ENES Data Space: an open, cloud-enabled data science environment for climate analysis

F. Antonio[1], D. Elia[1], A. Giannotta[1], A. Nuzzo[1], G. Levavasseur[2], A. Ben Nasser[2], P. Nassisi[1], A. D'Anca[1], S. Fiore[3], S. Joussaume[2], G. Aloisio[1]

[1] *Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy*
[2] *Institut Pierre Simon Laplace (IPSL), Centre National de Recherche Scientifique (CNRS), France*
[3] *University of Trento, Trento, Italy*

EGU General Assembly 2022
Vienna & Online | 23-27 May 2022
ESSI3.3 | EGU22-7330

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

The European Open Science Cloud (EOSC)

EOSC

- An environment for hosting and processing research data to support **EU science**
- **FAIR Data and Services** for science in Europe

EUROPEAN OPEN SCIENCE CLOUD

EGI-ACE: Advanced Computing for EOSC

- Implement the **Compute Platform of the European Open Science Cloud**
- Deliver **computing platforms**, **data spaces** and **tools** as an integrated solution

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

IS-ENES provides the EU contribution to the ESGF

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

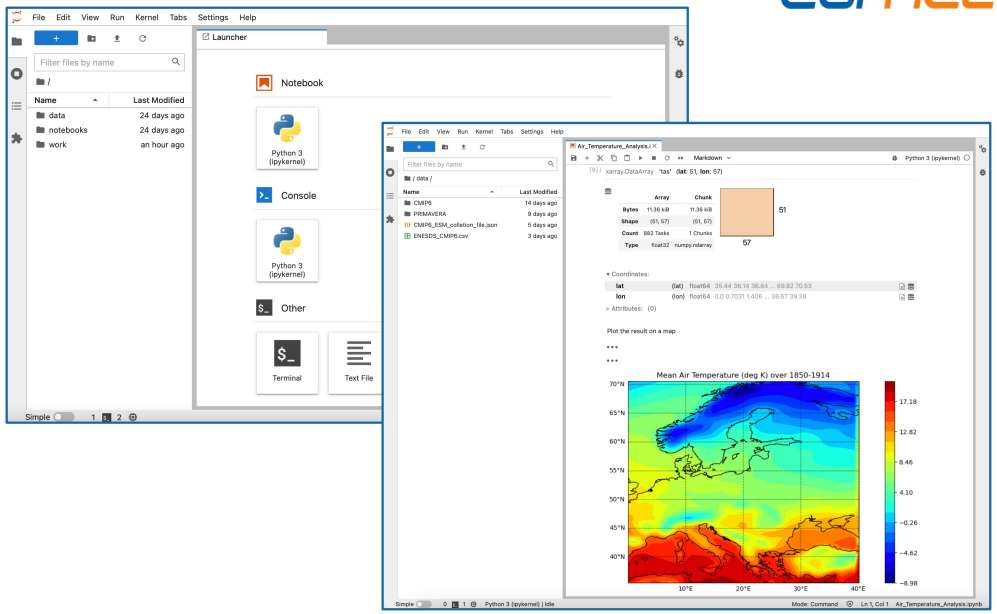
Goal**: Deliver an open, scalable and cloud-enabled **data science environment** for **climate data analysis** on top of the **EOSC Compute Platform

The **ENES *Data Space*** aims at providing an entry point to:

- **Datasets** → 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

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** (xarray, Dask, matplotlib, cartopy, ...)
- **CMIP variable-centric** collections from the **ESGF federated data archive**
  - ~ 8000 datasets, ~ 30 TB
  - **Synda** community tool to download and (one-way) synchronize local data pool

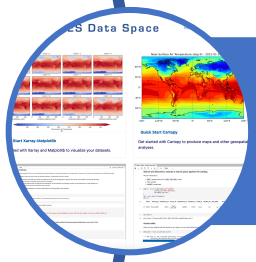


	index	activity_id	institution_id	source_id	experiment_id	member_id	table_id	variable_id	grid_label	version	time_range	start_year	end_year	path
0	5103	CMIP	CMCC	CMCC-CM2-SR5	historical	r1i1p1f1	Amon	hurs	gn	v20200616	185001-201412	185001	201412	/home/jovyan/data/CMIP6/CMIP/CMCC/CMCC-CM2-SR5...
1	5104	CMIP	CMCC	CMCC-CM2-SR5	historical	r1i1p1f1	Amon	huss	gn	v20200616	185001-201412	185001	201412	/home/jovyan/data/CMIP6/CMIP/CMCC/CMCC-CM2-SR5...
2	5105	CMIP	CMCC	CMCC-CM2-SR5	historical	r1i1p1f1	Amon	pr	gn	v20200616	185001-201412	185001	201412	/home/jovyan/data/CMIP6/CMIP/CMCC/CMCC-CM2-SR5...
3	5106	CMIP	CMCC	CMCC-CM2-SR5	historical	r1i1p1f1	Amon	tas	gn	v20200616	185001-201412	185001	201412	/home/jovyan/data/CMIP6/CMIP/CMCC/CMCC-CM2-SR5...
4	5107	CMIP	CMCC	CMCC-CM2-SR5	historical	r1i1p1f1	Amon	uas	gn	v20200616	185001-201412	185001	201412	/home/jovyan/data/CMIP6/CMIP/CMCC/CMCC-CM2-SR5...

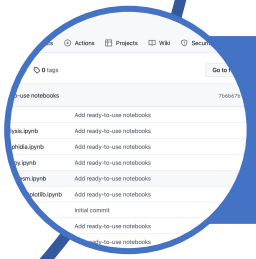
Climate research workflow

ENES Data Space

Data Search & Discovery

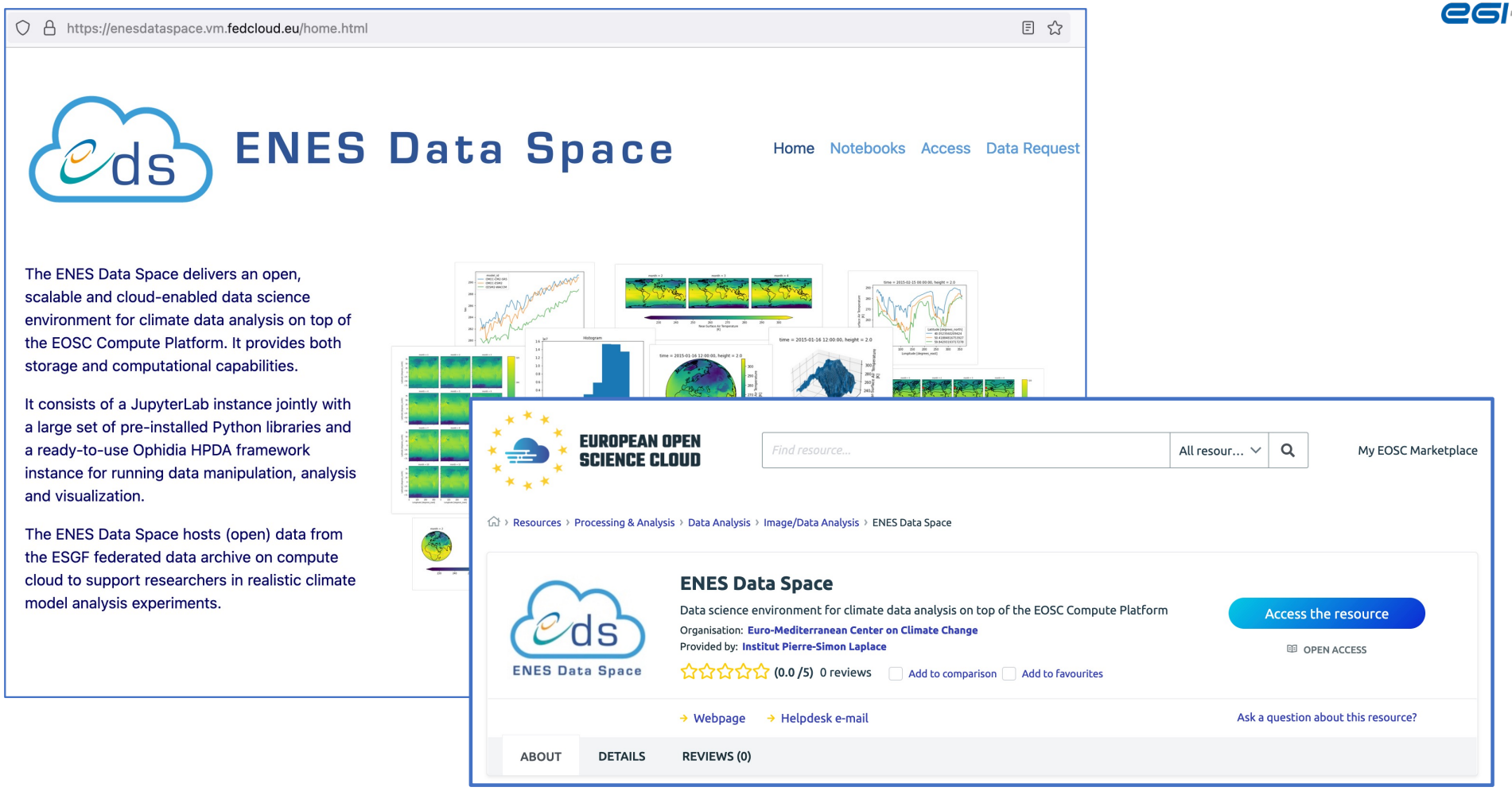


Data Analysis & Visualization



Reusability & Sharing

ENES Data Space portal



https://marketplace.eosc-portal.eu/services/enes-data-space/

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/

EGI-ACE *receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017567.*

IS-ENES3 *has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824084*

Thank you!

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

EGI Foundation

@EGI_eInfra

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.