



**C-SCALE**

Copernicus - eoSC AnaLytics Engine

# C-SCALE: Developments towards a federated compute and data infrastructure for Copernicus

Charis Chatzikyriakou, EODC ([charis.Chatzikyriakou@eodc.eu](mailto:charis.Chatzikyriakou@eodc.eu))

Zdeněk Šustr, CESNET ([sustr4@cesnet.cz](mailto:sustr4@cesnet.cz))

EODC Forum 2022 | 15.06.2022 | Vienna, Austria

# Background



- **EU Copernicus programme**: key global source for high resolution EO data
  - Copernicus is the largest producer of EO data in the world
  - Significant contribution to the *digital twin Earth* vision of EU
- There is **no single European processing back-end** that serves all datasets of interest
  - **limits the integration** of these data sources in science and monitoring applications
- **Big (Copernicus) Data Analytics** require a federated infrastructure with a core cloud computing and storage architecture optimised for very large data handling and fast user query response.



# The C-SCALE Project



## C-SCALE: Copernicus - eoSC AnaLytics Engine

- Project duration: Jan 2021 – June 2023
- Consortium of 11 partners with pan-European coverage



Enhance EOSC Portal with pan-European **federated data** and **computing infrastructure** for Copernicus



## 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 project results



## Compute and Data Federation

Uniform access to a federation of computing and data providers to execute Copernicus and Earth Observation workloads



openEO Platform

## openEO Platform

Versatile cloud-based processing and analytics environment for Earth Observation data on infrastructures supporting openEO API



## Metadata Query Service

One-step discovery of Copernicus data in data archives across the C-SCALE Data Federation and across Europe



## Workflow solutions

Easy deployment of workflows supporting monitoring, modelling and forecasting of the Earth system



C-SCALE Community

Enabling Copernicus Big Data Analytics through EOSC

<https://c-scale.eu/> [@C\\_SCALE\\_EU](https://twitter.com/C_SCALE_EU)

[community@c-scale.eu](mailto:community@c-scale.eu)

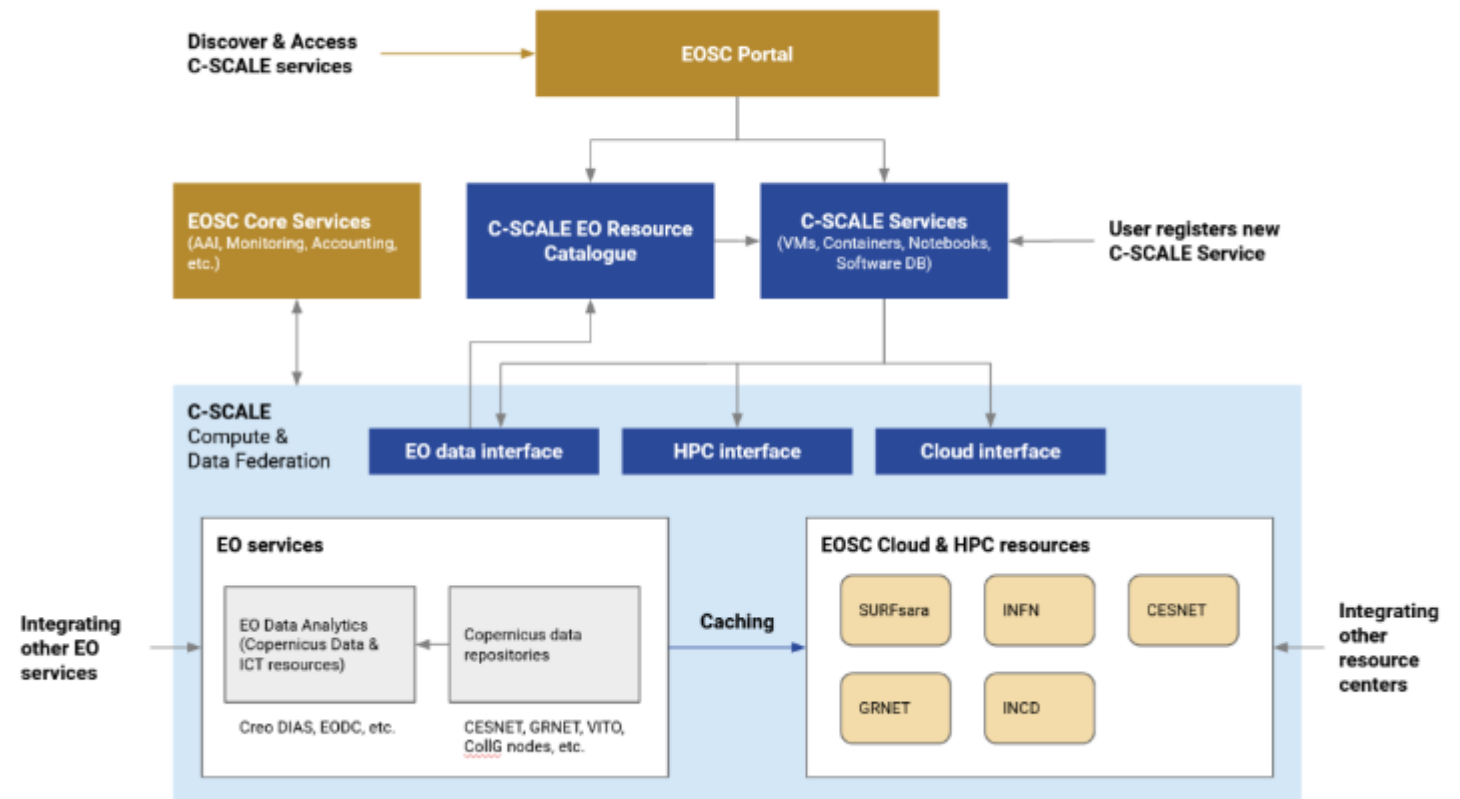
## C-SCALE Community

Functional co-design through a user forum for Copernicus users accompanied by a documentation site

# C-SCALE Federation



- Data, compute, platforms
- Federate **distributed** resources
  - Archives
  - Cloud
  - Supercomputing (HPC/HTC)
- In a uniform way, with a single identity



# Federating around Copernicus Data



- Partial archives all around us
  - ESA CollGS, DIAS, national institutes
  - Each covers a certain part of the total (area-wise, time-wise, ...)
- Let's make their data **discoverable**
  - Expose a standardized interface (STAC)
  - Offer a single point of contact to query all relevant (Metadata Query Service)
    - Registration of participating sites and their data policies
- Let's make sure user can **access** the data in a uniform way
  - Support remote access (HTTP) everywhere
  - Cannot avoid authentication altogether, but we can interate with what already exists

# Federating around Compute



- Both **cloud** and **batch** computing models are supported
- Infrastructures already integrated, under EOSC
  - Cloud based on EGI Federated Cloud
  - Batch (HTC/HPC) federating under SRAM
- Seamless transitions within the federation
  - Logging in once is enough

# Analytics Platforms



- **Batch processing: OpenEO in Cloud or Batch**
  - Managed or self-service deployment
  - Infrastructure Manager single click
  
- **Interactive analysis: Jupyter**
  - Reference implementation
  - Custom instances for custom needs



# Functional co-design

- C-SCALE Federation driven by **user requirements** to guarantee an environment that **satisfies user needs**.
- User requirements are derived from **use cases** deployed or **its usability** and **functional design**.

**E.g. On demand seasonal**

For each use case:

- **Scope** of the challenge
- **Functionality**
- **Workflow**
- **Dependencies**

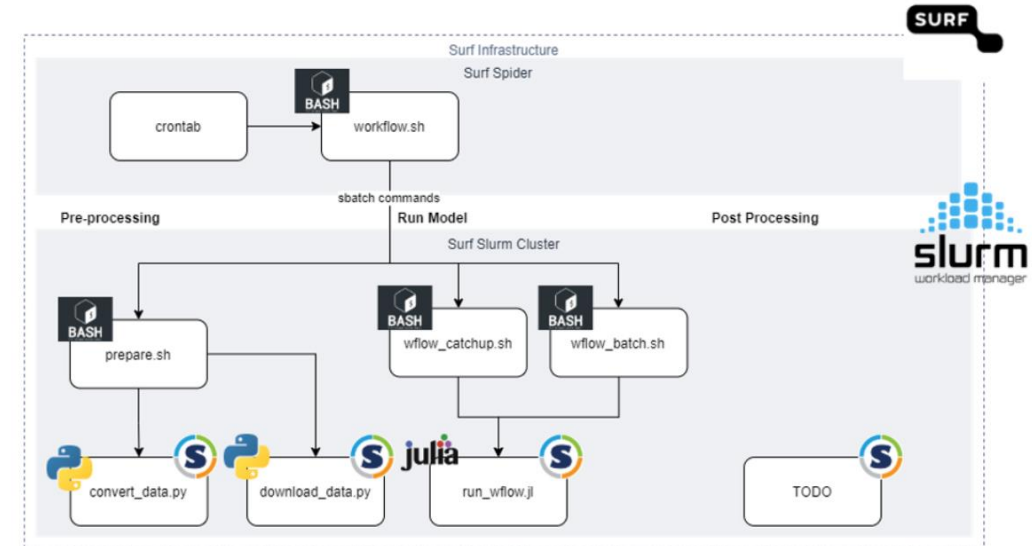
**Challenge:** It compares the performance of ensemble forecasting on HT resources.

**Functionality:** The user can easily deploy a workflow that produces a monthly high resolution, seasonal, ensemble river discharge forecast for a river basin of interest.

**Workflow:**

**Dependencies:**

- Hydrological model – WFLOW
- Data – Hydro-MERIT, HydroLAKES, etc.
- Workflow components – crontab, bash, Python etc.



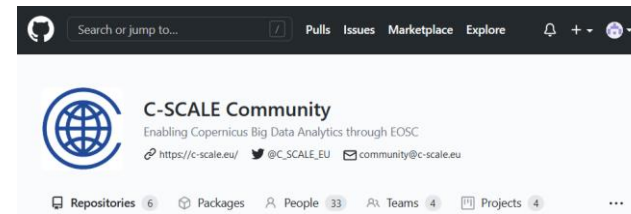
# C-SCALE use cases

#	Internal Use Cases
1	<b>Aqua Monitor:</b> mapping surface water changes globally
2	<b>WaterWatch:</b> quantification of water resources
3	<b>HiSea:</b> High-Resolution Copernicus-Based Information Services at Sea for Ports and Aquaculture
4	<b>HLSDA:</b> High-resolution Land Surface (Drought) Analysis
5	<b>RETURN:</b> Monitoring tropical forest recovery capacity
6	<b>Wetland Water Stresses:</b> Impact of water stress on wetland areas



#	External Use Cases
1	On-demand semantic EO data cubes
2	Data-driven forecasting of global shorelines
3	Development and application of advanced processing chains and data standard for exploitation of Sentinel-1 and other SAR data
4	Benchmarking GPU accelerated SAR ARD generation on cloud infrastructure

# 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

- <https://wiki.c-scale.eu/C-SCALE>

### Welcome to the C-SCALE documentation

C-SCALE's purpose is to federate Earth Observation service providers and offer users the possibility to access, process, analyse and share Copernicus data, tools, resources and services through the [EOSC Portal](#). Please visit the [C-SCALE](#) website to learn more about the mission and vision.



[C-SCALE Users](#)



[Data Inventory](#)



[C-SCALE Service Providers](#)

# C-SCALE Open Call



Are you interested in piloting a new federated infrastructure for Copernicus and Earth Observation data? **Apply now!**

The C-SCALE offering:

- 12 PB months of storage
  - 18 million Cloud CPU hours
  - 3.1 million HPC/HTC CPU hours
  - 6,000 GPU hours
- **Free at the point of use** enabled through EC's **Virtual Access** mechanism.

**C-SCALE**

About Services Work Packages Documentation Partners Community

**Our Call for Use Cases is now open!**

[Submit your application](#)

**Copernicus - eoSC Analytics Engine**

**Enabling Copernicus Big Data Analytics through EOSC**



Deltares



Copernicus - eoSC AnaLytics Engine

# Thank you for your attention.

Charis Chatzikyriakou, EODC ([charis.Chatzikyriakou@eodc.eu](mailto:charis.Chatzikyriakou@eodc.eu))

Zdeněk Šustr, CESNET ([sustr4@cesnet.cz](mailto:sustr4@cesnet.cz))



[info@c-scale.eu](mailto:info@c-scale.eu)



<https://c-scale.eu>



[@C\\_SCALE\\_EU](https://twitter.com/C_SCALE_EU)

EODC Forum 2022 | 15.06.2022 | Vienna, Austria



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