



cmcc

Centro Euro-Mediterraneo
sui Cambiamenti Climatici

www.cmcc.it

Insights into Climate Computing

EGI2024

Conference Opening

1 October 2024

*Fabrizio Antonio¹, Paola Nassisi¹, Sandro Fiore²,
Donatello Elia¹*

¹ Advanced Scientific Computing Division, CMCC Foundation

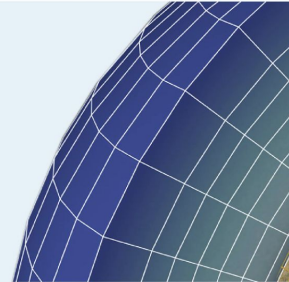
² Department of Information Engineering and Computer Science, University of Trento

ENES & IS-ENES

European Network for Earth System Modelling



Welcome to ENES



The European Network for Earth System modelling, **ENES, was launched in 2001**

Community working on **Earth's climate system modelling** with the aim to **accelerate progress** in this field.

Strongly involved in the assessments of the **Intergovernmental Panel on Climate Change (IPCC)**. It provides those predictions, on which **EU mitigation and adaptation policies** are elaborated.

The European Network for Earth System Modelling

The European Network for Earth System modelling, ENES, was launched in 2001. It gathers the community working on Earth's climate system modelling with the aim to accelerate progress in this field. This community is strongly involved in the assessments of the Intergovernmental Panel on Climate Change (IPCC) and provides those predictions, on which EU mitigation and adaptation policies are elaborated.

This portal, established and maintained by IS-ENES (Infrastructure for ENES), aims to provide information on ENES projects.

It also provides information on ENES research infrastructure. Services are provided through the IS-ENES providing access to key model data and software, complemented by ESIVACE projects with services on enhancing performance on High-performance computers.



About ENES

[More >](#)



Access to services

ENES Services on climate model data and services models, tools and high-performance computing

[More >](#)



The ENES projects

[More >](#)



<http://is.enes.org/>

IS-ENES infrastructure projects

IS-ENES (2009-2013)

IS-ENES2 (2013-2017)

IS-ENES3 (2019-2022)

ENES Infrastructure Strategy (2024-2033)

IS-ENES3 Deliverable D2.1

Infrastructure Strategy for Earth System Modelling for 2024-2033

What is needed to sustain large-scale European earth system modelling infrastructure from 2024 and beyond

Reporting period: 01/01/2022 – 31/03/2023

Authors: Bryan Lawrence (UREAD-NCAS), Fanny Adloff (DKRZ), Sylvie (CNRS-IPSL)

Reviewers: Michael Lautenschlager (DKRZ), Sophie Valcke (CERFA)
Release date: 22/05/2023

ABSTRACT

This document provides a set of recommendations which arise from an analysis of European large-scale climate modelling needs carried out by the European Network for Earth System Modelling. There are detailed recommendations made under each of: Infrastructure Development, Collaboration, Diagnostics, Data Systems, and Workforce.



Research
infrastructure



There are seven headline recommendations:

1. The HPC community must continue to provide both CPU and GPU machines; a lot of climate codes will not be sensibly deployed on GPU machines in the near future (even if they can be made to run on them).
2. There is a need for a more operational aspect to some aspects of climate science.
3. The community should continue to invest in managing and sustaining shared infrastructure.
4. Model development takes a long time and is resource intensive. Modellers will have to pay attention to the choices between Performance, Portability, and (scientific) Productivity; in the new world, we can only have two!
5. Large expensive modelling projects need to be treated like satellite missions, well publicised and documented.
6. The community should continue to invest in the necessary underpinning diagnostic tools and libraries.
7. Storage and data systems need to support a variety of use-cases.

Recommendations on «Storage and Data Systems»

3.6 Data Systems

25. It will be desirable that trans-national access to archives and compute systems be sustained in such a way as to minimise unnecessary data movement and data replication – and where possible support access to scientists from the global south in accordance with WCRP goals.
26. Sustaining access will need not only shared infrastructure, but shared infrastructure development.
27. Data volumes will continue to grow, and simulations will continue to be carried out on multiple platforms. It will be necessary to maintain distributed catalogue systems and methods to replicate data to national and international archives with co-located analysis compute.
28. Archive planning should cover transient (cache) and persistent (curated) use cases, recognizing that not all data products will be suitable for long-term curation, and different storage formats might be suitable for different use cases.

archives, not all data will be collocated for all workflows, and so port distributed analytics will need to be developed and integrated

a systems, catalogues and data analytics will continue to demand data storage and metadata. Modellers should continue to use and fast conventions to maximise data re-use in accordance with FAIR

31. The growing demand for climate services will lead to the need for data sharing across communities, not just within the research component of the earth system modelling community. The climate community will need to work with these other communities to ensure the appropriate services and information are available via commonly understood protocols.
32. Data users will also continue to need appropriate documentation as to how and why data were produced, and to be able to discover and report issues with the data after simulations have concluded. Systems to streamline the production and use of such information will need to be improved and maintained.

Large scale, community experiments

The **Coupled Model Intercomparison Project (CMIP)** represents a first-class large-scale global experiment for climate change research

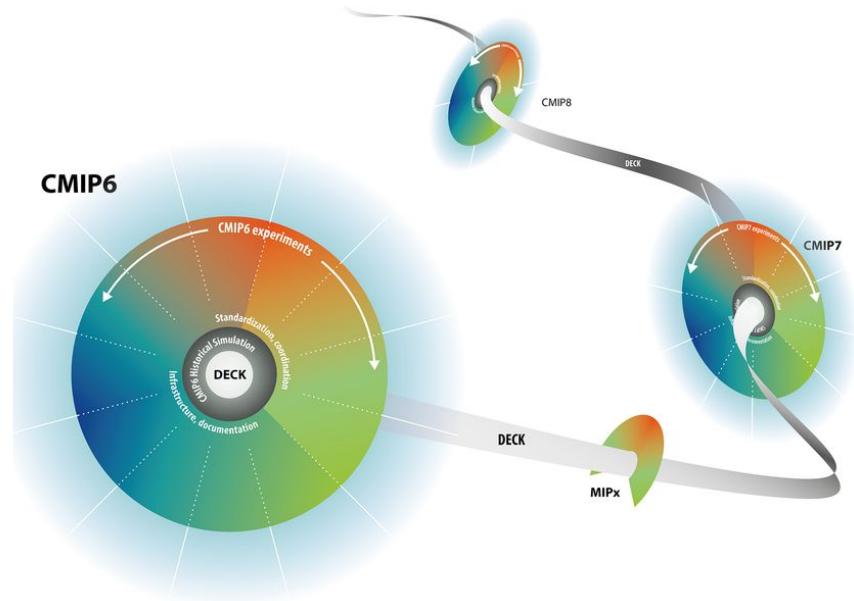


<https://doi.org/10.5194/egusphere-2024-453>
Preprint. Discussion started: 19 February 2024
© Author(s) 2024. CC BY 4.0 License.

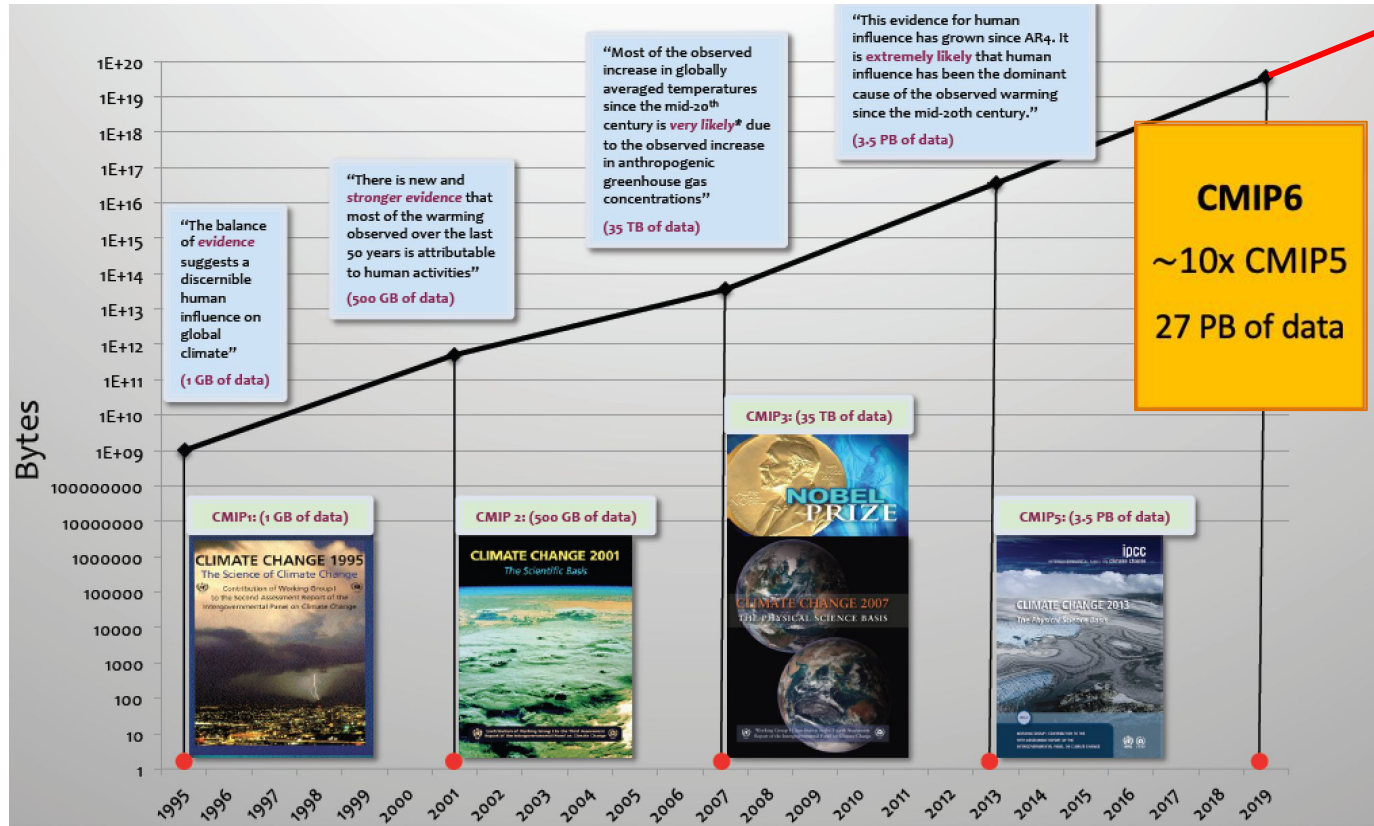


Bringing it all together: Science and modelling priorities to support international climate policy.

- Colin G. Jones¹, Fanny Adloff², Ben B. Booth³, Peter M. Cox⁴, Veronika Eyring^{5,6}, Pierre Friedlingstein^{7,8}, Katja Frieler⁹, Helene T. Hewitt³, Hazel A. Jeffery¹, Sylvie Jousseaume¹⁰, Torben Koenigk^{11,12}, Bryan N. Lawrence¹³, Eleanor O'Rourke¹⁴, Malcolm J. Roberts³, Benjamin M. Sanderson¹⁵, Roland Séférian¹⁶, Samuel Somot¹⁶, Pier Luigi Vidale¹³, Detlef van Vuuren^{17,18}, Mario Acosta¹⁹, Mats Bentsen^{20,21}, Raffaele Bernardello¹⁹, Richard Betts^{3,22}, Ed Blockley³, Julien Boé²³, Tom Bracegirdle²⁴, Pascale Braconnot¹⁰, Victor Brovkin²⁵, Carlo Buontempo²⁶, Francisco Doblas-Reyes^{19,27}, Markus Donat¹⁹, Italo Epicoco^{28,29}, Pete Falloon^{3,30}, Sandro Fiore³¹, Thomas Frölicher^{32,33}, Neven S. Fučkar^{34,35}, Matthew J. Gidden³⁶, Helge F. Goessling³⁷, Rune Grand Graversen³⁸, Silvio Gualdi³⁹, José M. Gutiérrez⁴⁰, Tatiana Ilyina⁴¹, Daniela Jacob⁴², Chris D. Jones^{3,43}, Martin Juckes^{1,44}, Elizabeth Kendon^{3,43}, Erik Kjellström¹¹, Reto Knutti⁴⁵, Jason Lowe^{3,46}, Matthew Mizielinski³, Paola Nassisi²⁸, Michael Obersteiner⁴⁷, Pierre Regnier⁴⁸, Romain Rochrig¹⁶, David Salas y Méjia¹⁶, Carl-Friedrich Schuessner⁴⁹, Michael Schulz⁵⁰, Enrico Scoccimarro³⁹, Laurent Terray²³, Hannes Thiemann⁵¹, Richard A. Wood³, Shuting Yang⁵², Sönke Zaehle⁵³



CMIP data history and its big data evolution



CMIP7
coming soon

Why does CMIP data matter?



The Intergovernmental Panel on Climate Change

The Intergovernmental Panel on Climate Change (IPCC) is the United Nations body

ipcc

AR6 Synthesis Report Change 2022

The IPCC finalized the Synthesis Report for the Sixth Assessment Report during the Panel's 58th Session held in Interlaken, Switzerland from 13 - 19 March 2023.

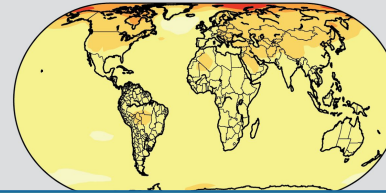
READ THE REPORT

CORE WRITING TEAM

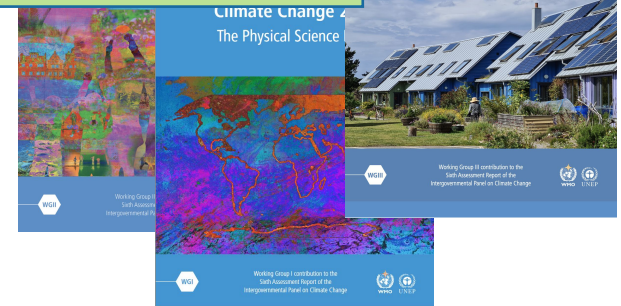
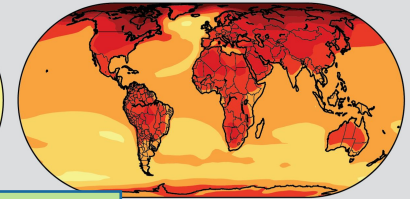
IPCC reports cover "the *scientific, technical and socio-economic* information relevant to understanding the scientific basis of risk of **human-induced climate change**, its **potential impacts** and options for **adaptation and mitigation**".

Projected Change in Average Annual Temperature

Rapid Emissions Reductions (RCP 2.6)



Continued Emissions Increases (RCP 8.5)



Earth System Grid Federation

The **Earth System Grid Federation (ESGF)** is a globally distributed peer-to-peer network of data servers using a common set of protocols and interfaces to archive and distribute **Earth system model outputs**



The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data, Cinquini L., Crichton D., Mattmann C., Harney J., Shipman G., Wang F., Ananthkrishnan R., Miller N., Denvil S., Morgan M., Pobre Z., Bell G.M., Doutraux C., Drach R., Williams D., Kershaw P., Pascoe S., Gonzalez E., Fiore S., Schweitzer R., Future Generation Computer Systems, Volume 36, 2014, Pages 400-417, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2013.07.002>

ESGF and IS-ENES

- ESGF seamlessly joins **climate science data archives** and **users** around the world
- **Data providers** make data available to the federation by publishing to one of two-dozen **ESGF node portals**
- Data can be **replicated** at other ESGF node sites for **backup**, to improve **ease of use** or to **exploit site resources**

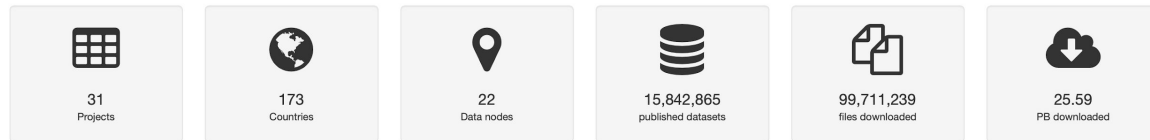
IS-ENES provides the EU contribution to ESGF



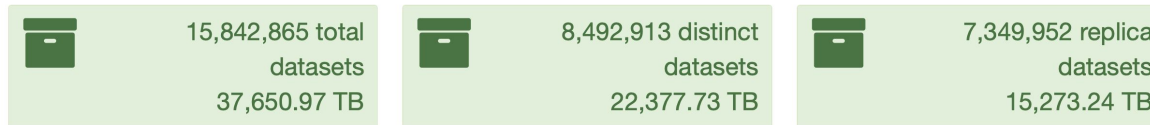
ESGF by the numbers

<http://esgf-ui.cmcc.it/esgf-dashboard-ui/>

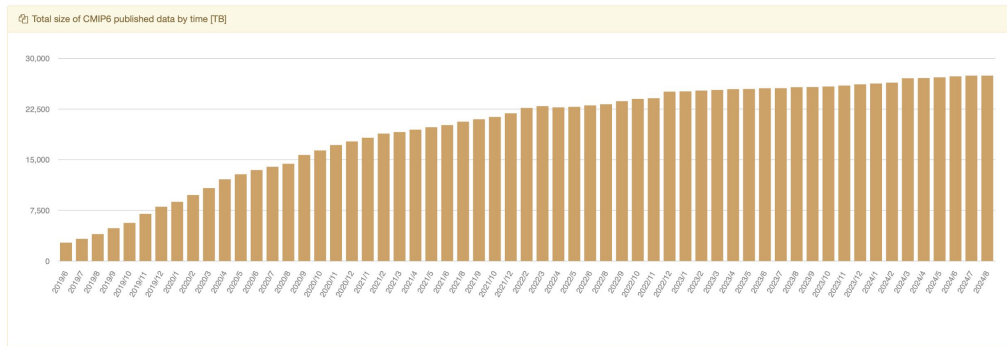
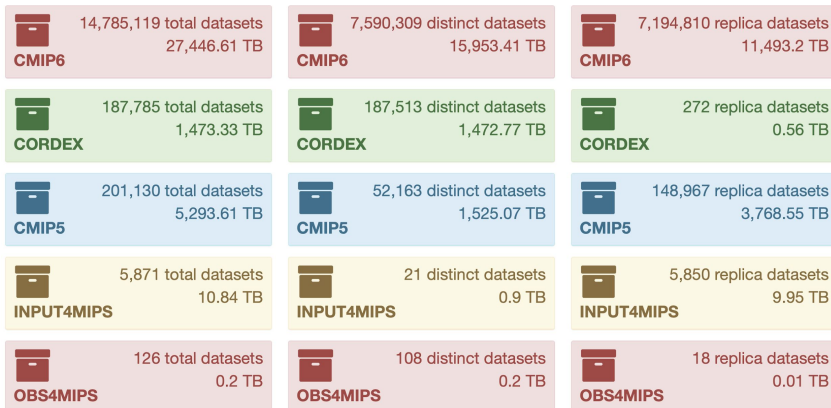
ESGF data usage and data publication metrics



Data published over the federation



Top projects

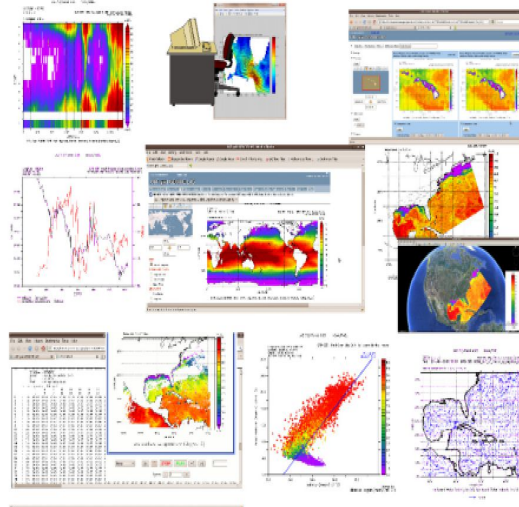


Total size of CMIP6 published data by time from 2PB in June 2019 to more than 25PB in 2024

From «climate data» to «climate computing»



The screenshot shows the ESGF (Earth System Grid Federation) website. The top navigation bar includes 'Home', 'Search', 'Tools', and 'Login'. Below this, there are sections for 'Current Submissions' and 'Search & Catalogue'. A search bar is visible with the text 'Search All Sites'. The main content area displays a map of the United States with various data points and a search form. The search form includes fields for 'Search type' (with options 'Enhanced' and 'Basic'), 'Search address', and 'Data fields'. The map shows a color-coded distribution of data points across the United States.



The new computational platform [...] will support **parallel and distributed computing** tasks by including **OpenMPI, Map/Reduce** and streaming computing models. The new compute node will allow for **large-scale manipulation and analysis of data** [...] We intend to fully explore the possibility of providing a configurable and scalable ESGF environment that can be easily deployed on the **cloud** [...] to meet requirements such as **high availability and elastic allocation of computing processes**.



The image shows the cover of the journal 'Future Generation Computer Systems'. The cover features the journal title at the top, a central image of a globe, and the Elsevier logo. The journal homepage is listed as www.elsevier.com/locate/fgs. The cover also includes the text 'Contents lists available at ScienceDirect' and 'Future Generation Computer Systems'.

The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data

Luca Cinquini^{ab,*}, Daniel Crichton^{ab}, Chris Mattmann^{ab}, John Harney^c, Galen Shipman^c, Feiyi Wang^c, Rachana Ananthakrishnan^{de}, Neill Miller^{de}, Sebastian Denvil^f, Mark Morgan^f, Zed Pober^g, Gavin M. Bell^h, Charles Doutriaux^h, Robert Drach^h, Dean Williams^h, Philip Kershaw^h, Stephen Pascoe^h, Estanislao Gonzalez^h, Sandro Fiore^h, Roland Schweitzer^h

- ^a Jet Propulsion Laboratory (JPL), 4800 Oak Grove Drive, Pasadena, CA 91109, USA
- ^b California Institute of Technology, 263 South Chester Avenue, Pasadena, CA 91106, USA
- ^c Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA
- ^d University of Chicago, USA
- ^e Argonne National Laboratory (ANL), Argonne, IL, USA
- ^f Institut Pierre Simon Laplace (IPSL), Paris, France
- ^g Goddard Space Flight Center (GSFC), Greenbelt, MD, USA
- ^h Lawrence Livermore National Laboratory (LLNL), Livermore, CA, USA
- ⁱ STFC Rutherford Appleton Laboratory, United Kingdom
- ^j NCAS/BMRC, Didcot, Oxfordshire, United Kingdom
- ^k German Climate Computing Centre (DKRZ), Hamburg, Germany
- ^l Institute of Meteorology and Climatology Berlin, Germany
- ^m Euro-Mediterranean Centre on Climate Change (CMCC), Lecce, Italy
- ⁿ Pacific Marine Environmental Laboratory (PMEL), Seattle, WA, USA

HIGHLIGHTS

- ESGF is a global infrastructure to support climate change research.
- ESGF nodes around the world are serving tens of thousands of users.
- ESGF includes services for data discovery, access, analysis and visualization.
- ESGF is supporting operationally the CMIP5 global distributed archive (3PB).
- ESGF includes model output, observations, and reanalysis data.

ARTICLE INFO

Article history:
Received 10 January 2013
Received in revised form 10 June 2013
Accepted 17 July 2013
Available online 17 September 2013

ABSTRACT

The Earth System Grid Federation (ESGF) is a multi-agency, international collaboration that aims at developing the software infrastructure needed to facilitate and empower the study of climate change on a global scale. The ESGF's architecture employs a system of geographically distributed peer nodes, which are independently administered yet united by the adoption of common federation protocols and application programming interfaces (APIs). The cornerstones of its interoperability are the peer-to-peer messaging that is continuously exchanged among all nodes in the federation; a shared architecture and API for search

* Corresponding author at: Jet Propulsion Laboratory (JPL), 4800 Oak Grove Drive, Pasadena, CA 91109, USA.
E-mail address: luca.cinquini@jpl.nasa.gov (L. Cinquini), daniel.j.crichton@jpl.nasa.gov (D. Crichton), chris.mattmann@jpl.nasa.gov (C. Mattmann), harney@ornl.gov (J. Harney), gshipman@ornl.gov (G. Shipman), fwang@ornl.gov (F. Wang), rananthakrishnan@ornl.gov (R. Ananthakrishnan), neillmiller@ornl.gov (N. Miller), sebastian.denvil@ipm.jpl.nasa.gov (S. Denvil), markmorgan@ipm.jpl.nasa.gov (M. Morgan), zed.pober@ornl.gov (Z. Pober), gavinm.bell@ornl.gov (G.M. Bell), charles.doutriaux@ornl.gov (C. Doutriaux), rdrach@ornl.gov (R. Drach), williams@ornl.gov (D. Williams), philip.kershaw@psu.ac.uk (P. Kershaw), stephen.pascoe@psu.ac.uk (S. Pascoe), estanislao.gonzalez@met.rwth-berlin.de (E. Gonzalez), sandro@euro-mediterranean.eu (S. Fiore), Roland.Schweitzer@llnl.gov (R. Schweitzer).

Climate Computing and the collaboration with EGI

Key collaboration with EGI Foundation:

- Represent the **ENES climate community** in the **EOSC ecosystem**
- Contribute in **strategies** and **open solutions** for advanced climate data analytics
- Explore novel open data and **cloud-based technologies** in the context of EOSC
- Design **AI-based solutions** and **components** for challenging climate sciences applications
- Bridge **climate community** needs



Developments in Data Space area (EGI-ACE project)

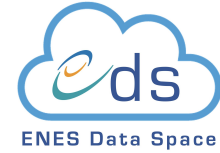
- ENES Data Space*

The screenshot shows the ENES Data Space interface. At the top, there is a navigation bar with the 'eds' logo and the text 'ENES Data Space'. Below the navigation bar, there is a grid of data visualizations, including line graphs, heatmaps, and maps. To the left of the visualizations, there is a text box that reads: 'The ENES Data Space delivers an open, scalable and cloud-enabled data science environment for climate data analysis on top of the EOSC Compute Platform. It provides both storage and computational capabilities. It consists of a JupyterLab instance jointly with a large set of pre-installed Python libraries and a ready-to-use Ophidia HPDA framework instance for running data manipulation, analysis and visualization. The ENES Data Space hosts (open) data from the ESGF federated data archive on compute cloud to support researchers in realistic climate model analysis experiments.'

Future planned activities:

- Strategies for distributed data management in the Cloud and infrastructure design (**GlobalCoast** initiative, **RI-SCALE** project)

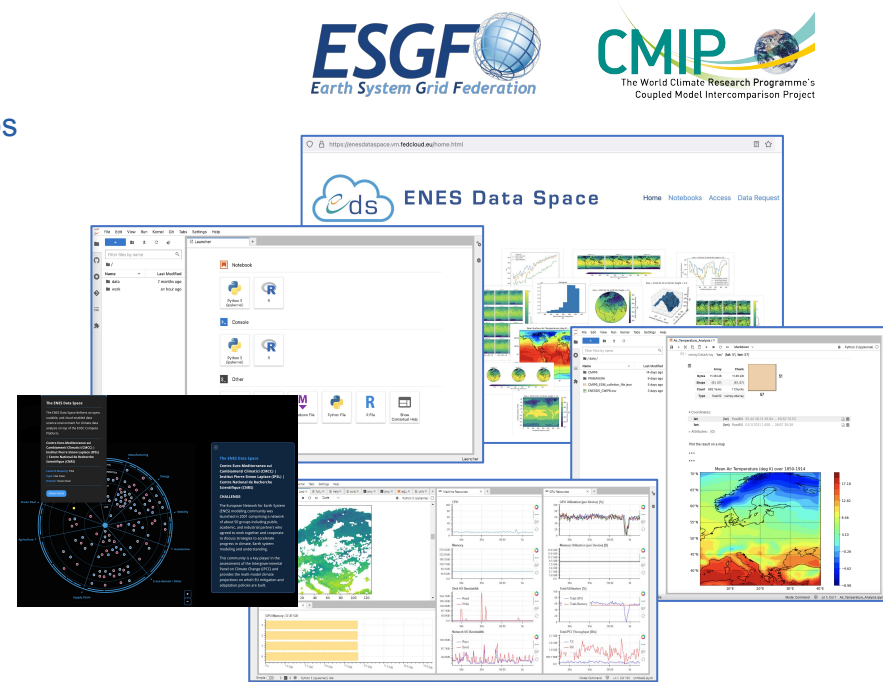
The «ENES Data Space»



- **Motivation:** Tackle key challenges and practical issues related to large-scale climate analysis
- **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**

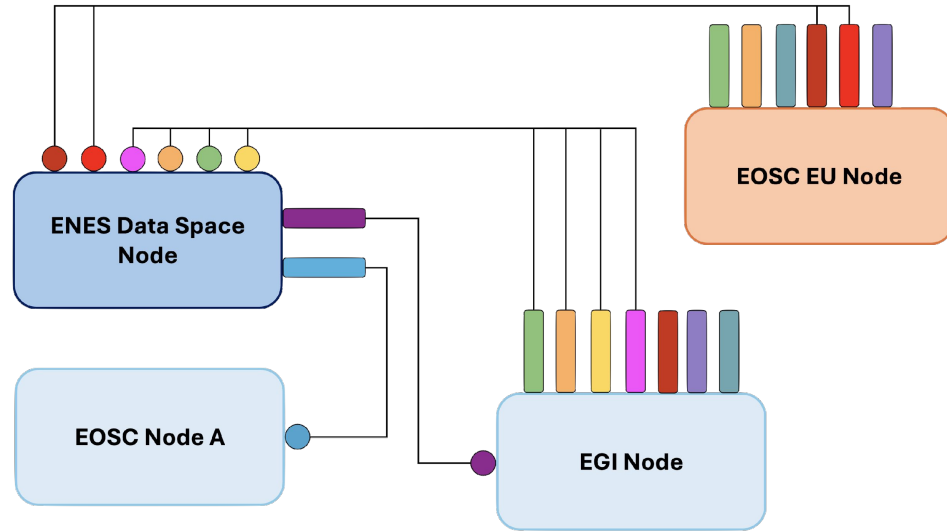
D. Elia et al., "A Data Space for Climate Science in the European Open Science Cloud", Comput Sci Eng. 2023; 25(1): 7–15. DOI: 10.1109/MCSE.2023.3274047



EOSC Beyond: advancing innovation and collaboration for research

- From the **EOSC Platform** to a **Network of EOSC Nodes**
- Nodes provide entry points for users to **access the full EOSC Federation**
- **ENES Data Space**, one of the pilot nodes in EOSC Beyond targeting climate community
- Integrate and validate the new EOSC Beyond **Next Generation Core services**
- Two new capabilities planned to be offered to the Federation:

- **Provenance service**
- **Thematic Data Catalogue**



Fabrizio Antonio et al.

ENES Data Space: an EOSC Beyond thematic node for the climate community
Carlo V, 2 Oct 2024, 15:15-15:30

yProv: a Cloud-enabled Service for Multi-level Provenance Management And Exploration in Climate Workflows
San Martino, 3 Oct 2024, 9:40-10:00

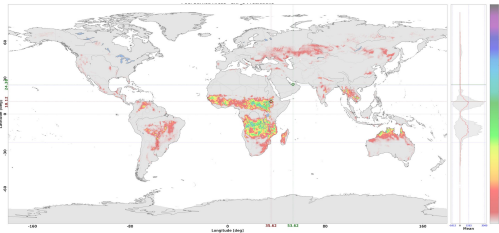
interTwin

interTwin aims at creating a **prototype** of a **Digital Twin Engine**:

- Based on a **DTE Blueprint Architecture** (interdisciplinary)

Partners from the **ENES community** (CMCC, CERFACS, IPSL and UNITN) involved in:

- Development of **thematic** and **core** components for **environmental DTs** on **extreme weather events** (droughts, wildfires, Tropical Cyclones)
- Exploiting **CMIP6 projection data**
- **Provenance tracking** in ML experiments

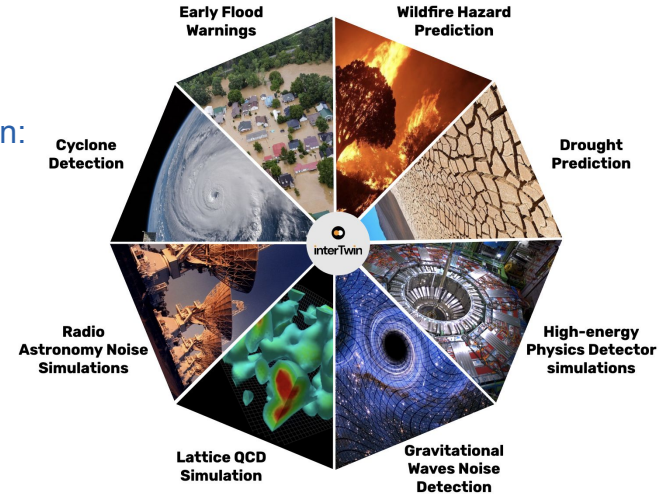


Emanuele Donno et al.

*ML4Fires: A Digital Twin Component for Wildfire Danger Analysis via
Global Burned Areas Prediction on Climate Projection Data*
Carlo V, 3 Oct 2024, 10:20-10:30



interTwin



<https://www.intertwin.eu/use-cases/>

Some of the key challenges:

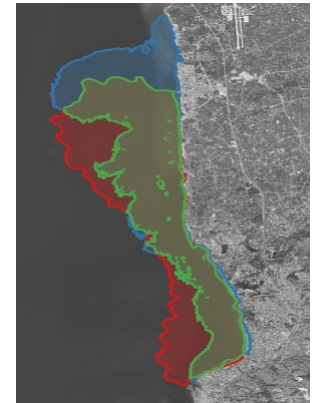
- **Workflows** for supporting **ML applications** and **data processing**
- Integration of **different data sources** (reanalysis, observations, simulation)
- Exploration and definition of **AI-based DT applications**

- **Goal:** provide a portfolio of **image datasets** and **AI-based** high-performance **image analysis tools** in aquatic sciences
 - Deliver a framework for AI model development, training, and deployment
- **Scientific use case:** **oil spill detection** through a data-driven approach
- **CMCC contribution:**
 - **enhance** the existing oil spill monitoring and forecasting system by establishing an operational service on the **iMagine platform**
 - improve the oil spill modelling service by assimilating **satellite observations**
 - **Bayesian optimization approach** for improving numerical model simulations

DEFAULT NUMERICAL
MODEL SIMULATION



ML-BASED NUMERICAL
MODEL SIMULATION



Marco Mariano De Carlo et al.

A Bayesian Optimization workflow for improving oil spill numerical simulations

Carlo V (Hilton Garden Inn), 3 Oct 2024, 09:50-10:00

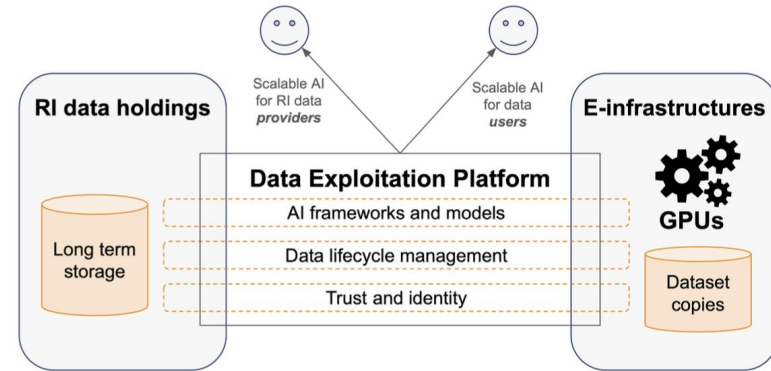


UNIVERSITÀ
DI TRENTO



What's next? RI-SCALE

- **RI-SCALE**: Unlocking RI potential with Scalable AI and Data
- It will deliver **Data Exploitation Platforms** (DEPs) and **scalable environments** to co-host scientific data with preconfigured **AI frameworks** and **models** on powerful compute resources
- **4 RIs** involved: **ENES**, EISCAT, BBMRI and Euro-Biolmaging
- Integration of **data** and **services** from the **ENES RI**:
 - Data from ESGF (e.g., CMIP6)
 - ESGF Data Statistics service
 - ESGF Search service
 - ENES representatives: CMCC, DKRZ, UKRI, UNITN
- Development of AI-based use cases for **environmental sciences**



Alessandra Nuzzo et al.

A data statistics service for data publication and usage metrics in the climate domain

Barocco, 1 Oct 2024, 16:00-16:15

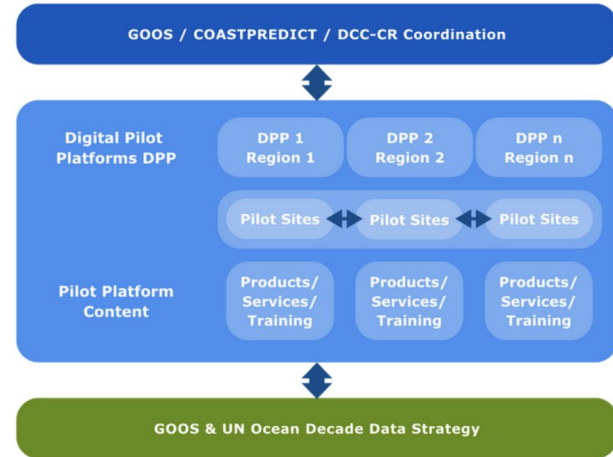
GlobalCoast: the CoastPredict Global Coastal Ocean Experiment

- **GlobalCoast**: a central framework for coordination and practical implementation of the **CoastPredict Programme**
- **Goal**: provide advanced **knowledge**, innovative **products** and **services** to support **coastal community resilience**
- More than **124 Pilot Sites** in **66 countries**
- **CMCC** and **EGI** have been collaborating to the technical design of the **cloud-based GlobalCoast infrastructure**

Nadia Pinardi, Giovanni Coppini

GlobalCoast Cloud: enabling equitable coastal resilience for the Future

Carlo V, 2 Oct 2024, 09:45-10:05



Conclusions

- **Climate change** is one of the major challenges of our time
- **ENES** provides an umbrella organisation for the **European climate modelling community** working on understanding and predicting **climate variability** and **change**
- The ENES community is facing **new challenges**
 - Increased **model complexity**
 - Ever-increasing **volumes of data** for addressing climate change societal challenges
- Use of **AI** is growing rapidly in climate applications
- Need for suitable and advanced **data** and **compute infrastructures**
- Collaboration between **ENES** and **EGI** is key to further integrate climate data services into the **European e-infrastructure** landscape



www.cmcc.it

