



Forum **TERATEC** **23**

Unlock the future

31 MAI & 1^{er} JUIN 2023 • Au Parc Floral, Paris

Un événement organisé par

 **infoprodigital**





Building the European Digital Twin Ocean

Quentin Gaudel – Mercator Ocean international
quentin.gaudel@mercator-ocean.eu

A blurred background image of a modern office or public space with people walking. Overlaid on this are two white rectangular boxes containing the text "Unlock the future" in blue.

**Unlock
the future**

The logo for Forum TERATEC 23 features a stylized circular icon on the left, composed of blue and yellow segments. To its right, the word "Forum" is in a blue sans-serif font, "TERATEC" is in a larger, bold blue sans-serif font, and "23" is in a large, bold yellow sans-serif font.

Forum TERATEC 23

Summary

1. Why a European Digital Twin of the Ocean?
2. What is the Public European Digital Twin Ocean?
3. Technical insight of the functionalities and architecture of **EDITO-Infra**



Why a European Digital Twin of the Ocean?

A core infrastructure, conceived as a public good and service, to support the implementation of the marine Green Deal objectives

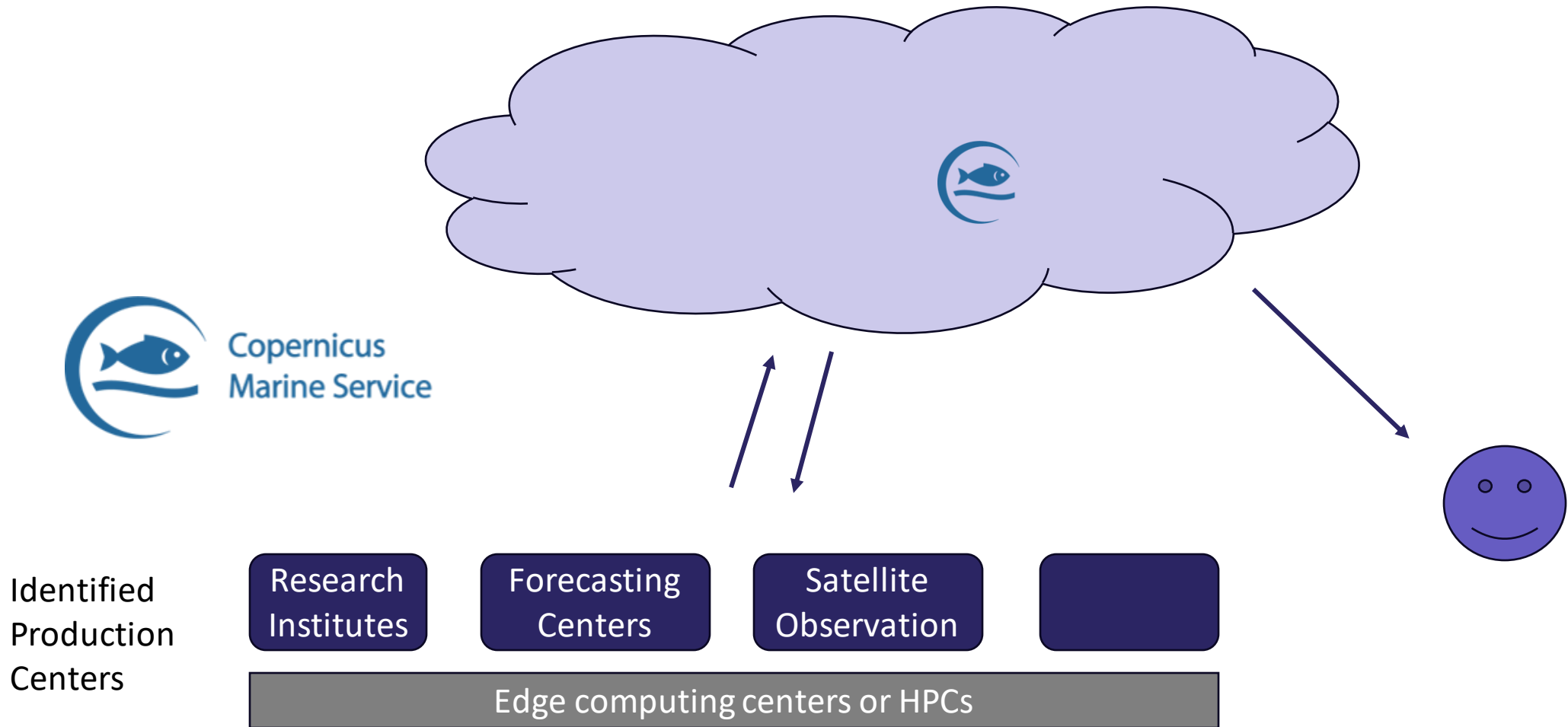
At European level: hundreds of initiatives, Regional, National and European marine projects and infrastructures, Private initiatives, Monitoring services, Citizen science campaigns.... are committed to study and protect the marine ecosystem.



The complexity is to help them to work together and with other international initiatives.

Need for user-driven powerful tools, fit for digital age, to strengthen ocean knowledge and ocean management: by integrating and connecting wide range of data and models (from physics to socio-economics) with cloud infrastructures, HPC, AI and services.

An evolution of Marine Data Portals



Facilitate the scientist work and reduce costs

1. **Limit data transfer** as datasets become bigger and bigger with **data resolution**
2. **Ease HPC access** as computer protocols and networks is not earth science

→ Enable **effective data subsets** and **near-data computing**

Facilitate the contribute to the ocean science

1. **Improve FAIRness (Findable, Accessible, Interoperable, Reusable)** of data and processes
2. **Easily involve more parties** instead of identified institutes

→ Offer **efficient tools** in an **open and collaborative platform**

ENTER THE **EUROPEAN DIGITAL TWIN OF THE OCEAN**

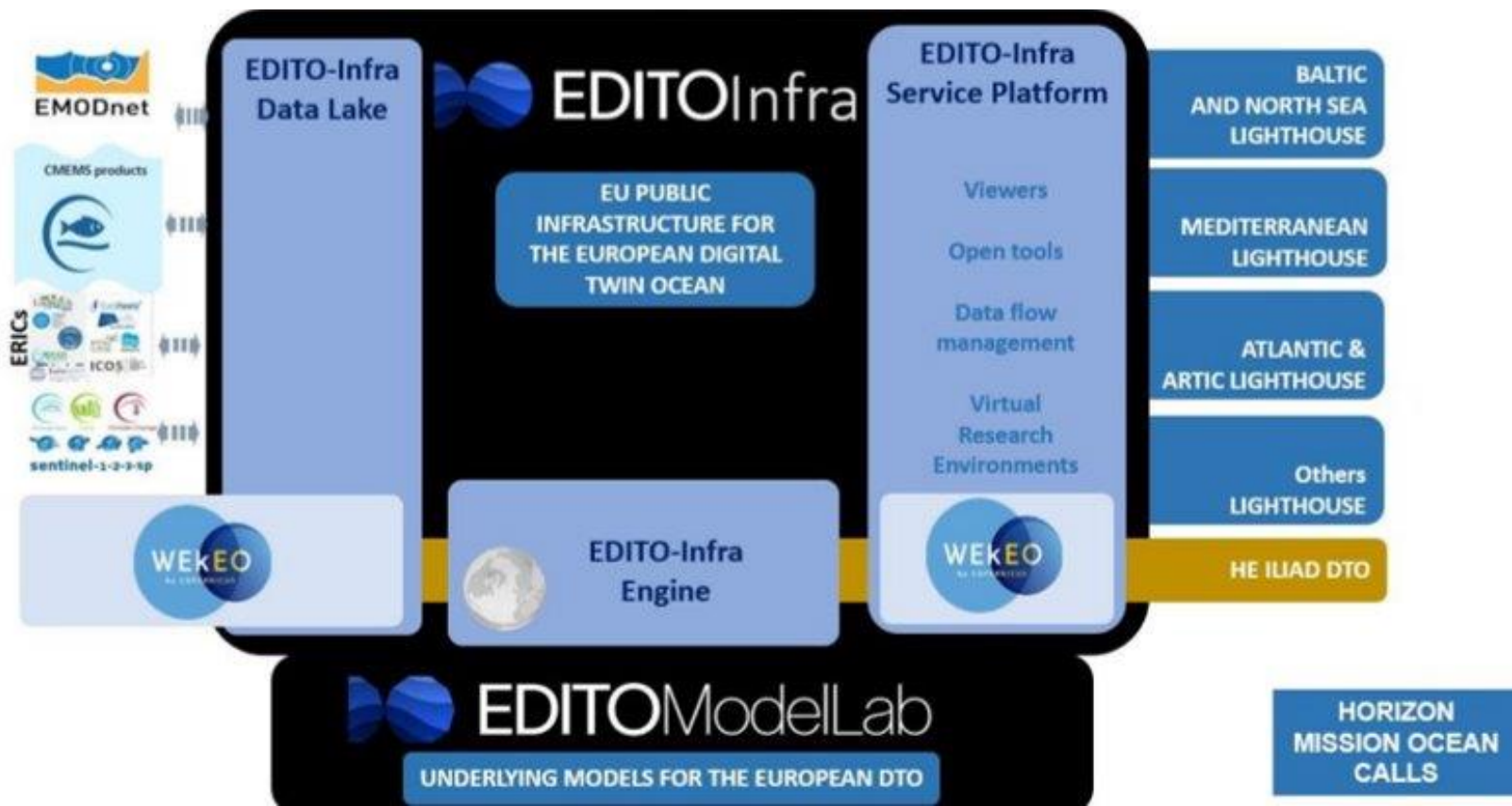


What is the Public European Digital Twin Ocean?



EDITO-Infra: 2-years grant, ~3 M€

EDITO-Modellab: 3-years project: ~7 M€



What is the Public European Digital Twin Ocean?

Integrate data from other European platforms, such as **Copernicus Marine Service** and **EMODnet**.

An exhaustive **reference** to all **ocean data products** available

STACAPI supporting **metadata** based and **semantic search** queries

EDITO-Infra is a **European** platform that provides a **single point of entry** to **search**, **explore** and **exploit products** and knowledge to support **digital twins of the ocean**.

User can add new data to the platform either directly (upload) or by deriving existing data to create new data using computational functions.

All processes provided by the platform are **exposed** through **OGC API processes** standard for a **seamless third-party integration**



Explore	Contribute	As-a-service
<ul style="list-style-type: none"> • Browse and access ocean data (CMEMS, EMODnet, etc.) seamlessly with adaptative resolution • Create your own ocean data catalog • Extract data demand or launch pre-defined ocean analysis and model forecasts on the browsing data • Launch pre-defined ocean what-if-scenarios 	<p>Explore, and:</p> <ul style="list-style-type: none"> • Upload your data • Launch virtual near-data R&D environments (Jupyter, RStudio, IDE, Python, R, etc.) or other pre-defined services • Exploit cloud computing clusters (Dask, Spark) • Define new remote processes to extract, run analysis and forecast and launch what-if-scenarios • Launch and configure automation systems to industrialize your data pipeline • Integrate your contributions (data, environments, processes, pipelines, etc.) with HPC • Share your contributions with your communities or to all 	<p>Contribute, and:</p> <ul style="list-style-type: none"> • Build new web applications that implement or integrate your services • Integrate your applications with HPC • Share your services with your communities or to all • Host your business services

EDITO-Infra platform added values

Digital Ocean-oriented specialized services
Core domains

Ocean catalog (CF convention, etc.)
Ocean data (ARCO format)
Ocean experience (UI/UX, viewer)
Smart-request (data orchestration)

Automatic data cataloguing
Automatic data check

Standard services
Supporting domains

STAC API metadata catalog
S3 data storage
Processes API registry
Distributed computing clusters (CPU/GPU)
Virtual R&D Environments deployment
HPC/AI workflow managers deployment
Container deployment

Infrastructure/Cloud Provider



Standard (OGC) compliant

Open Collaborative platform
Resources controlled via Identity and Access Management IAM



PANGEO-inspired

EDITO-Infra platform added values



Digital Ocean-oriented specialized services
Core domains

Ocean catalog (CF convention, etc.)
Ocean data (ARCO format)
Ocean experience (UI/UX, viewer)
Smart-request (data orchestration)

Automatic data cataloguing
Automatic data check

Standard services
Supporting domains

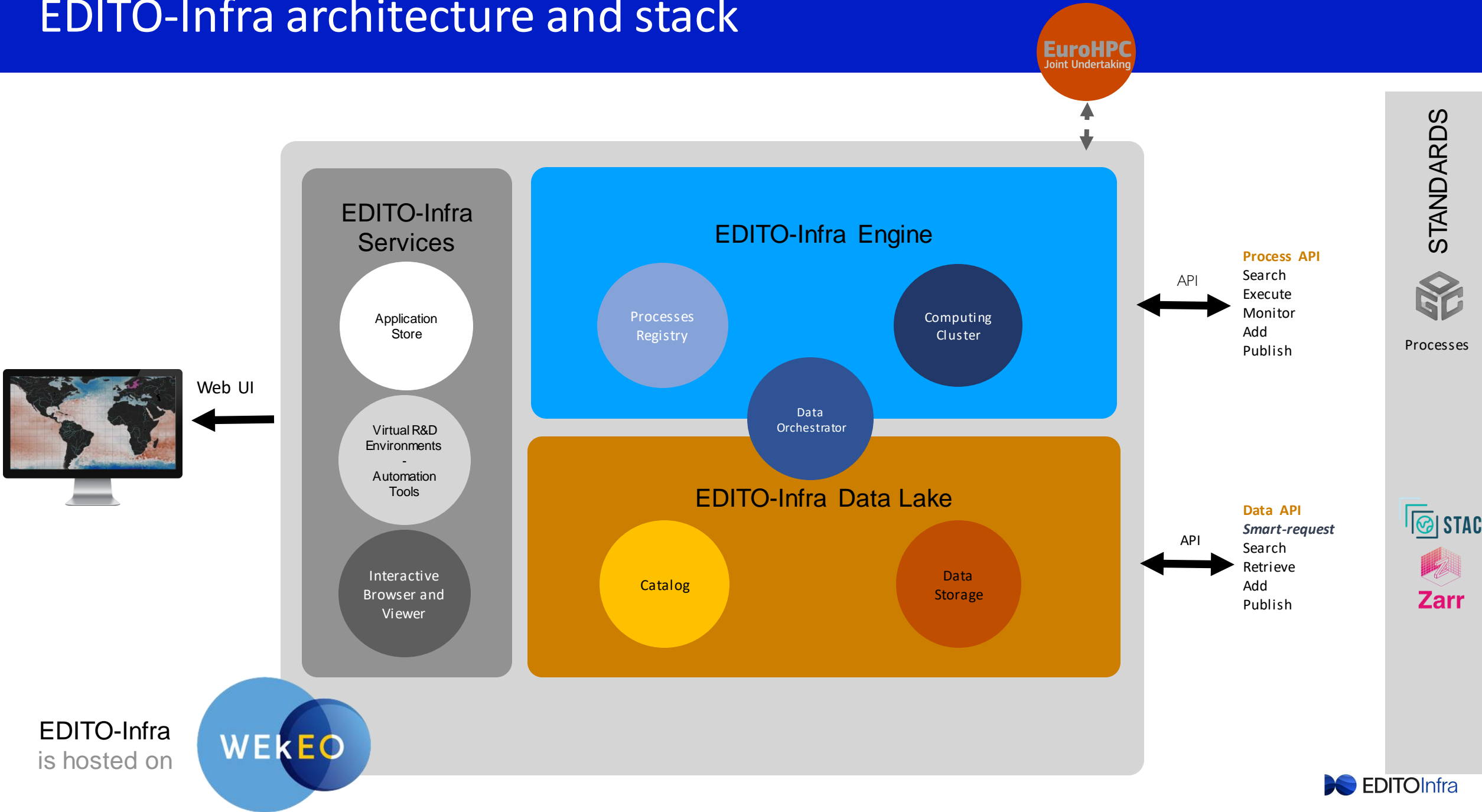
STAC API metadata catalog
S3 data storage
Processes API registry
Distributed computing clusters (CPU/GPU)
Virtual R&D Environments deployment
HPC/AI workflow managers deployment
Container deployment

Infrastructure/Cloud Provider

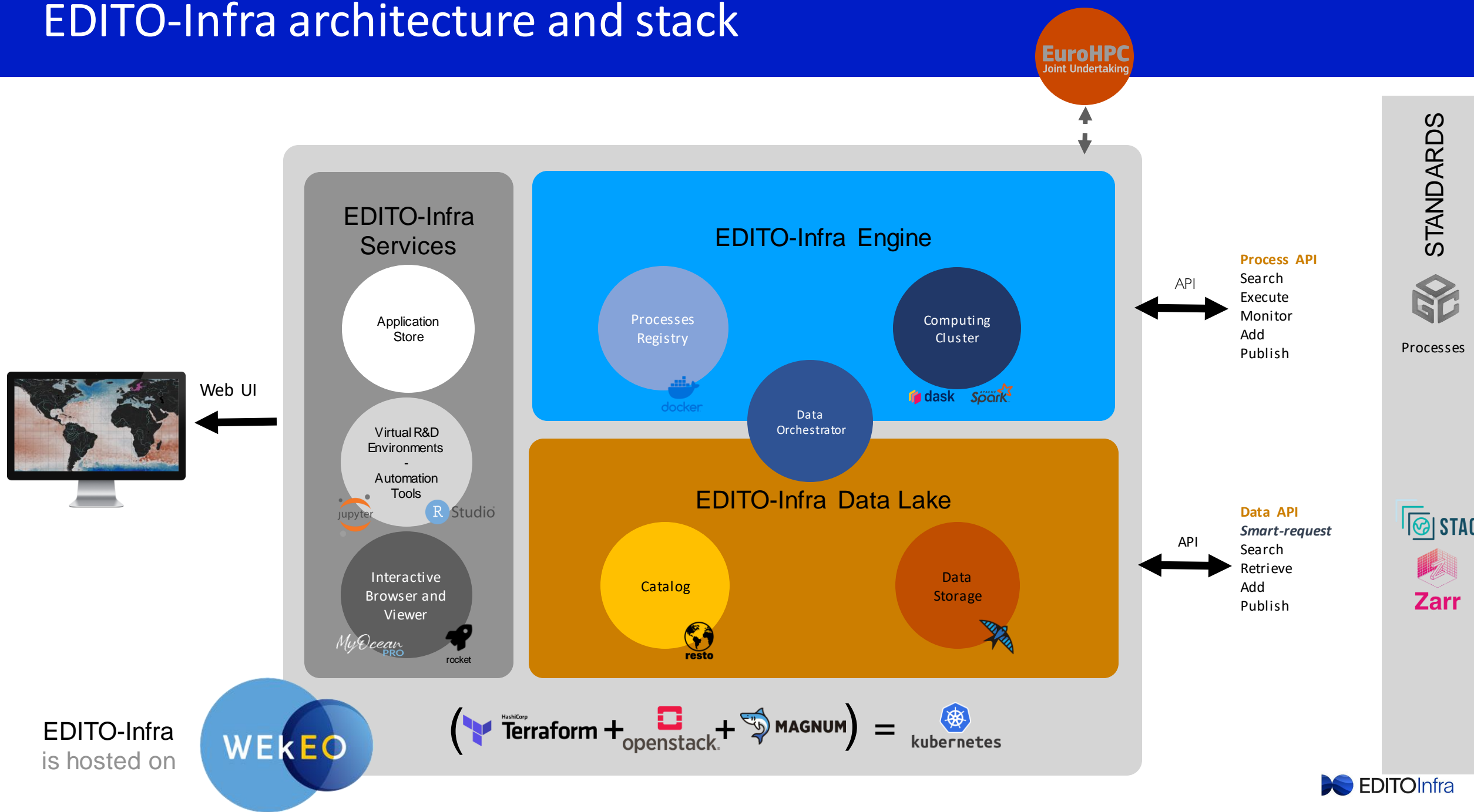
Layer to adapt depending on infrastructure providers



EDITO-Infra architecture and stack



EDITO-Infra architecture and stack



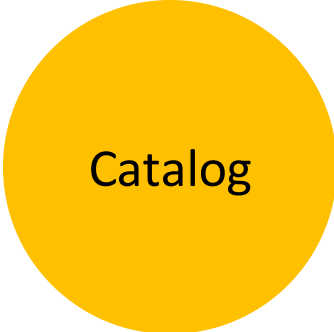
EDITO-Infra is hosted on



$$(\text{HashiCorp Terraform} + \text{openstack} + \text{MAGNUM}) = \text{kubernetes}$$

EDITO-Infra Data Lake

- 🔒 Search data
- 🔒 Reference external data
- 🔒 Create catalogs



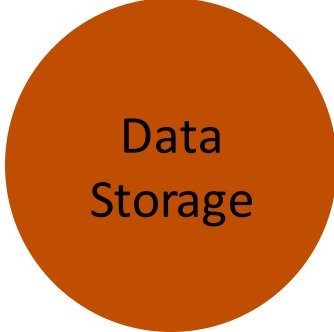
Catalog

Based on **resto** 
Shipped with **CMEMS** and **EMODNET** metadata
Search API conforms to **STAC** 



Data are automatically **referenced** within the Catalog

- 🔒 Push data
- Retrieve data



Data Storage

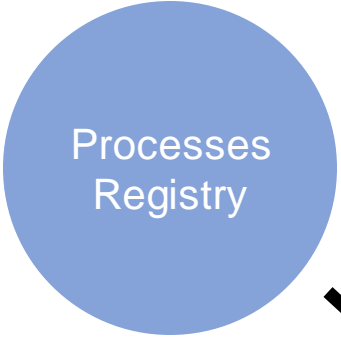
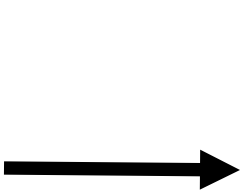
S3 object storage based on **OpenStack SWIFT**
EMODNET data in **ARCO** format
CMEMS data stored in **ZARR** format



🔒 Identity and Access Management

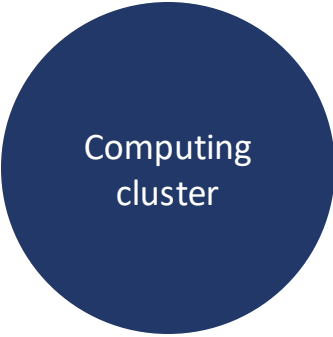
EDITO-Infra Engine

- Search process
- Execute a process (i.e., job)
- Monitor execution
- Add a new process



Provides a **catalog** of all available **processes**
Processes are **remotely callable**
Conforms to **OGC API processes**

Dispatch operations

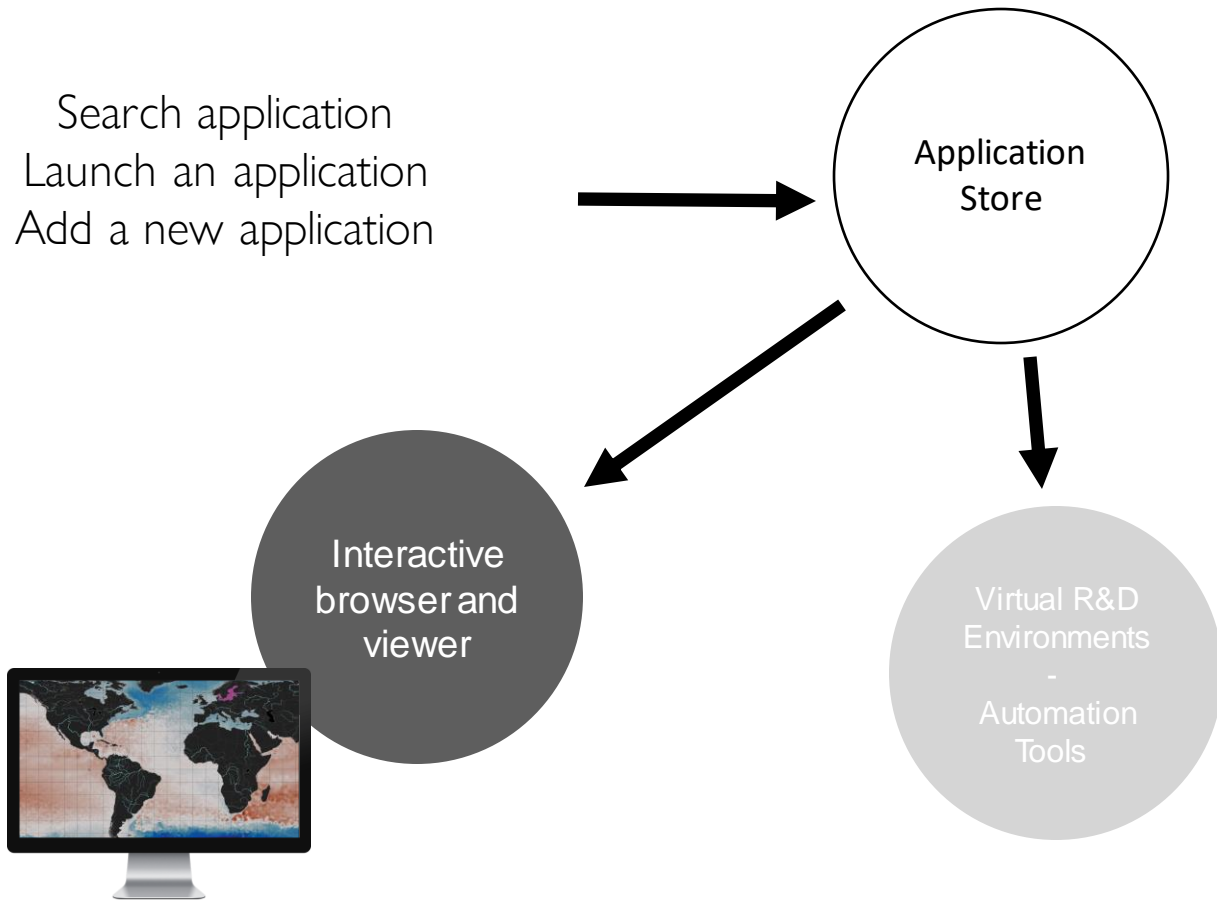


Powered by **Kubernetes**
Pangeo-inspired (Dask, Spark, etc.)



EDITO-Infra Services

- 🔒 Search application
- 🔒 Launch an application
- 🔒 Add a new application



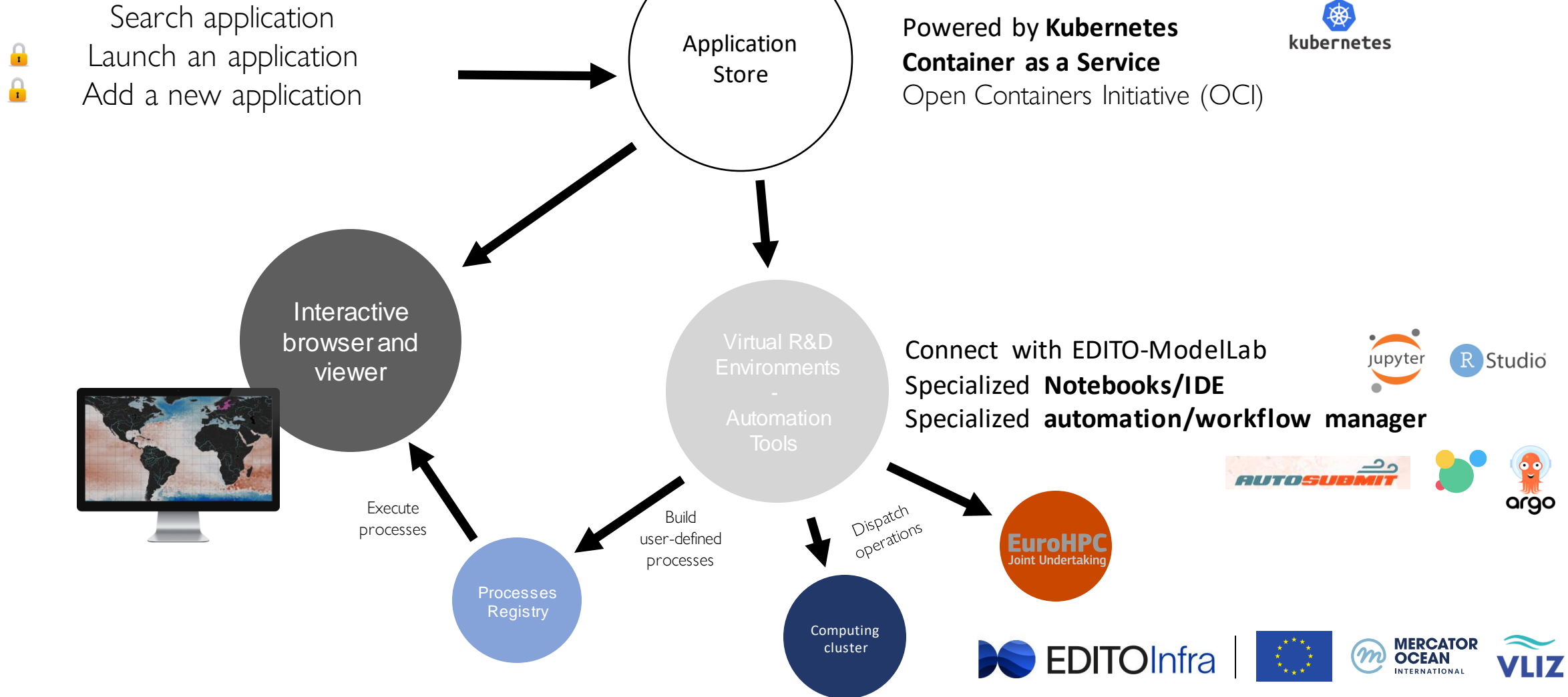
Powered by **Kubernetes**
Container as a Service
Open Containers Initiative (OCI)



Connect with EDITO-ModelLab
Specialized **Notebooks/IDE**
Specialized **automation/workflow manager**

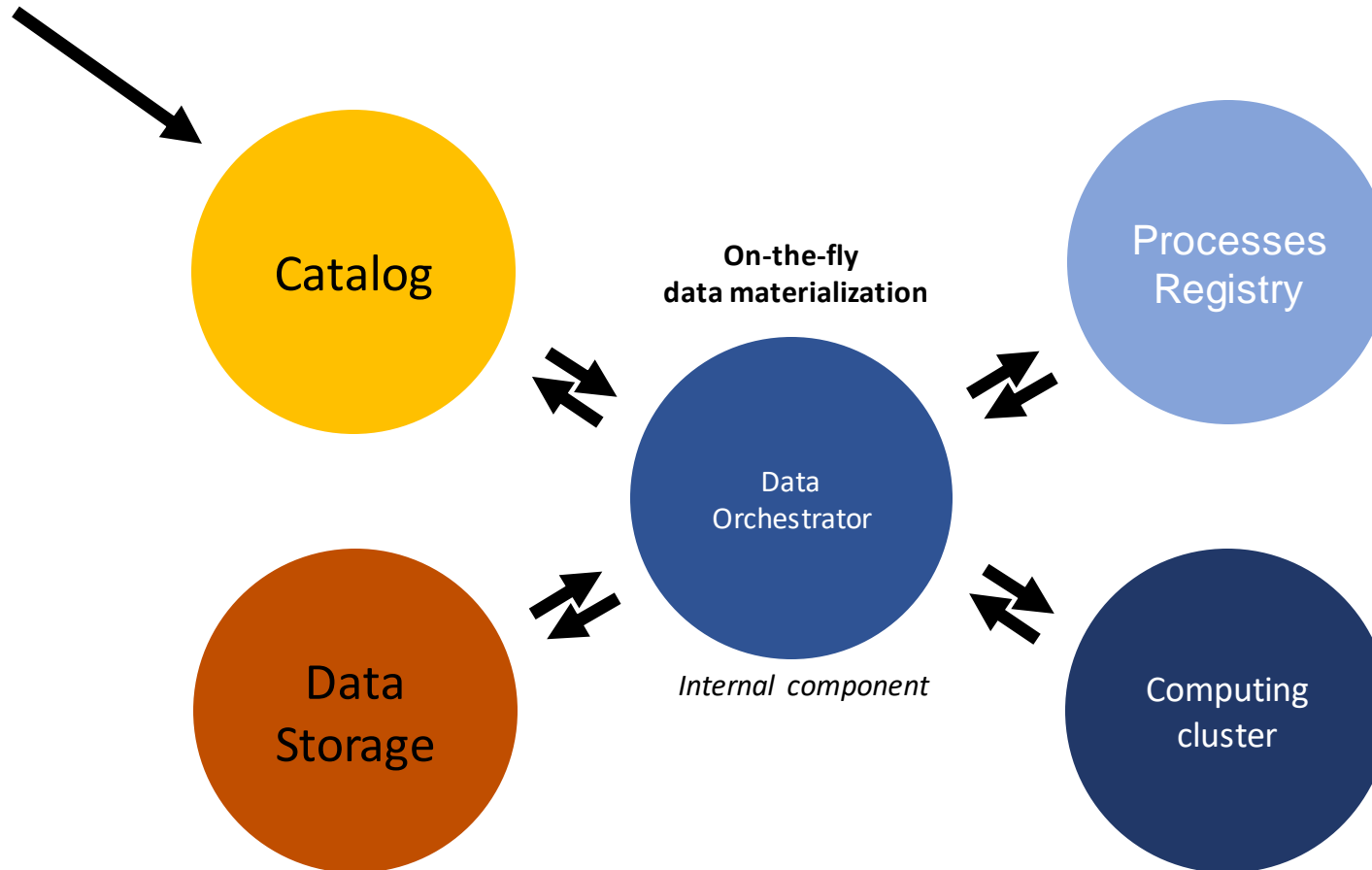


EDITO-Infra Services



Smart-request and asset materialization

Smart request





Thank you!

EDITO-Infra website: <https://edito-infra.eu/>

EDITO-Modellab website: <https://edito-modellab.eu/>

Mercator Ocean international website: <https://www.mercator-ocean.eu/>

European Commission website: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters/european-digital-twin-ocean-european-dto_en