The ENES Data Space

Period 1 Review meeting, 24/05/2022

Fabrizio Antonio
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)
On behalf of the ENES Data Space team

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

http://enes.org/

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

http://is.enes.org/

IS-ENES infrastructure projects
IS-ENES (2009-2013)
IS-ENES2 (2013-2017)
IS-ENES3 (2019-2022)

Support WCRP internationally coordinated climate model experiments
(CMIP & CORDEX)

Support sharing of expertise on
climate models, tools & HPC
https://is.enes.org

Slide courtesy: Sylvie Joussaume (IPSL)

The Coupled Model Intercomparison Project (CMIP) and the Earth System Grid Federation (ESGF) data archive
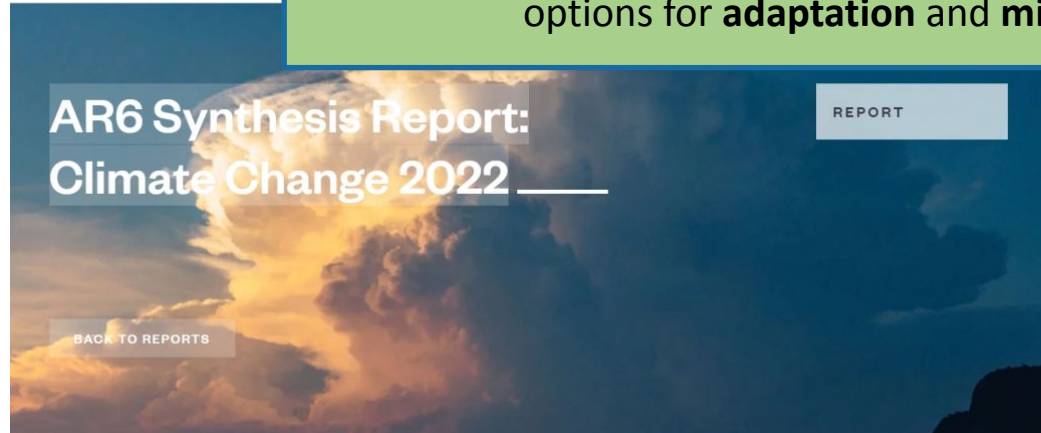

IS-ENES provides the EU contribution to the ESGF
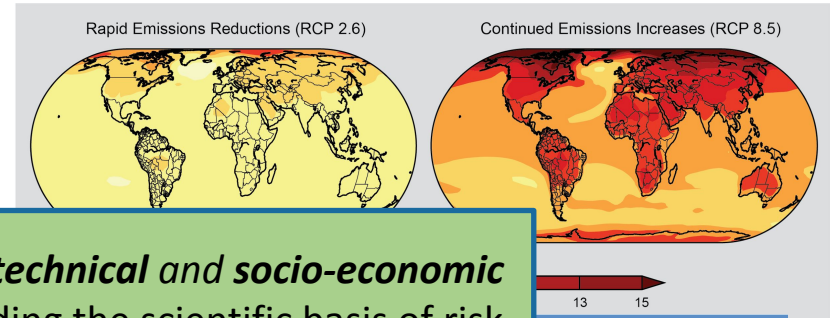
Why does this data matter?

Projected Change in Average Annual Temperature

Rapid Emissions Reductions (RCP 2.6)

Continued Emissions Increases (RCP 8.5)

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**".

Climate analysis challenges & issues

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

- Input data from **multiple models**

- **Data download is a big barrier** for climate scientists

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

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

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

ENES Data Space: main objectives

*__Goal__: Deliver an open, scalable and cloud-enabled **data science environment** for **climate data analysis** on top of the **EOSC Compute Platform***
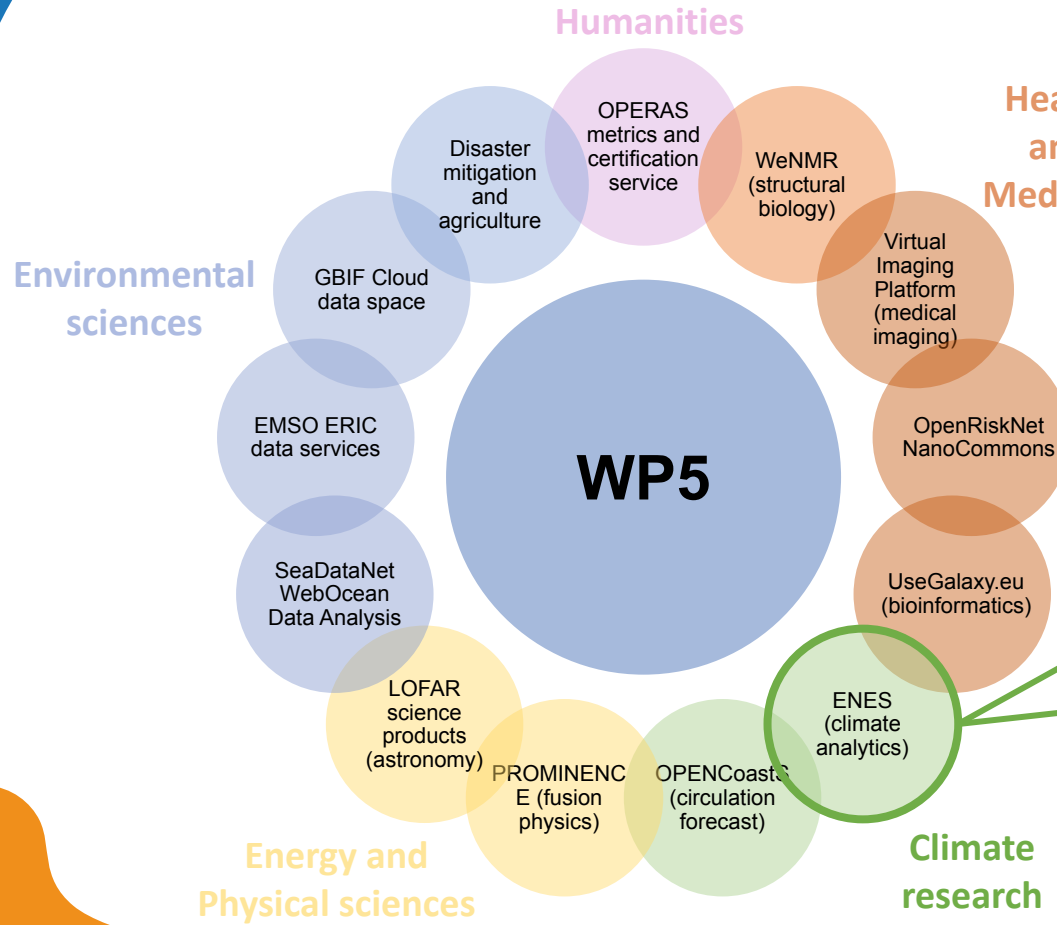
The **ENES *Data Space*** aims at providing an entry point to:

- **Datasets** (e.g. CMIP6) → most relevant from ESGF; pre-staged

- **Storage** & **Compute** resources → provided by EGI

- **Data Science Software Stack** → to address a wide spectrum of analysis needs

- **IDE** and **Applications** → to devel/share/(re-)use apps → FAIR principles

- **Cloud-Enabled** → SaaS for end-users applications; PaaS for data analysis service
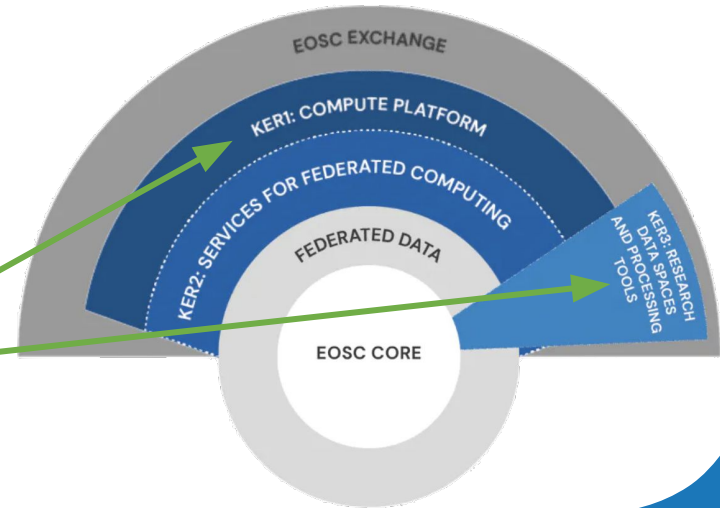
→ thus promoting **Democratization of eScience**
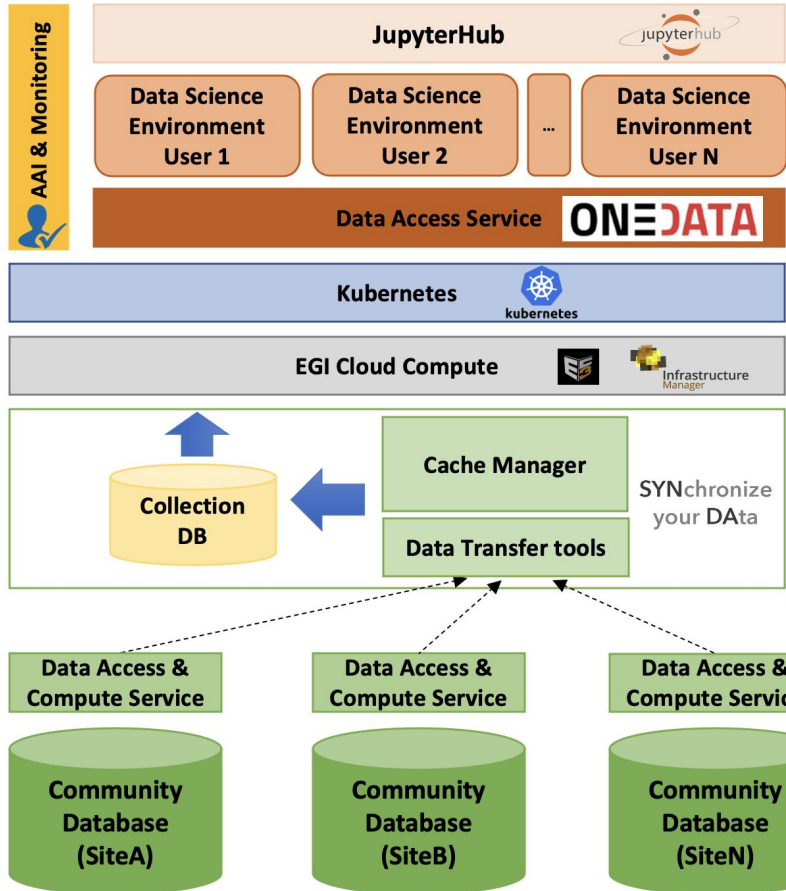
Contribution to the EGI-ACE project
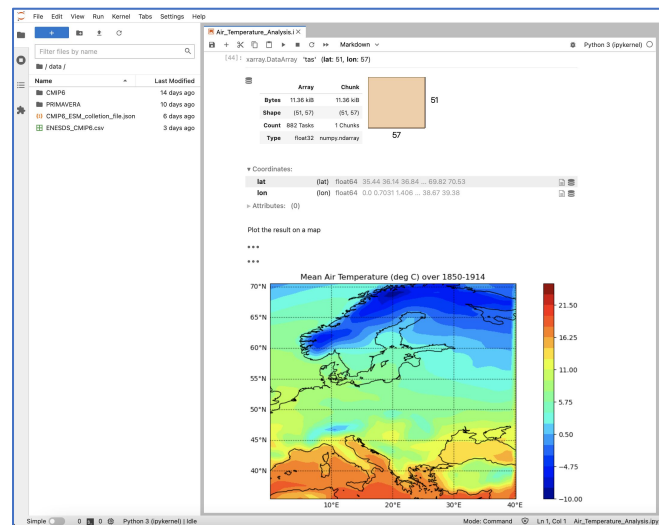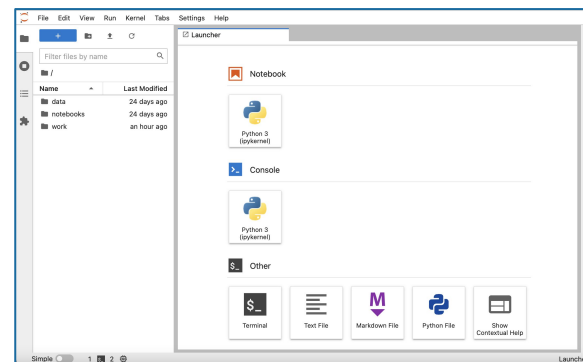

Image courtesy: EGI

ENES Data Space architecture

Jupyter-based Data Science environment

- **JupyterHub** as entry point to computational environment and resources

- **JupyterLab** instance equipped with a set of **open source Python modules** and **computing frameworks**

- **CMIP variable-centric** collections from the **ESGF federated data archive**

  - **Synda** community tool to download and (one-way) synchronize local data pool

  - **Data Search & Discovery** capabilities

Resource Capacity, Services & Data pool

Cloud Compute

- Number of virtual CPU cores: 160

- Memory [GB]: 160

- Cluster size: 10 WNs

Online storage

- Guaranteed storage capacity [TB]: 300

Accounting

- Through the EGI Accounting Portal

Monitoring

- Service Level Monitoring through ARGO service

Data pool

- Total datasets: ~ 8000

- Total size [TB]: ~ 30

- Variables: 22
  - air temperature, precipitation flux, relative and specific humidity, ...

- Frequencies : yearly, monthly, daily, 6-hourly

- Experiments: historical, future scenario based on SSPs

- Climate Models: ~ 60
  - CMCC-ESM2, CMCC-CM2-HR4, CMCC-CM2-VHR4, IPSL-CM6A-LR, IPSL-CM6A-ATM-HR, EC-Earth3, ...

DEMO

Conclusions and future plan

Conclusions

- The **ENES Data Space** represents a domain-specific implementation of the data space concept

- It provides a single entry-point to **compute capabilities** co-located with a **local data store** hosting a specific data selection from ESGF

Future plan

- Keep enhancing the ENES Data Space
  - **Additional datasets** based on users needs
  - Integration of **new community tools and libraries**
  - Improve **Reusability & Sharing**
- Support scientific **use cases** selected from open calls
- **Training** and **dissemination** activities

Useful links

EGI-ACE: https://www.egi.eu/projects/egi-ace/

ENES Data Space: https://enesdataspace.vm.fedcloud.eu/

ENES portal: https://portal.enes.org/

Thank you!

Contact: **egi-ace-po@mailman.egi.eu**
Website: **www.egi.eu/projects/egi-ace**

EGI Foundation

@EGI_eInfra