



Copernicus - eoSC AnaLytics Engine

FedEarthData: Federated Earth System Simulation and Data Processing Platform

Enol Fernández, EGI Foundation – Raymond Oonk, SURF

EGI Conference 2022 | 21th September 2021



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

Motivation



C-SCALE aims at providing an adequate IT infrastructure to allow researchers to easily discover, access and process the massive streams of high resolution EO data from Copernicus and related initiatives.

- Avoid lock-in and use a distributed infrastructure → federation
- Flexible approach to computing → support Cloud, HTC/HPC, PaaS & interactive analysis
- Free at point of use → Virtual Access

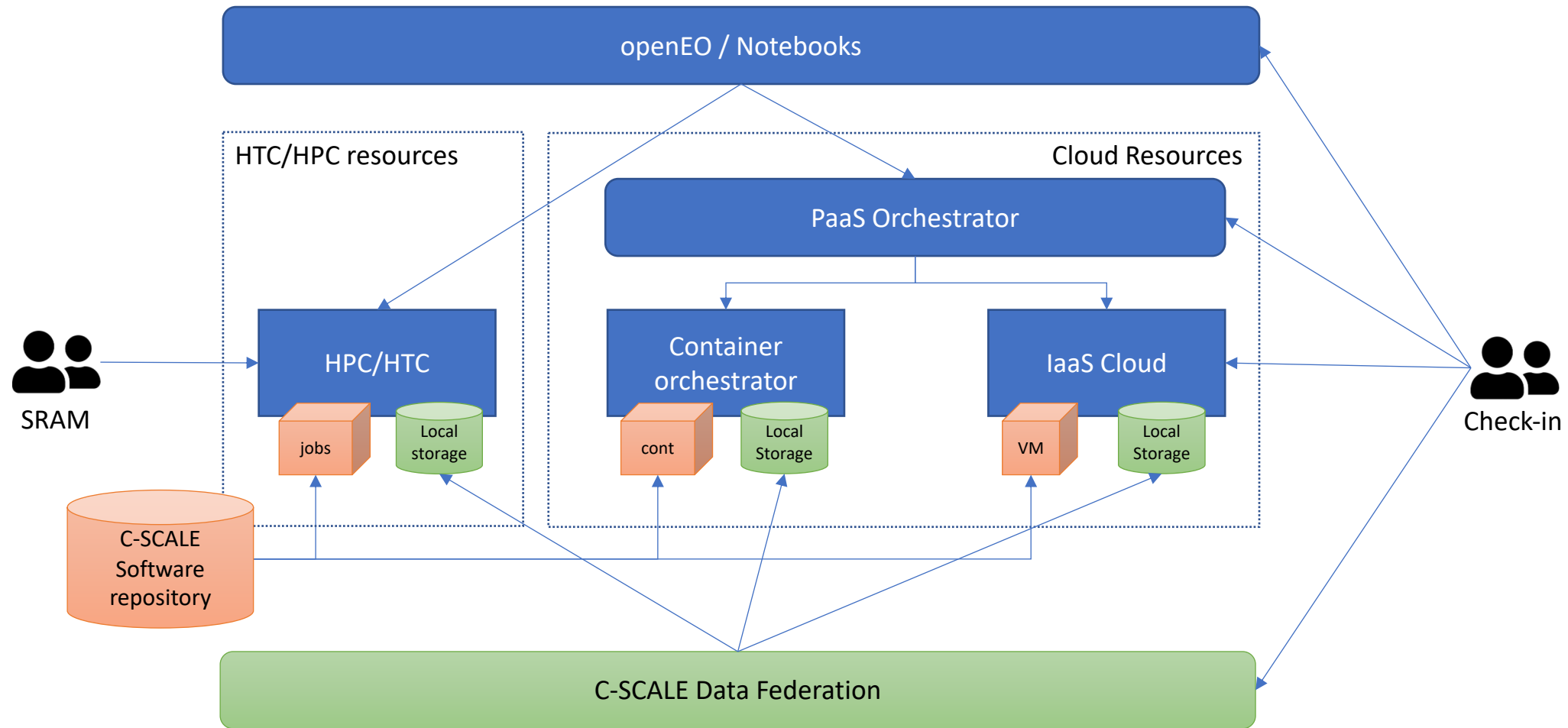
The FedEarthData service



- A distributed infrastructure of data and compute providers to support the execution of Earth System Simulation and Data Processing workflows at scale
- Flexible computing capacity
 - Cloud IaaS
 - HTC & HPC
 - PaaS Orchestration
 - Notebooks
 - openEO
- With access to a large collection of EO (Copernicus) datasets



FedEarthData



FedEarthData: providers

- **HPC/HTC**

- EODC/TUW (AT)
- GRNET (GR)
- SURF (NL)

- **Cloud**

- CESNET (CZ)
- CloudFerro (PL)
- EODC (AT)
- INCD (PT)
- INFN (IT)
- VITO (BE)



Accessing the federation: AAI

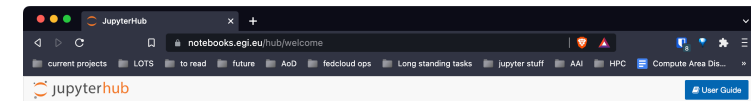


- AAI (Authentication and Authorization Infrastructure) provides federated login into the distributed providers
 - Build on the existing work around EOSC AAI and AARC
 - Using mature products: Check-in, Perun & SRAM
 - Proven technology: OpenID Connect and LDAP
- Clear responsibility model
 - Each community has a (set of) manager(s) that decide who is entitled to be a member of the community
 - Providers take authorisation decisions locally based on membership to communities of users (and potentially other attributes)

Notebooks



- C-SCALE interactive environment is Jupyter
 - Open Source, browser accessible data analytics platform
 - Support for Python, R, Julia and many other languages
 - Ready-to-use deployments at C-SCALE providers: EODC, CESNET, VITO, CloudFerro
- EGI Notebooks for evaluation of C-SCALE
 - Running on CESNET, wide range of data analytics libraries available
 - Now populating EO-specific samples for users



Notebooks is an environment based on Jupyter and the [EGI cloud service](#) that offers a browser-based, scalable tool for interactive data analysis. The Notebooks environment provides users with notebooks where they can combine text, mathematics, computations and rich media output.

The notebooks are limited to 2 CPU, 4GB RAM and 20GB of persistent storage per user. Access requires a valid EGI account and enrolling to the [vo.notebooks.egi.eu VO](#).

[Start your notebooks!](#)

User communities/advanced users can have their customised EGI Notebooks service instance. EGI offers consultancy and support, as well as can operate the setup. Order a community notebooks instance via the Marketplace.



[Privacy policy](#) | [Terms of use](#)

- C-SCALE batch processing environment is **openEO**.
 - Intuitive API to process a wide variety of earth observation datasets
 - Data access and processing on multiple infrastructures based on data-cubes
- Using openEO on C-SCALE:
 - openEO platform - <https://openeo.cloud/>, coming soon to EOSC marketplace – **featured as demo this morning!**
 - Provider-managed: deployment and configuration of openEO for a set of users but provider (not yet) part of openEO platform
 - Self-service: user-managed deployment of the openEO stack on cloud provider of choice → PaaS orchestrator

Cloud Federation

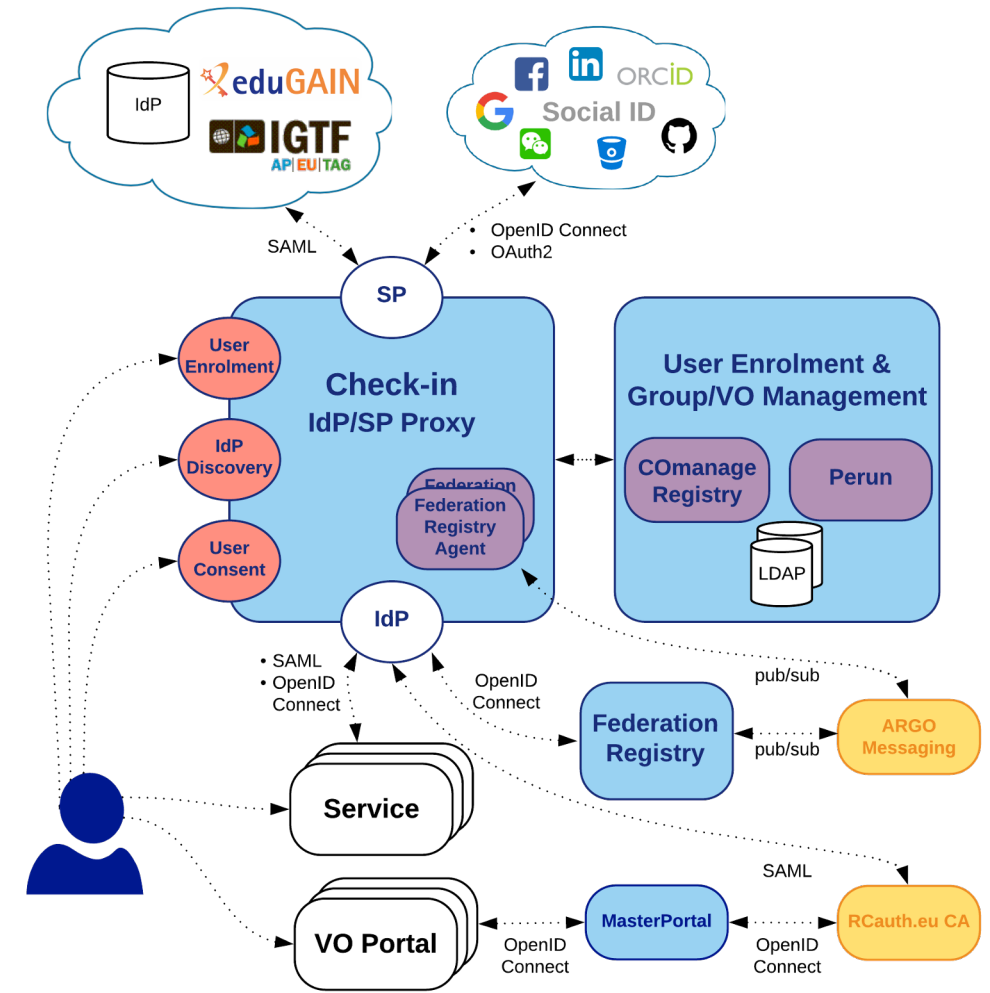


- C-SCALE cloud builds on the EGI Federated Cloud...
 - IaaS providers with federated identity and operational integration (accounting, monitoring, shared VM images)
 - Well-established operational model (production since 2014)
 - Allows to run a wide range of workloads
- ...and goes beyond with PaaS Orchestration
 - Automated deployment of applications across different providers/infrastructures (VM + containers + potentially others) with data aware deployment
 - With templates for commonly used applications (e.g. kubernetes, spark, **openEO**)

Check-in



- Check-in provides authentication, authorization and user management for C-SCALE cloud & data federation
- Standards based:
 - SAML 2.0 / OpenID Connect 1.0 / OAuth 2.0 / LDAP
- Interoperable:
 - AARC and EOSC AAI compliant
 - Support for legacy X.509 services via MasterPortal
- Community management:
 - Comanage and Perun supported
 - Other Community AAI's pluggable



AppDB



- AppDB provides a catalogue of VM images for cloud providers
 - Community curated
 - Automatic sync with providers
 - Non-VM software also supported
- Discovery of providers

The screenshot displays the 'Applications Database' website. The header includes a logo and navigation links: Home, Software Marketplace, Cloud Marketplace, and People. The breadcrumb trail shows 'Home > Virtual Organizations > aquamonitor.c-scale.eu'. The main content area features a globe icon and the organization name 'aquamonitor.c-scale.eu' with a permalink and ID (15655). It lists details such as Disciplines (Earth sciences), Scope (Global), Validated On (2021-05-11 11:53:12.0), and Middlewares (n/a). A description states: 'This VO supports the operation activities of the Aquamonitor use case of the C-SCALE project. [read more](#)'. Below this are links for 'Homepage', 'Enrollment', 'EGI Operations Portal link', and 'VMCatcher Image List'. A 'Virtual Organization Resources' section lists: RAM/1386 core, RAM/x86_64 core, Scratch space for jobs, Max CPU time for jobs, and Max wall clock time for jobs. A 'Related items' section shows filters for All (3), Software (0), Virtual Appliances (1), Software Appliances (0), and Resource Providers (2). Three items are displayed: 'EGI Ubuntu 20.04' (Ubuntu 20.04 LTS, 249 visits), 'INFN-CLOUD-BARI' (INFN-CLOUD-BARI), and 'NCG-INGRID-PT' (NO CENTRAL GRID - LIP/FCCN/LNEC - Lisbon - Portugal).

Federation Operations



The screenshot shows the ARGO dashboard with a sidebar on the left containing links to Dashboard, Reports, Trends, Status, Issues, Metrics, Profile Details, and Recomputation. The main content area displays a table of metrics for C-SCALE. The table has columns for Month (2022-05 to 2022-09) and rows for providers (CESNET and INFN-CLOUD-BARI). Each cell contains 'Av' and 'Re' values, both of which are 100, indicating high availability and reliability.

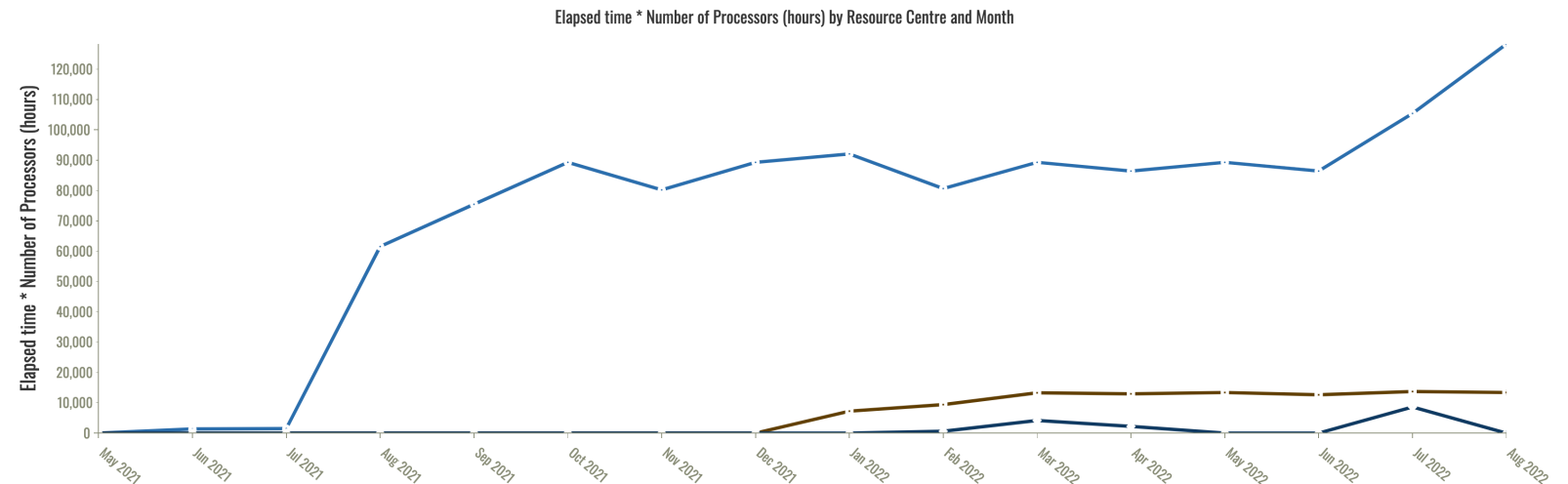
Month	2022-05		2022-06		2022-07		2022-08		2022-09	
	Av	Re	Av	Re	Av	Re	Av	Re	Av	Re
CESNET	100	100	100	100	100	100	100	100	100	100
INFN-CLOUD-BARI	100	100	100	100	100	100	100	100	100	100

Monitoring

- Availability/Reliability metrics of each federated provider, customised to the service types

Accounting

- Gather usage information of communities on available providers



HTC/HPC Federation

FedEarthData on EOSC



- C-SCALE compute federation is already operational
 - 10+1 different use cases supported
 - Covering cloud/HTC/HPC
- FedEarthData service being onboarded into EOSC portal
 - Rely on EOSC helpdesk
 - Finalising service management details with all providers
 - Preliminary entry already registered

The screenshot shows the user interface of the "Federated Earth System Simulation and Data Processing Platform". At the top left is a circular logo with concentric rings and a central diamond shape. To its right, the title "Federated Earth System Simulation and Data Processing Platform" is displayed, followed by the subtitle "Easy processing of Copernicus data" and "Provided by: EGI Foundation". A blue button labeled "ACCESS THE RESOURCE" is in the top right, with "OPEN ACCESS" text below it. Below the header, there are links for "Webpage" and "Helpdesk e-mail", and a link to "Ask a question about this Resource?". A navigation bar at the bottom of the header section contains "ABOUT" and "DETAILS" tabs. The main content area on the left contains a paragraph describing the platform's distributed infrastructure and its use of Copernicus data. On the right, there are two sections: "SCIENTIFIC CATEGORISATION" with a list of categories (Natural Sciences, Engineering & Technology, Agricultural Sciences) and "CATEGORISATION" with a list of categories (Compute).

Federated Earth System Simulation and Data Processing Platform
Easy processing of Copernicus data
Provided by: EGI Foundation

[Webpage](#) [Helpdesk e-mail](#) [Ask a question about this Resource?](#)

ABOUT **DETAILS**

The Federated Earth System Simulation and Data Processing Platform provides a distributed infrastructure of data and compute providers to support the execution of Earth System Simulation and Data Processing workflows at scale. It offers a flexible cloud-based data processing capacity to create and scale data processing pipelines that run on optimised execution environments near the data. Jupyter Notebooks and openEO API offer user friendly and intuitive processing of a wide variety of Earth Observation datasets on these computing providers, including the ability to integrate these data with modelling and forecasting workflows leveraging specialised compute resources. Providers of the Copernicus Data Processing Platform already count with an extensive collection of Copernicus datasets, managed according to the FAIR principles, and may be further extended with new datasets requested by users of the platform.

SCIENTIFIC CATEGORISATION

- Natural Sciences
- Natural Sciences
- Engineering & Technology
- Engineering & Technology
- Agricultural Sciences

CATEGORISATION

- Compute
- Compute

Joining as a provider



- Well-defined procedures to join
 - Technical integration – configure your system to allow federated identity
 - Non-technical integration: contacts, AUP, Privacy Policies...
- We will support you through the whole process
- Does it sound too complex? Feedback is welcomed!

Next steps



- openEO improvements
 - HPC / Cloud support
 - Data management in the C-SCALE federation
- C-SCALE Software collection
 - Rely on AppDB as a repository to publish software (containers)
- PaaS Orchestrator improvements
 - Support for new providers and new platforms
 - Integration with data discovery (see details on discovery next)
- EOSC integration
 - Get FedEarthData on EOSC portal
 - Accounting/Monitoring improvements



Thank you for your attention.