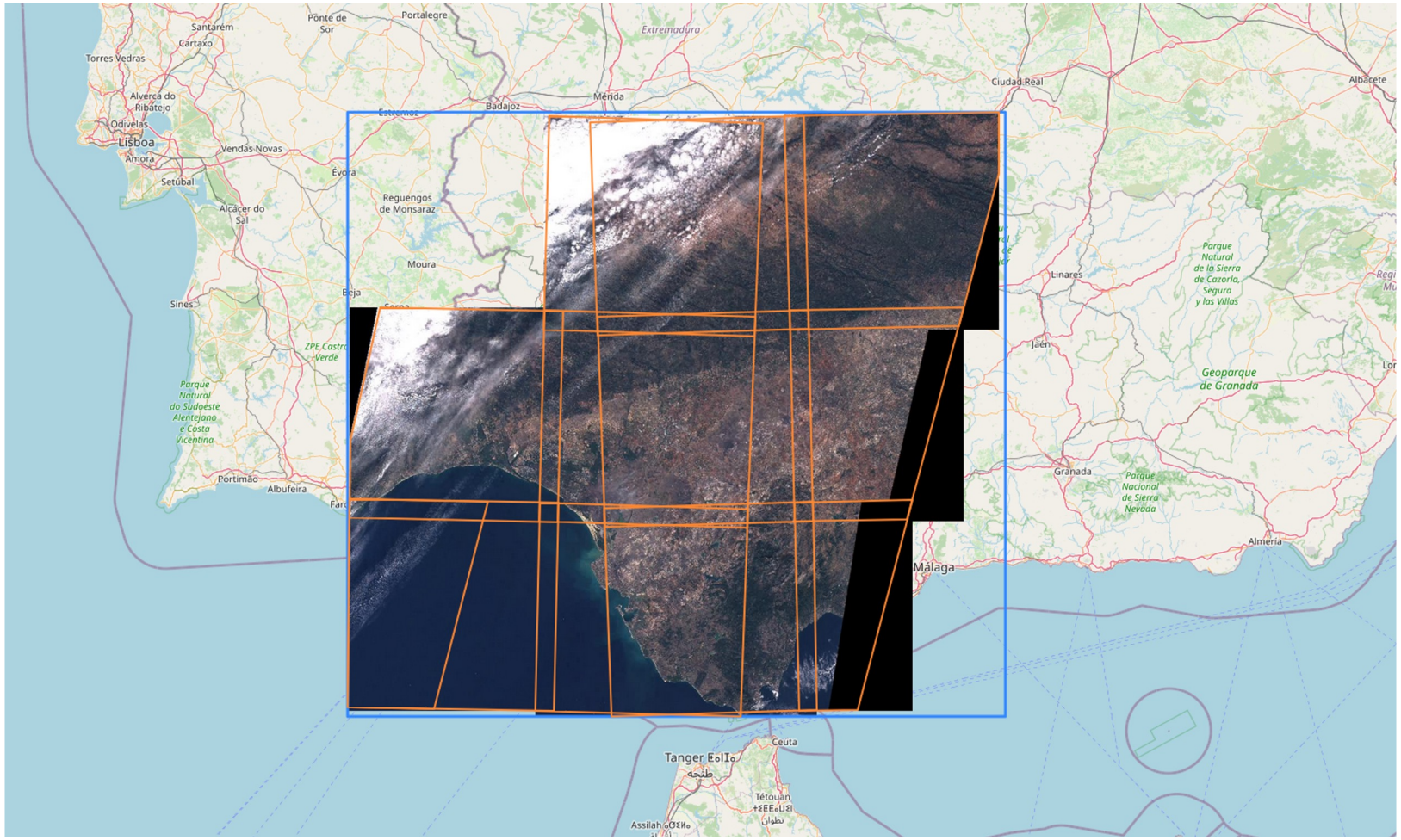
Sharing of Sentinel data;

Creation of a local Stac catalog instance;

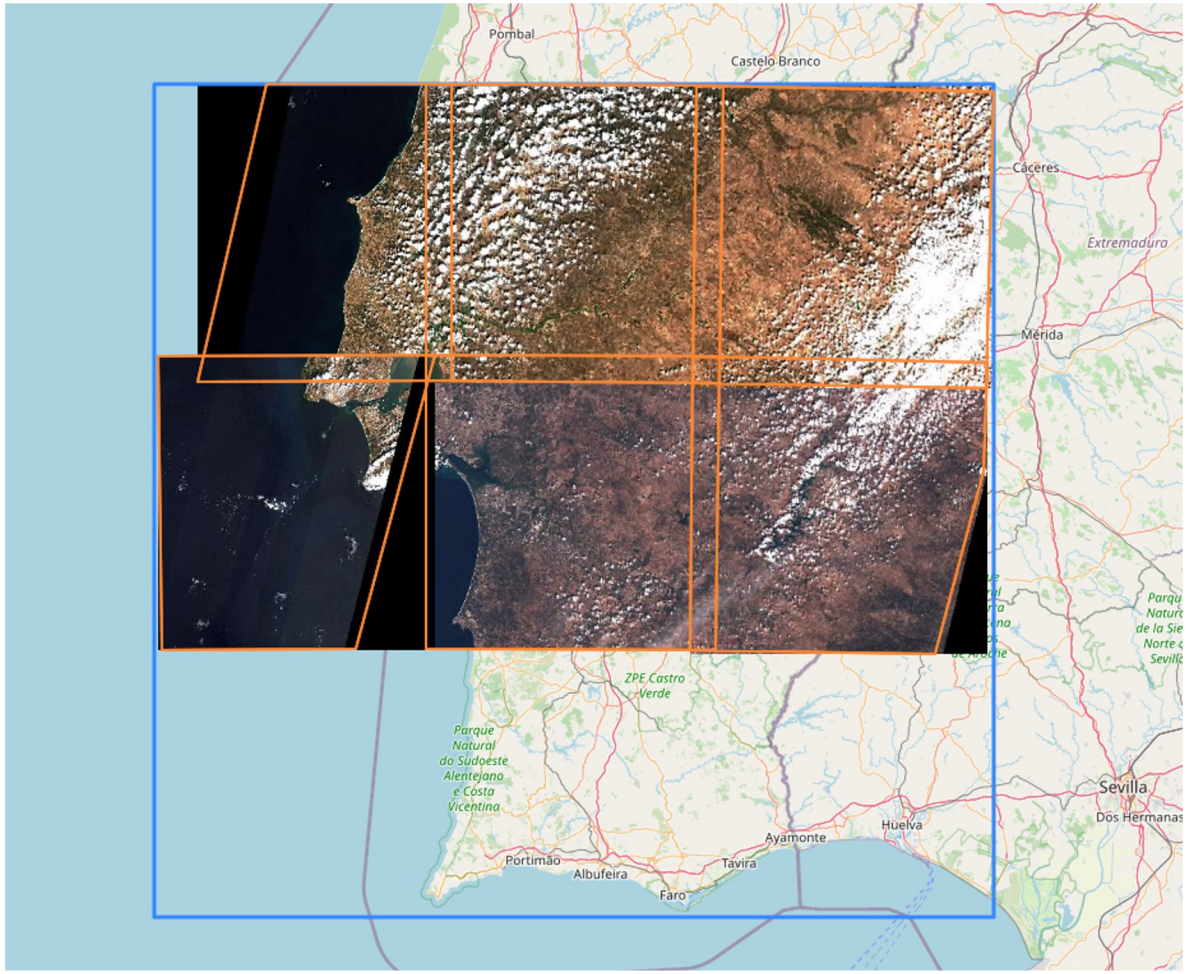
Creation of local Stac collections;

Registration of metadata at INCD Stac Catalog.

Satellite imagery data - OPENEO



Satellite imagery data - AGDATA



Next steps - Agriculture of data pilots

Make relevant data easily available to the applications

Improve applications to exploit the combined data

- Meteorological
- Sentinel 1 and 2
- Exploit higher image resolutions

C-SCALE

Copernicus - eoSC AnaLytics Engine

Thank you for your attention.

Xavier Salazar (EODC), Zacharias Benta (INCD) and the C-SCALE Team

Xavier.Salazar@egi.eu , zacarias@lip.pt

contact@c-scale.eu

https://c-scale.eu

@C_SCALE_EU

https://www.youtube.com/@c-scale_project

Ibergrid 2023, Benasque, Spain & Online | 28 Sept 2023

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

- Any questions?
- You can find me at zacarias@lip.pt