Interactive and Flexible Environment for Climate Data Analysis

Christian Pagé
CERFACS, Toulouse, France

Abel Aoun
CERFACS Toulouse, France

Alessandro Spinuso, Mats Veldhuizen, Ian van der Neut
KNMI, De Bilt, Netherlands

Klaus Zimmermann, Lars Bärring
SMHI, Norrköping, Sweden

Phillip Kershaw
NCEO/CEDA, STFC, Didcot, UK

14th International Workshop on Science Gateways
15th-17th June 2022, Trento, Italy

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

Climate Data Distribution through the ESGF

ESGF represents a multinational effort to securely **access**, **monitor**, **catalog**, **transport**, and **distribute** reference **data** for **climate** research experiments and observations.

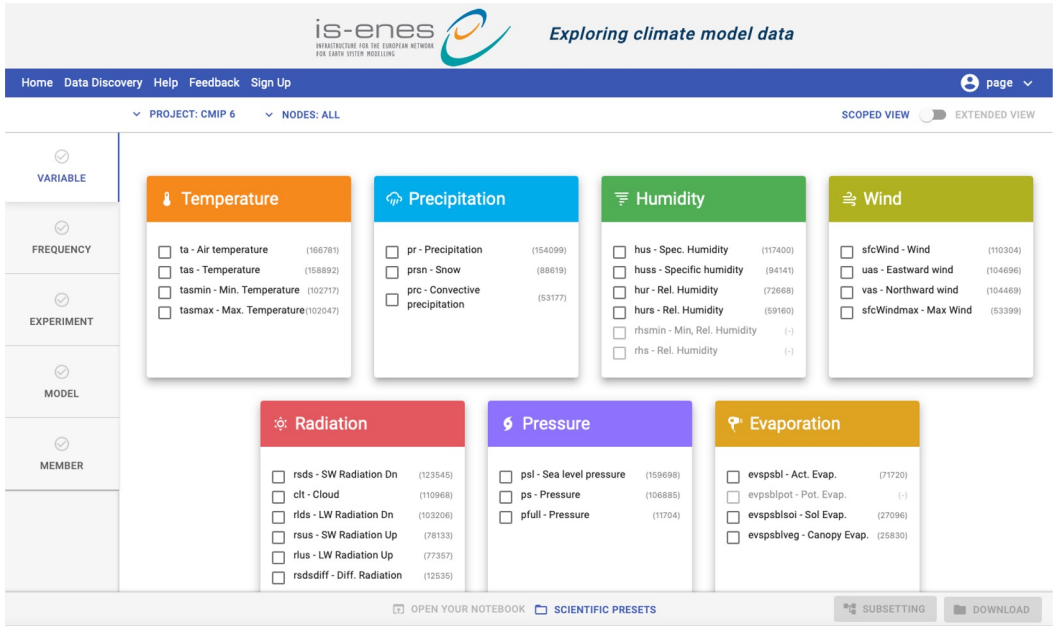
ESGF
Earth System Grid Federation

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

What is the climate4impact (C4I) portal? (1/2)

Platform for researchers to explore climate data and perform analysis

- Front-end to climate data infrastructure (ESGF)
- Tailored Search Interface with view modes
- Jupyter-Lab enhanced environment
- Notebooks gallery
- Flexible analysis features
  - Climate indices calculation: **icclim**
  - Data Staging/Reduction Workflows
  - Personal store for processing outcome
- Automated reproducibility mechanisms and documentation (Data/Analysis)
- Modular Deployment & Decoupled Architecture
- Pages for Models Performance Comparison (**ESMValTool**)



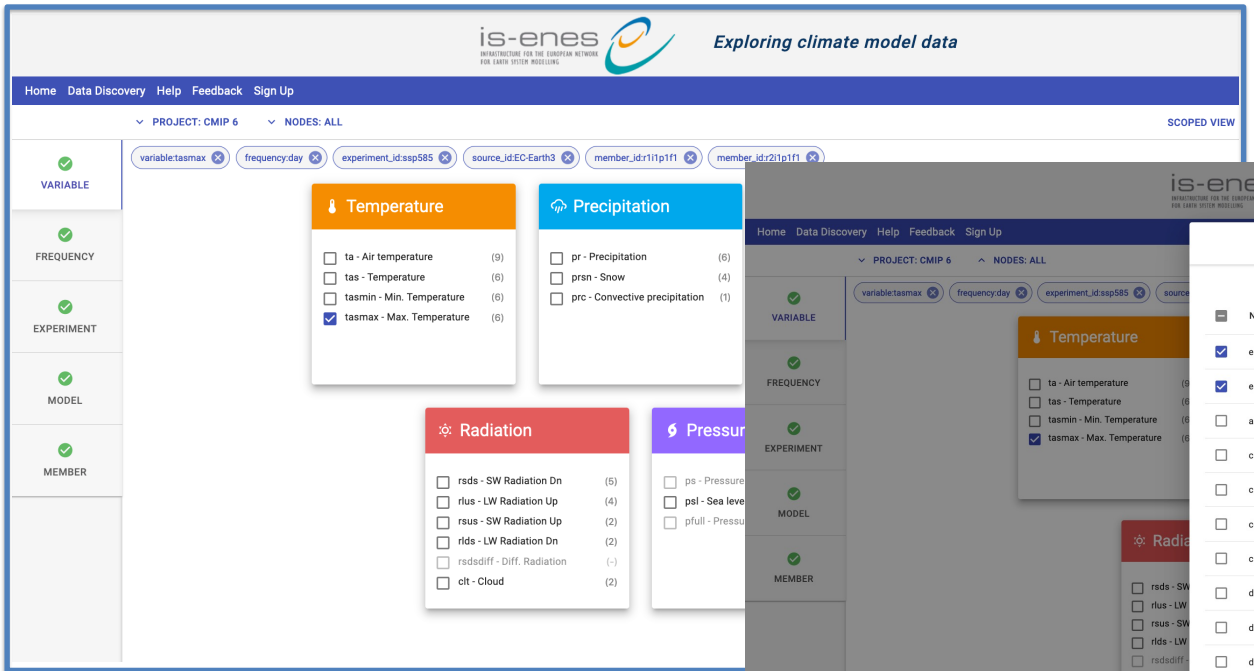
V2: Complete Redesign from V1
Beta version available https://dev.climate4impact.eu

Findable Accessible Interoperable Reusable

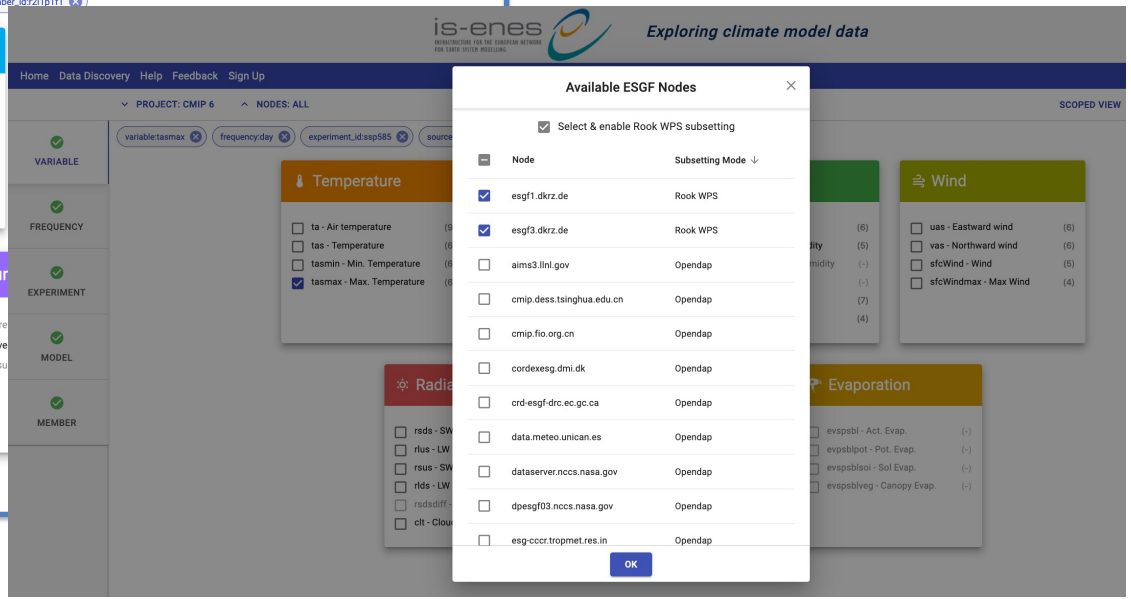
IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Climate4Impact (v2) Search & Nodes

Search Parametrisation made easier



Nodes Selection by Service

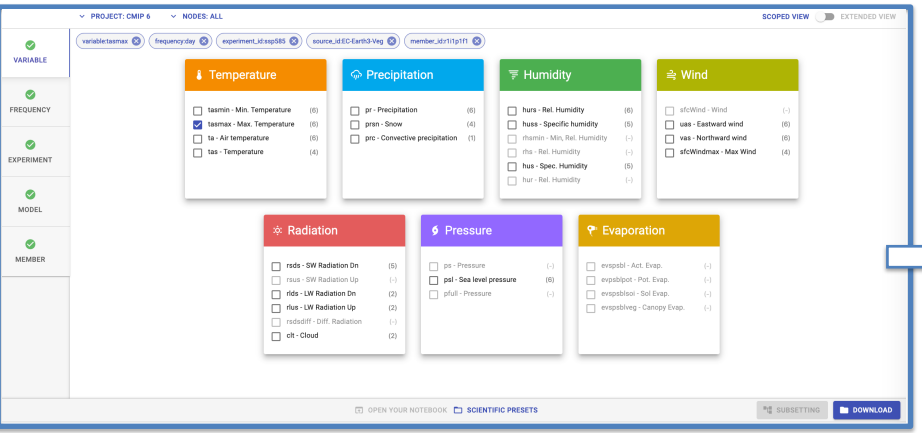


IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

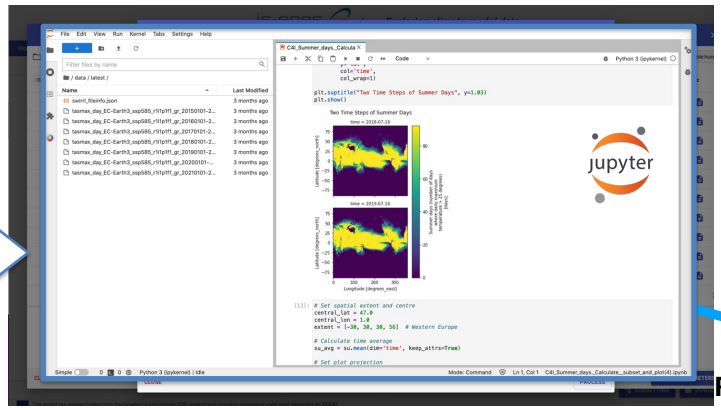
Climate4Impact (v2) Workflows & Workspaces

Climate4Impact Search for CMIP5/6 CORDEX Data

https://dev.climate4impact.eu



Workflows for data staging & reduction onto iccim powered Notebooks



Save/Share Progress to Git

Reduced Data

MyBinder Reproduce

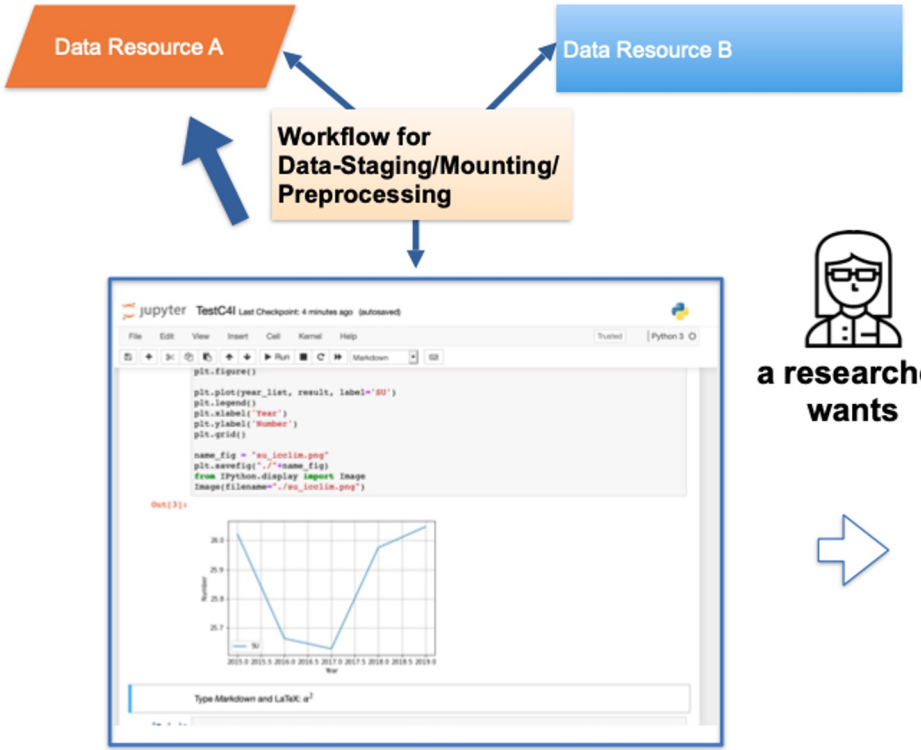
- Trace Changes to Restore, Recover Software and/or Data

W3C PROV

CERFACS

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

C4I Workspace Use Case

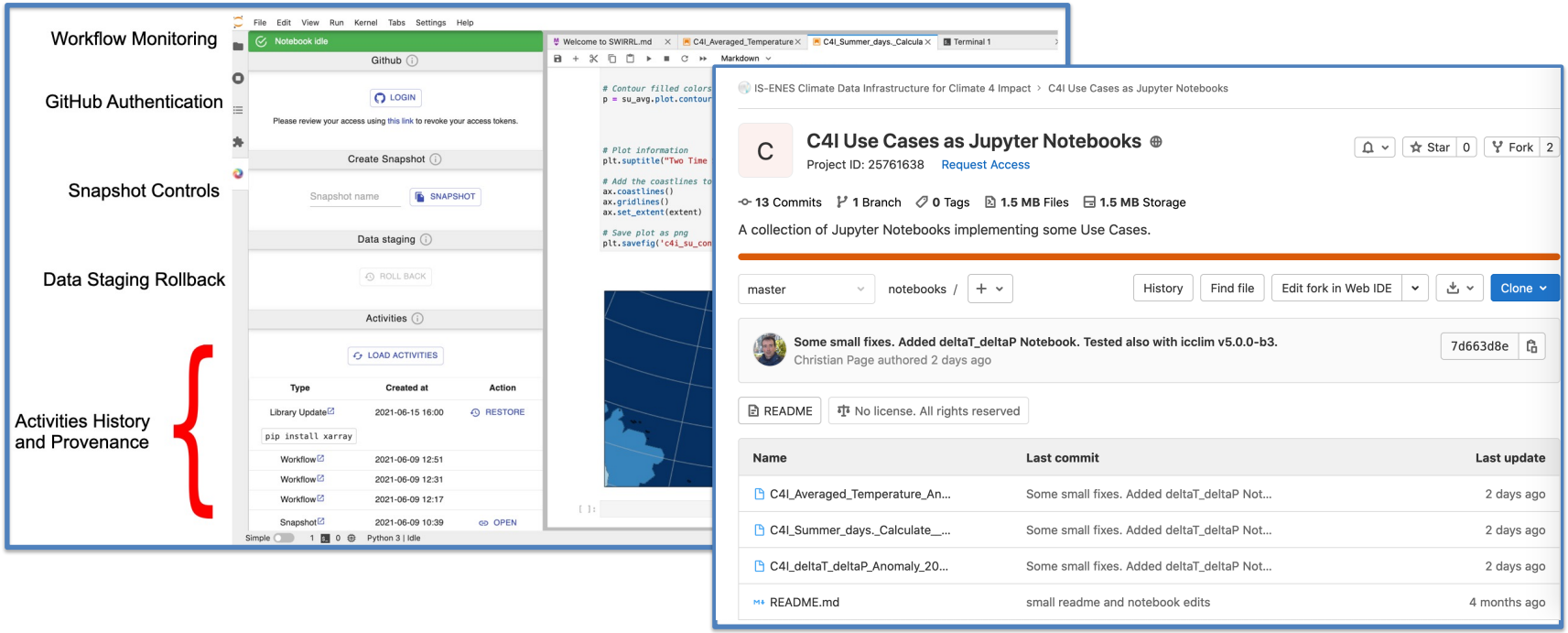


a researcher wants

- **access distributed raw data**
- **develop, document and reuse** methods for processing and visualisation.
- **update/extend** raw data and software
- **Track changes and rollback** (Traceability/Recovery)
- **keep old versions of the data** after updates (Reproducibility)
- **snapshot and restore** the state of a workspace software (Reproducibility)

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

SWIRRL JupyterLab Extension and Sample Notebooks based on icclim (climate indices)

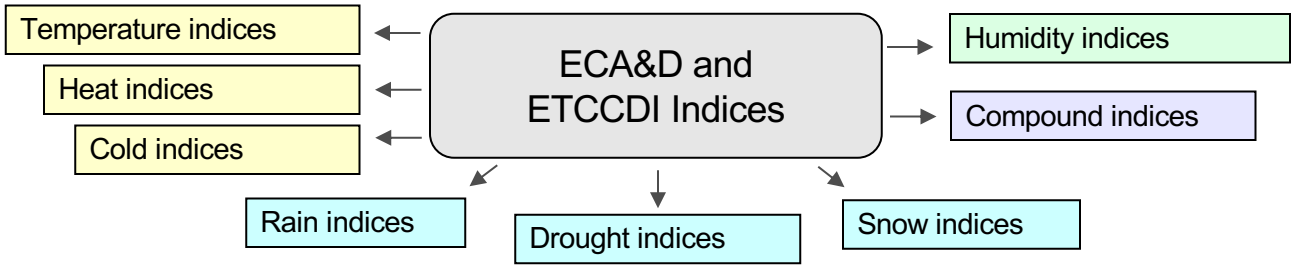


https://gitlab.com/is-enes-cdi-c4i/notebooks

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

On-demand calculations

Climate indices calculation in climate4impact: **icclim**



- Intra-period extreme temperature range [°C] - **ETR**
- Warm days (days with mean temperature > 90th percentile of daily mean temperature) - **TG90p**
- Summer days (days with max temperature > 25 °C) - **SU**
- ...

- Python code developed at Cerfacs since September 2013
- Funded by EU FP7 IS-ENES2, FP7 CLIPC and H2020 IS-ENES3
- Generic and modular approach, can be reused in other environments
- New V5 completely rewritten and using underlying xclim functions, based on xclim, xarray and dask
- I/O interface is structured for optimal performance
- Implement the proper percentile indices calculations when calculation period overlaps reference period (called bootstrapping method)

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

icclim: climate indices

Documentation: https://icclim.readthedocs.io/en/latest/python_api.html
Source code: https://github.com/cerfacs-globc/icclim
Current Version 5.3: https://github.com/cerfacs-globc/icclim/releases/tag/5.3

`icclim.index(**kwargs)`

Parameters:

- **in_files** (*str | list[str] | Dataset | DataArray,*) – Absolute path(s) to NetCDF dataset(s), including OPeNDAP URLs, or path to zarr store, or xarray.Dataset or xarray.DataArray.
- **index_name** (*str*) – Climate index name. For ECA&D index, case insensitive name used to lookup the index. For user index, it's the name of the output variable.
- **var_name** (*str | list[str] | None*) – `optional` Target variable name to process corresponding to `in_files`. If None (default) on ECA&D index, the variable is guessed based on the climate index wanted. Mandatory for a user index.
- **slice_mode** (*str*) – Type of temporal aggregation: {"year", "month", "DJF", "MAM", "JJA", "SON", "ONDJFM" or "AMJJAS"}. Default is "year". See slice_mode for details.
- **time_range** (*list[datetime.datetime]*) – `optional` Temporal range: upper and lower bounds for temporal subsetting. If `None`, whole period of input files will be processed. Default is `None`.
- **out_file** (*str | None*) – Output NetCDF file name (default: "icclim_out.nc" in the current directory). Default is "icclim_out.nc". If the input `in_files` is a `Dataset`, `out_file` field is ignored. Use the function returned value instead to retrieve the computed value. If `out_file` already exists, icclim will overwrite it!
- **threshold** (*float | list[float] | None*) – `optional` User defined threshold for certain indices. Default depend on the index, see their individual definition. When a list of threshold is provided, the index will be computed for each thresholds.

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

SWIRRL-API

SWIRRL hides the complexity of orchestrating Workspaces in a target Cloud resource based on Kubernetes Cluster

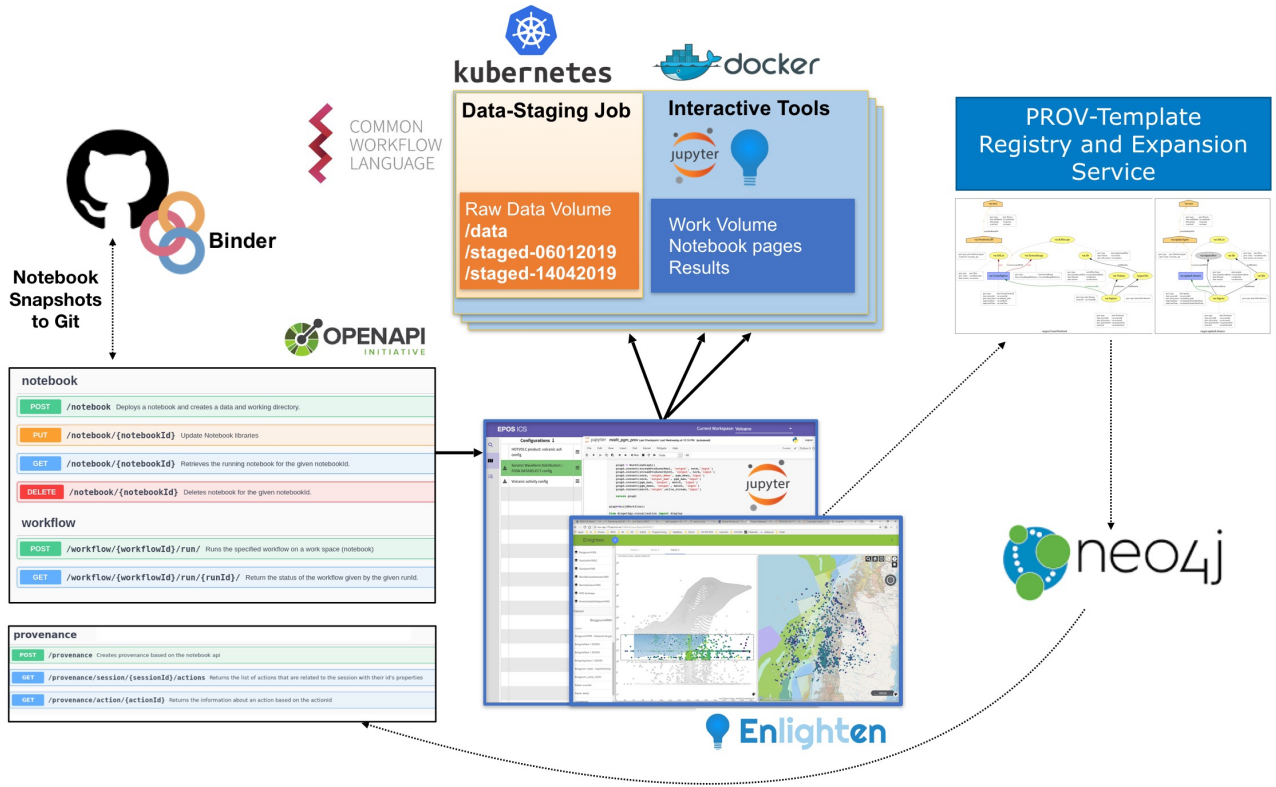
Integrates Interactive Tools (Notebooks, Workflows, GIT, Binder)

Offers a REST Web API

Manages Metadata (Provenance)

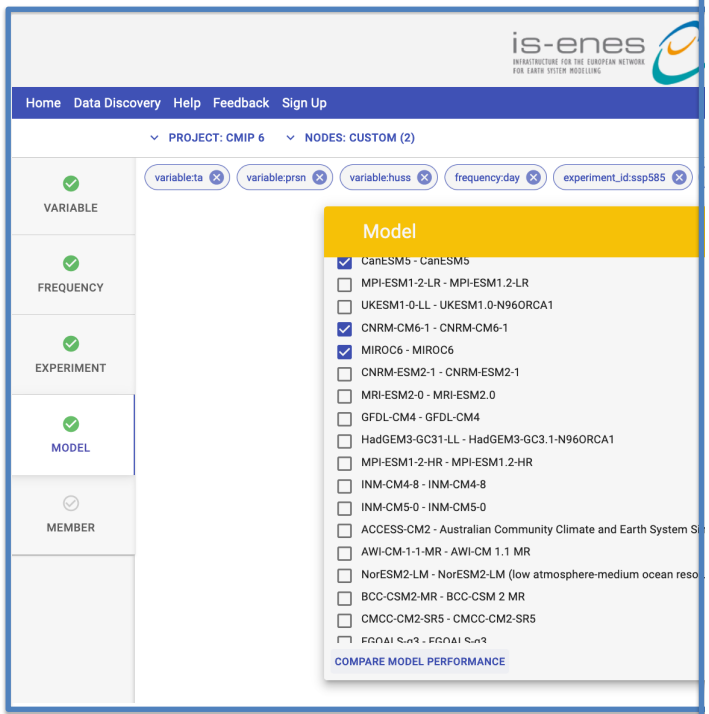
Findable Accessible Interoperable Reusable

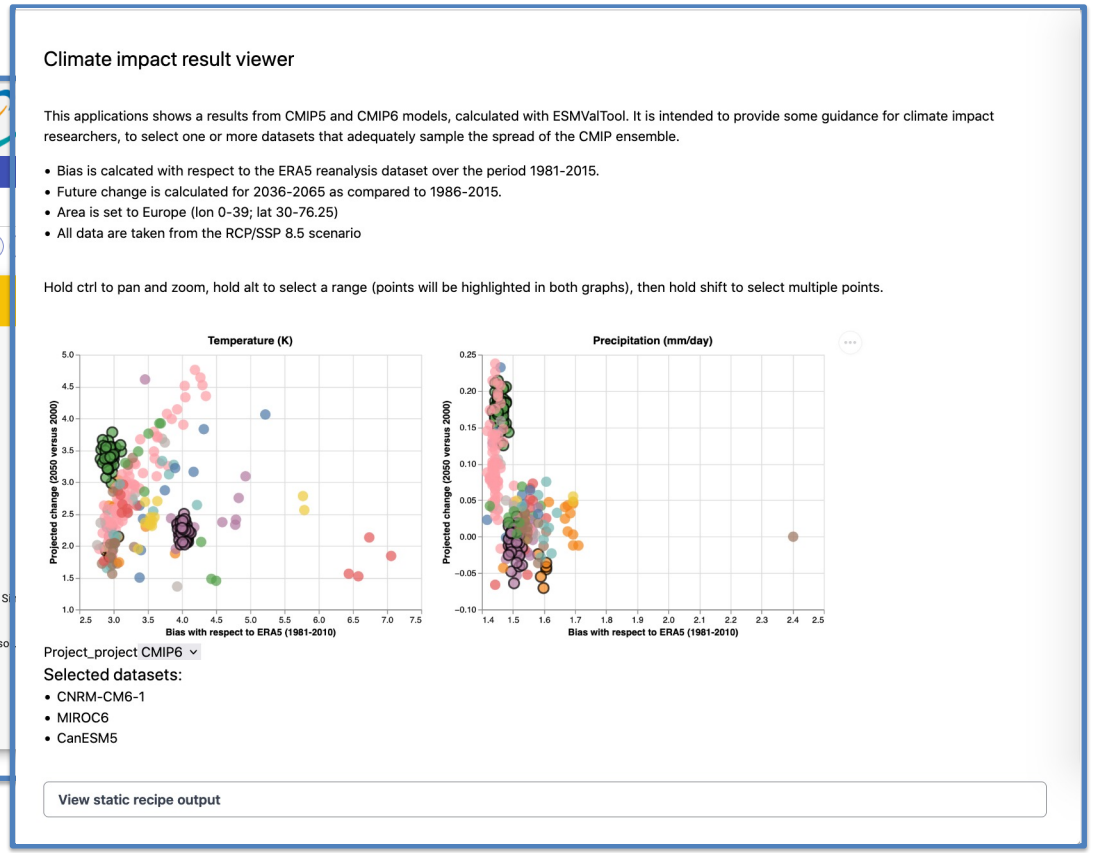
Supports Reproducibility



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

ESMValTool in C4I
Models' Performance Comparison

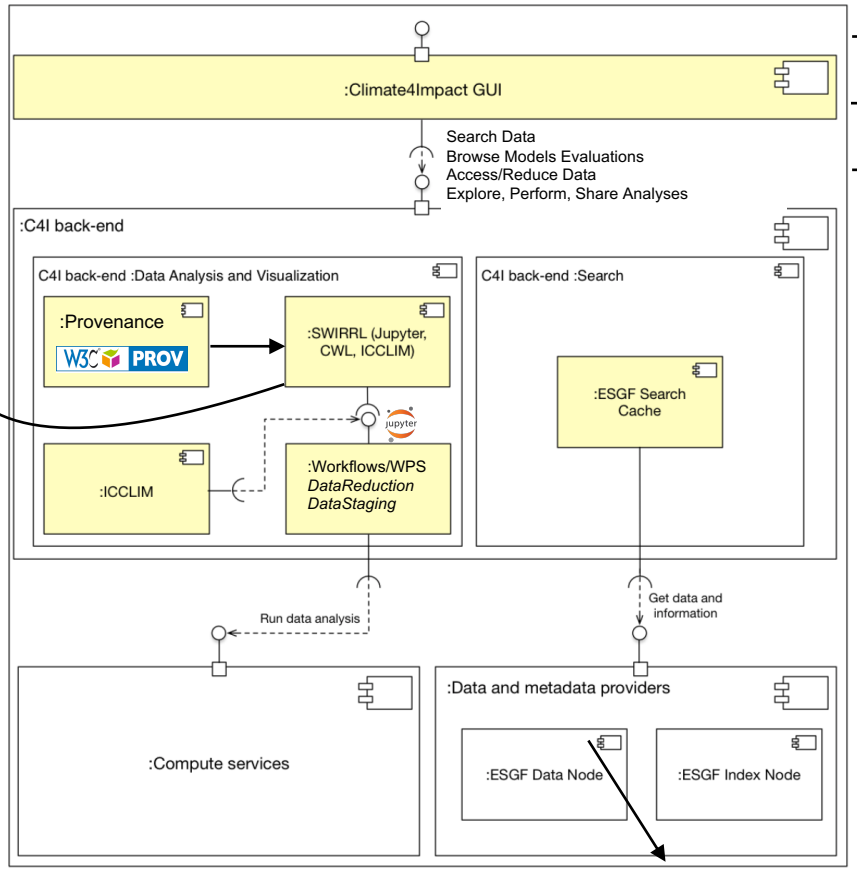




Netherlands eScience Center

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

Climate4Impact (v2) - ENES CDI



Git Store/Share Analyses

Binder Reproduce

→ **SSO with IDeA AAI**

→ **ESMValTool Evaluation pages**

→ **Link to DOI Pages ES-DOC**

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

Thanks !

On behalf of the climate4impact and icclim teams
https://dev.climate4impact.eu
https://icclim.readthedocs.io/en/stable/

Christian Pagé christian.page@cerfacs.fr
Alessandro Spinuso alessandro.spinuso@knmi.nl
Abel Aoun abel.aoun@cerfacs.fr

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

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

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