

**Authors:** Vernet Marine<sup>1</sup>, Harscoat Valérie<sup>1</sup>, Queric Antoine<sup>1</sup>, Maudire Gilbert<sup>1</sup>, Antonio Fabrizio<sup>2</sup>, Kulüke Marco<sup>3</sup>, Kindermann Stephan<sup>3</sup>, Sudre Joël<sup>4</sup>, Pierkot Christelle<sup>4</sup>, Jourdain Cédric<sup>5</sup>, Donvito Giacinto<sup>6</sup>.

1 – IFREMER, 2 – CMCC Foundation, 3 – DKRZ Deutschen Klimarechenzentrum, 4 – CNRS Centre national de la recherche scientifique, 5 – CINES, 6 – INFN Istituto Nazionale di Fisica Nucleare

## Challenges

Studying the Earth system requires heterogeneous, voluminous data of different types, formats and sources, stored in distributed domain-dependent repositories. In this context, how to:

- ★ Facilitate the discovery of datasets of interest, regardless of domains and types?
- ★ Facilitate and speed up remote access to heterogeneous distributed data?
- ★ Facilitate on-demand cross-domain or cross-sources analysis of Earth observation and environmental Big data?

## On-demand data analysis & processing

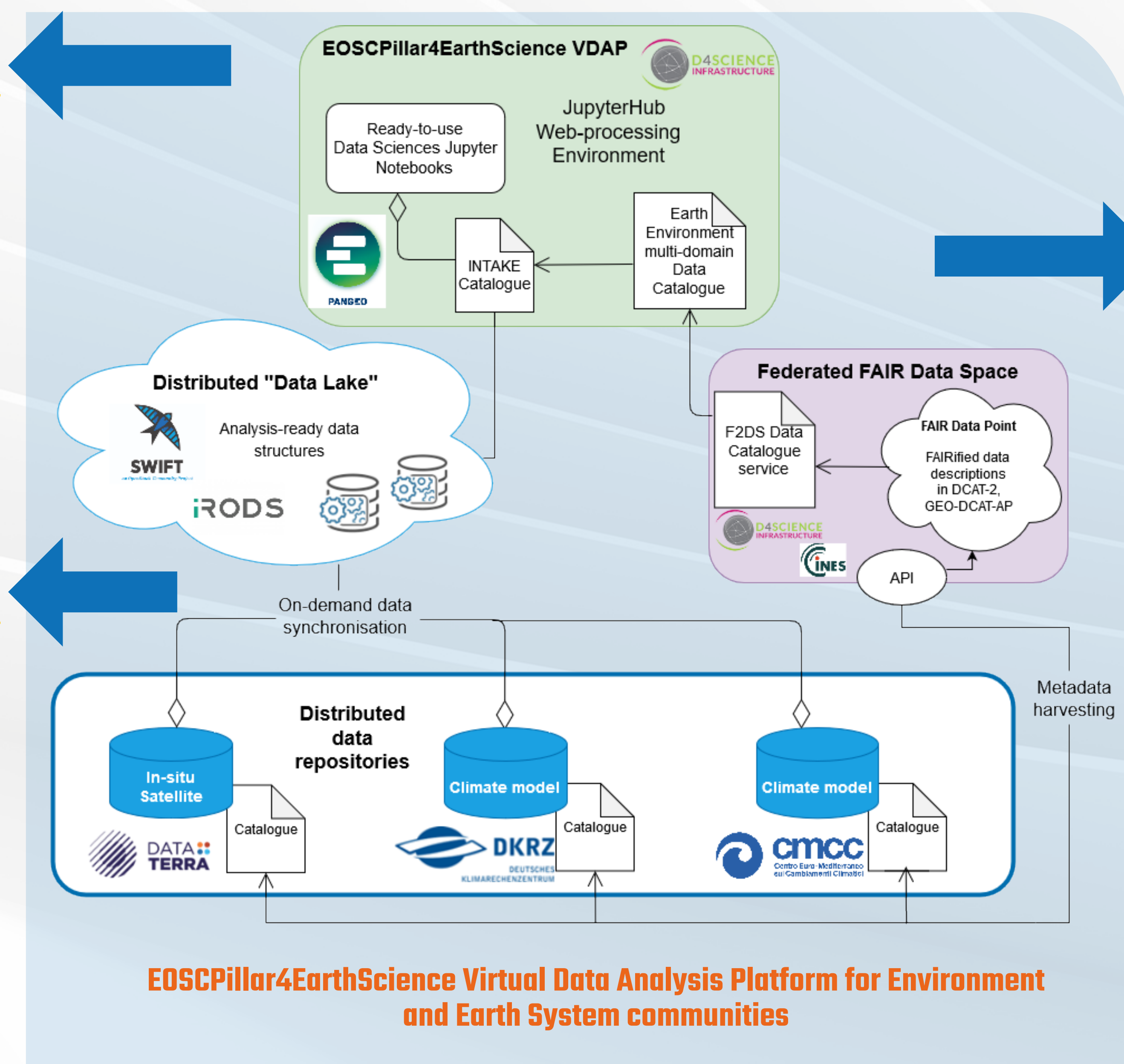
- ★ Through Data Sciences Notebooks powered by PANGEO software stack
- ★ On remote IT facilities

## Remote data access & conversion

- ★ On-demand, from the domain-dependent data repositories to the VDAP
- ★ Through Swift and iRODS services

## EOSC-Pillar services

- ★ D4Science VRE service and associated IT resources (storage, CPU, RAM)
- ★ FFDS FAIR API and multi-domain FAIR data catalogue
- ★ Data services from UC's partners: Swift, iRODS, Ophidia



## Data Discovery

- ★ Through one cross-domain data catalogue : FFDS
- ★ Connected with Earth domain-specific catalogues
- ★ Common metadata model with geospatial information

## Results

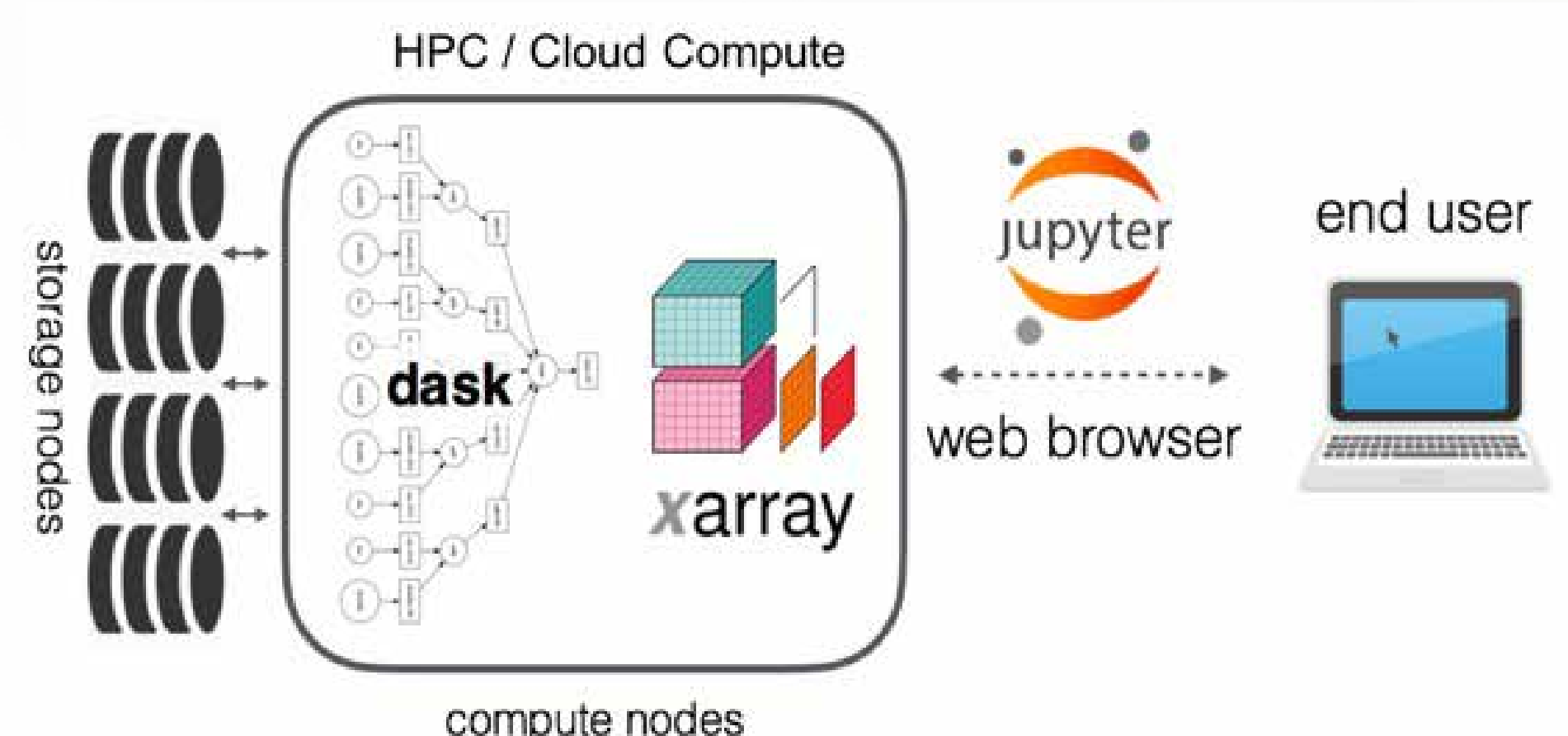
- ★ Easier data discovery through connection with FFDS
- ★ Data analysis and processing Notebooks powered by PANGEO ecosystem and local services, made available for all users of the VDAP
- ★ Transparent and fast data access for the user through Intake catalogue and Swift data service. Scalability of iRODS service to be improved to match infrastructure security measures
- ★ Faster, easier data processing through data conversion into analysis-ready format (Zarr, Parquet)
- ★ VDAP infrastructure to improve to match user needs regarding parallel computing

## For more details

- ★ EOSC-Pillar UC2 webpage : <https://www.eosc-pillar.eu/use-cases/agile-fair-data-environment-and-earth-system-communities>
- ★ EOSC-PILLAR4EarthScience VDAP demonstrator: <https://eosc-pillar.d4science.org/group/eoscpillar4earthscience>
- ★ EOSCPILLAR4EarthScience Quick Start guide: <https://doi.org/10.5281/zenodo.7128695>

## What is PANGEO ?

A community platform for Big Data Geosciences  
Offering relevant open-source scientific Python technologies  
<https://pangeo.io>



## Participating Institutions

