C-SCALE
Copernicus - eoSC AnaLytics Engine

User-driven co-design of a state-of-the-art digital infrastructure for Copernicus Big Data Analytics:

Opportunities to get involved

Björn Backeberg, Deltares

bjorn.backeberg@deltares.nl

SURF Seminar | Enabling Copernicus Big Data Analytics through European Open Science Cloud | Oct 2021

Objectives

- C-SCALE plans to deliver
    - A **federated compute and data infrastructure** offering Copernicus/EO data
    - A **seamless user experience** where the complexity of Copernicus data, compute and storage resource provisioning and orchestration is hidden from the end-user
    - Access to **optimized** low level **data** and higher level **analysis ready data**
    - **On-demand solutions** to generate analysis ready data where these are not readily available

- The **research communities**, through use cases, will **co-design, test, pilot, refine** and ultimately **help create** a federated infrastructure that delivers data and platform services that are **useful for the community**

Federation principles

- Services accessible through **homogeneous and standard interfaces**
- Ensure **FAIR** EO data across the providers
- Make it **easy** for providers to **join the federation**
- Follow **EOSC policies** and operational and technical **requirements**
- Leverage basic and operational features (AAI, accounting, etc.) available through **EOSC core services**
- Maximise **interoperability** with other EOSC services
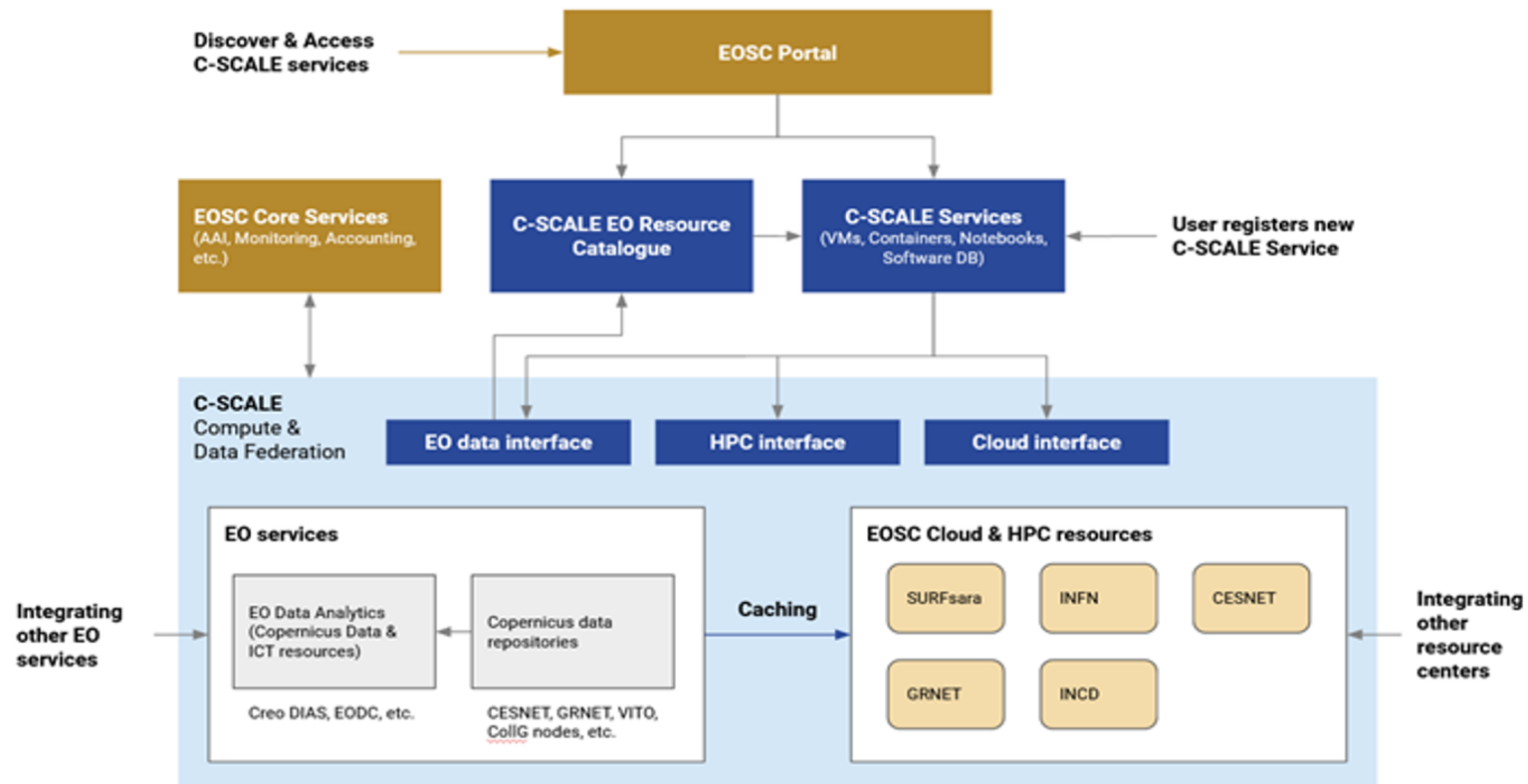
What does that mean for the end-users?

- Enable **users to quickly and easily generate meaningful results**

- Avoid that users have to deal with the technical infrastructure details to get data processing and analytics pipelines to work.

- **Complexity** of using Copernicus, compute and storage resource provisioning and orchestration is **abstracted away from the end-users**

- **Homogenous access to resources**

Co-designing the C-SCALE federation

- C-SCALE Federation design is **driven by user requirements** to guarantee the delivery of an environment that **satisfies user needs**

- User requirement are derived from deploying mature applications (**use cases**) on the C-SCALE federated infrastructure to **test its usability and functional design.**

- Provide **feedback to the infrastructure** providers on:
  - ease of use of the resources,
  - effectiveness of support,
  - appropriateness of the technology,
  - speed of access to resources and data,
  - resultant usability of the application running on the federated infrastructure
  - missing functionality/resources,
  - satisfaction of the service/resource

Get involved!
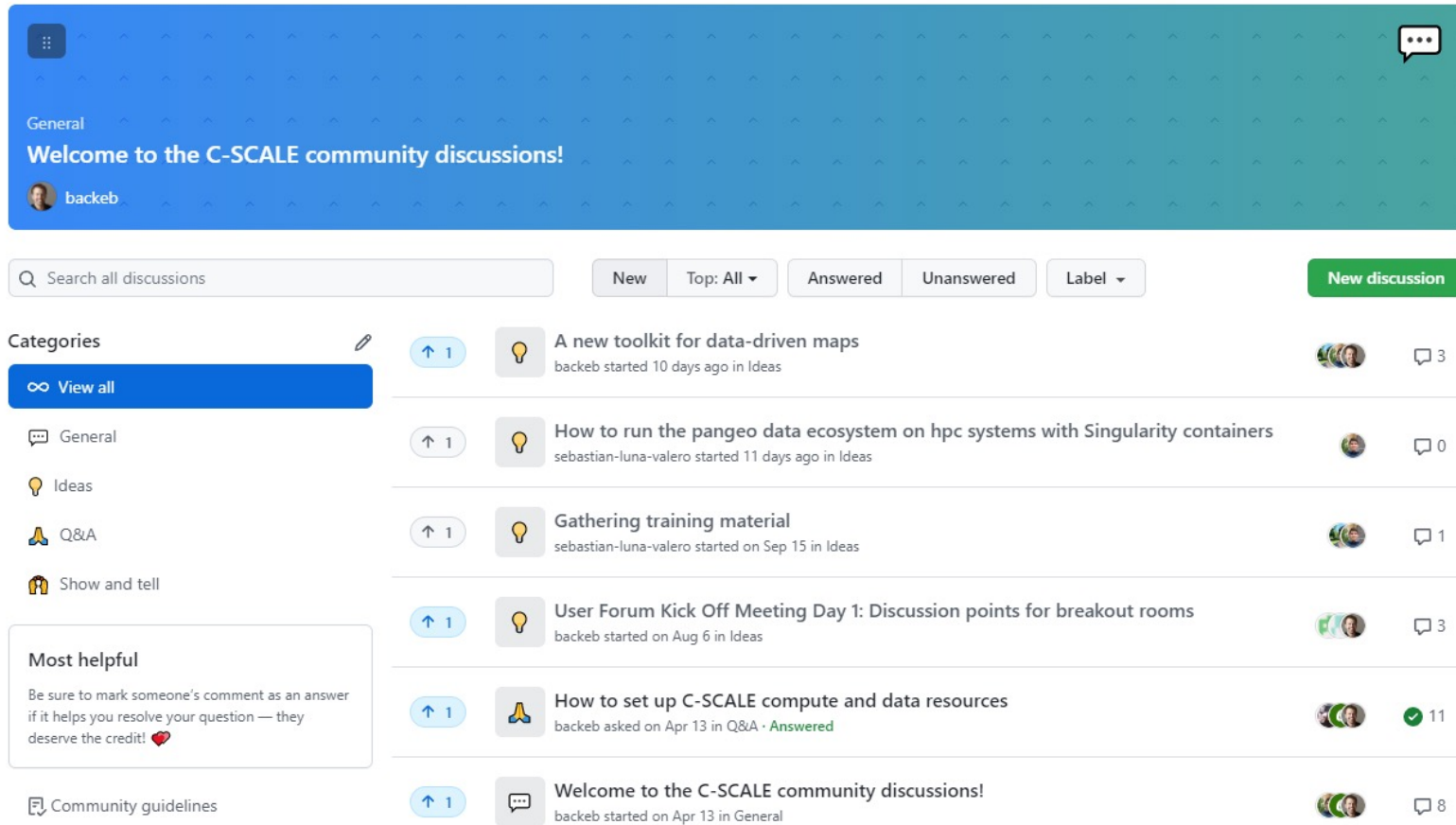
Help us design and create a compute and data infrastructure with platform services that are useful for the community

How?

Join the discussions!

https://github.com/c-scale-community/discussions/discussions

How?

Submit a use case

https://c-scale.eu/call-for-use-cases/
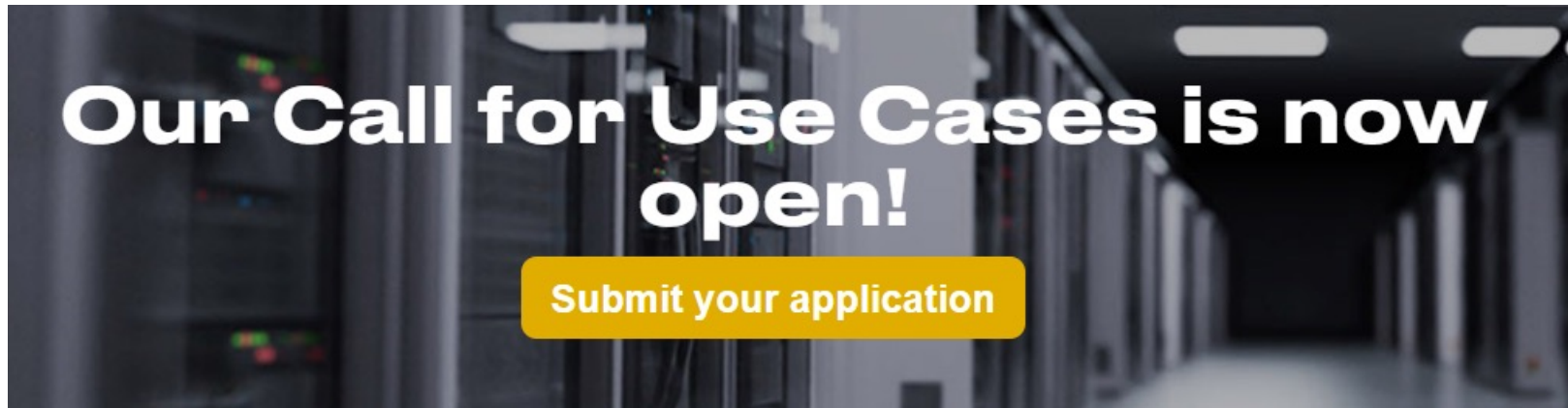
What's in it for you?

- Receive access to **compute (including GPU) and storage resources** serving Copernicus data

- Benefit from **new technologies and platform services**

- Increase your application's TRL by **collaborating with infrastructure providers**

- Make applications more **cloud agnostic** and **reduce vendor lock-in**

- **Scale applications** to achieve planetary-scale analytics

- Leverage **cross-disciplinary services and solutions** from the EOSC landscape

- Receive **support** from infrastructure providers to deploy the application

- Register application/services in the EOSC Portal to **engage a broader group of stakeholders**

What's in it for you?

Did I mention…

(… sponsored by the project)

The offering

- Access to **compute and storage resources** hosting Copernicus data

  - 12 PB months of storage
  - 18 million Cloud CPU hours
  - 3.1 million HPC/HTC CPU hours
  - 6,000 GPU hours.

  Total capacity available from the C-SCALE project between 2021-2023

- Provided by eodc cesnet INFN SURF vito grnet CloudFerro INCD

- **Federation services** to manage user access, to monitor service availability-reliability, to measure compute and storage usage.

The offering

Access to…

- **Scientific tools, applications and virtual machine images** that can be used and customised on the infrastructure resources for your analysis

- **Platform services** to
  - manage complex workloads,
  - train AI models,
  - deploy automated clusters,
  - run containerised applications,
  - perform interactive analysis with Notebooks.

- **Data management services** to transfer, to catalogue, to cache and, federate data across multiple compute and data sites.

- **Network of experts** to help port the application to the infrastructure
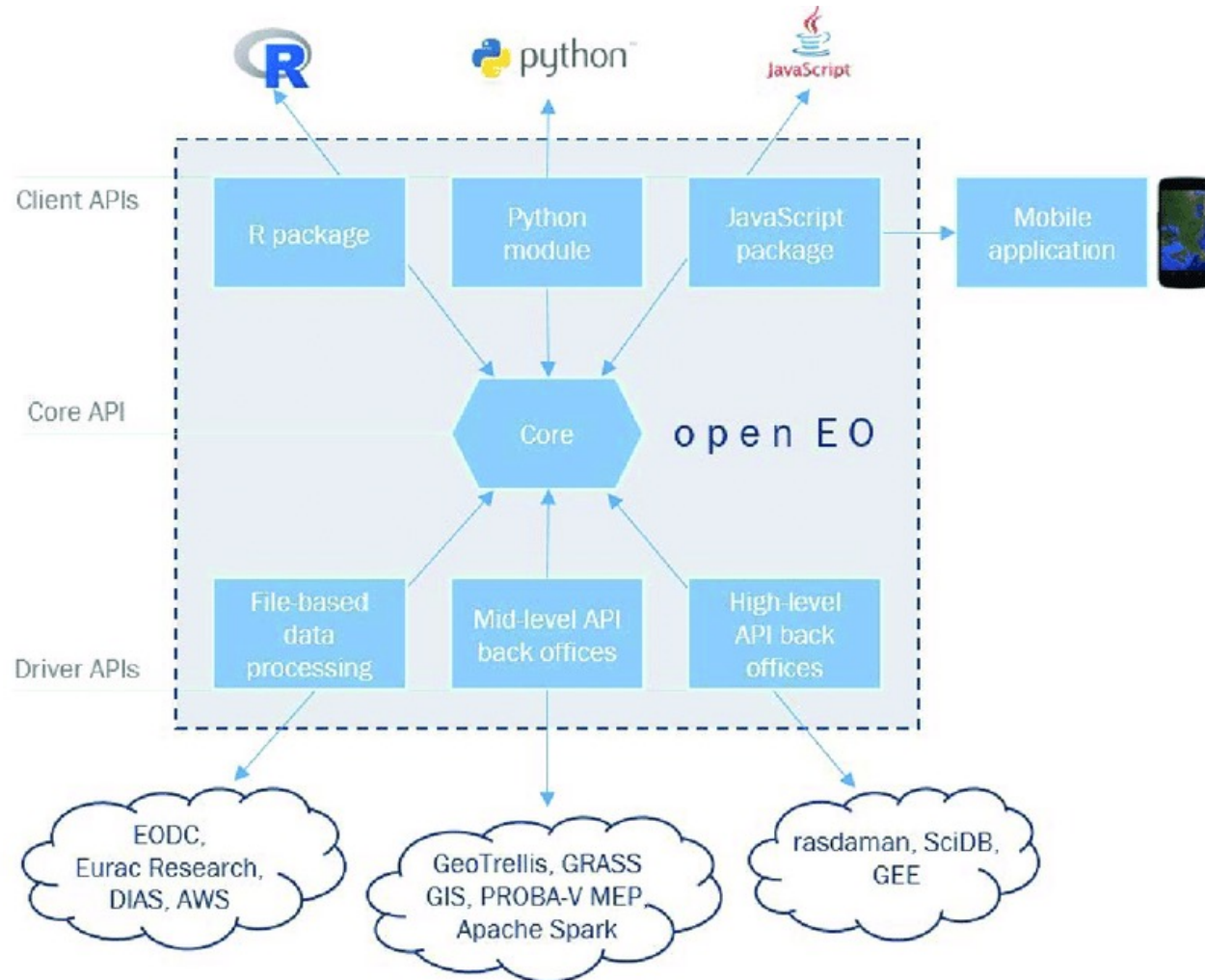
Aquamonitor example

① Compute Surface Water Changes (Algorithms)

② Satellite > Composite Image Generation

③ Render Map Tiles

④ Front-end (Users)

https://aqua-monitor.appspot.com/

https://github.com/Deltares/aqua-monitor
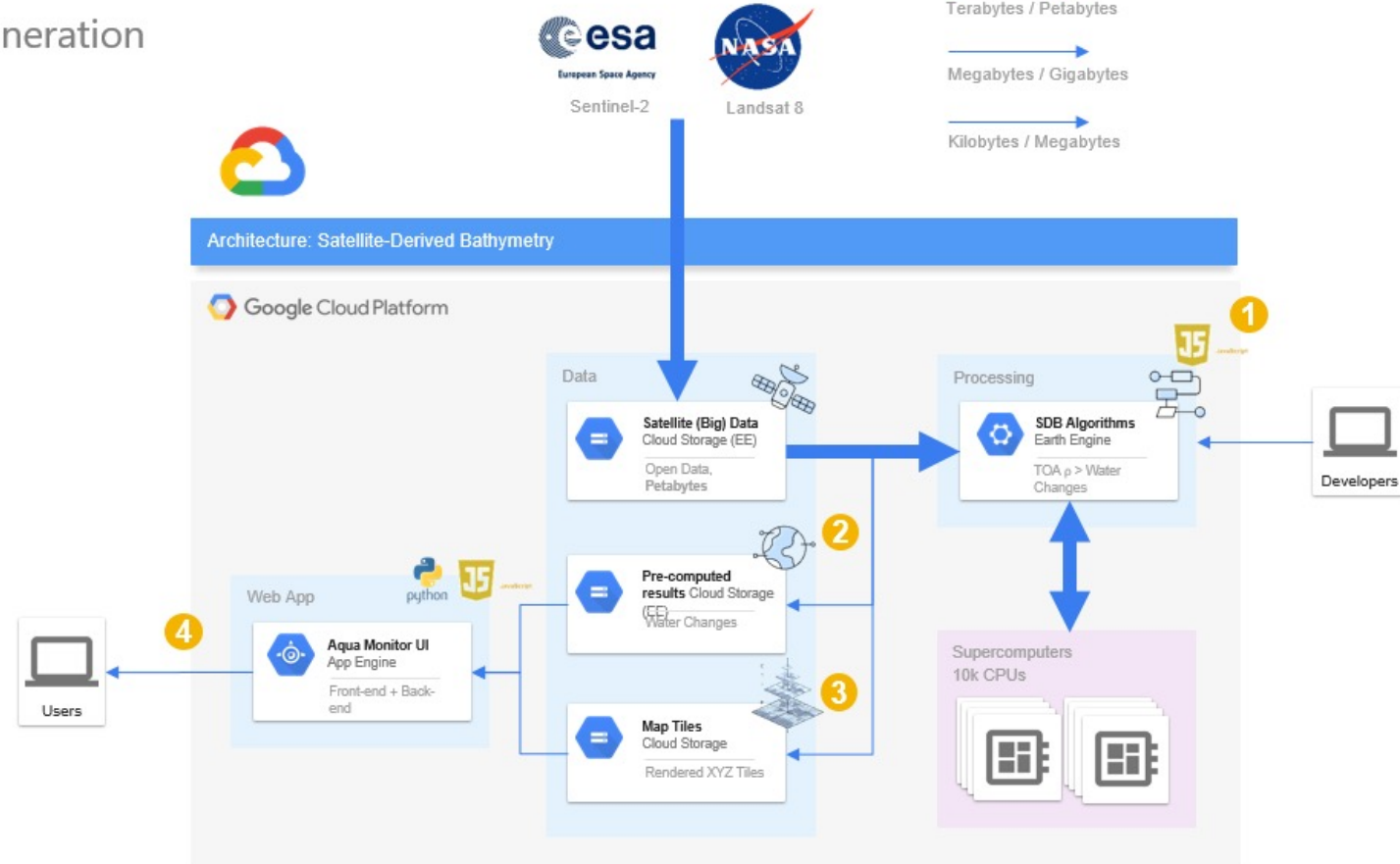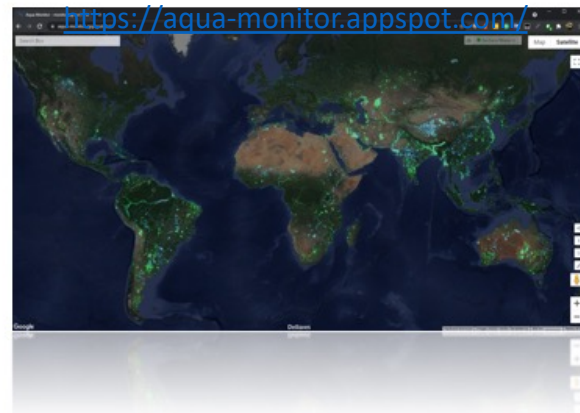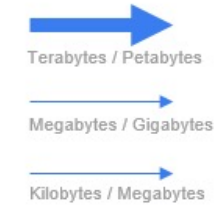
# Porting from GEE to C-SCALE: OpenEO Platform

- OpenEO to get an open-sourced "Big EO Data" platform to support Aquamonitor requirements

- Ideal for the "User" perspective:

  - No worries about infrastructure provisioning

  - One vision on definition of EO Data

  - Minimal parallelization effort

  - Easy to communicate to- and collaborate with infrastructure providers
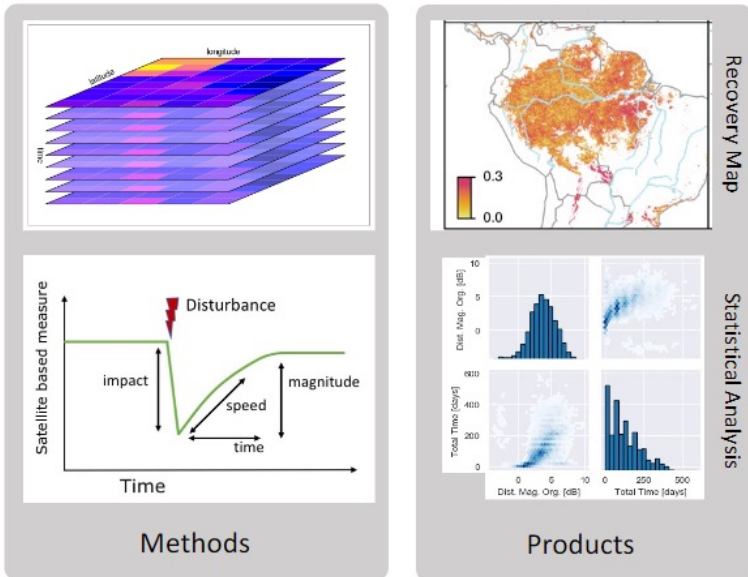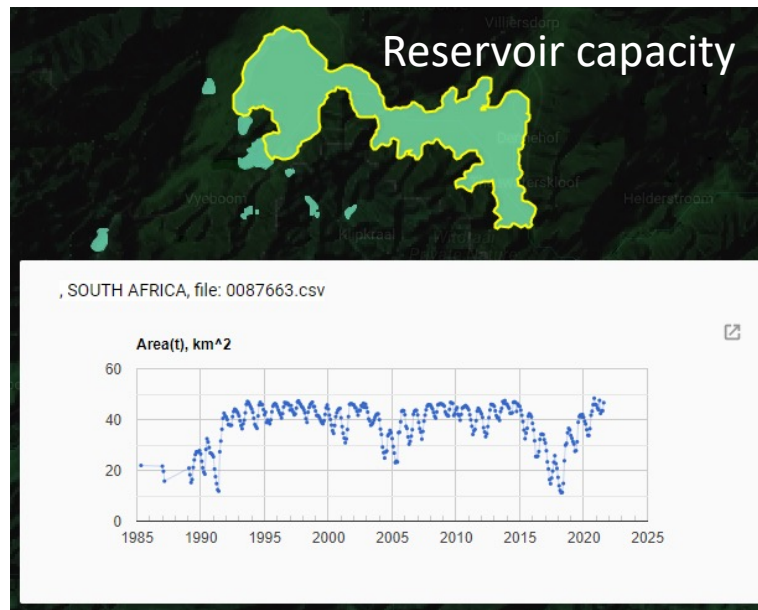
Use cases

Drought forecasts
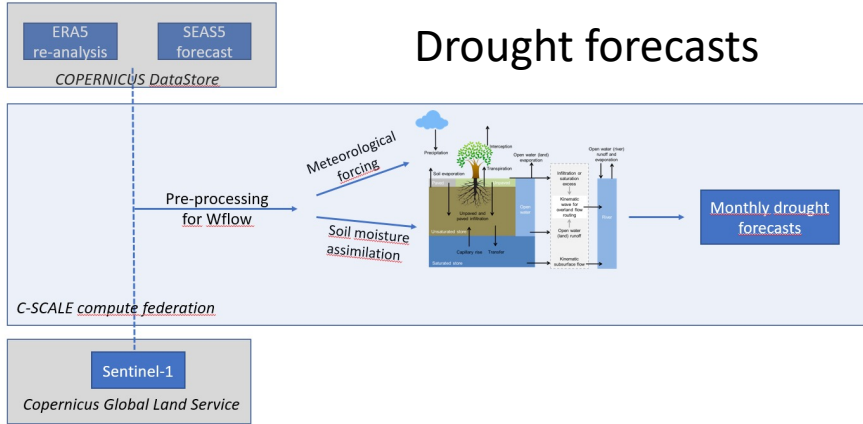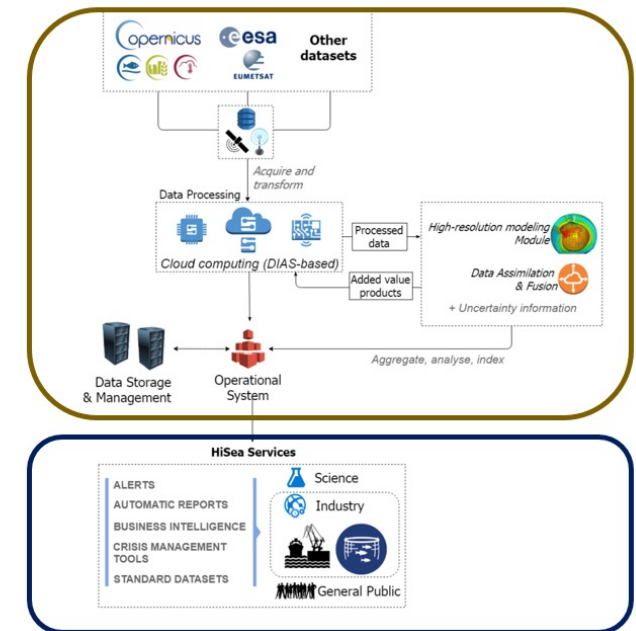
Land-surface changes
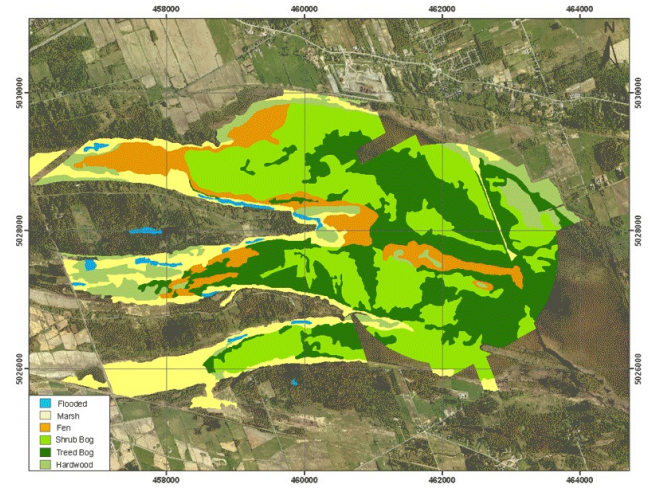
Wetland water stress classification

Forest recovery capacity

Reservoir capacity

, SOUTH AFRICA, file: 0087663.csv

Ports and Aquaculture early warning

Community-populated PaaS layer

In deploying the use cases on the infrastructure in collaboration with the infrastructure providers we will **create deployment recipes and (TOSCA) templates** for easier reuse later, thus **populating the PaaS layer**.

Thank you for your attention.

Björn Backeberg, Deltares

bjorn.backeberg@deltares.nl

SURF Seminar | Enabling Copernicus Big Data Analytics through European Open Science Cloud | Oct 2021