



French Data Analysis Platform for ENES (aka. ESPRI)

Levavasseur G.

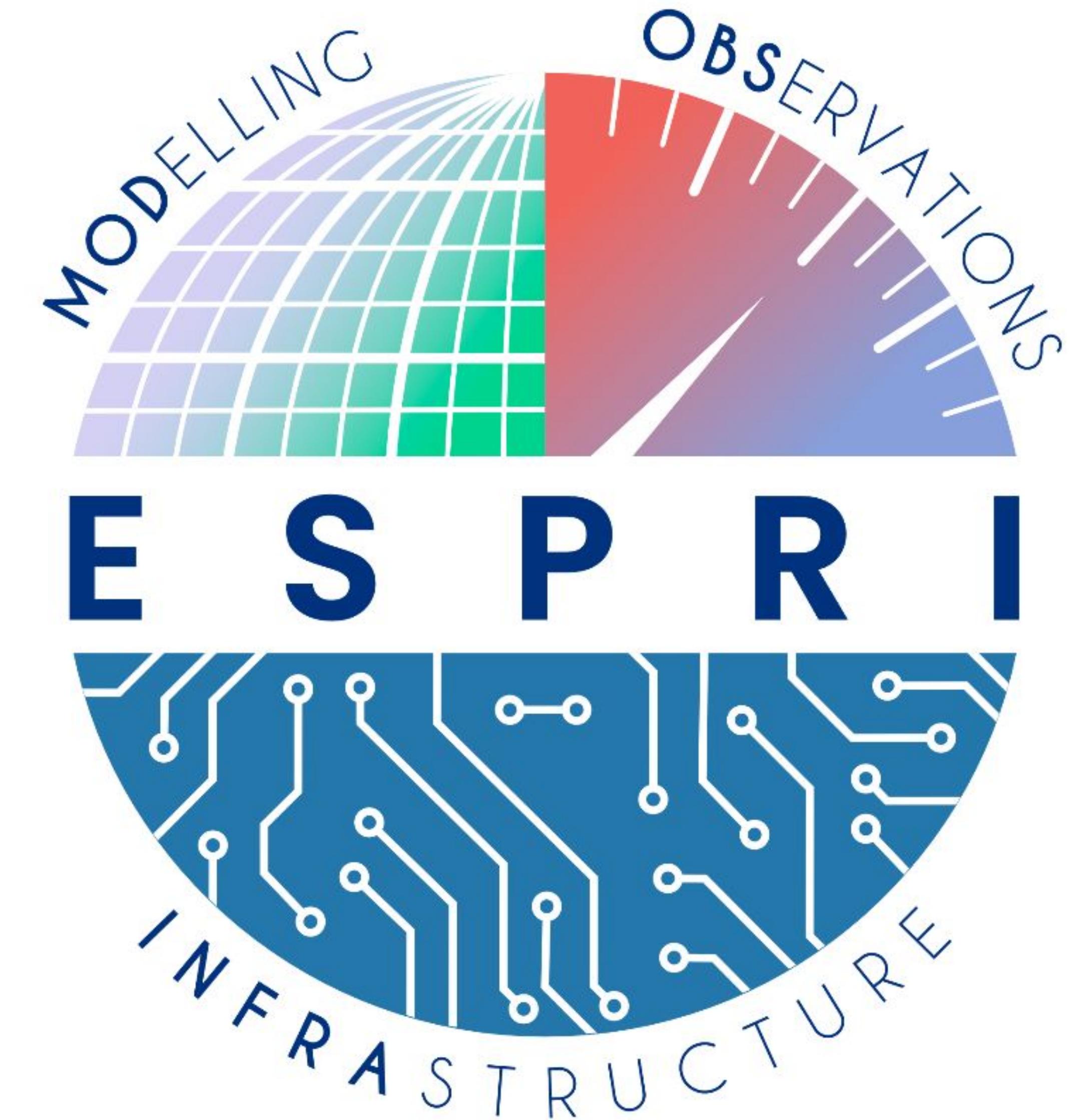
What's “ESPRI”?



The Institut Pierre-Simon Laplace provides its laboratories with coordination resources and services to help develop major projects and disseminate the results.

The **IPSL Computing and Data Centre** was born out from several initiatives at the IPSL and its laboratories, wishing to share IT resources and joint projects through a numerical facility for research.

For more than 20 years, the IPSL computing and data centre has been providing **Ensemble of Services for (tr. “Pour”) Research** at the IPSL – **ESPRI**.



Community infrastructure



IPSL laboratories



INSU partners



OSU Ecce Terra



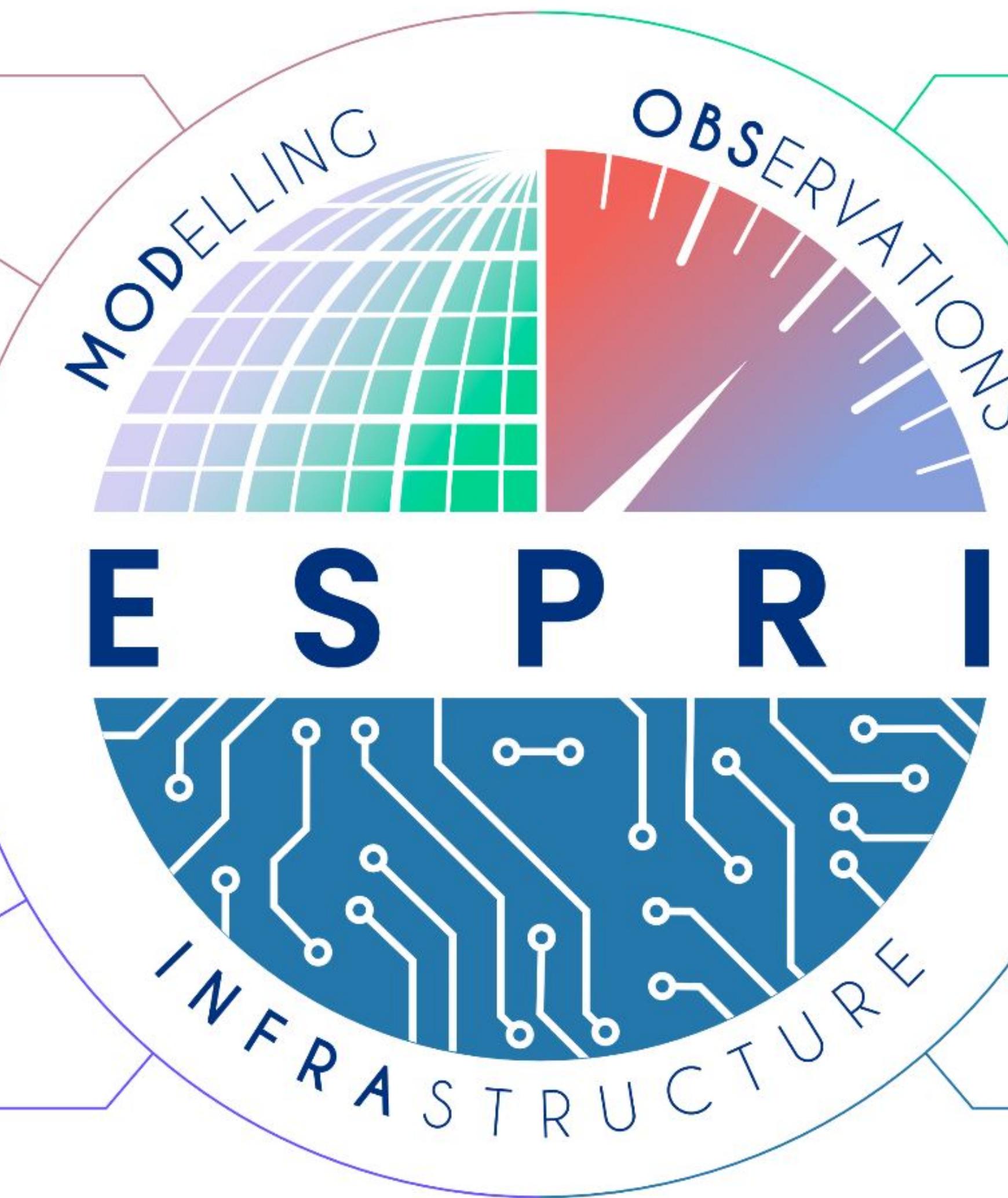
Data Terra



ClimERI-France



GAIA-DATA



Levels: ● Regional ● National ● European ● International



ENES



ACTRIS



Copernicus



Horizon Europe



ESGF



RDA-France

ESPRI-Mod community infrastructure

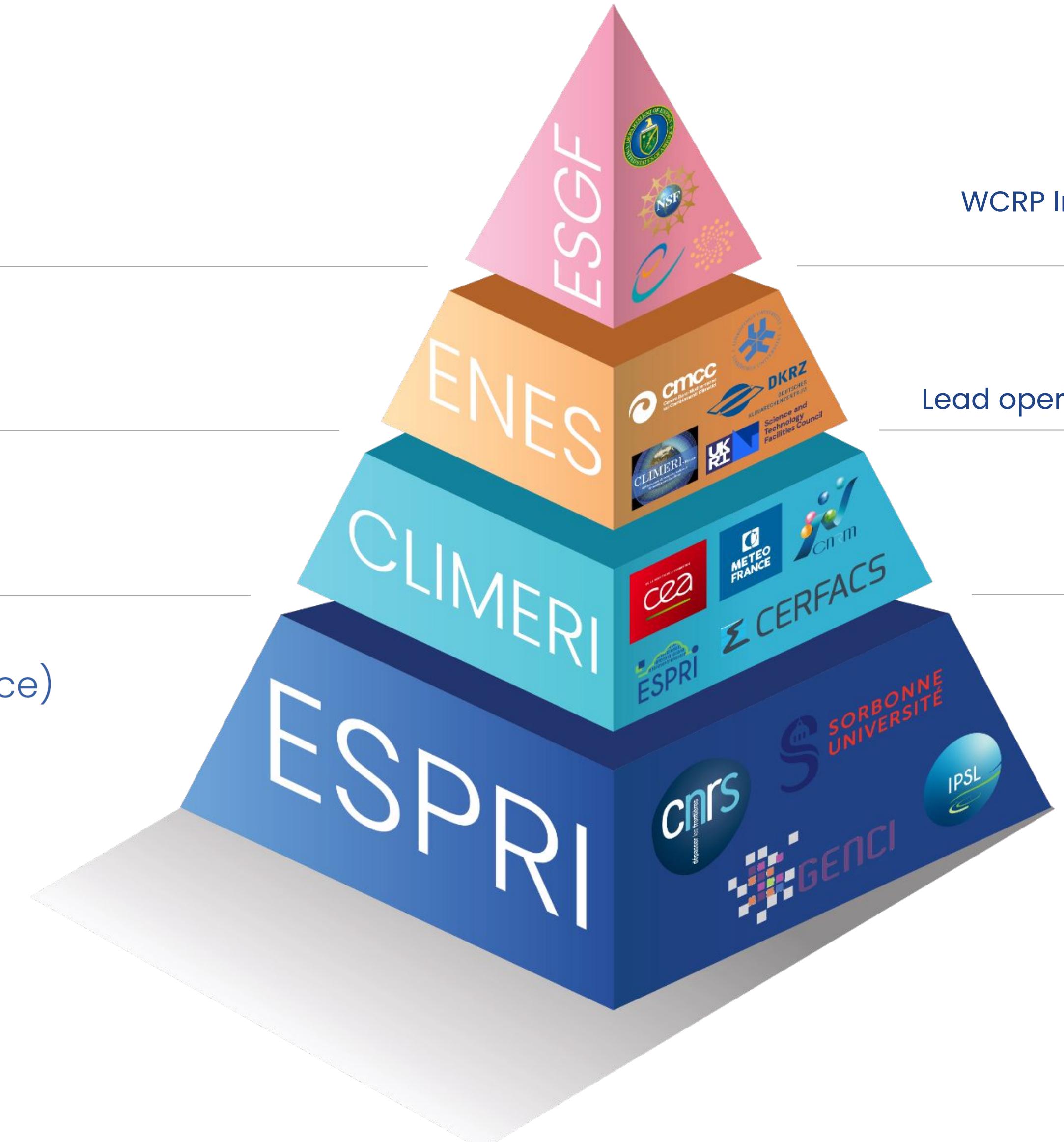


International

European

National (France)

Regional (Ile-de-France)



Largest contributor ($> 1,2M$ datasets $\approx 1.7\text{PB}$)

Tier 1 ESGF server

WCRP Infrastructure Panel contributor (Task Team co-leader)

ES-DOC governance member (Principal Investigator)

Coordination of the ENES infrastructure

Feeding national needs into large-scale projects

Maintain and consolidate the ESGF

Lead operator of the climate data archive for the C3S and EOSC

Supporting the French community

Coordination with HPC partners

Replication & multi-model caching

Response to climate services

Distribution of French climate simulations

Analysis & software environments

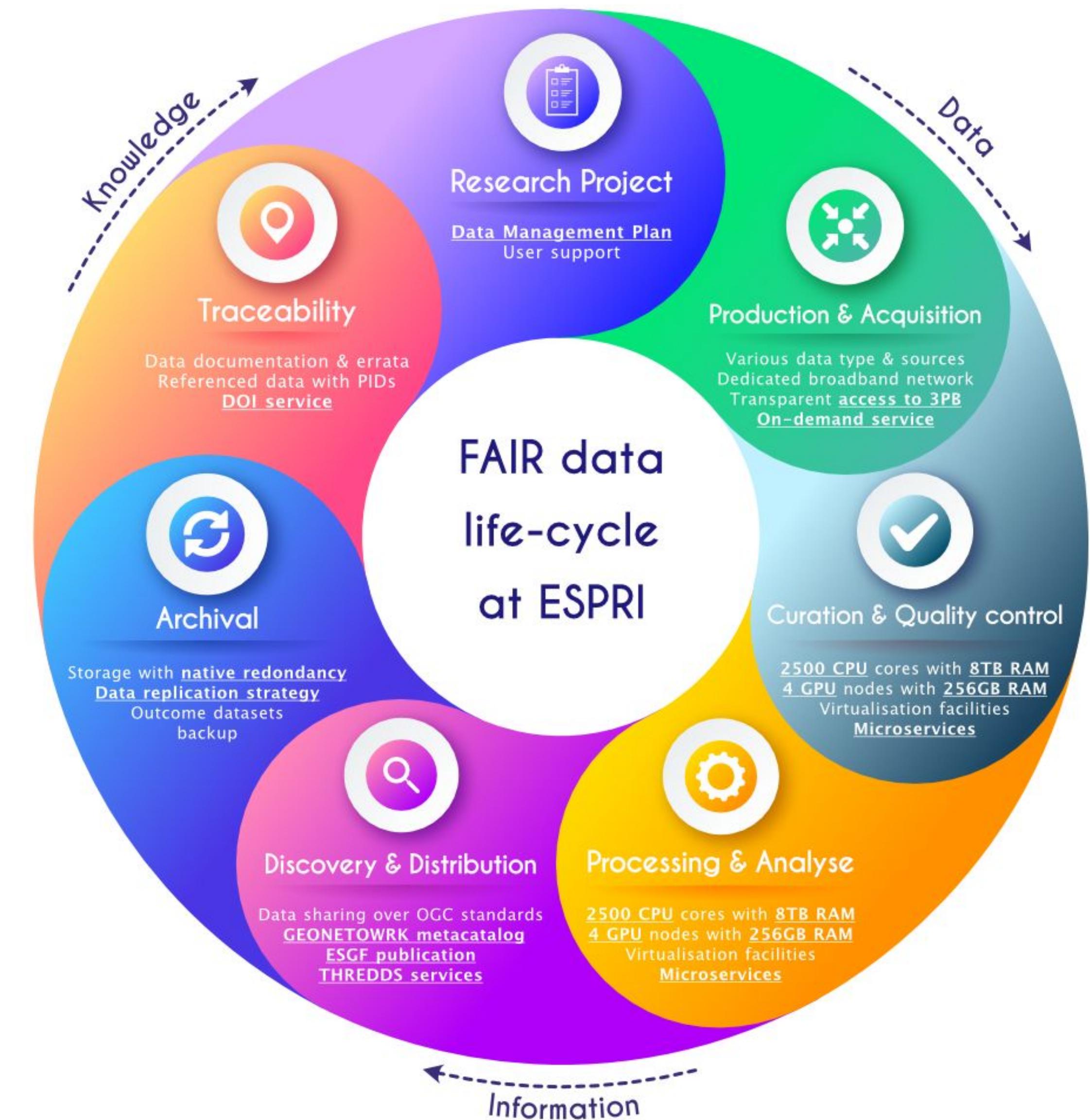
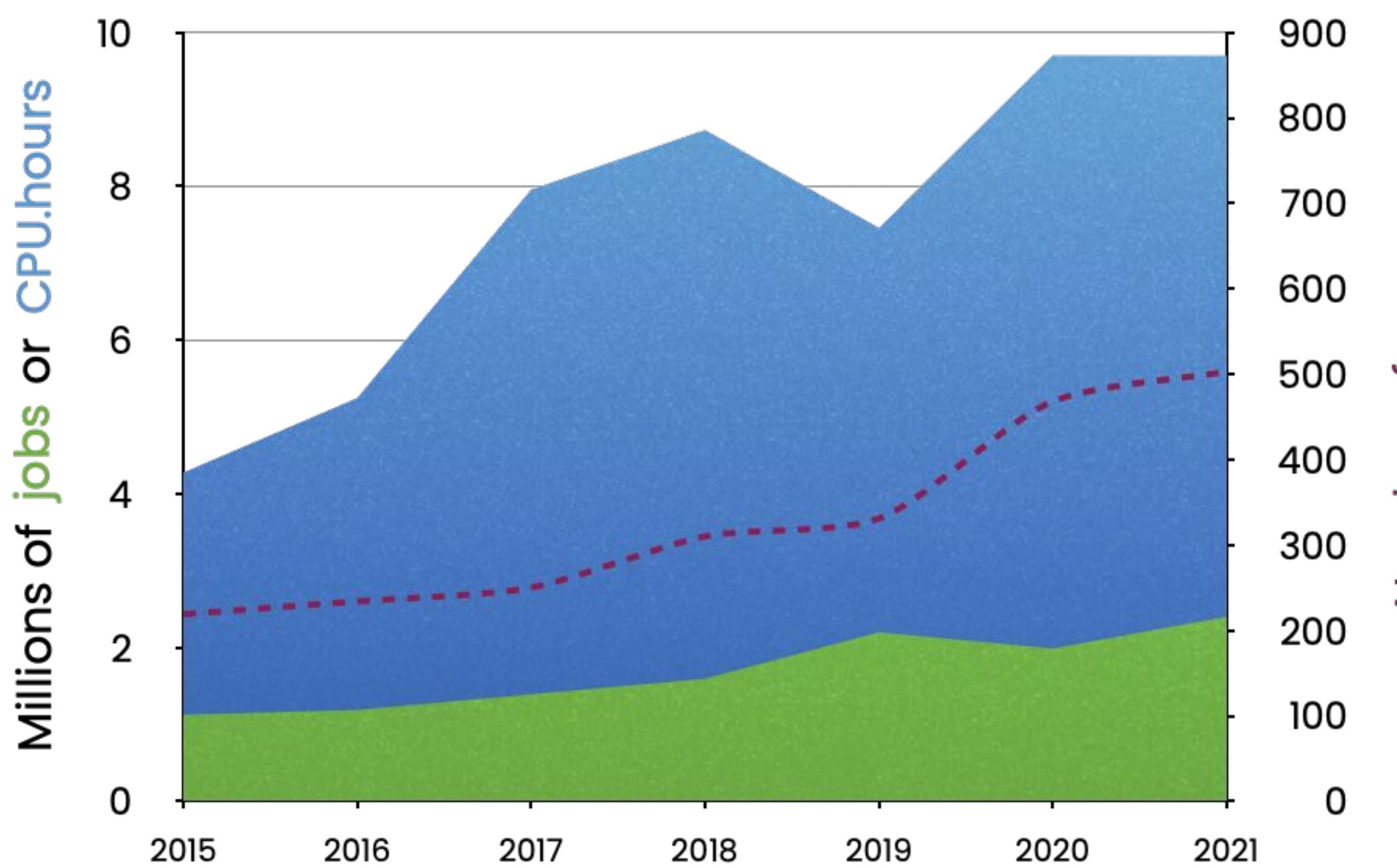
Processing workflow (bias correction, etc.)

Documentation & traceability of simulations

Mission & key figures



ESPRI aims to provide a distributed numerical infrastructure for research at IPSL with "FAIR" (**F**indability, **A**ccessibility, **I**nteroperability, **R**eusability) data services.





Free access

- IPSL community
- IPSL partners from national IR
- Non-EU partners through ENES VA calls

Other access:

- Project funded access
- Quote on request according to the ESPRI pricing

Registration & recommended contact to get started

- <http://mesocentre.ipsl.fr>
- Guillaume.Levavasseur@ipsl.fr

Login through SSH

Expertise and support:

- Data analysis (discovery, access, dissemination, etc.)
- ML/DL (through ESPRI-IA)
- Bias correction





Research environment

Computing

2 500 CPU cores with 8TB shared RAM

36 GPU cores

Data archives

~3Po of various climate data

Dedicated high bandwidth access

Storage: 20GB on /home + 1TB on /data

Compilers: GNU, Intel, etc.

Librairies: HDF5/netCDF4/openMPI, etc.

Products: Python2/3, CDO/NCO, Ferret/GraDs, Matlab,
ncl, R, SciLab, Git/SVN

Shared python environments: ClimAF, "PANGEOT-like"

On-demand data sharing

Group shared space

THREDDS shared space with OGC protocols

<http://thredds-su.ipsl.fr>

GridFTP

DOI/PIDs



SPIRIT cluster - Sorbonne Université (Paris)



SPIRITX cluster - Polytechnique (Palaiseau)



Earth observations

Soil & in-situ data

Campaigns (CAL/VAL MT, balloons, etc.)

Airborne compounds from 17 stations

SIRTA measurements

Radio soundings (ARSA, TIGR)

Satellite

Level 1 to 4 products (POLDER, PARASOL, CFMIP-obs)

Numerical outputs (INDOEX, AMMA, HyMex ChArMEx)

Reanalyses

ERA

MERCATOR-OCEAN

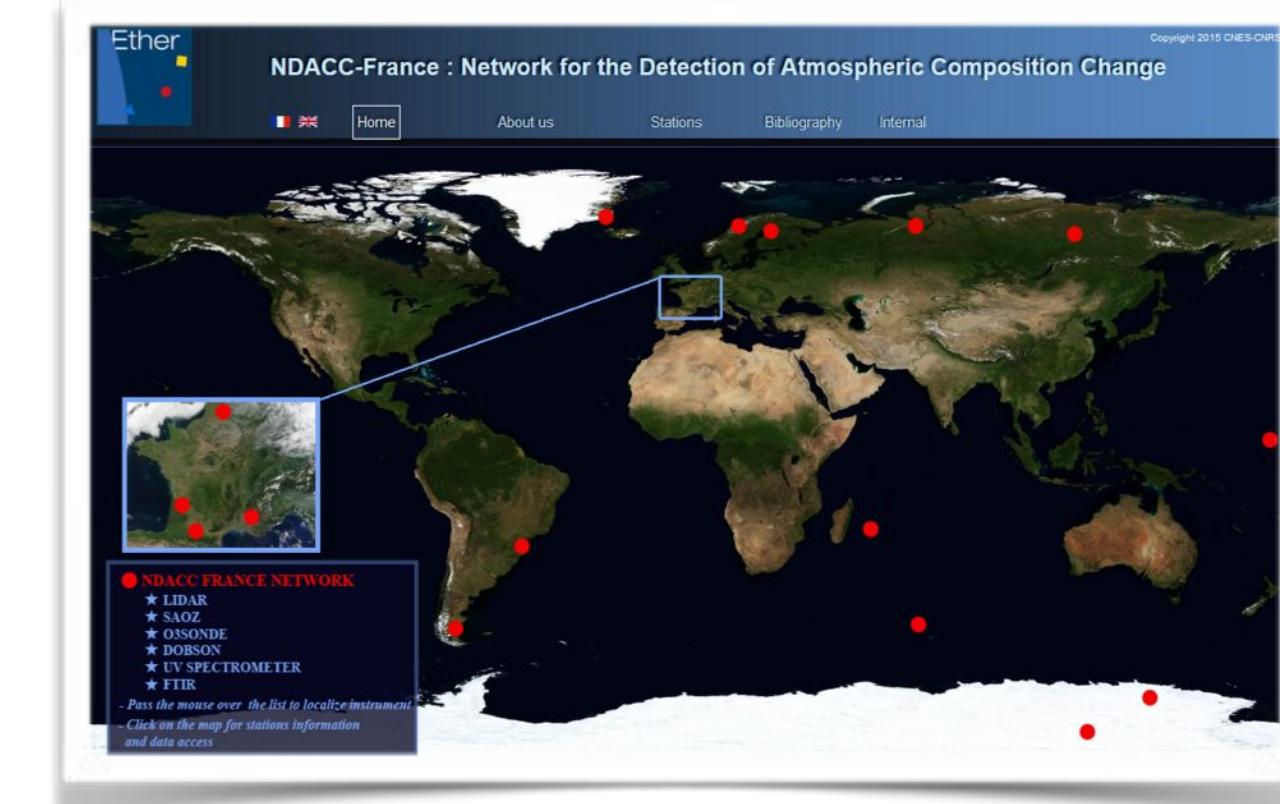
NCEP

FCDR (AMSI, SSMI, GridSat)

Atmospheric model outputs

50 airborne compounds from REPROBUS

Potential vorticity and temperature from MIMOSA



Satellite data

IASI level 1C (METOP-A-B)

IASI level 2 (O3, CO, SO2, CH4, HCOOH, NH3)

AMSUA-MHS-HIRS4 level 1C (METOP-A-B)

GOME2 level 1B (METOP-A-B)

GOME2 level 2 (METOP-A-B)

GOSAT level 1B / FTS/CAI

GOSAT level 2 / FTS/CAI

SAGE II, UARS, SPOT3, SPOT4, ODIN, ENVISAT



Climate simulations

Coupled Model Intercomparison Project (CMIP)

CMIP3, CMIP5, CMIP6

Tied projects (PMIP3, GeoMIP, etc.)

With bias correction

Coordinated Regional Climate Downscaling Experiment (CORDEX)

Several geographical domains

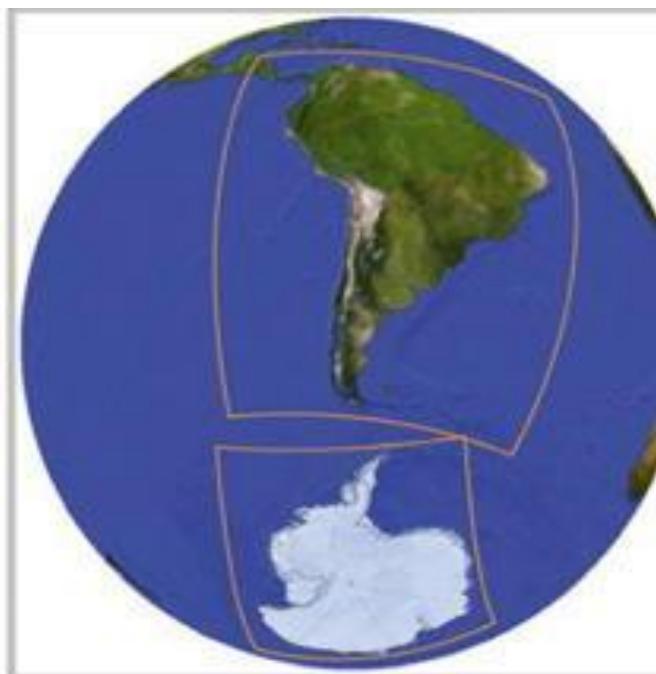
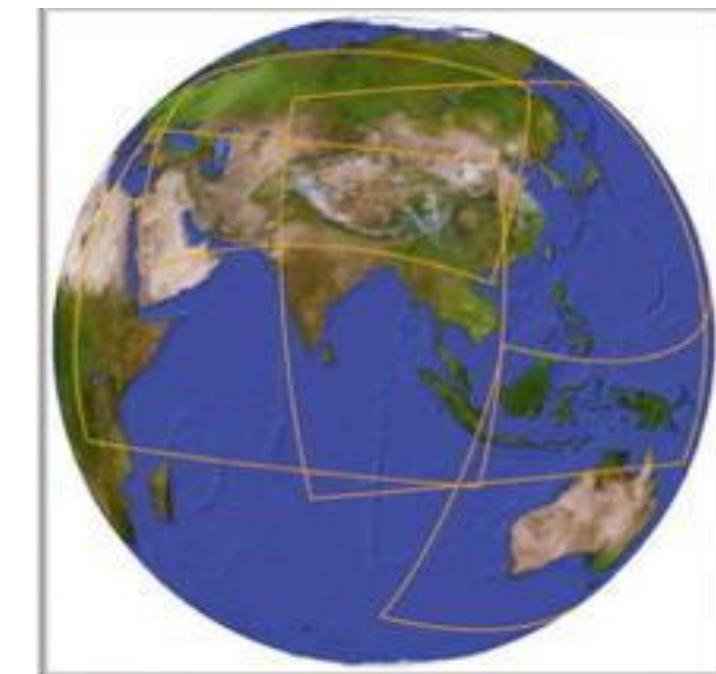
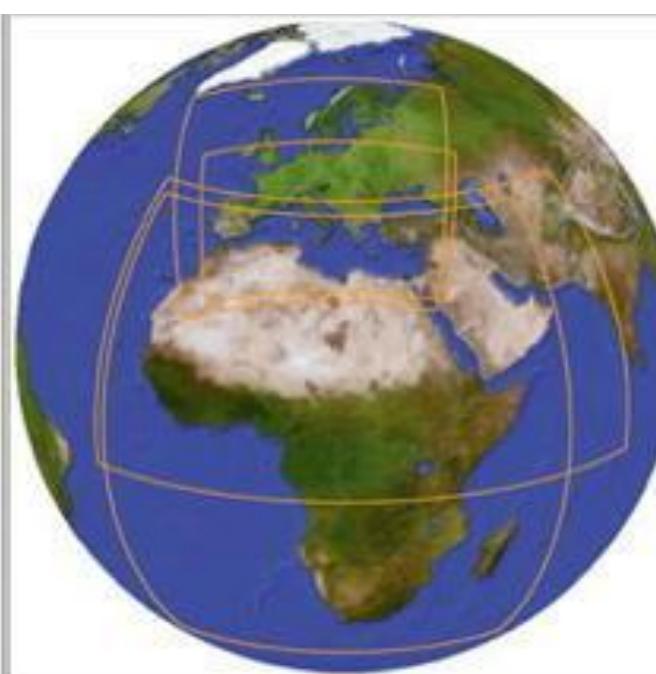
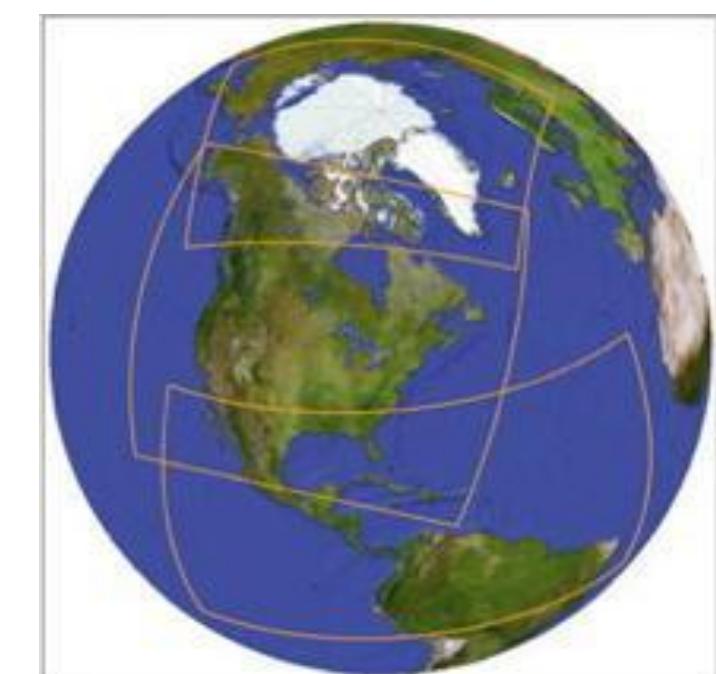
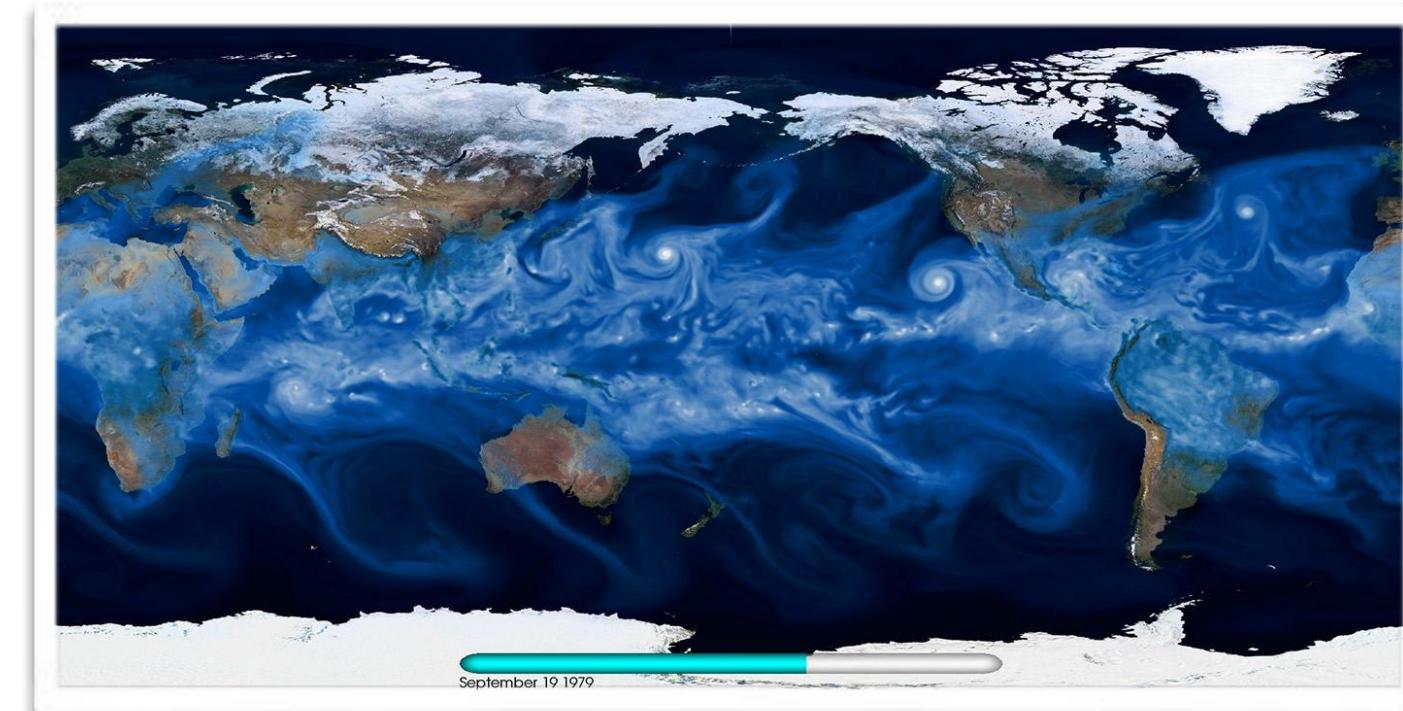
Several RCMs

With bias correction

Observations for Model Intercomparison Project (obs4MIP)

Input Datasets for Model Intercomparison Project (input4MIP)

Climate projections for Copernicus Climate Data Store (including IPCC Climate Atlas)



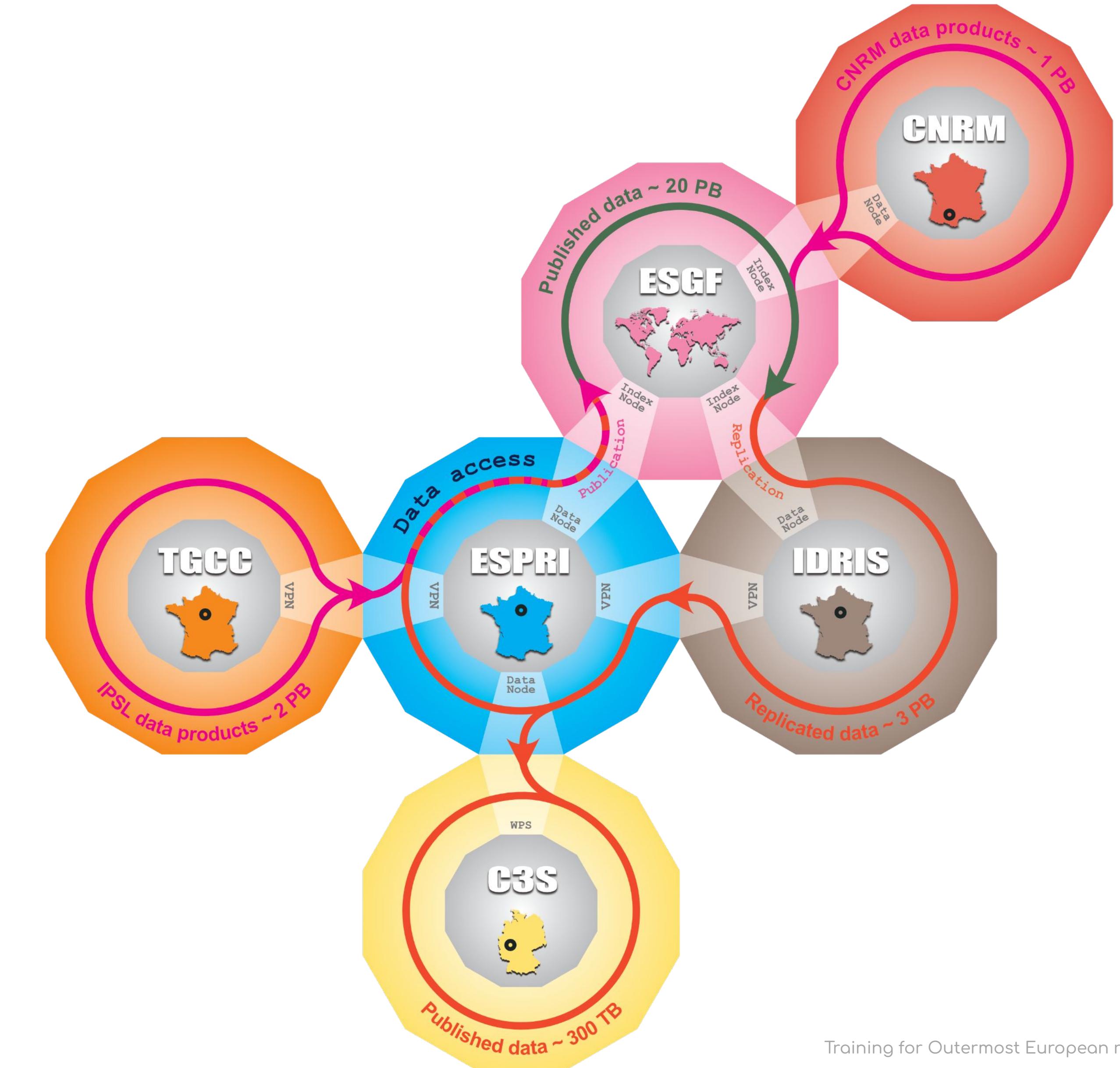
ESPRI centralizes ClimERI-France data archives



All IPSL-CM production (CMIP5/6 + CORDEX)

Replica pools:

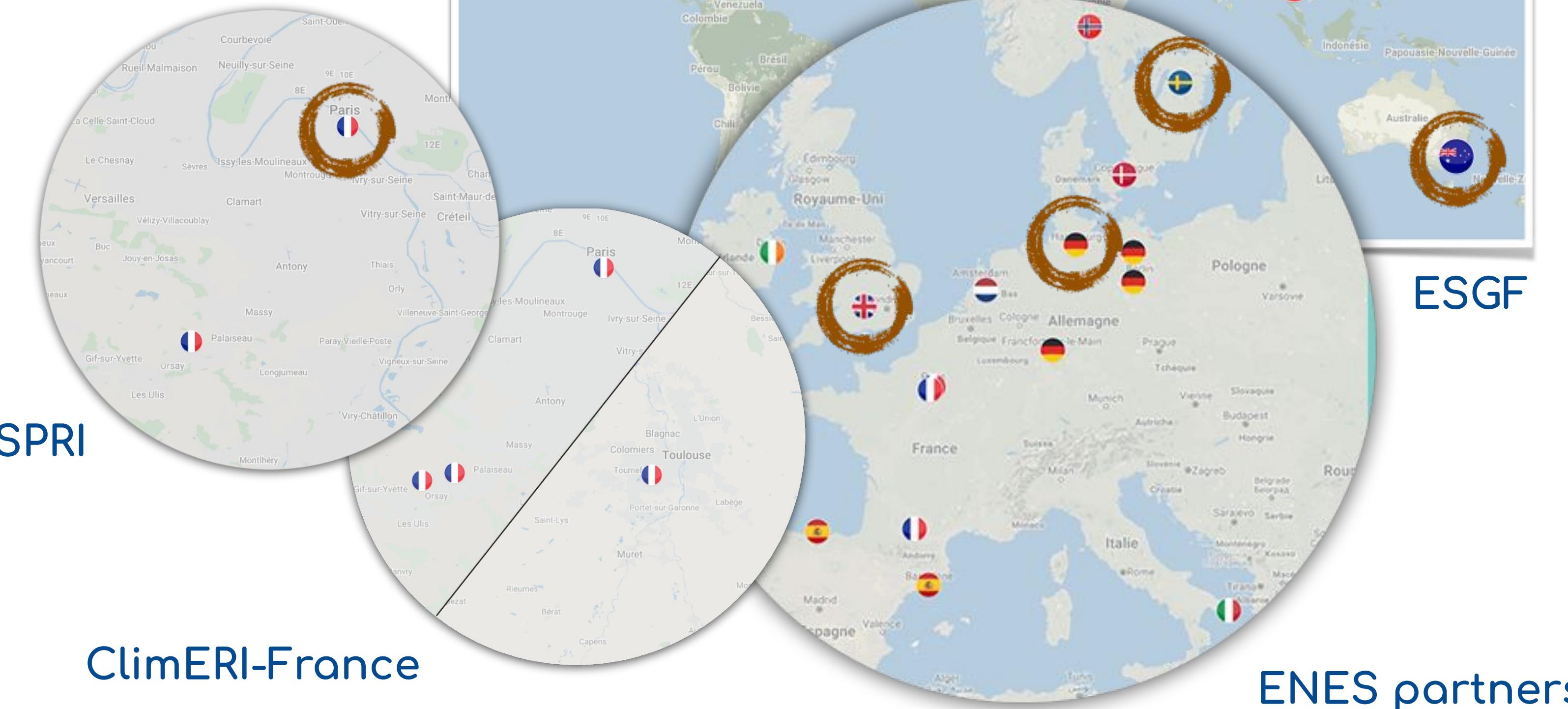
Project	Thousands of files	Volume
CMIP6	4 916	1.5 PB
CMIP5	811	360 TB
CORDEX	458	223 TB
input4MIPs	8	3 TB
C3S-CMIP5	174	25 TB
C3S-CORDEX	582	275 TB
C3S-CMIP6	235	18 TB
C3S-IPCC-Atlas	(coming soon)	
Total	<u>7 184</u>	<u>2.6 P</u>





ESPRI hosts the French ESGF Tier 1 node

The **Earth System Grid Federation** is a decentralized and federated nodes network with international collaboration. It's a system of distributed nodes that interact dynamically through a Peer-To-Peer (P2P) paradigm.



INSTITUT PIERRE-SIMON LAPLACE

Training for Outermost European regions

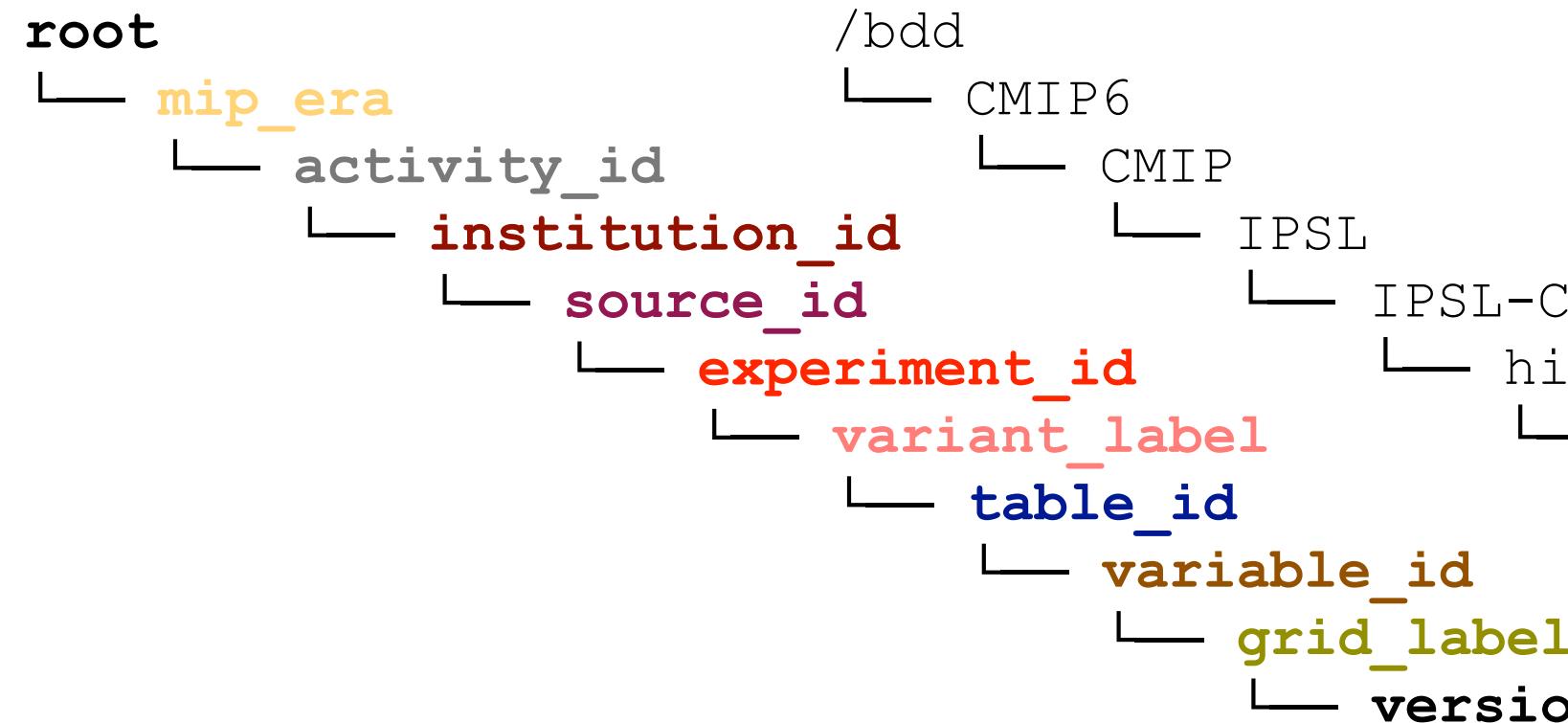
ESPRI provides data services for CLIMERI-France and ENES



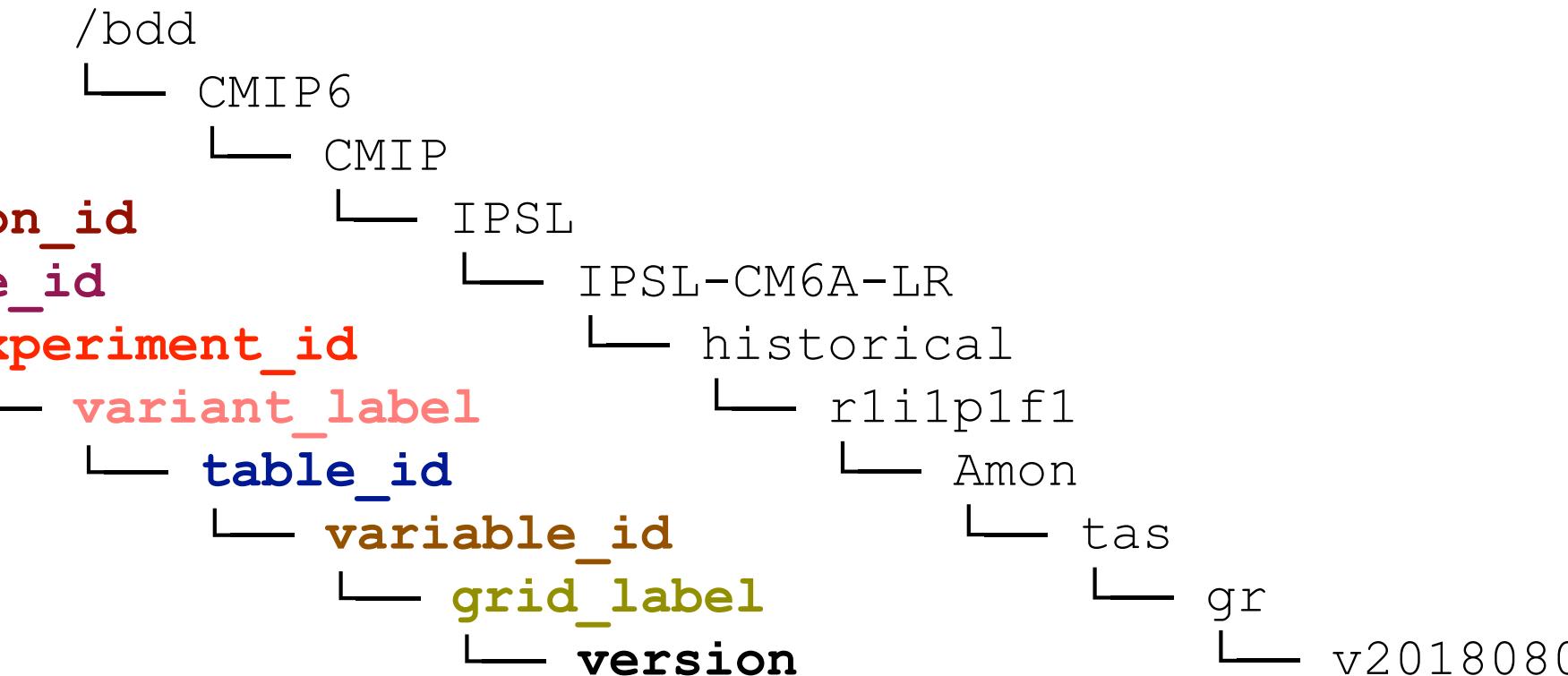
Local data discovery through single point of access: **/bdd**

CMIP6 directory structure

Pattern



Example



CMIP6 file naming

Pattern

<variable>_<table_id>_<source_id>_<experiment_id>_<variant_label>_<grid_label>[_<period_start>-<period_end>] .nc

Example

tas_Amon_IPSL-CM6A-LR_historical_r1i1p1f1_gr_185001-201412.nc

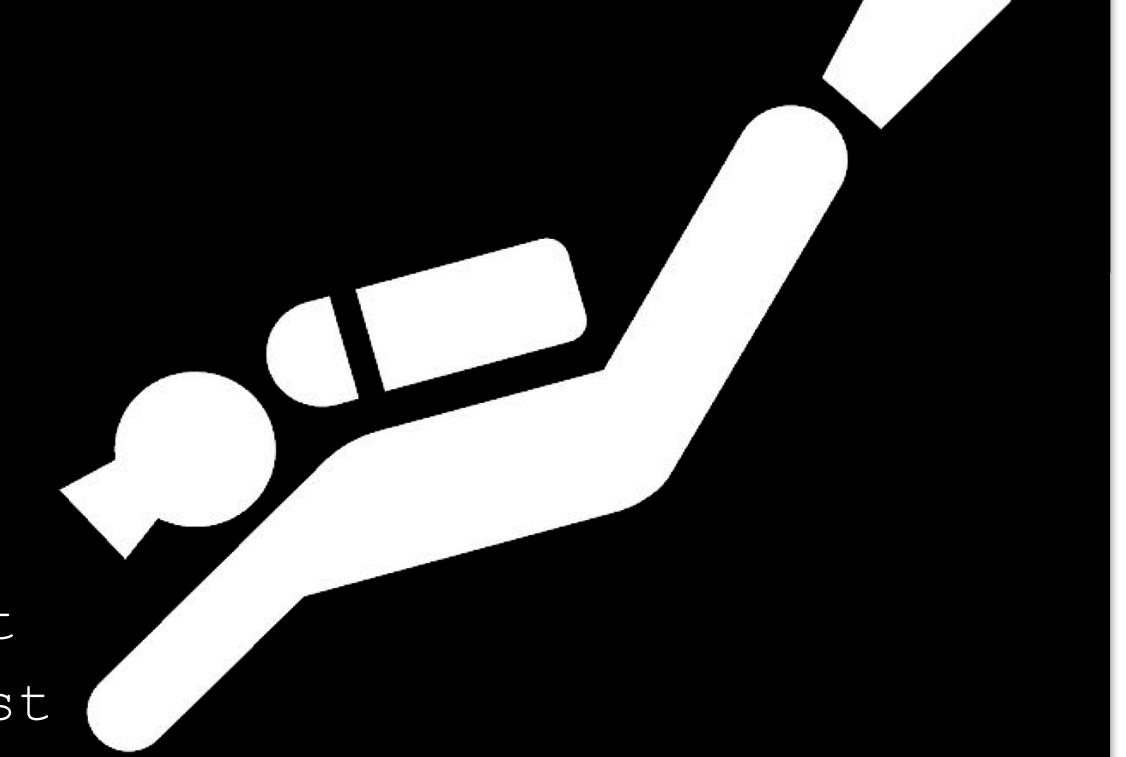
Constraints

Users must know the directory structure of data archives.

Users must know the controlled vocabulary of data archives.

Referenced data

```
$> ls /bdd/CMIP3
$> ls /bdd/CMIP5
$> ls /bdd/CMIP6
$> ls /bdd/CORDEX
$> ls /bdd/obs4MIPs
$> ls /bdd/input4MIPs
$> ls /bdd/CMIP5-Adjust
$> ls /bdd/CORDEX-Adjust
$> ls /bdd/C3S-CMIP5
$> ls /bdd/C3S-CORDEX
$> ls /bdd/C3S-CMIP6
```



```
00 000000
0000110101000100
01110111011001100000
01001111100100 111010100000 0
001 001111101010100010001000 0 000
000101000010000010100011010 0000 0010
00111011010 1010101001010 0110100010 0
00 01000 010010 0100 101010 111000 0101000
00 00110000101100 011 01011 11000 0001100 000
0000110 001011100 1 1011111100 000110 000000
0001 1 0000110001 11011 101000011 000 010 0
0001 0000111 10 1111 1101100 00010 100100 0
00 000011000100 001 10 1110 00 010 10 00 0000
011 00110 1 010 11 11010101000 0010
01100 11 11110 11010 10 10100
0000000 100 1111 1110 000110 001 00
000 10000000111 1101011000
011100010 110
11110
11010
1101
1000
10010
0110110110
01110000011101101110000100
```

ESPRI provides data services for CLIMERI-France and ENES



Local data discovery through `intake-esm` catalogs

```
import intake

catMaster = intake.open_catalog("/modfs/catalogs/master.yml")
catMaster

master:
  args:
    path: /modfs/catalogs/master.yml
  description: ClimERI-France master catalog for all data pool catalogs available
  driver: intake.catalog.local.YAMLFileCatalog
  metadata: {}

subcats = list(catMaster.items())
subcats

[('CMIP6', <CMIP6 catalog with 13482 dataset(s) from 7089602 asset(s)>),
 ('CMIP5', <CMIP5 catalog with 3678 dataset(s) from 938400 asset(s)>),
 ('CORDEX', <CORDEX catalog with 1942 dataset(s) from 428146 asset(s)>)
 ('CORDEX', <CORDEX catalog with 1942 dataset(s) from 428146 asset(s)>)]
```

ESPRI provides data services for CLIMERI-France and ENES



Local data discovery through `intake-esm` catalogs

```
catCMIP6 = catMaster["CMIP6"]
catCMIP6
```

CMIP6 catalog with 13482 dataset(s) from 7089602 asset(s):

	unique
path	7089602
project	1
activity_id	19
institution_id	40
source_id	107
experiment_id	237
member_id	738
table_id	39
variable_id	1034
grid_label	12
version	867
init_year	63
period_start	26094
period_end	32849
climatology	2
latest	2

ESPRI provides data services for CLIMERI-France and ENES



Local data discovery through `intake-esm` catalogs

```
mydata = catCMIP6.search(  
    experiment_id=["historical", "ssp585"],  
    table_id="Amon",  
    variable_id="tas",  
    latest=True)  
mydata
```

CMIP6 catalog with 111 dataset(s) from 15882 asset(s):

	unique
path	15882
project	1
activity_id	2
institution_id	33
source_id	63
experiment_id	2
member_id	201
table_id	1
variable_id	1
grid_label	3
version	185
init_year	0
period_start	258

ESPRI provides data services for CLIMERI-France and ENES



Local data discovery through `intake-esm` catalogs

```
mydata.serialize("myCat")
```

```
Writing catalog with 15882 entries into: myCat.json  
Writing ESM collection json file to: myCat.json
```

Main features

- User-friendly search interface
- Record your own catalog from the referenced ones
- Directory structure agnostic
- netCDF driver (with xarray)
- Aggregation over time, simulation members and initial year (decadal predictions)
- And more at <https://intake-esm.readthedocs.io/>

Constraints

- Users must know the controlled vocabulary of data archives.



Interactive research environment (Shell based - default)

Famous gridded data analyzing tools

NCO/CDO operators

Ferret/GraDs

Shared Python environments

```
$> module load modtools-python2  
$> module load modtools-python3  
$> source activate analyse_3.6  
# xarray, intake-esm, ipython-notebook, Dask, etc.
```

Specific climate data analysing tools

CliMAF - <https://climaf.readthedocs.io/>

ESMValTool (latest version, thanks to Klaus Zimmerman and Stephane Senesi)

KNMI Climate Explorer (deployment under consideration)

```
$> module load climaf  
$> module load esmvaltool
```

ESPRI provides data services for CLIMERI-France and ENES



A screenshot of the JupyterHub launcher interface. The top navigation bar includes File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. Below this is a file browser sidebar with a search bar and a list of files under the directory '/'. The main area is titled 'Launcher' and contains three sections: 'Notebook' (Python 3, panel [?], Python [conda env:root], R [conda env:root]), 'Console' (Python 3, Python [conda env:root], R [conda env:root]), and 'Other' (Terminal, Text File, Markdown File, Show Contextual Help).

A screenshot of a Jupyter Notebook cell titled 'Untitled.ipynb'. The cell contains the following code and output:

```
[1]: ls /bdd/CMIP6/PMIP
[2]: AWI@      CSIRO@      IPSL@      MRI@      NCC@ 
CASE@     EC-Earth-Consortium@  MIROC@    NASA-GISS@  NERC@ 
CNRM-CERFACS@ INM@       MPI-M@    NCAR@     NIUST@ 

[3]: import dask
import intake_esm
import xarray
import matplotlib
```

Remote research environment through JupyterHub

Access: <https://data.ipsl.fr/jupyter>

Status: Production since April 2022

Documentation: <https://documentations.ipsl.fr/>

Will be completed and available through the new incoming ESPRI website.

Python environment: PANGEOT + useful modules.

Data access: All open data from /bdd

ESPRI provides data services for CLIMERI-France and ENES



REQUEST

Search criteria called facets are used to select which files to download. They can be set on command line or using a template.



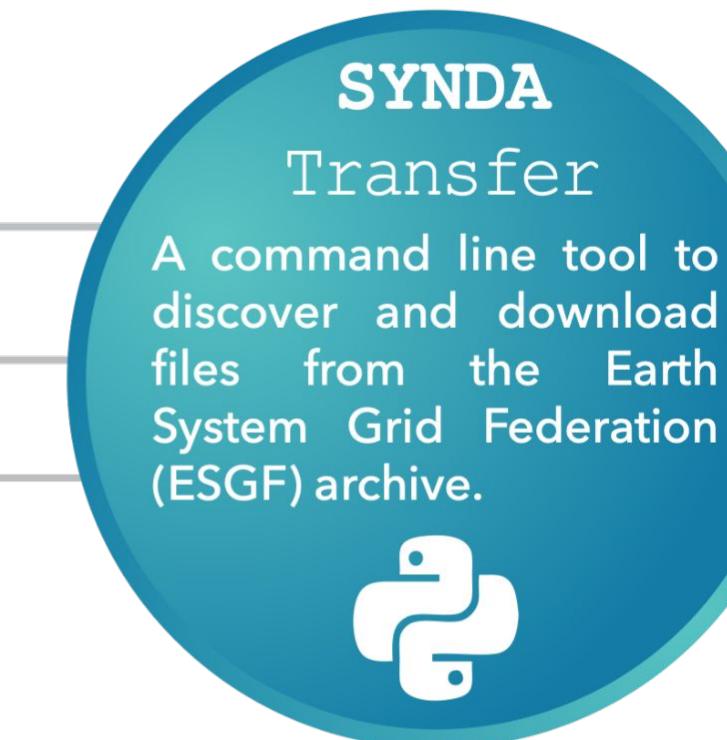
ESGF NODES

SDT retrieves the certificates and builds the HTTP requests to Solr corresponding to the search criteria.



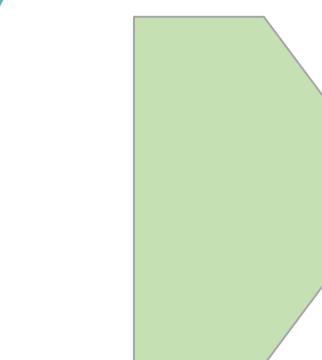
FILESYSTEM

ESGF files are downloaded using the HTTP or GridFTP protocol and managed on the local filesystem following the Data Reference Syntax.



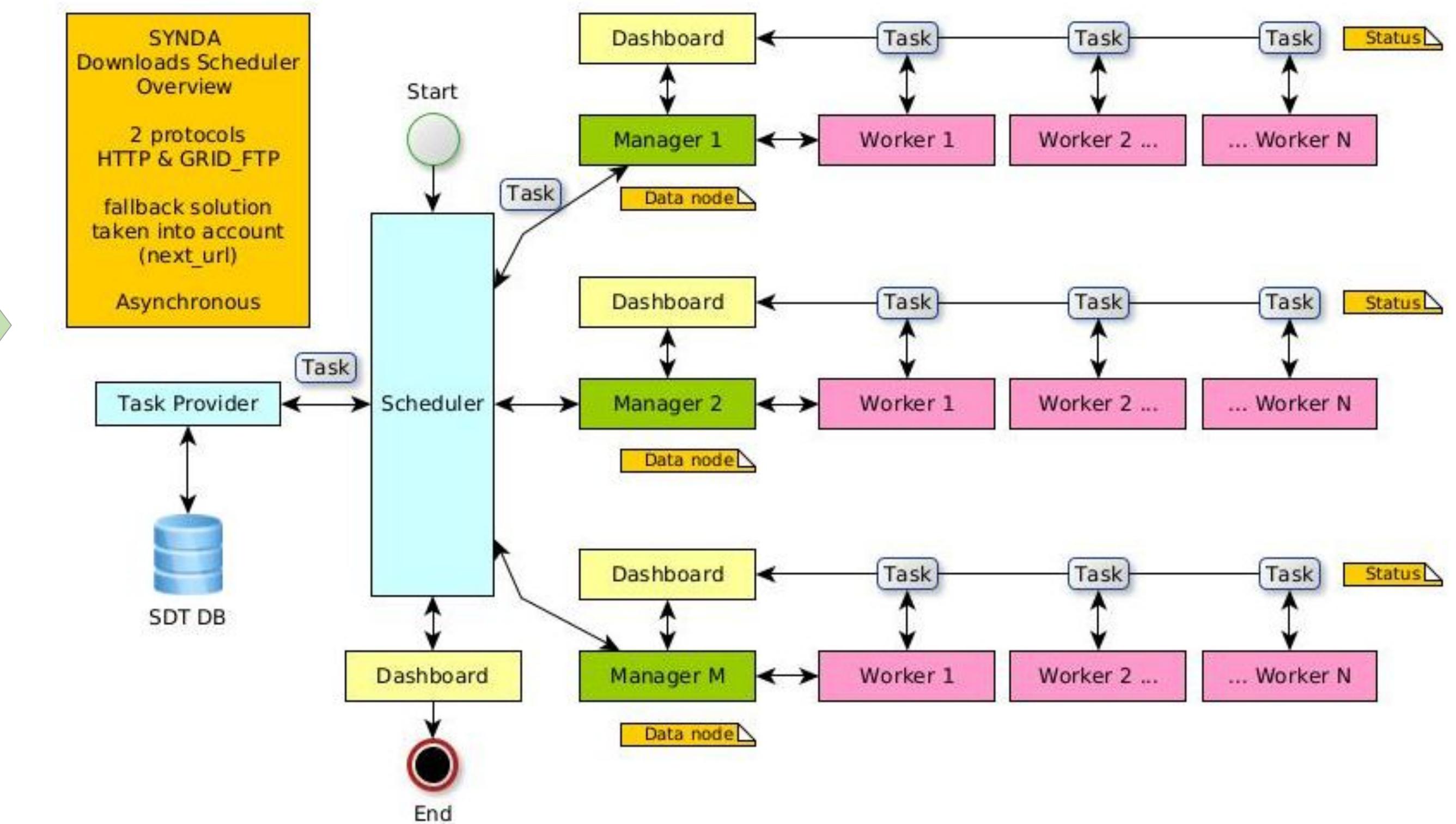
SDT DATABASE

A SQLite database records each downloaded file and dataset. A complete dataset triggers a "dataset_complete" event, which informs the SDP module to start the pipeline.



On-demand data replication service

New paradigm implemented to perform **parallel downloads asynchronously**. Once each workers have their download tasks given out by the scheduler, they are able to asynchronously carry out their duties without having to wait for each other. Our tool is able to replicate data from ESGF at a download rate of 4TB/day.



ESPRI provides data services for CLIMERI-France and ENES



On-demand data replication service

How to request additional referenced data?

1. Explore the existing database in `/bdd`
2. If data you need are not locally available, explore the ESGF catalogue: <https://esgf-node.ipsl.upmc.fr>
 - a. If data you need are **not available** on the ESGF: data does not exist or have not been published.
 - b. If data you need is **available** on the ESGF: create a `.txt` file with the following syntax

```
$> vi my_template.txt
```

```
#login@ipsl.fr
project=CMIP6
experiment=historical amip
model=all
ensemble=r1i1p1f1
variable[day]=clt tas
variable[Omon]=sic evap
```

3. Send your file to espri-mod@listes.ipsl.fr
4. Your request will be examined quickly (space required vs. free space)
 - a. We validate your request and run data replication.
 - b. We specify together your expectations to satisfy the storage spaces.
5. We notify you as soon as your data is available
6. And of course, we support you!

ESPRI provides data services for CLIMERI-France and ENES



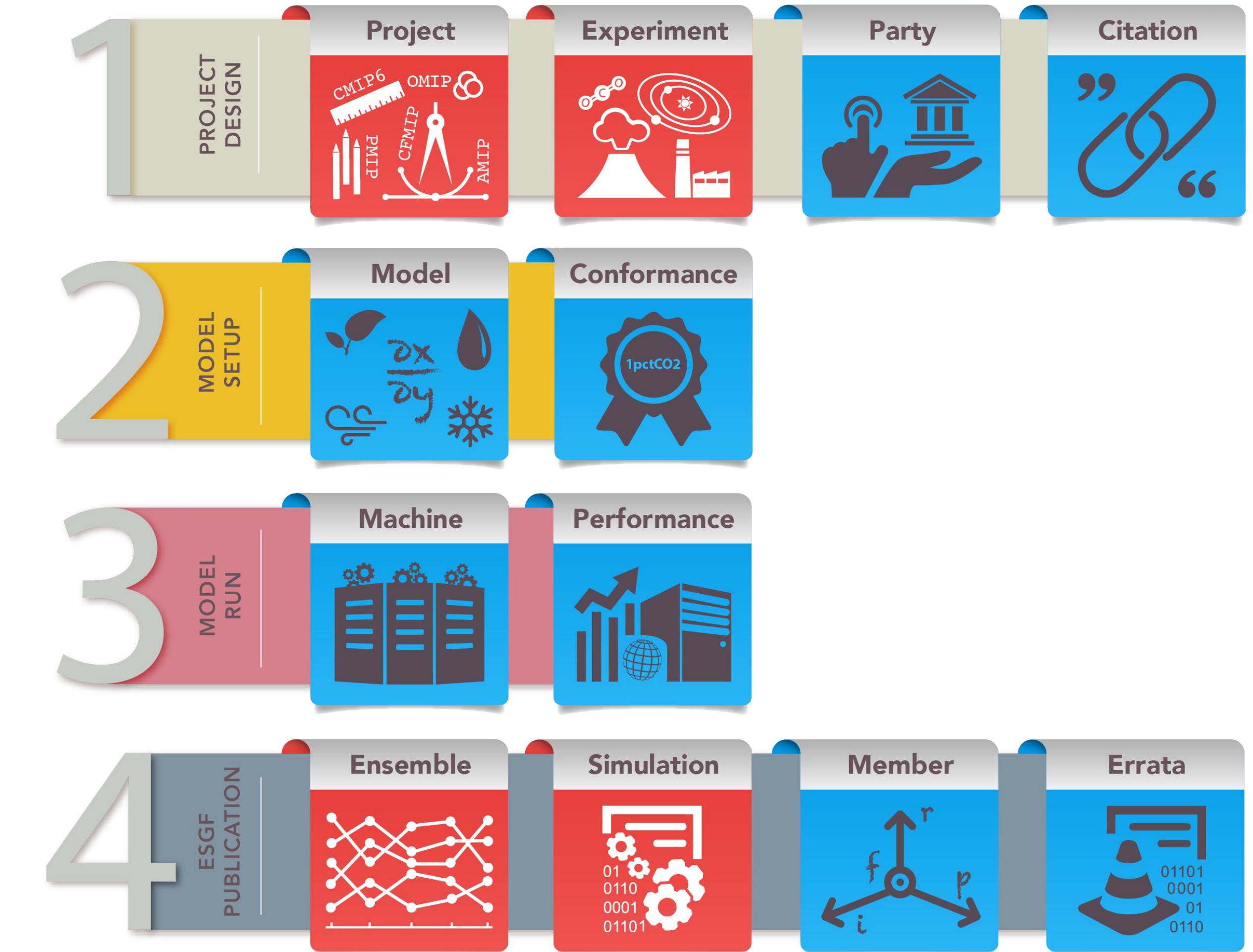
Expertise and support in data documentation and citation

The **Earth System Documentation** (ES-DOC) aims to nurture an ecosystem of tools and services in support of Earth System documentation creation, analysis and dissemination. Such an ecosystem enables the scientific community to better understand and utilize Earth system model data. ES-DOC is coordinated with other community efforts such as CMIP and ESGF via the World Climate Research Programme work group on Climate Modelling (WGCM) and its Infrastructure Panel (WIP).

Home page
<https://es-doc.org/>

Documentation service
<https://search.es-doc.org/>

Errata service
<https://errata.es-doc.org/>



ESPRI provides data services for CLIMERI-France and ENES



Expertise and support in data documentation and citation

WARNING: Not all models include a variant "r1i1p1f1", and across models, identical values of variant_label do not imply identical variants. To learn which forcing datasets were used in each variant, please check modeling group publications and documentation provided through ES-DOC.

Enter Text: [?] Search Reset Display 10 ↑ results per page [More Search Options]

Source ID Show All Replicas Show All Versions Search Local Node Only (Including All Replicas)

Search Constraints: **IPSL**

Total Number of Results: 606899
-1- 2 3 4 5 6 Next >>

Add all displayed results to Data Cart Remove all displayed results from Data Cart

Expert Users: you may display the search URL and return results as XML or return results as JSON

1. CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [Show Citation] [PID] [Further Info]

DOI for 'CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos'
doi:10.22033/ESGF/CMIP6.5142

General Information Creators Editors

General Information

Name CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos
Abstract Coupled Model Intercomparison Project Phase 6 (CMIP6) data sets.
These data includes all datasets published for 'CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos' according to the Data Reference Syntax defined as 'mip_era.activity_id.institution_id.source_id.experiment_id.member_id.table_id.variable_id.grid_label.version'.

Cite this data

Citation Boucher, Olivier; Denvil, Sébastien; Caubel, Arnaud; Foujols, Marie Alice (2019). IPSL IPSL-CM6A-LR model output prepared for CMIP6 DCPP dcppC-amv-ExTrop-pos. Version YYYYMMDD^[1].Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.5142>

[1] Please use the latest dataset version or if not available the latest data download date as version in your data citation.

A unique and immutable DOI for each CMIP6 simulation registered at
<https://cera-www.dkrz.de>

ESPRI provides data services for CLIMERI-France and ENES



Expertise and support in data documentation and citation

WARNING: Not all models include a variant "r1i1p1f1", and across models, identical values of variant_label do not imply identical variants. To learn which forcing datasets were used in each variant, please check modeling group publications and documentation provided through ES-DOC.

Enter Text: [?] Search Reset Display 10 results per page [More Search Options]

Source ID Show All Replicas Show All Versions Search Local Node Only (Including All Replicas)

Search Constraints: IPSL

Total Number of Results: 606899
-1- 2 3 4 5 6 Next >>

Add all displayed results to Data Cart Remove all displayed results from Data Cart

Expert Users: you may display the search URL and return results as XML or return results as JSON

1. CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [Show Citation] **[PID]** Further Info

Add to Data Cart

Source ID IPSL (606899)
Institution ID
Source Type
Nominal Resolution
Experiment ID
Sub-Experiment

Dataset CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr

General Information

Dataset Id	CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr
Persistent identifier	hdl:21.14100/770c4a29-6b16-3191-a94a-aad019633ea
Version	20190110

Data host(s)

vesg.ipsl.upmc.fr Original
aims3.llnl.gov Replica

Files belonging to this dataset

intvaw_Emon_IPSL-CM6A-LR_dcppC-amv-ExTrop-pos_r13i1p1f1_gr_185001-185912.nc hdl:21.14100/cba4eeaf-586f-402a-8d3d-6d6f13b5a926

A unique and immutable file identifier, called Persistent IDentifier (PID) or Handle, permanently stored even if the data is removed or deleted. It provides a landing page that collect information about:

- The dataset ID and PID
- Version genealogy with file PIDs
- Links to the errata
- Access to replicas

ESPRI provides data services for CLIMERI-France and ENES



Expertise and support in data documentation and citation

WARNING: Not all models include a variant "r1i1p1f1", and across models, identical values of variant_label do not imply identical variants. To learn which forcing datasets were used in each variant, please check modeling group publications and documentation provided through ES-DOC.

Enter Text: [?] Search Reset Display 10 results per page [More Search Options]

Search Constraints: IPSL Show All Replicas Show All Versions Search Local Node Only (Including All Replicas)

Total Number of Results: 606899
-1- 2 3 4 5 6 Next >>
Add all displayed results to Data Cart Remove all displayed results from Data Cart
Expert Users: you may display the search URL and return results as XML or return results as JSON

1. CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [Show Citation] [PID] [Further Info]
 Add to Data Cart

WCRP
World Climate Research Programme

CMIP6 Further Information v1.1.2 Support Help

Further Info URL: <https://furtherinfo.es-doc.org/CMIP6.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.none.r13i1p1f1>

ES-DOC Documentation

MIP Era	CMIP6
Institution	IPSL
Model	IPSL-CM6A-LR
Experiment	dccppC-amv-ExTrop-pos
Ensemble Description	N/A
Machine Performance	N/A

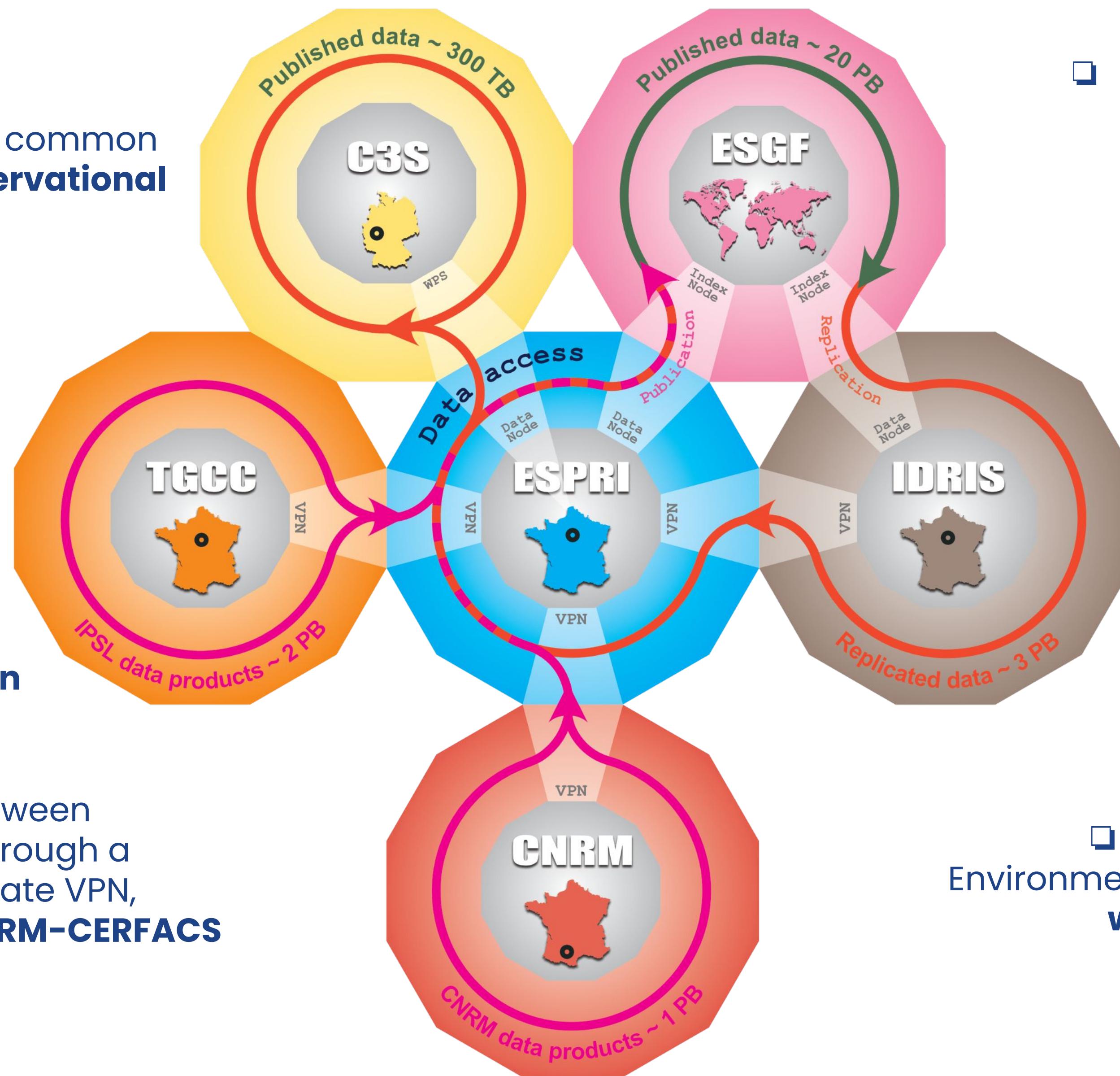
A dedicated and unique URL, called “further info URL” that collects available information on each CMIP6 simulation:

- Links to the documentation of the model and the experiment
- Links to the errata
- Links to the citation

2023 and beyond with GAIA-Data



- ❑ Towards **STAC catalogs** with common **search engine** between **observational** and **modelling** data sets.
- ❑ Extending and automating **DOI services to codes**.
- ❑ Implement a **web-based thesaurus** that **guides user in data discovery**.
- ❑ Network mounting point between **CNRS-CERFACS** and **ESPRI** through a **10Gbps** RENATER link and private VPN, allowing **direct access to CNRM-CERFACS climate simulations**.



- ❑ Strengthening **on-demand bias correction** processing based on **workflow managers**.
- ❑ Getting the **CoreTrustSeal label** from the Research Data Alliance (RDA).
- ❑ Extending a multi-thematic **storage at IDRIS**.
- ❑ Strengthening Virtual Research Environment based on **Jupyter Notebooks** with **PANGEOT** and **AI4GEO** suites.





Thank you for your attention

Institut Pierre-Simon Laplace