Empowering Open Science: EGI Community's Impact on EOSC

ENES Data Space: an EOSC Beyond thematic node for the climate community

F. Antonio[1], S. Fiore[2], D. Elia[1], A. Shuvo[2], P. Nassisi[1]

[1] Advanced Scientific Computing Division, CMCC Foundation
[2] Department of Information Engineering and Computer Science, University of Trento

02 | 10 | 2024 by Fabrizio Antonio

Several **key challenges** and **practical issues** related to **large-scale climate analysis**

- Input data from **multiple models**

- **Data download** from central repositories to desktop machines

- **Client-side** & **sequential approaches**

- Several **data analysis tools** and **libraries** needed

- **Strong requirements** in terms of **computational** and **storage resources**

A Data Space for Climate Science in the European Open Science Cloud

Goal: Deliver an **open**, **scalable** and **cloud-enabled** data science environment for **climate analysis** on top of the **European Open Science Cloud platform**

- Access to **climate variable-centric** collections from **ESGF**

- **Jupyter-based gateway** to develop and re-use climate apps

- **Data Science software stack** for climate data analysis, visualization, and AI/ML use cases

- **Storage** & **Compute** resources from EGI

- Collaborations with European and international initiatives: **EGI**, **IDSA**, **ENES RI AISBL**
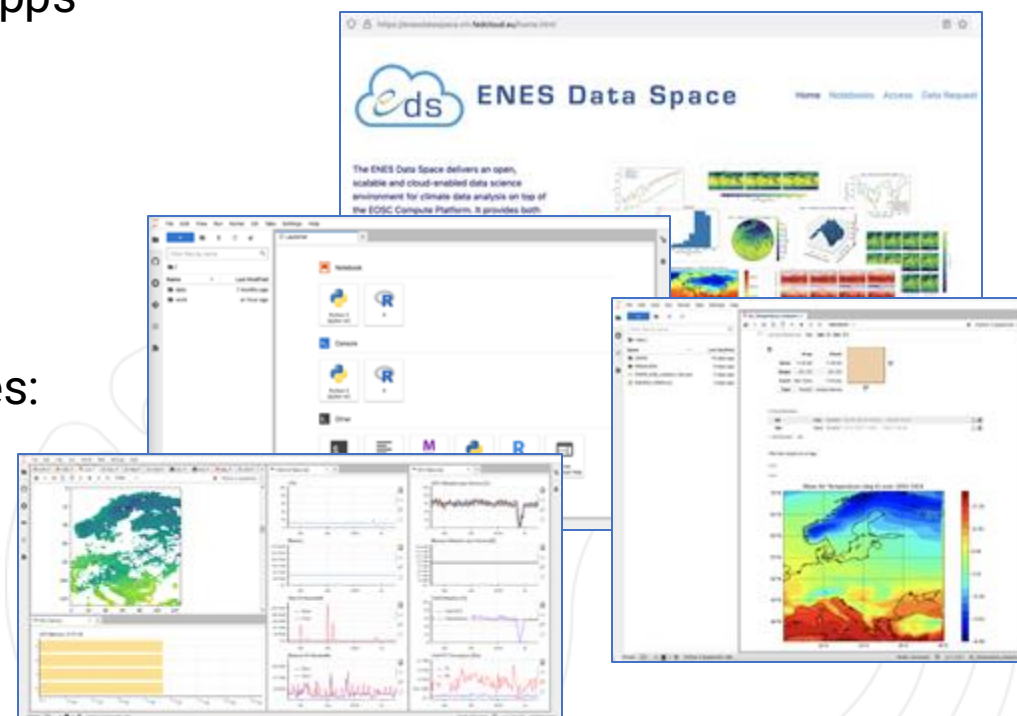
Ultimate goal: promote **Open Science** for **data** and **services**

ENES & IS-ENES

European Network for Earth System Modelling

A network of European groups in climate/Earth system modelling
Launched in 2001 (MOU)

Ca 50 groups from academic, public and industrial world

Main focus :
discuss strategy
to accelerate progress in climate/
Earth system modelling and
understanding

enes
EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

http://enes.org/

http://is.enes.org/

Slide courtesy: Sylvie Joussaume (IPSL)

The Coupled Model Intercomparison Project (**CMIP**) and

the Earth System Grid Federation (**ESGF**) data archive

IS-ENES provides the EU contribution to the ESGF

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING

Target community (II)

IPCC reports cover "the **scientific**, **technical** and **socio-economic** information relevant to understanding the scientific basis of risk of **human-induced climate change**, its **potential impacts** and options for **adaptation** and **mitigation**".

User stories
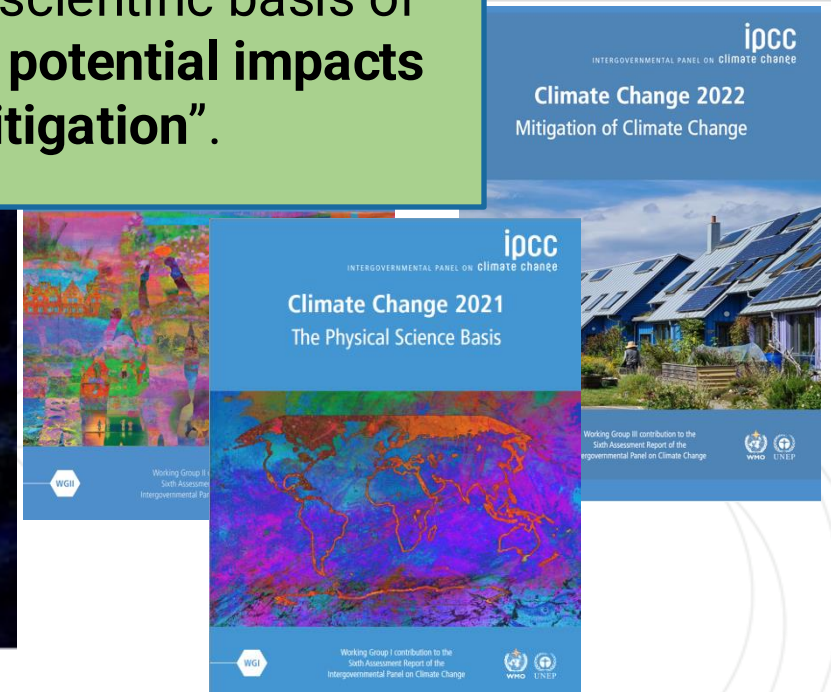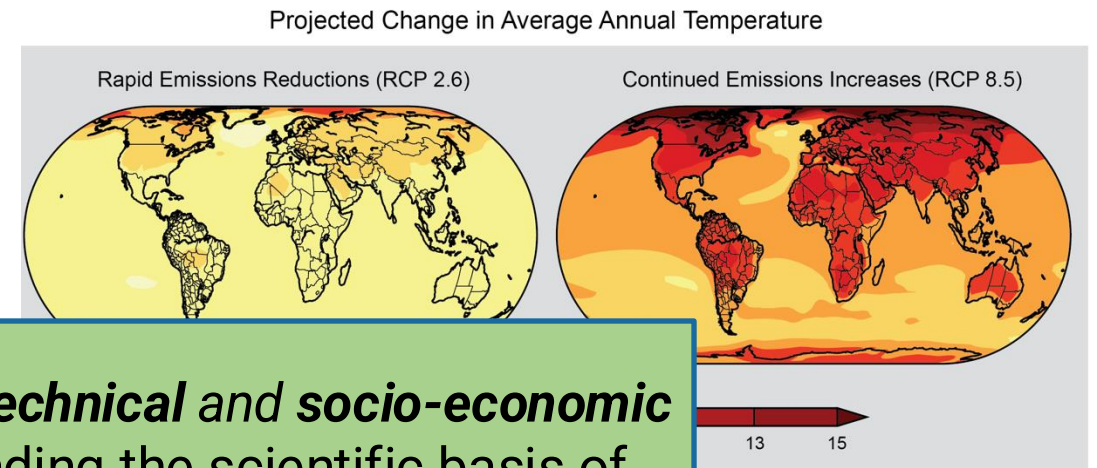
As a **scientist**, I want to get access to **computing resources** so that I can run **interactive analysis**, **AI/ML-based applications**, and **visualization** on climate data.

Scientist

As a **research division**, I want to expose my **data collections** and make them **searchable**, **discoverable** and **accessible** through a **federated data catalogue.**

Research Division

As a **scientist**, I want to run my **analytics workflow** and publish and/or manage **provenance information** at different levels of granularity.
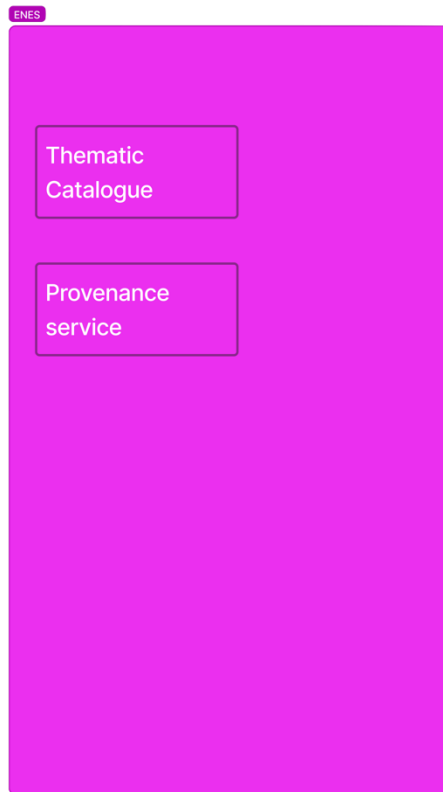
Scientist

Integration

EOSC Beyond core services planned to be integrated:

- **AAI**: Access to service via AAI proxy

- **Service Accounting**: Tracking usage of consumable services

- **Service Monitoring**: Monitoring of service performance (availability, reliability)

- **Orchestration Service:** To provision and configure cloud resources (e.g., Infrastructure Manager)

- **PID Service**: To enhance findability and citability of research products

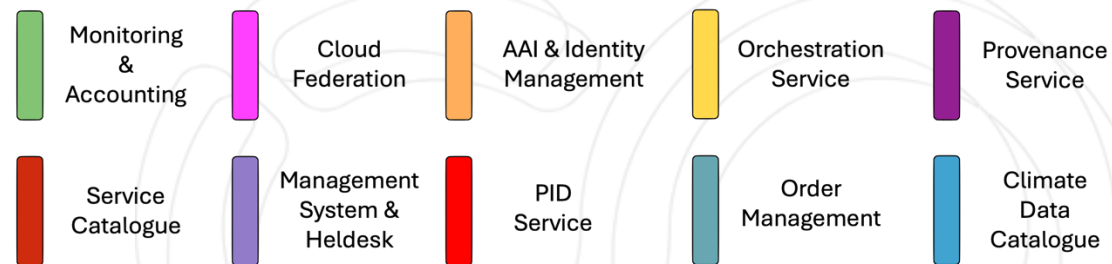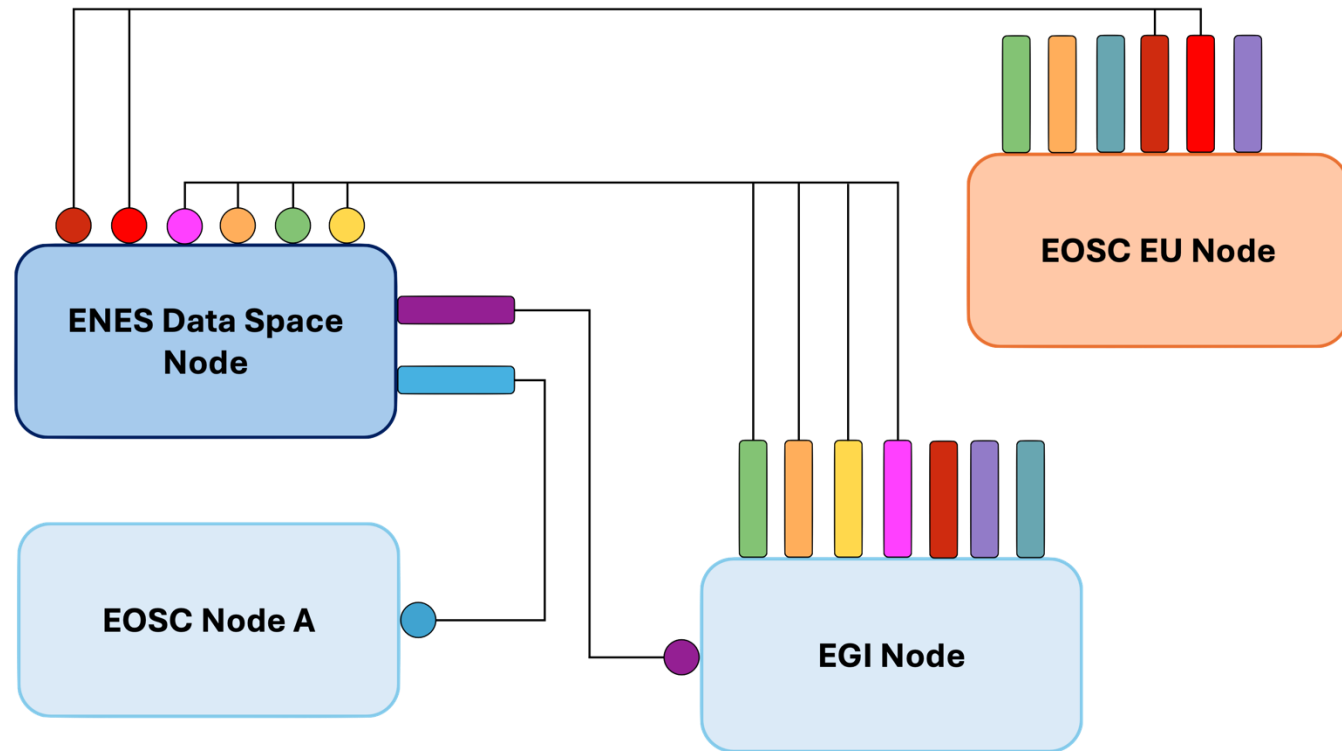- **Service catalogue**: Onboarding to the EOSC Resource Catalogue

ENES Data Space high-level architecture
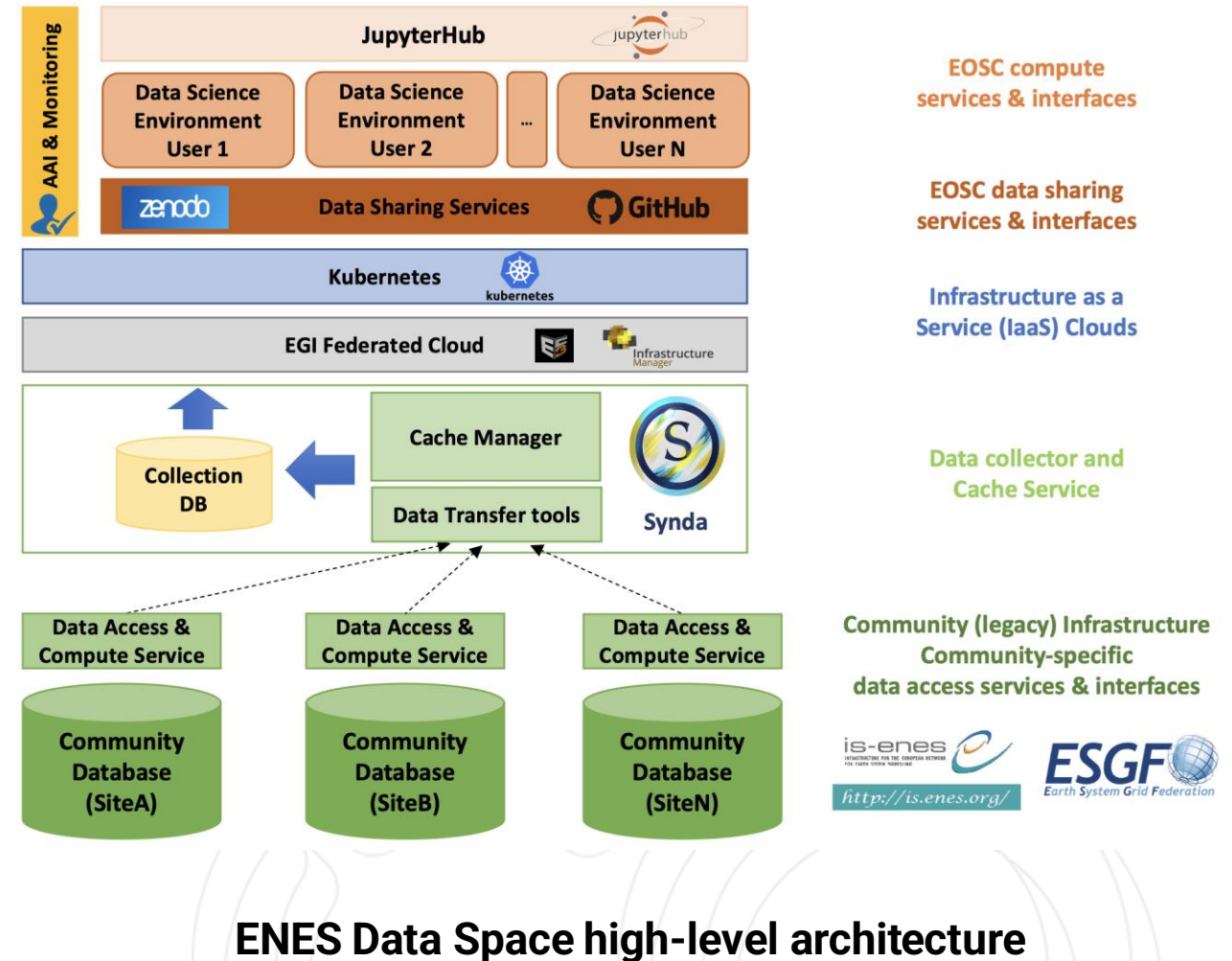
Federation

ENES as a Consumer – Capabilities **to take** from the Federation

ENES as a Supplier – Capabilities **offered** to the Federation

Federating Capabilities

Provenance Service https://github.com/HPCI-Lab/yProv

- **yProv**, an **interoperable service** for provenance management within end-to-end scientific workflows

- Compliance with the **W3C PROV Family of Standards**

- Tracking provenance in complex **AI-based scientific workflows** jointly with a set of **key metrics**

- Foster **provenance exploration** and **analysis opportunities**

- **Co-PIs**: S. Fiore (UniTrento) and F. Antonio (CMCC)

Fabrizio Antonio et al.

yProv: a Cloud-enabled Service for Multi-level Provenance Management And Exploration in Climate Workflows

San Martino, 3 Oct 2024, 9:40-10:00

Thematic Data Catalogue

- **Data catalogue** built on top of climate collections
  - Standardize the way geospatial asset metadata is structured and queried
  - Improve data search and discovery
  - **STAC** (SpatioTemporal Asset Catalog) family of specifications to be explored
- **Object Storage** and **Zarr** format to improve data access

- The **ENES Data Space** represents a domain-specific implementation of the **data space concept**, supporting scientific communities towards a **more sustainable**, **effective** and **FAIR** use of data

- It aims to offer **core services** and **capabilities** relevant to the **climate community**

  o Data from well-known initiatives (e.g., CMIP)

  o Compute & Storage resources

  o Infrastructural components for deployment and orchestration of services

  o Software solutions supporting researchers and institution departments in realistic scenarios

- Contribution to **EOSC Beyond**:

  o Provenance service and thematic data catalogue as federated capabilities to be offered

  o Help test and validate the new EOSC Core capabilities

![EOSC Beyond logo]

Get in touch with us

Thank you!