

C-SCALE

Copernicus – eoSC AnaLytics Engine

Chasing EO Data around Europe

Zdeněk Šustr, CESNET
sustr4@cesnet.cz

EGI Conference | 20. October 2021



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

C-SCALE Data Federation: Motivation

Before C-SCALE (as seen from an analytics platform in EOSC)

- No single provider has all data
 - Different agencies, different (satellite) platforms
 - Different geographical areas of interest
 - Different retention policies
- Personal experience or Google search required to **find data**
- Separate registrations necessary to **access data**
- Different mechanisms used to **stage data**

We won't be able to fix all in **C-SCALE** but we will try to help

Objectives I

Problem: **No single provider has all data**

👉 This will remain but we will try to improve things.

- Providers catalogue (**EGI** GOCDDB)
 - Where they are
 - How to reach them
- Describe their data collections & policies (ISO 19115?)
 - Custom fields in GOCDDB
 - ▶ Type of data
 - ▶ Geographic regions
 - ▶ Retention policy

Result: There is still no single warehouse of all data, but at least users (**and robots!**) know where it is worth looking.

Objectives II

Problem: **Personal experience or “Google” search required**

- Introduce a common query interface
 - **MQS**: MetaData Query Service
- Select a common query language/protocol
 - STAC-API
- GOCDDB (previous slide) already tells us who keeps what
 - ⇒ Re-format and **pass** the query to sites who *might* have answers

Result: MQS makes your search across the federation for you.
... and even aggregates the results.

Objectives III

Problem: **Separate registration required at data sites**

- Implement support for EOSC AAI (OIDC everywhere)
 - Contribute to existing services
 - ▶ ESA/SERCO DHuS
 - ▶ CreoDIAS
 - Simple accessor for products stored on disk (e.g., EODC)
 - ▶ Also reusable for use cases building their own caches
- Interactive access as well as **delegation** for PaaS/HTC/HPC
 - The job only needs to carry a token
- Authorization will remain site-specific (default: none)

Result: Federated or at least federation-negotiated access everywhere

Objectives IV

Problem: **Different mechanisms used to stage data**

- All EO products downloadable over HTTPs
 - Fine and understandable for end users
 - Also fine for robots (`curl`, `wget`, `rclone`)
- Always think about dual access
 - Manual download by end users
 - Batch access by machines in the PaaS/HTC/HPC layer
- Consider also product upload/import
 - Bilateral agreement between producer and receiving data site
 - ▶ Potential topic for G-SCALE?

Result: Not only users but **even machines** know how to stage data

The World after C-SCALE

- Can you do it manually?
⇒ You can also delegate it on a platform in EOSC.
- Accessing EO data is a two-step process:
 1. Search for data
 - ▶ Single point for queries to be passed on to relevant sites
 - ▶ Matching products returned
 2. Stage data
 - ▶ EOSC federated identity accepted
- Redundancy is avoided wherever possible
 - Only sites and their properties are catalogued
 - Data and metadata remain where they were

Where are we now

Partners: CESNET, EODC, GRNET, VITO, CloudFerro

→ Representing ESA CollGS, DIASes & institutional sources

Architecture done, starting development

- **Out of Sync** with use cases 😞
 - They could use now what we are only starting to develop
 - ▶ At least it confirms our intent

Culture shock

- **Un**teach users to pre-download whole countries
 - The network is faster than your algorithm, surely
- **Not all** providers consider bandwidth infinite
- It will pass ...

Thank you for your attention!

Want more detail? Read *C-SCALE Copernicus Data Access and Querying Design*.



Zdeněk Šustr, CESNET
sustr4@cesnet.cz