C-SCALE
Copernicus – eoSC AnaLytics Engine

FedEarthData
Federated Earth System Simulation and Data Processing Platform

Zdeněk Šustr, CESNET
sustr4@cesnet.cz,
Enol Fernández, EGI
enol.fernandez@egi.eu

High Level C-SCALE Motivation

- Adequate IT infrastructure for
    - **discovery**,
    - **access**,
    - **processing** of high-resolution EO data from Copernicus and related initiatives.
- Avoid lock-in and and use distributed infrastructure
    - $\implies$ **federation**
- Flexible approach to computing
    - $\rightarrow$ support cloud, HTC/HPC, PaaS & interactive analysis
- Free at point of use
    - $\rightarrow$ virtual access

Federation

C-SCALE

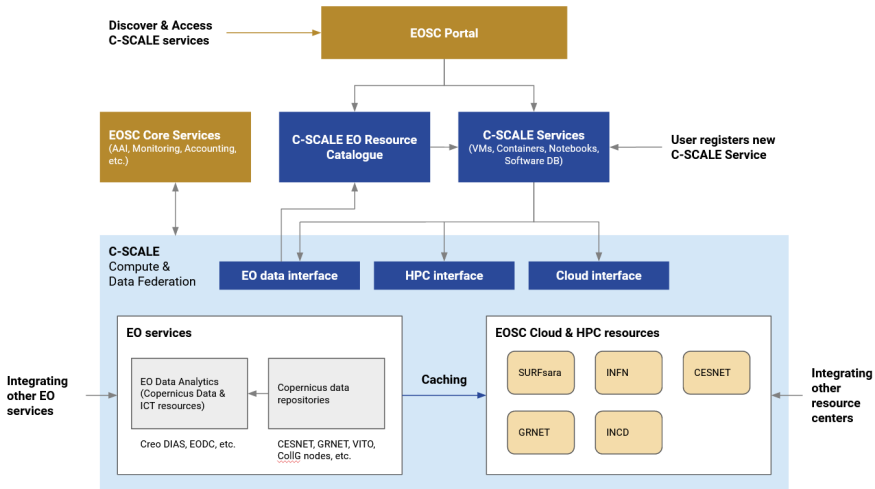## Federation = **Agreement**

$\implies$ Finding common ground
- Standards / technologies
  - ▶ OIDC, STAC, TOSCA, HTTPS, . . .
- Interfaces
  - ▶ OpenEO, STAC-API, SSH, . . .
- Naming and terminology
  - ▶ Product? Collection? Access?

The FedEarthData service

- Distributed infrastructure of data and compute providers
  - execute Earth system sim. 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
  - ESA Collaborative Ground Segment
 - DIAS
  - Area-specific resources

C-SCALE

Discover & Access C-SCALE services → **EOSC Portal**

EOSC Core Services (AAI, Monitoring, Accounting, etc.)

C-SCALE EO Resource Catalogue

C-SCALE Services (VMs, Containers, Notebooks, Software DB)

User registers new C-SCALE Service

C-SCALE Compute & Data Federation

EO data interface   **HPC interface**   **Cloud interface**

EO services

EO Data Analytics (Copernicus Data & ICT resources)

Copernicus data repositories

Creo DIAS, EODC, etc.

CESNET, GRNET, VITO, CollG nodes, etc.

Caching

Integrating other EO services

EOSC Cloud & HPC resources

SURFsara   INFN   CESNET

GRNET   INCD

Integrating other resource centers

FedEarthData: Map I

C-SCALE
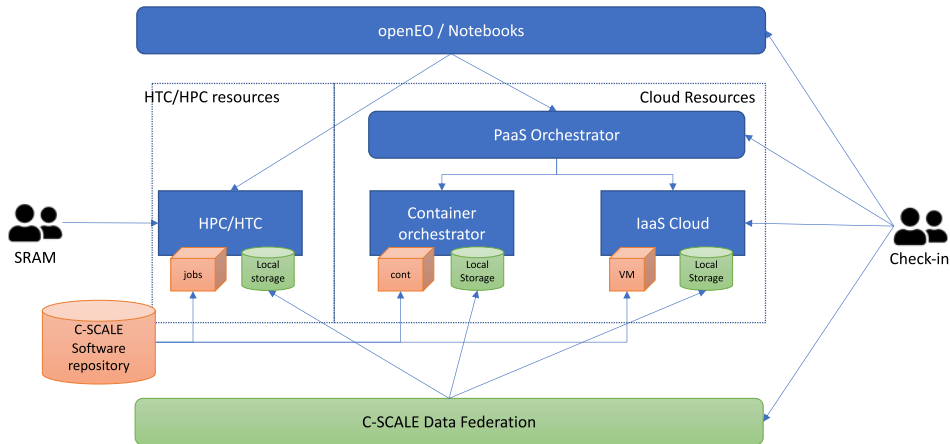
HPC/HTC
- EODC/TUW (AT)
- GRNET (GR)
- SURF (NL)

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

Data
- CESNET (CZ)
- CloudFerro (PL)
- EODC (AT)
- GRNET (GR)
- VITO (BE)
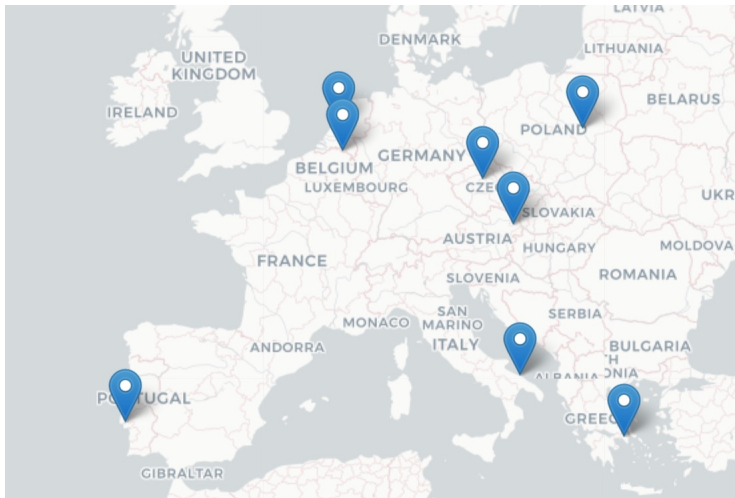
FedEarthData: Map II

Accessing the federation: AAI

- AAI – federated authentication to distributed providers
  - Build on existing work around EOSC AAI and AARC
  - Use mature products: Check-in, Perun & SRAM
  - Proven technology: OpenID Connect and LDAP
- Single sign-on – move **seamlessly** among resources
- 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**
  - Browser accessible data analytics platform
  - Python, R, Julia, 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
  - EO-specific notebook in development
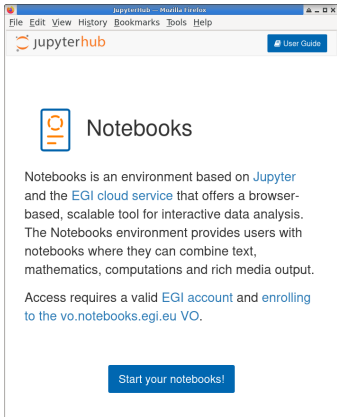
OpenEO

- C-SCALE **batch processing environment**
  - Intuitive API to process a variety of EO datasets
  - Data access and processing on multiple infrastructures based on datacubes
- Using openEO on C-SCALE:
  - openEO platform – `https://openeo.cloud/`, coming soon to EOSC marketplace
  - Managed
    - ▶ onboarding in progress
  - Self-service
    - ▶ multiple providers
    - ▶ IM / 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-stablished 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**)
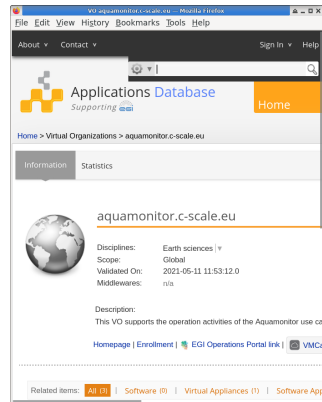
Data Federation

- Technologically compatible with the Cloud Federation
- **Seamlessly** accessible from user environments
    - Notebooks
    - OpenEO
    - VMs deployed in the Compute federation
    - HTC/HPC workflows (with an extra step)
- **Reach** for data from anywhere within compute (C-SCALE, EGI …)
    1. Use MQS to locate data if you don't know where they are
    2. Use your regular identity to access the data
        - Find your token already injected into your VM, notobook, …

Check-in

- AAI and user management for C-SCALE cloud & data federation
- SAML 2.0 / OIDC 1.0 / OAuth 2.0 / LDAP
- Interoperable:
  - AARC and EOSC AAI compliant
  - Legacy X.509
- Community management
  - Perun, Comanage
    - ▶ Specific attributes on enrollment
  - Other Community AAIs pluggable

AppDB
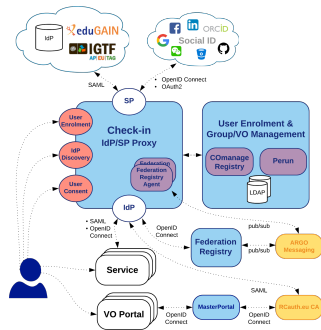
- Catalogue of VM images for cloud providers
  - Community curated
  - Automatic synchronization with providers
  - Non-VM software also supported
- Provider discovery

Federation Features

C-SCALE

Monitoring

- Availability/Reliability metrics for providers


Accounting

- Usage information breakdown

HPC in C-SCALE – SRAM

C-SCALE

- Batch workflows, scaling out local workflows
- Managed batch environment
- Federation is the greatest challenge, similar to Cloud & Data
  - ⟹ Federated HPC prototype for EO
    - SURF Research Access Mgmt.
    - AARC BPA compliant AAI
    - Supports LDAP, OIDC, SAML
    - Manages user attributes
    - IdP validated users
    - CO connects to service(s)
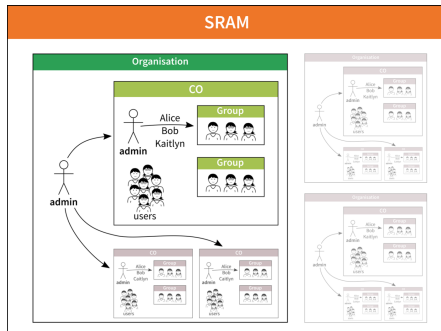
Joining as a Provider

Any branch of the federation

- Well-defined procedures to join
  - Technical integration
    - ▶ configure your system to allow federated identity
    - ▶ register in catalogues
  - Non-technical integration: contacts, AUP, Privacy Policies…
- Support **is** available

Next Steps

C-SCALE

- OpenEO improvements
  - HPC/Cloud support
  - Data management in the C-SCALE federation
- C-SCALE Software collection
  - SW, container images (HPC, Notebooks)
  - Rely on AppDB as a repository to publish software (containers)
- Orchestration improvements
  - Support for new providers and new platforms
  - Integration with data discovery (follow data with compute)
- EOSC portal registration
- Integration of HPC and Cloud access

FedEarthData on EOSC

Currently onboarding

- Preliminary entry already registered
- Rely on EOSC helpdesk
- Finalising sustainability details with providers

Federated Earth System Simulation and Data Processing Platform

Easy processing of Copernicus data

Provided by: EGI Foundation

ACCESS THE RESOURCE

OPEN ACCESS

→ 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

Thank you

Questions any time

Zdeněk Šustr, CESNET
sustr4@cesnet.cz,
Enol Fernández, EGI
enol.fernandez@egi.eu

C-SCALE | 14 November 2022