



Copernicus - eoSC AnaLytics Engine

Enabling Copernicus Big Data Analytics through EOSC

Christian Briese, EODC

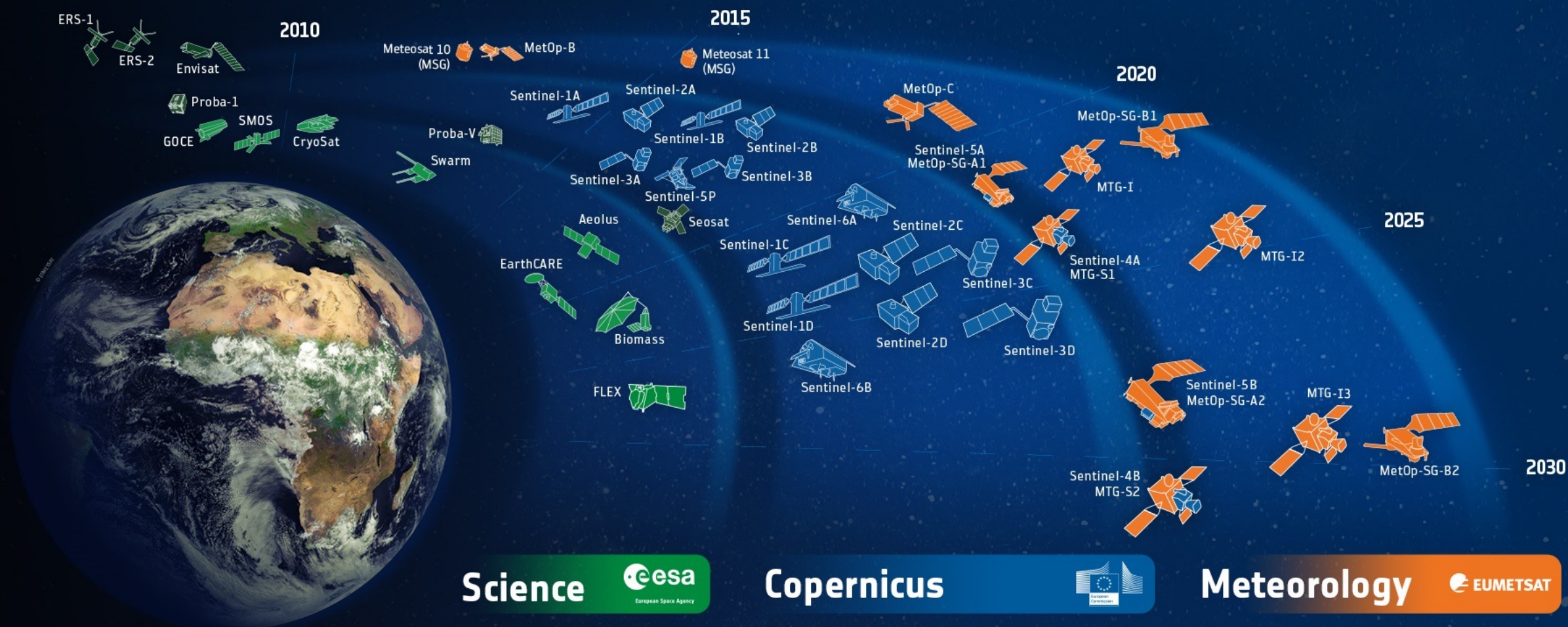
christian.briese@eodc.eu

Seminar: Enabling Copernicus Big Data Analytics through European Open Science Cloud

27/10/2021

Enormous increase in satellites and data

Taken from ESA Φ-WEEK, openEO platform Side Event, Patrick Griffiths, ESA



Copernicus | Problem Statement



- **EU Copernicus programme**: key global source for high resolution EO data
 - Copernicus is the largest producer of EO data in the world
 - Daily data volume: **20 TB per day from the Sentinel mission**
 - 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.



Project Introduction



C-SCALE (Copernicus - eoSC AnaLytics Engine) - Grant agreement ID: 101017529

- Funded by the EU
 - under the Programme: H2020-EU.1.4.1.3. - Development, deployment and operation of ICT-based e-infrastructures
 - and the topic: INFRAEOSC-07-2020 - Increasing the service offer of the EOSC Portal
- Project duration: January 2021 – June 2023 (30 months)
- Coordinated by EODC, 11 partners with pan-European coverage
- Overall budget: ~ 2M €

C-SCALE Consortium



The C-SCALE consortium brings together expertise from:

✓ the EO sector:



✓ e-Infrastructure:



Project Mission



Vision

To empower European researchers, institutions and initiatives to easily discover, access, process, analyse and share Copernicus data, tools, resources and services through the EOSC Portal in a way that can be seamlessly integrated into their processes and research practices.



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



Mission

- Enhance EOSC Portal with pan-European **federated data and computing infrastructure** for Copernicus
- **Integrate cross-/inter- disciplinary EOSC services**, ensuring **interoperability** between distributed data catalogues, computational tools and infrastructure
- **Increase the service offer of the EOSC Portal** providing state-of-the-art research enabling services to its users.

Service provisioning

The C-SCALE federation will make available a suite of three services in through the EOSC Portal:

C-SCALE EO Data Archive

- Access and download data through the C-SCALE EO Data archive



C-SCALE Compute Services

- Access large C-SCALE compute services through standard interfaces near to the data



C-SCALE Analytics Platforms

- Seamlessly deploy data analytics on top of the C-SCALE EO Data Archive and Compute Services



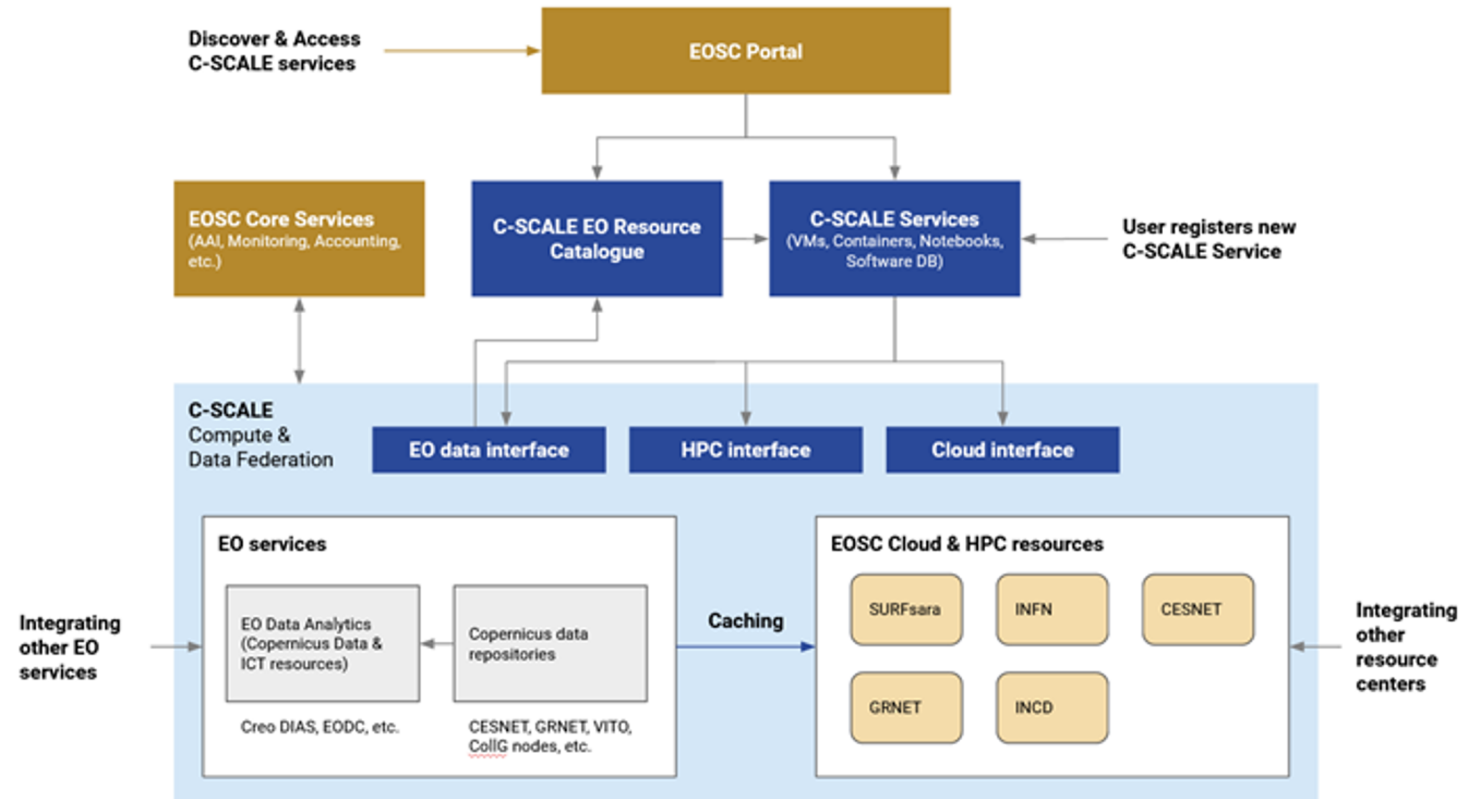
The C-SCALE federation services will be available in the EOSC Portal in the second half of 2022.

Federation of infrastructures and EO data



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



Use Cases



#	Use Case
1	Aqua Monitor
2	WaterWatch
3	HiSea
4	High-resolution Land Surface (Drought) Analysis
5	RETURN
6	Virtual European Sentinel Data Cubes

Role: **validate** and **optimise** the C-SCALE federation

Benefits:

- Cloud agnostic
- Independence from commercial, closed, non-EU providers
- Cross- / inter-disciplinary exposure
- FAIRer

User support & training



Training and user support to help them become proficient in using the C-SCALE federated data analytics infrastructure and platforms

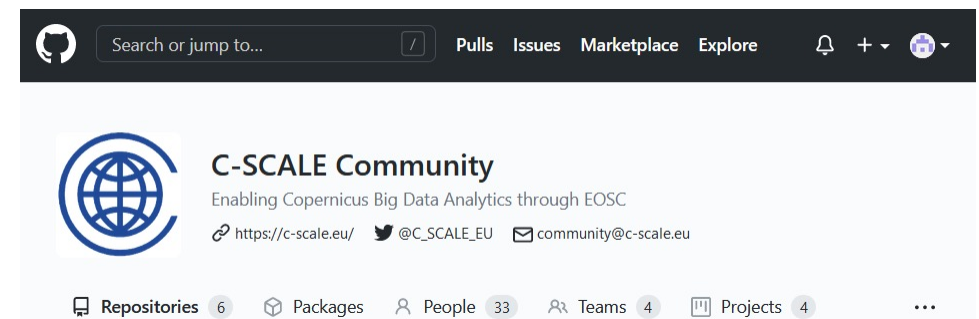
- Create **training material**
- Organise **webinars** and **tutorials**
- Participate and contribute to **events** within the EO community
- Additional, on-demand, support to ensure efficient and effective service delivery and user satisfaction



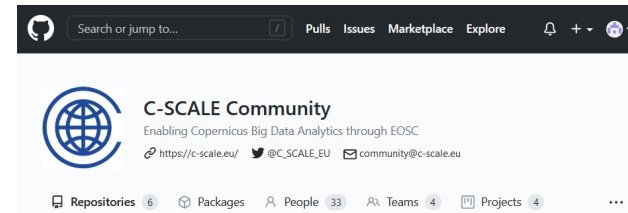
First trainings are planned to start in Q3 2021.

User Forum

- Creation of a **C-SCALE community**
 - <https://github.com/c-scale-community/discussions>
 - *“The C-SCALE community is a place for researchers, end-users, institutions and initiatives: in short anyone with an interest in Copernicus data, tools, resources and services”*
- Serves for **discussions**, **user support**, **community engagement**



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 organizations** invested in Copernicus services



Early Adopter Programme

- **Open call:** <https://c-scale.eu/call-for-use-cases/>
- **Aim:** expand the user community and further enable the co-design of C-SCALE components



Outreach activities

- increase **awareness** of Copernicus-related EOSC services and broaden the user community

C-SCALE Virtual Access resources



Installations (short name)	Service Category	Unit of access	Units
EODC EO Storage	Storage Service	TB year	150
EODC EO Cloud	Cloud Service	CPU year	620
EODC HPC	HPC/HTC Service	CPU Hour	600,000
CESNET MetaCentrumCloud - CPU	Cloud Service	CPU Month	2,850
INFN Bari Storage	Storage Service	TB Month	1,500
INFN Bari CPU	Cloud Service	CPU Month	3,840
SURFsara dCache Front end Storage	Storage Service	TB Year	154
SURFsara Spider (HTDP) storage	Storage Service	TB Year	160
SURFsara Data Processing Compute	HPC/HTC Service	CPU Hour	1,500,000
SURFsara MS4 Storage	Storage Service	TB Year	10
SURFsara MS4 SSD Storage	Storage Service	TB Year	96
SURFsara MS4 Compute	HPC/HTC Service	CPU Hour	147,600
VITO CVB (Storage)	Storage Service	TB Month	1,200
VITO CVB (Compute)	Cloud Service	CPU Hour	876,000
GRNET KNS-Storage Cloud	Storage Service	TB Hour	250,000
GRNET KNC Cloud	Cloud Service	CPU Hour	1,650,000
GRNET ARIS-Storage HPC	Storage Service	TB Hour	219,000
GRNET ARIS-Compute HPC	HPC/HTC Service	CPU Hour	1,000,000
CREODIAS – Storage	Storage Service	TB Month	2,762
CREODIAS - Compute	Cloud Service	vCore hour	1,500,000
CREODIAS – GPU	GPU service	GPU hour	6,000
INCD Lisbon (NCG) (Storage)	Storage Service	TB Month	450
INCD Lisbon (NCG) (Compute)	Cloud Service	CPU Day	4500

Total capacity of resources offered through VA:

- Cloud: 1648 CPU core/year
- HTC/HPC: 370 CPU core/year
- Storage: 1104 TB/Year;

Conclusions



- C-SCALE puts together **EO** and the **e-infrastructure partners** to:
 - Facilitate the exploitation of Copernicus data leveraging on large resources and advanced technologies from e-infras and EOSC
 - Make Copernicus resources easily accessible to new research areas and EOSC in general
- C-SCALE will deliver a **federation of EO** and **e-infrastructure services and resources**
 - Create a very large distributed repository of EO data close to compute resources
 - EO data will be FAIR through the federation
 - Federation services accessible through the EOSC Portal
- C-SCALE federation will be **co-designed with researchers**
- C-SCALE also provides a link to the recently launched **openEO platform**, see <https://openeo.cloud/>
- Large amount of resources available through the **Virtual Access** mechanism



Thank you for your attention.



info@c-scale.eu



<https://c-scale.eu>



[@C_SCALE_EU](https://twitter.com/C_SCALE_EU)