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

Final General Assembly

16-18 January 2023 Toulouse

ESGF: Main achievements for CMIP6, main evolutions and towards CMIP7

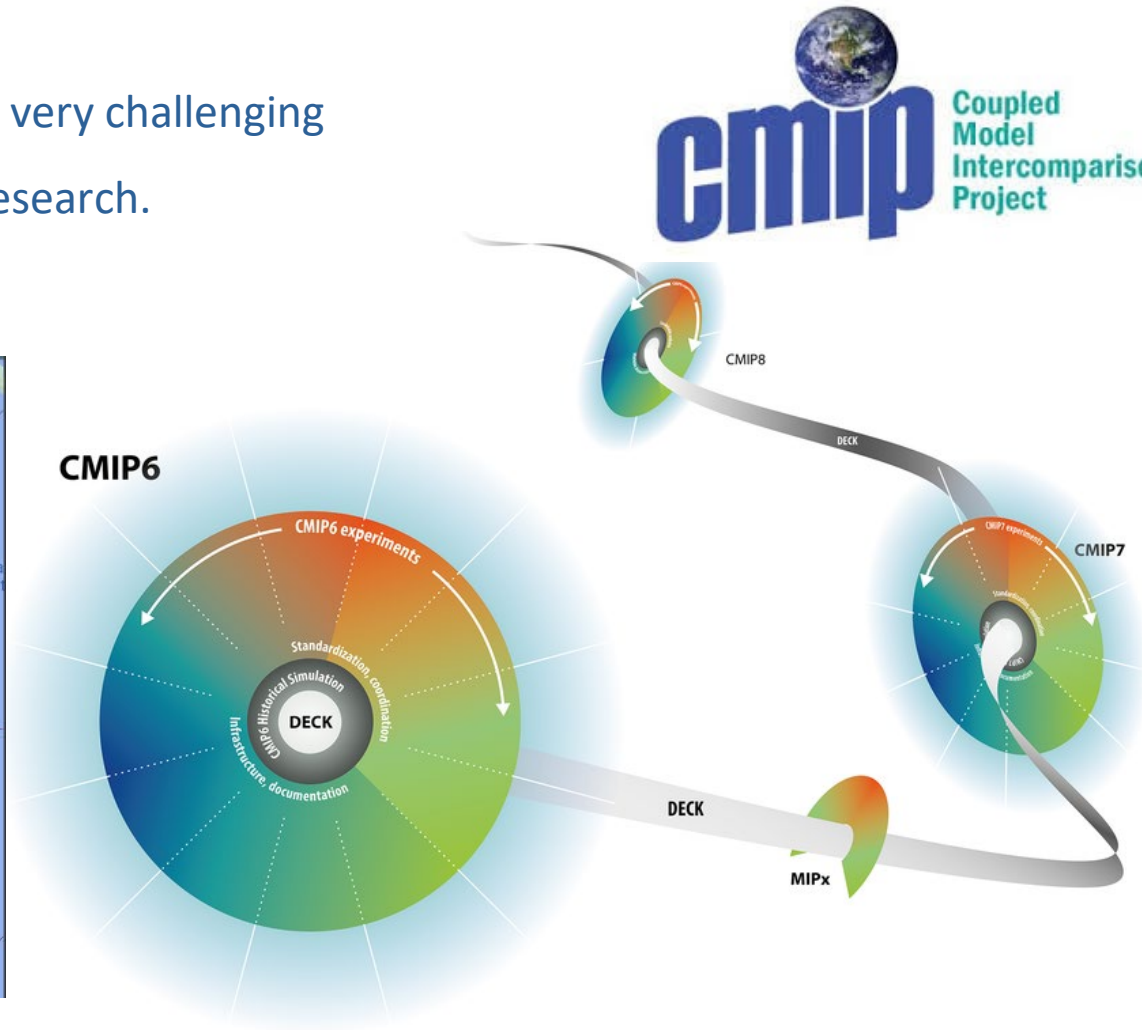
Philip Kershaw (CEDA)
Paola Nassisi (CMCC)

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

Large scale experiment: Coupled Model Intercomparison Project (CMIP)

The Coupled Model Intercomparison Project (CMIP) represents a very challenging and relevant large-scale global experiments for climate change research.

CMIP6 modeling institutes



The CMIP6 challenge

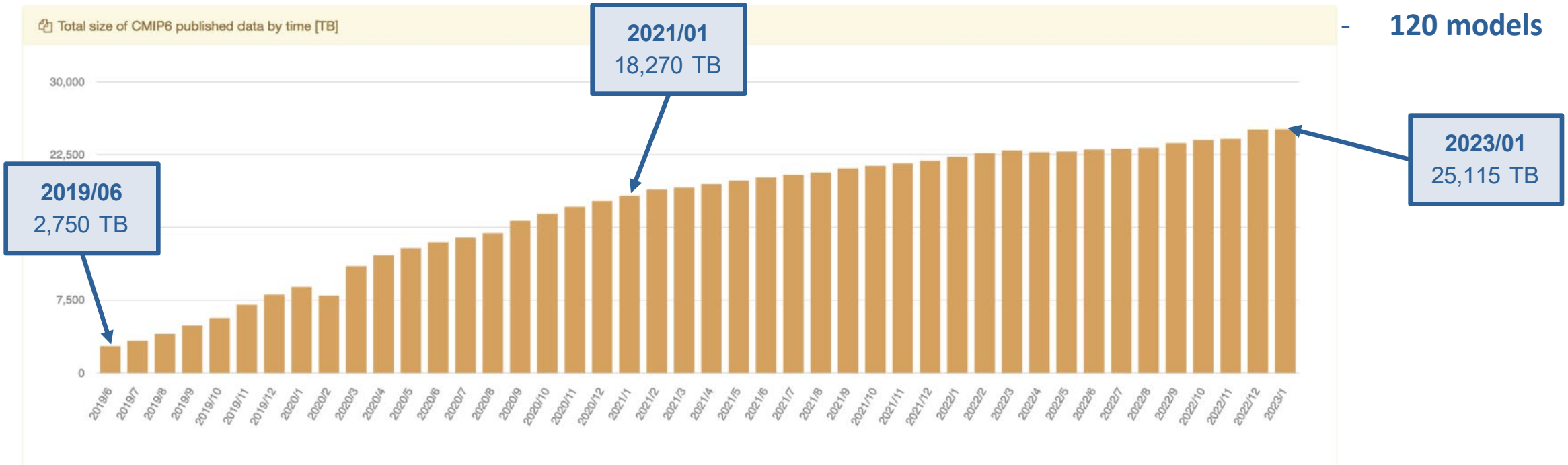
CMIP6 13,606,408 total datasets 25,212.14 TB

CMIP6 6,759,621 distinct datasets 14,295.37 TB

CMIP6 6,846,787 replica datasets 10,916.77 TB

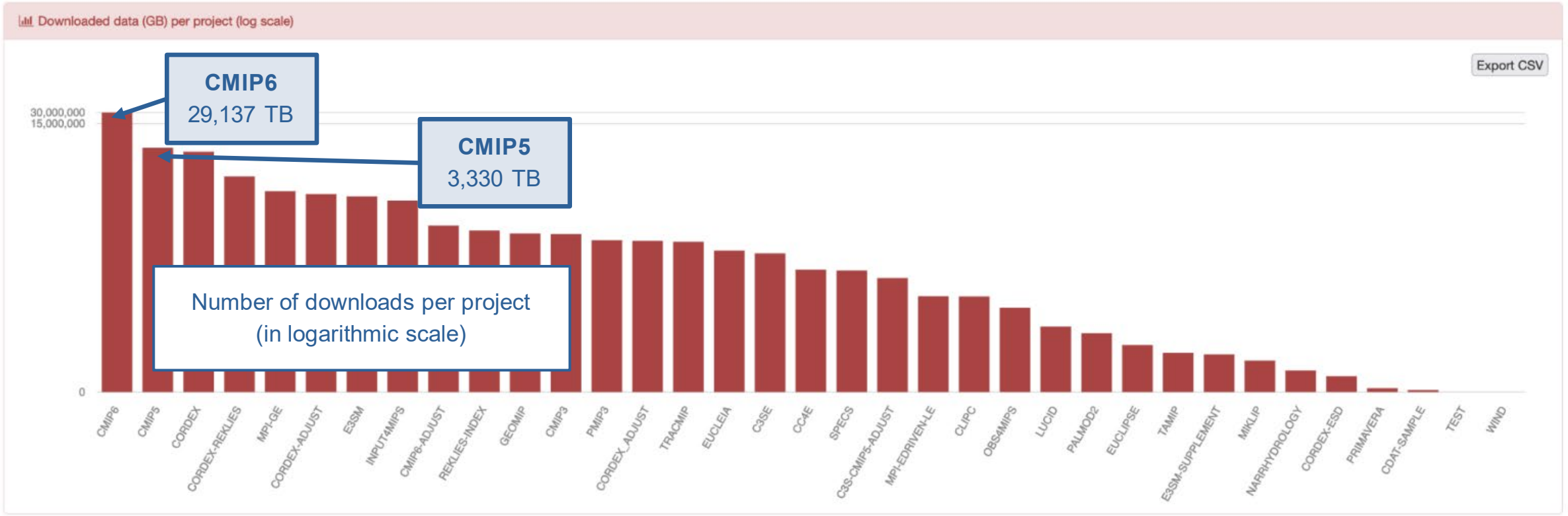
CMIP6 in figures:

- about **25 PB** of data
- more than **13 millions** of datasets
- **44 institutions**
- **120 models**



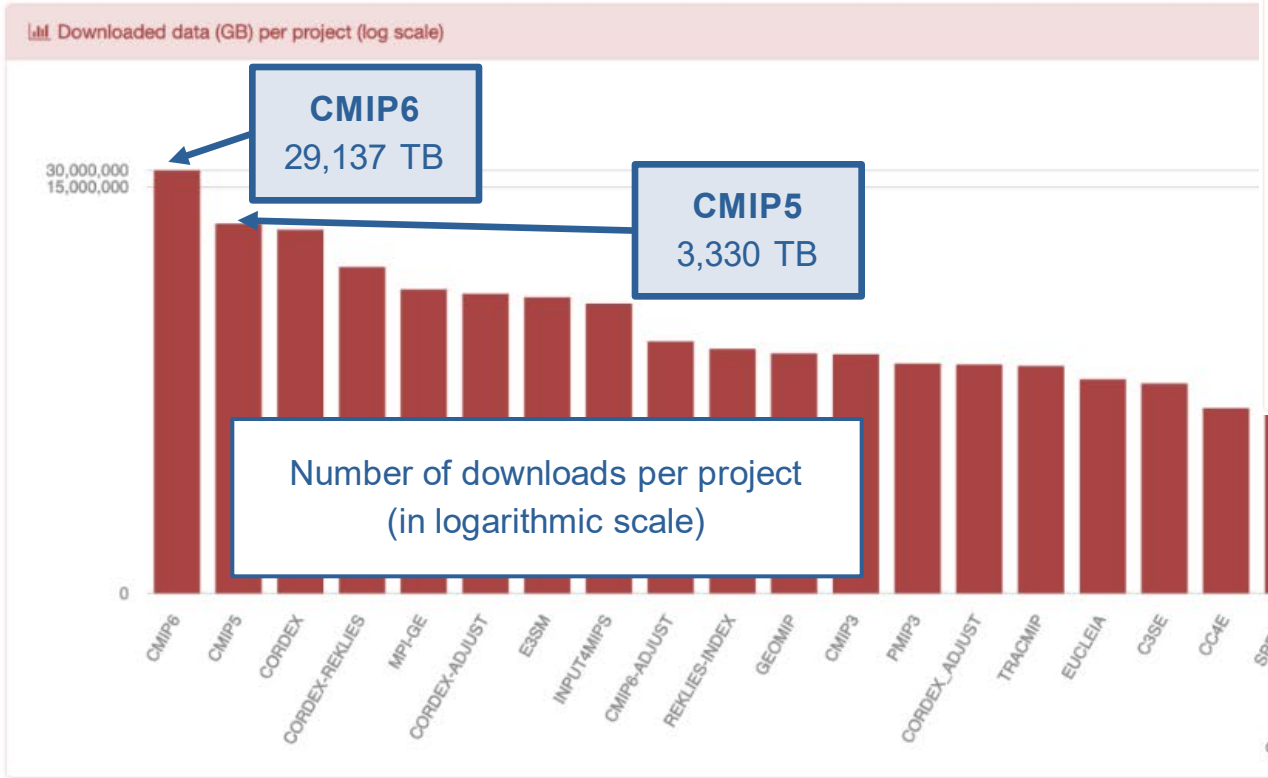
CMIP6 exploitation through ESGF

CMIP6 is the most downloaded project with almost 30 PB and over 1 billion files.

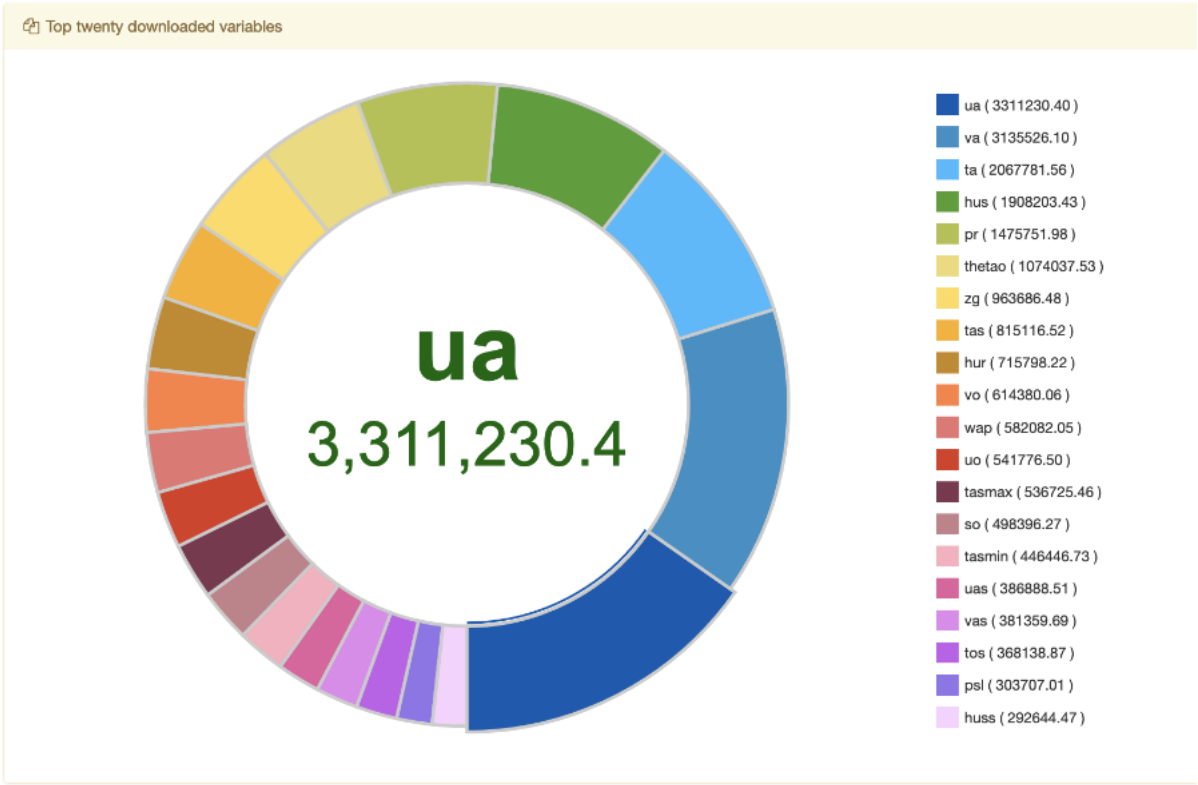


CMIP6 exploitation through ESGF

CMIP6 is the most downloaded project with almost 30 PB and over than 1 billion files.

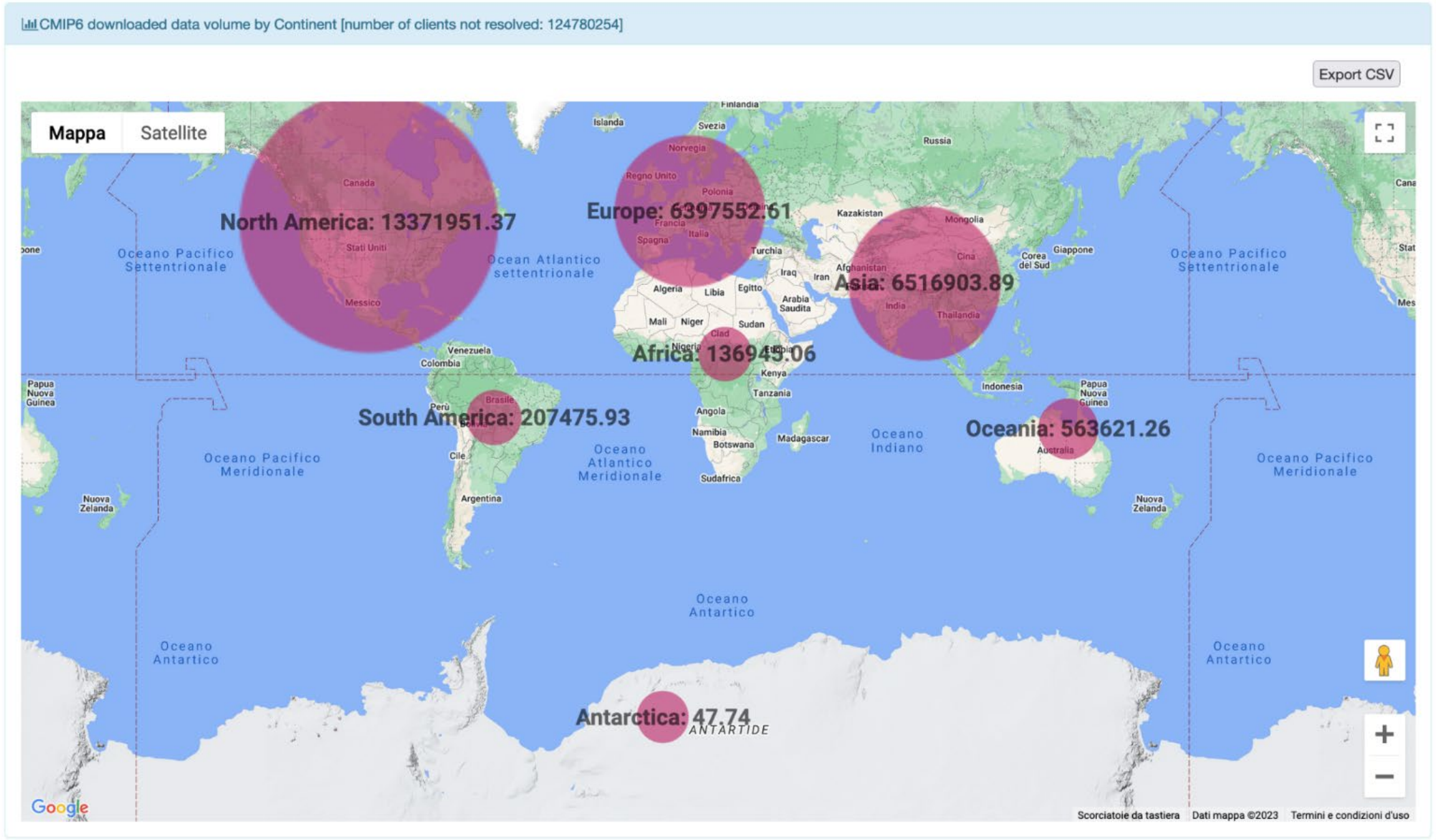


Number of downloads per project (in logarithmic scale)

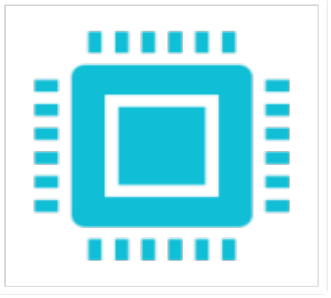


The CMIP6 most downloaded variable is the **Eastward Wind (ua) (3 PB),** followed by **Northward Wind (va)** and **Air Temperature (ta)**.

CMIP6 downloads geolocation



ESGF: Evolution and towards CMIP7

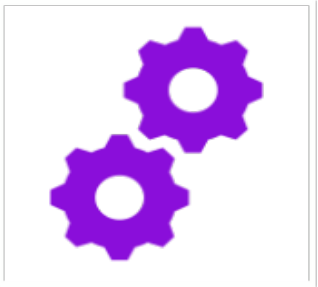


Installation and Systems Administration

Search Services

Identity and Access Management

New modes for data access ands storage



Compute Services

Metrics Collection

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

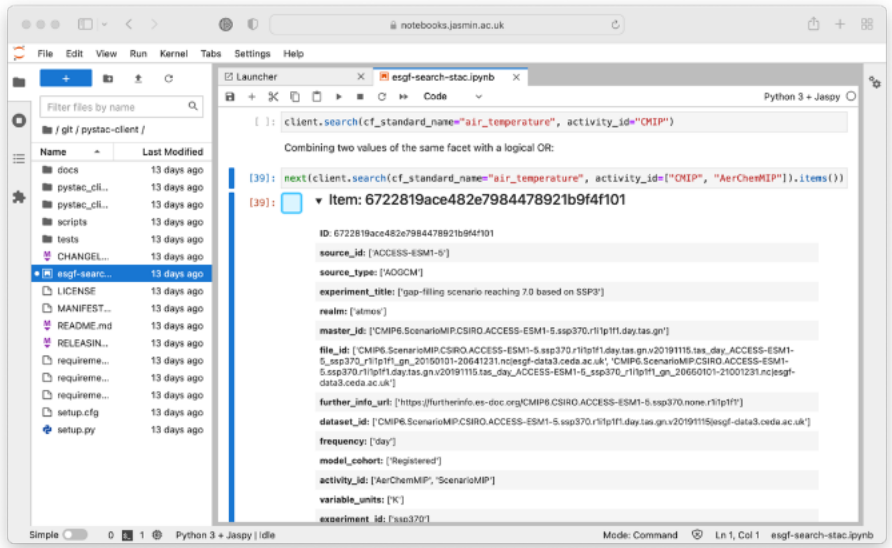
Installation and Systems Administration: Container -based Deployments

- Modular, easy to deploy and maintain: infrastructure-as-code approach

Institution	Deployment option	Status
CEDA (esgf.ceda.ac.uk)	Kubernetes	Production
GFDL	Kubernetes (AWS)	Production
DKRZ	Ansible Docker	Pre-production
LLNL	Kubernetes	Pre-production
ORNL	Kubernetes	Pre-production
NCI	Ansible Docker	Pre-production

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

Search Service: STAC and ElasticSearch



- **Components** - recap
  - STAC as community standard API
  - ElasticSearch as a scalable NoSQL database backend

- **Extensions to STAC Search API** - created:
  - Free-text query, asset search
  - Discover "queryables" based on current query - i.e. facet keys/values e.g. CMIP6:source_id = "HadGEM3-GC31-MM"

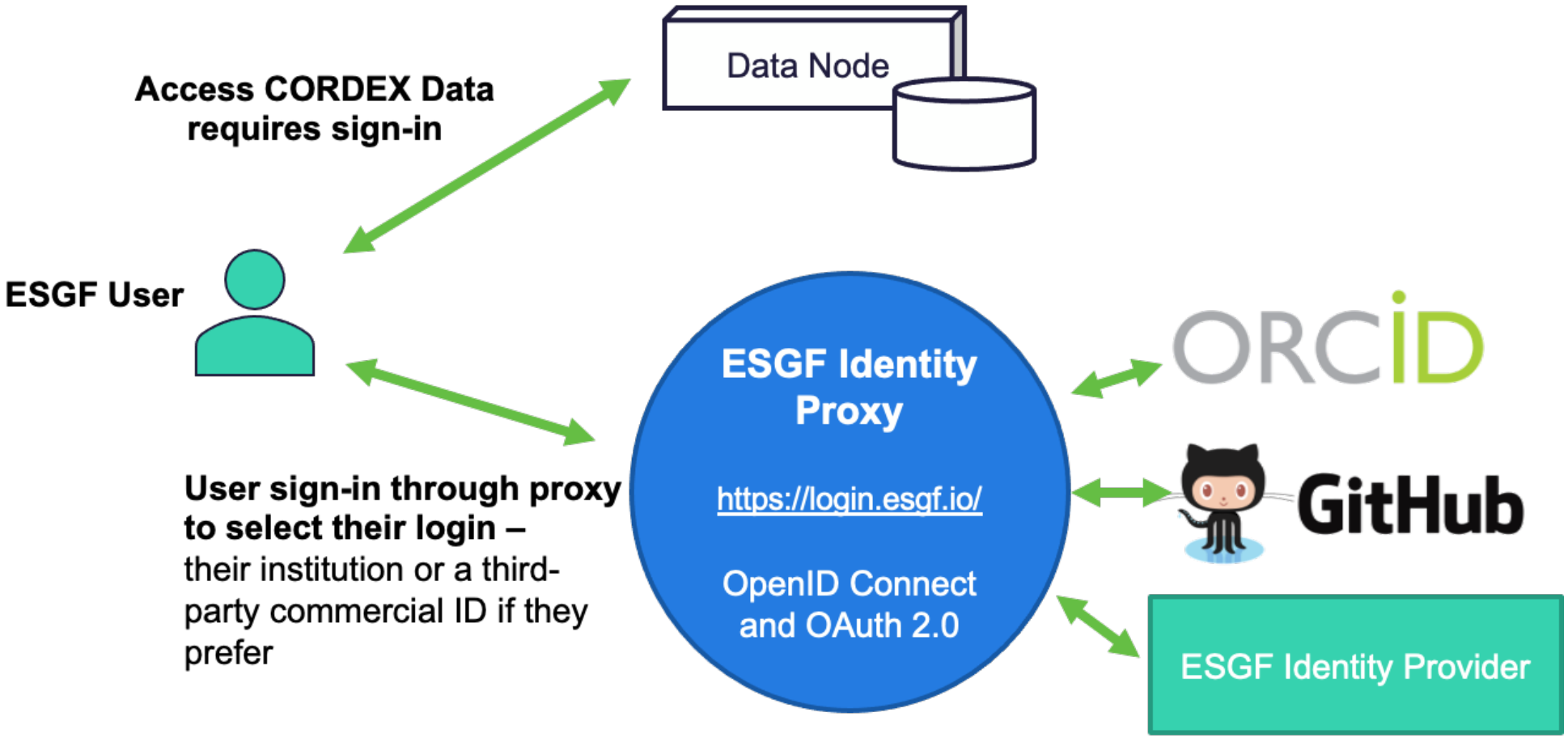
- **Client** - Extended Python STAC library ("pystac" and "pystac-client") + created a lightweight wrapper library:
  - https://github.com/cedadev/esgf-stac-client

- **Facets/Controlled Vocabularies** - created a framework for term mapping:
  - Allows search using general terms (e.g. "model") and project-specific terms (e.g. "source_id")

- **Catalogue generation** - proof-of-concept indexed:
  - Method: ingested Solr records and mapped to STAC content
  - CMIP6 datasets as STAC Items (all hosted at CEDA): 677,813
  - CMIP6 files as STAC Assets (all hosted at CEDA): 6,137,991

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

Identity and Access Management: Single Sign-On



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

Identity and Access Management: Single Sign-On

- Identity Provider deployed at https://login.esgf.io/
- … + integrated with new ESGF web frontend MetaGrid

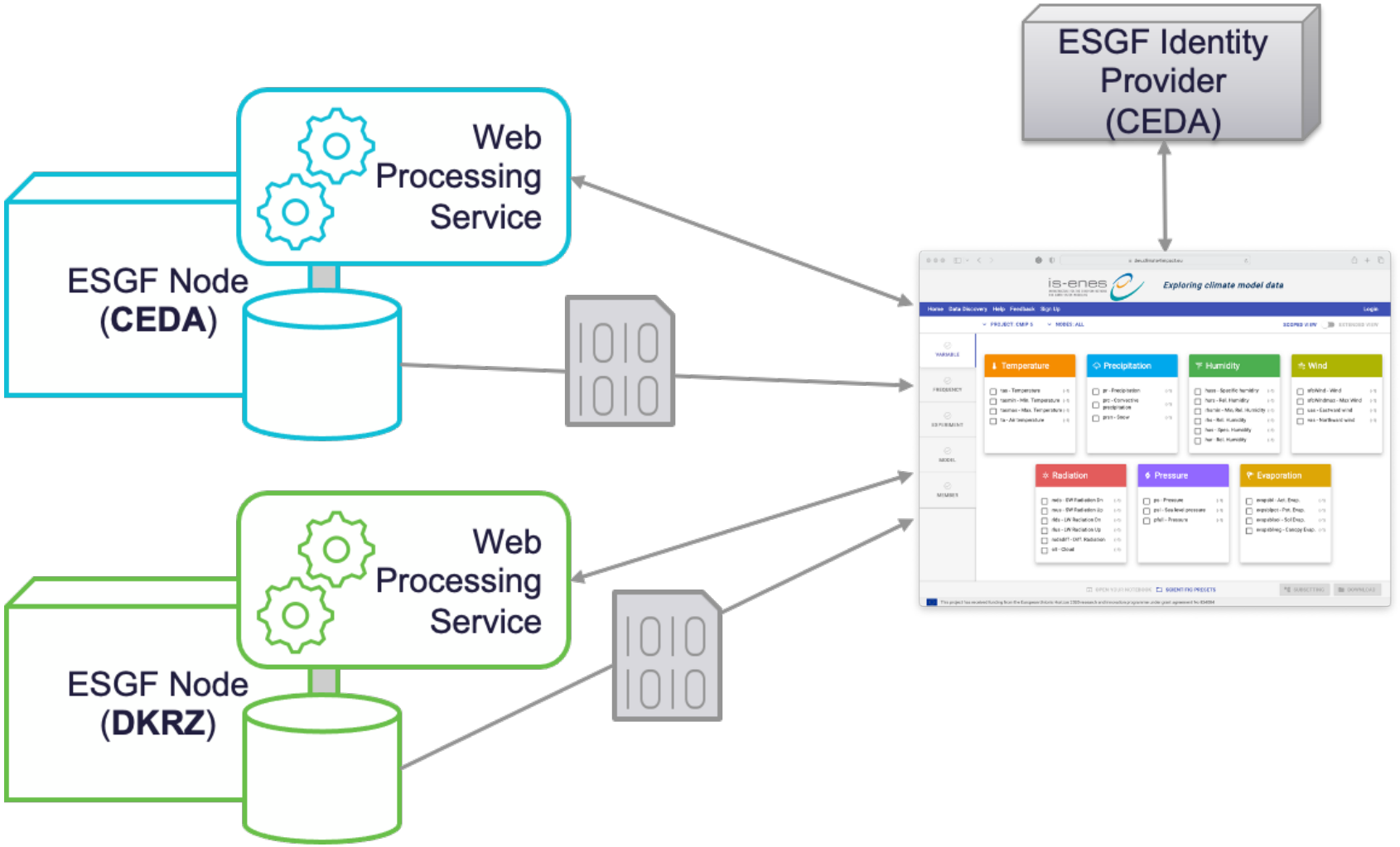
IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Identity and Access Management: Authorisation

- New Authorisation system being completed for testing CORDEX data access with Climate4Impact
  - Registration web UI - https://login.esgf.io/registration/
  - Authorisation system at CEDA using new OPA (Open Policy Agent system)

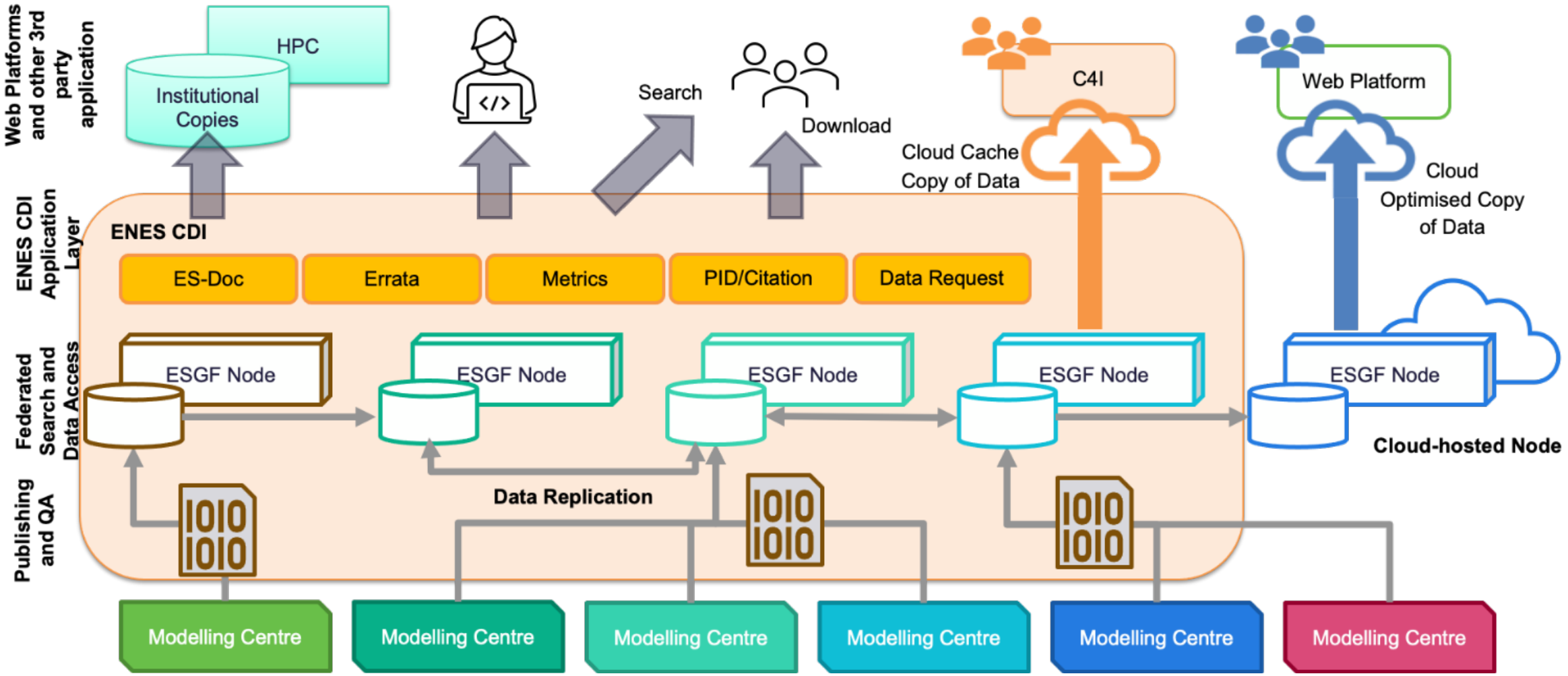
IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Compute Services



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

ESGF and the ENES RI: towards CMIP7



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

ESGF and the ENES RI: towards CMIP7

- Priority: complete replacements to re-establish baseline functionality
  - Agree ‘macro’ architecture pattern for search and identity management - centralised services in US and Europe that sync with one another
  - Complete new search service

- Compute and analytics
  - Establish operational baseline server-side compute
  - Data - interfaces - POSIX/S3, Kerchunk
  - Data - tiering - ESGF-level APIs for retrieval from cold or warm storage

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

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

THE CONSORTIUM

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

Our website
https://is.enes.org/

Follow us on Twitter !
@ISENES_RI

Contact us at
is-enes@ipsl.fr

YouTube
Follow our channel
IS-ENES3 H2020