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

IS-ENES3 Webinar
13 April 2023

Climate4Impact v2

Enhance the use of climate research data and methods

climate4impact.eu

Alessandro Spinuso, Mats Veldhuizen,
Ian van der Neut (KNMI),
Christian Pagé (CERFACS)

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

Climate4Impact v2

Online platform that combines **discovery of climate research data of the Earth System Grid Federation (ESGF), with experimentation workspaces,** offering **reproducible and sharable** environments for **climate-impact analysis**.

Climate4Impact v2

Online platform that combines **discovery of climate research data of the Earth System Grid Federation (ESGF), with experimentation workspaces,** offering **reproducible and sharable** environments for **climate-impact analysis**.

Delivered in the context of **the cooperation between IS-ENES and ESGF, which co-develop tailored tooling, infrastructure services and trainings**

Climate4Impact v2

Online platform that combines **discovery of climate research data of the Earth System Grid Federation (ESGF), with experimentation workspaces,** offering **reproducible and sharable** environments for **climate-impact analysis**.

Delivered in the context of **the cooperation between IS-ENES and ESGF, which co-develop tailored tooling, infrastructure services and trainings**

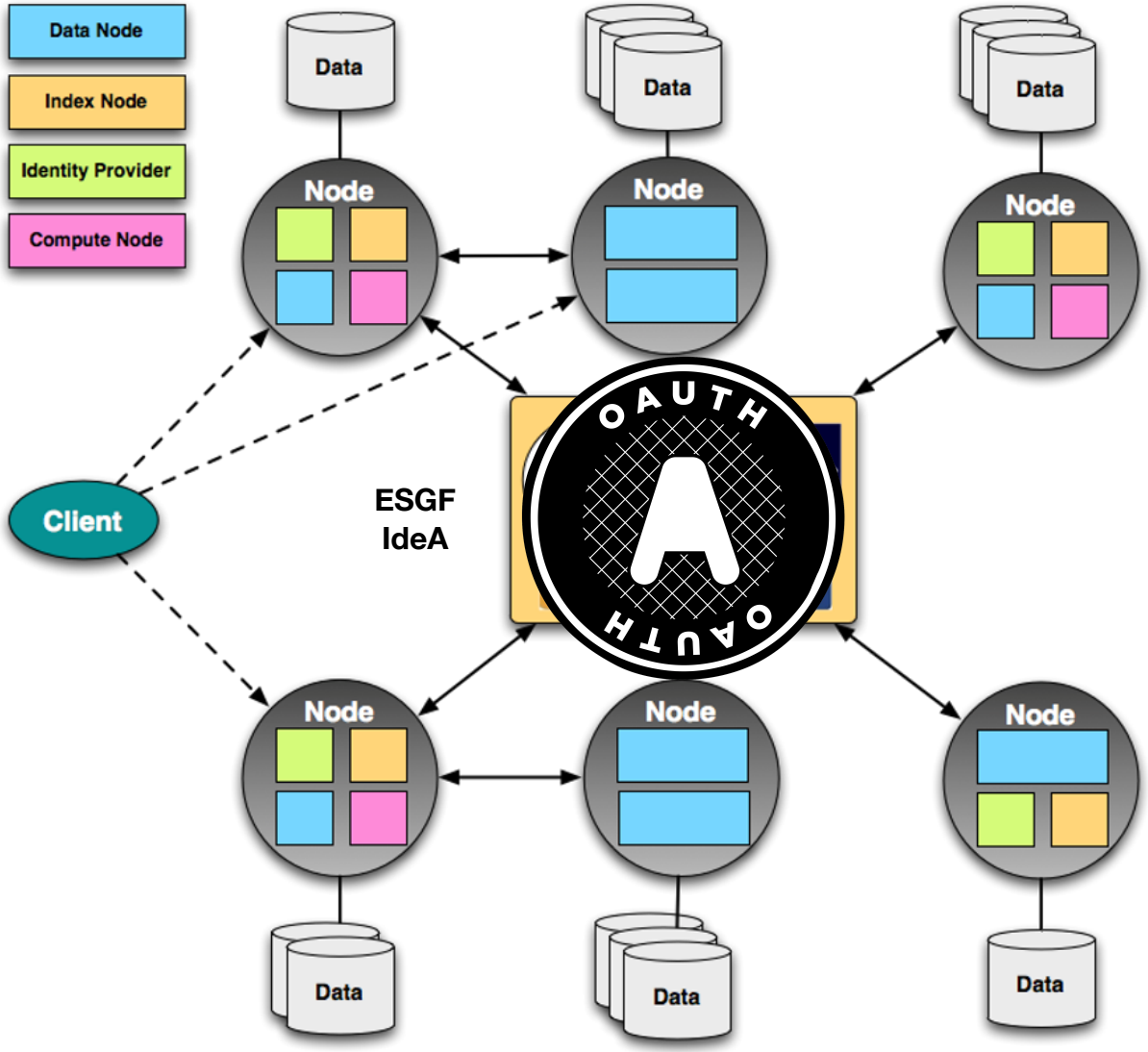
Aimed at supporting research on **predicting and mitigating climate change effects** on the whole Earth ecosystem

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

ESGF

ESGF is a **multinational** effort that manages climate and environmental data.
A network of federated and **standardised data discovery, access and analysis services.**
Data Volume, in the order of tens of PB, is constantly increasing and distributed across different projects, e.g.

- CMIP5-6.. **7 coming up**
- CORDEX
- Obs4MIPs



Climate4Impact v2 - Projects

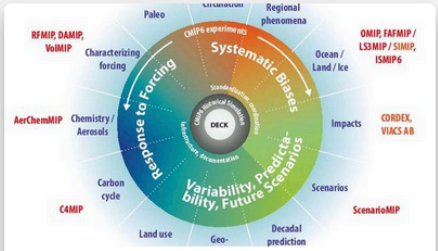
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Exploring climate model data

Home | Data Discovery | Help | Feedback | Register | Login

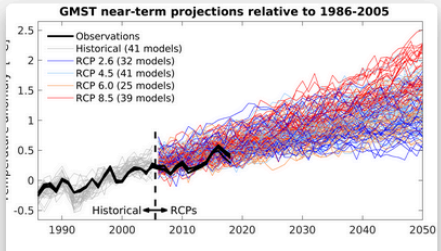
Select a project

You can search the Earth System Grid Federation for various kinds of climate data. Please select the project you are looking for.



CMIP6

Coupled model intercomparison experiment 6.



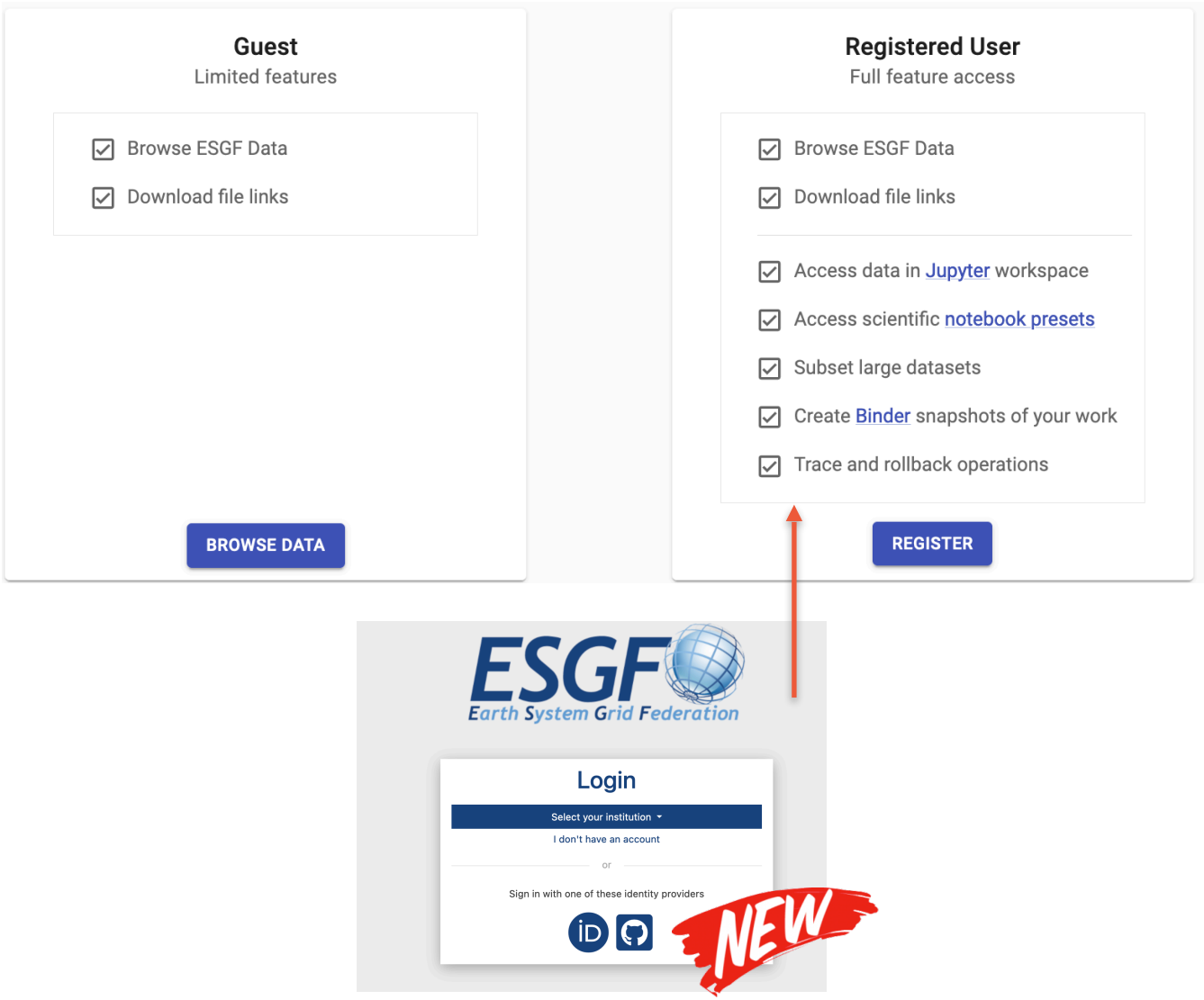
CMIP5

Coupled model intercomparison experiment 5.

CORDEX

Global

Climate4Impact v2 - Profiles and SSO



Two User Profiles

Guest: *Basic Search & Local Download*
Registered User: *Data Analysis & Subsetting (soon CORDEX support)*

Registration procedure

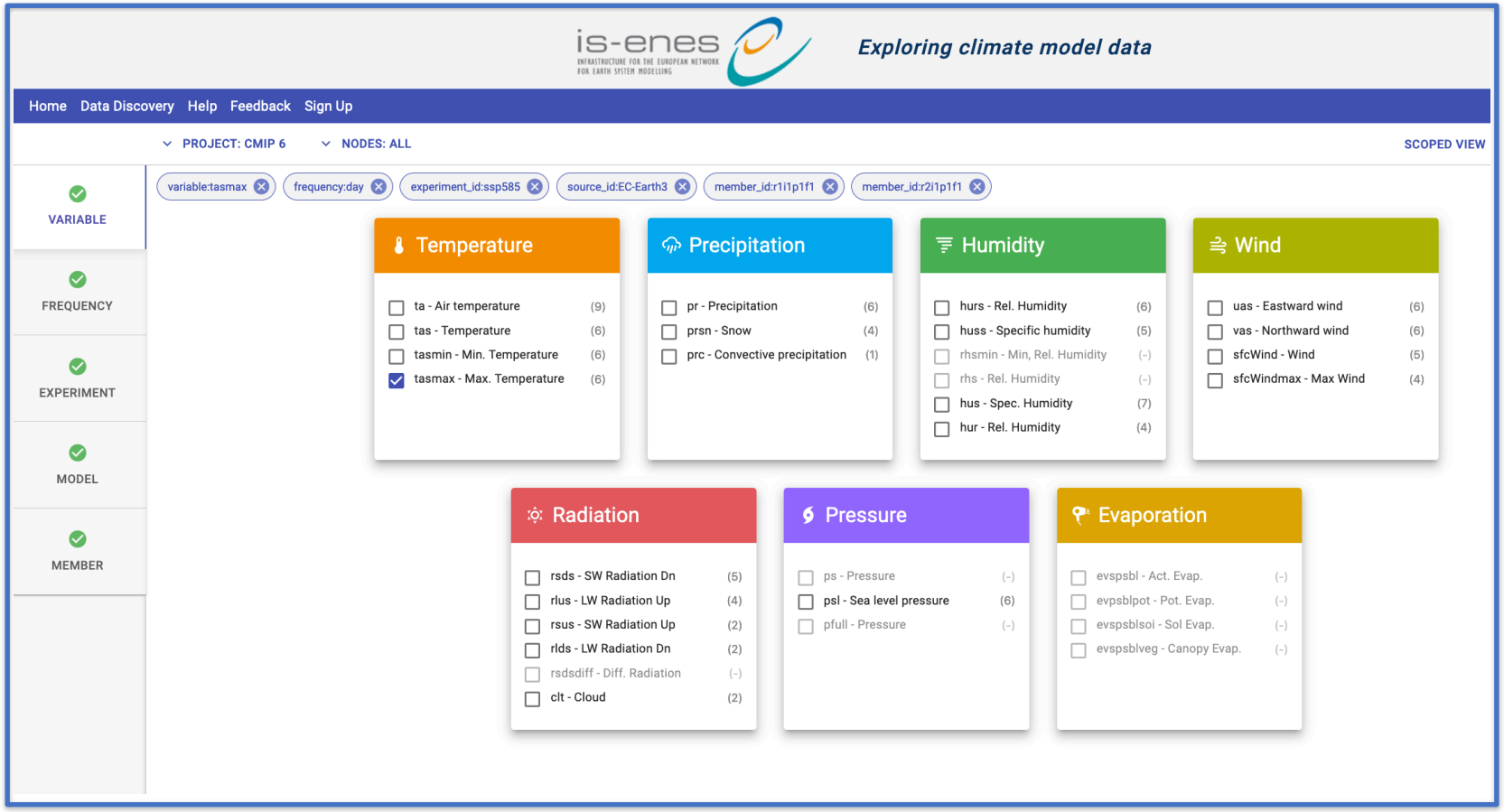
1 - Send an email to c4i@knmi.nl specifying affiliation and motivation to become a Registered User

2 - Make sure the email is associated with at least one of your GitHub, ORCID, ESGF-CEDA accounts

3 - Once approved, login to C4I via the ESGF SSO

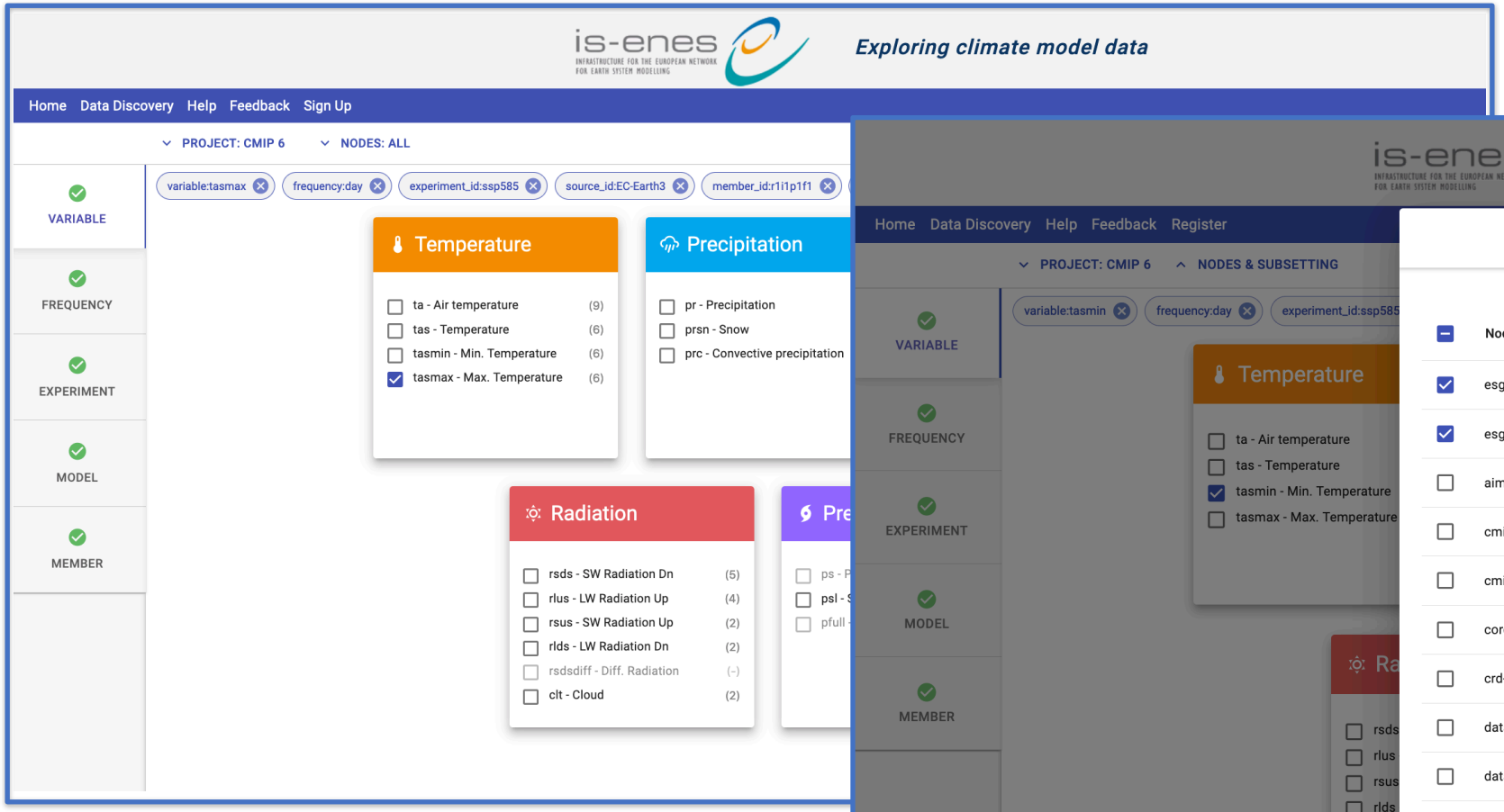
C4I v2 - Search and Nodes

Search Parametrisation made easier

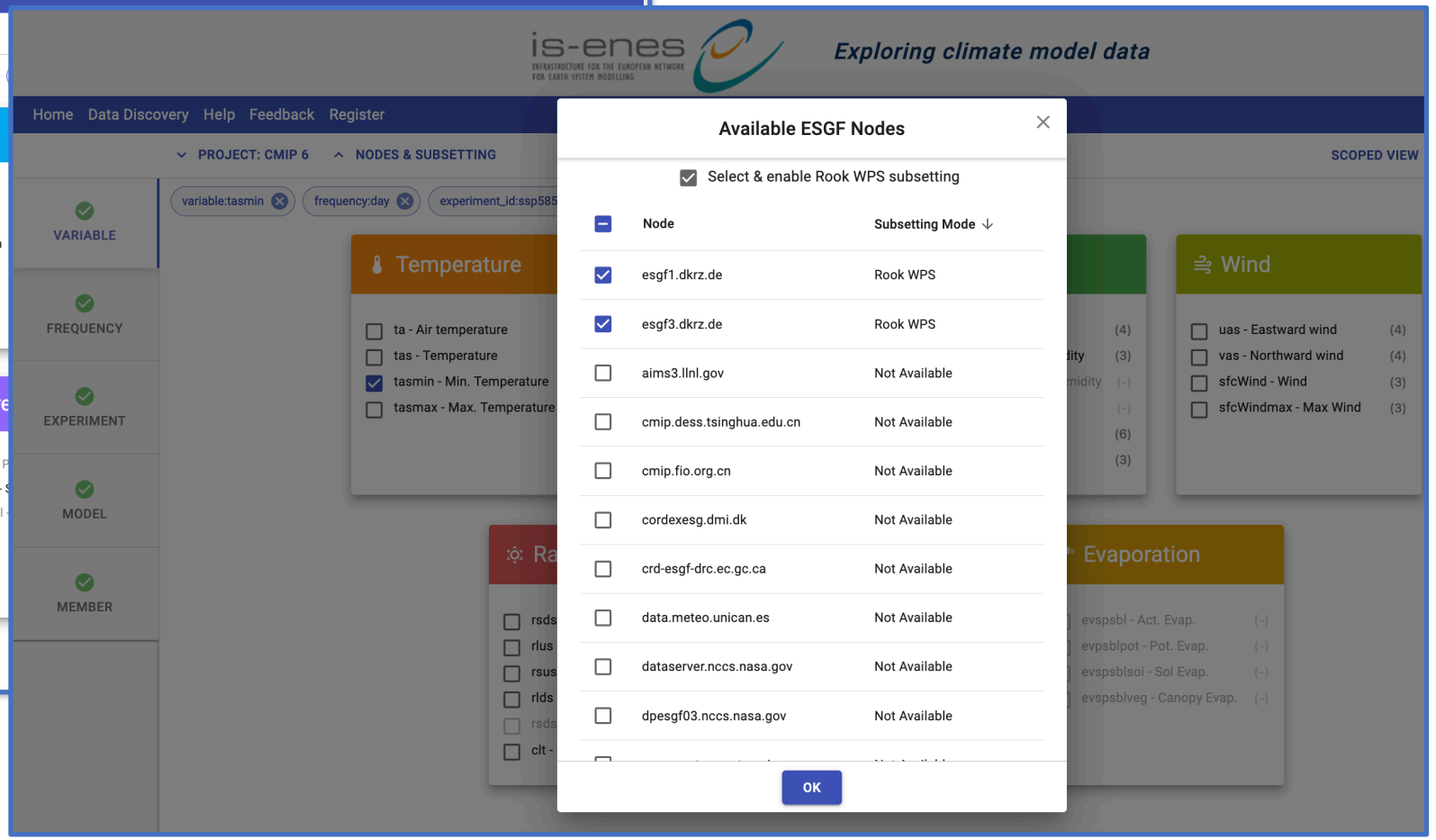


C4I v2 - Search and Nodes

Search Parametrisation made easier



Data Nodes Selection by Service



ESMValTool in C4I

Models' Performance Comparison

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

Exploring climate model data

Home Data Discovery Help Feedback Sign Up

PROJECT: CMIP 6 NODES: CUSTOM (2)

variable:ta variable:prsn variable:huss frequency:day experiment_id:ssp585 source_id:CanESM5

VARIABLE
FREQUENCY
EXPERIMENT
MODEL
MEMBER

Model

- [x] CanESM5 - CanESM5
- [ ] MPI-ESM1-2-LR - MPI-ESM1.2-LR
- [ ] UKESM1-0-LL - UKESM1.0-N96ORCA1
- [x] CNRM-CM6-1 - CNRM-CM6-1
- [x] MIROC6 - MIROC6
- [ ] CNRM-ESM2-1 - CNRM-ESM2-1
- [ ] MRI-ESM2-0 - MRI-ESM2.0
- [ ] GFDL-CM4 - GFDL-CM4
- [ ] HadGEM3-GC31-LL - HadGEM3-GC3.1-N96ORCA1
- [ ] MPI-ESM1-2-HR - MPI-ESM1.2-HR
- [ ] INM-CM4-8 - INM-CM4-8
- [ ] INM-CM5-0 - INM-CM5-0
- [ ] ACCESS-CM2 - Australian Community Climate and Earth System Simulator Climate Model
- [ ] AWI-CM-1-1-MR - AWI-CM 1.1 MR
- [ ] NorESM2-LM - NorESM2-LM (low atmosphere-medium ocean resolution, GHG concentrati
- [ ] BCC-CSM2-MR - BCC-CSM 2 MR
- [ ] CMCC-CM2-SR5 - CMCC-CM2-SR5
- [ ] FGOALS-g3 - FGOALS-g3

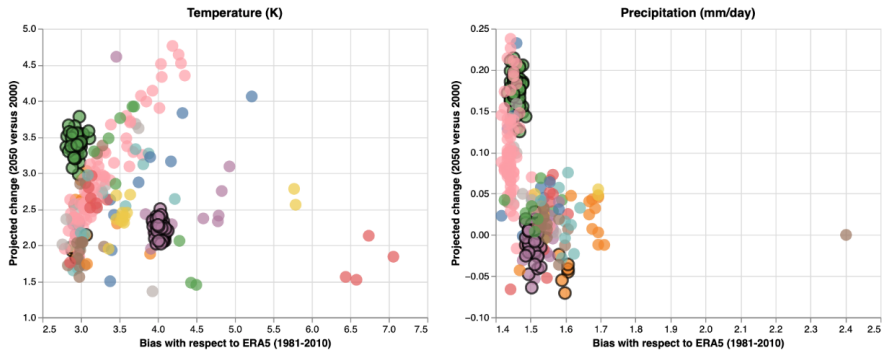
COMPARE MODEL PERFORMANCE

Climate impact result viewer

This applications shows a results from CMIP5 and CMIP6 models, calculated with ESMValTool. It is intended to provide some guidance for climate impact researchers, to select one or more datasets that adequately sample the spread of the CMIP ensemble.

- Bias is calcated with respect to the ERA5 reanalysis dataset over the period 1981-2015.
- Future change is calculated for 2036-2065 as compared to 1986-2015.
- Area is set to Europe (lon 0-39; lat 30-76.25)
- All data are taken from the RCP/SSP 8.5 scenario

Hold ctrl to pan and zoom, hold alt to select a range (points will be highlighted in both graphs), then hold shift to select multiple points.



Project_project CMIP6

Selected datasets:

- CNRM-CM6-1
- MIROC6
- CanESM5

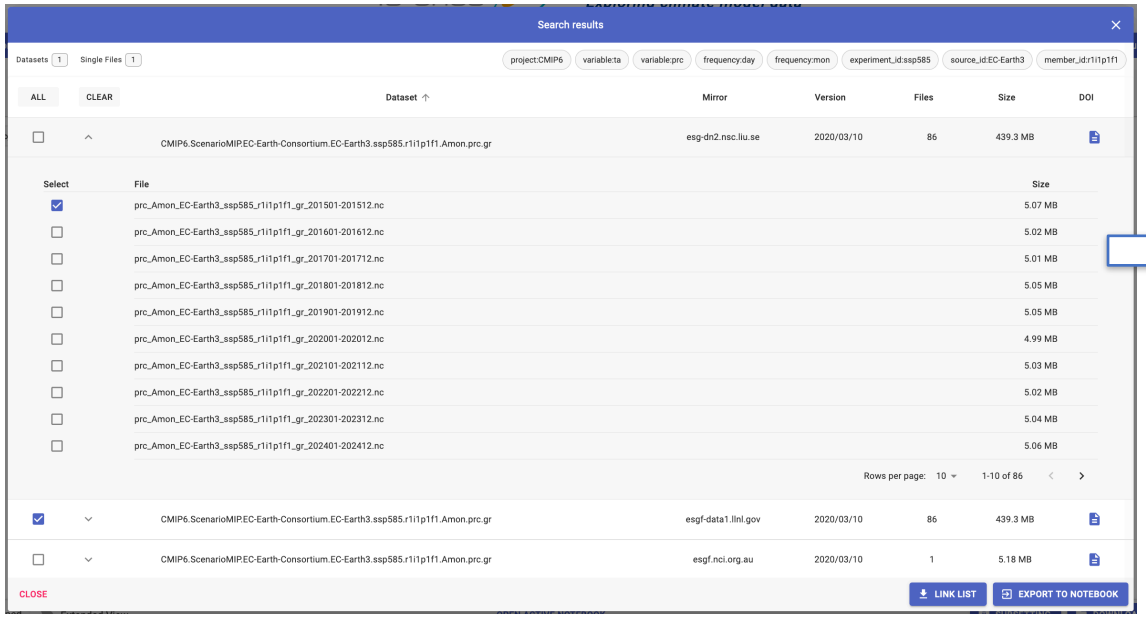
Netherlands eScience Center

View static recipe output

C4I v2 Local Download

Climate4Impact Search for CMIP5/6 (soon CORDEX) Data

https://climate4impact.eu



Local Download via Metalink

https://en.wikipedia.org/wiki/Metalink

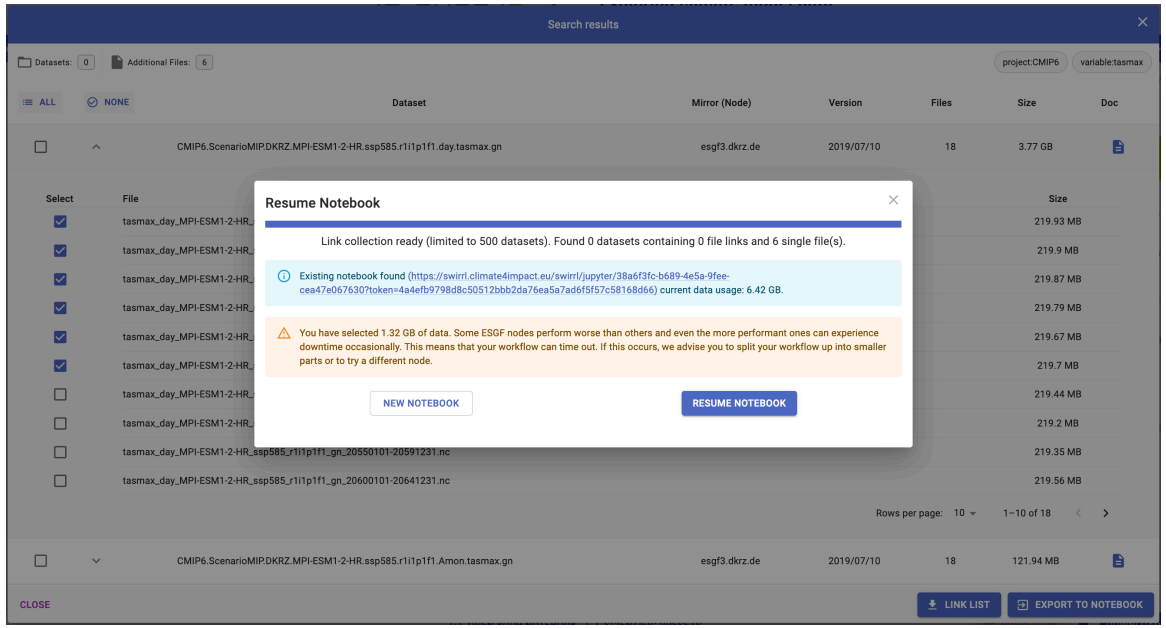
```
<metalink xmlns="urn:ietf:params:xml:ns:metalink">
    <file name="tasmin_day_EC-Earth3_ssp585_r1i1p1f1_gr_20150101-20151231.nc">
        <size>140066295</size>
        <hash type="SHA256">6c54d27ecc06771fd0f593bba68df5cde31e144058149914295208 8ae126213e</hash>
        <url priority="1">http://esgf3.dkrz.de/thredds/fileServer/cmip6/ScenarioMIP/EC-Earth-Consortium/EC-Ea
    </file>

    <file name="tasmin_day_EC-Earth3_ssp585_r1i1p1f1_gr_20160101-20161231.nc">
        <size>140536230</size>
        <hash type="SHA256">eb19230d74844f4b122ef144a51b8805c569f1f66e934824c8659bdbc09c0a2d</hash>
        <url priority="1">http://esgf3.dkrz.de/thredds/fileServer/cmip6/ScenarioMIP/EC-Earth-Consortium/EC-Ea
    </file>

    <file name="tasmin_day_EC-Earth3_ssp585_r1i1p1f1_gr_20170101-20171231.nc">
        <size>140142064</size>
        <hash type="SHA256">46420a196bd05aea989966bb49137ad0be93c784fef9aa3f64b815eb31984ed9</hash>
        <url priority="1">http://esgf3.dkrz.de/thredds/fileServer/cmip6/ScenarioMIP/EC-Earth-Consortium/EC-Ea
    </file>

</metalink>
```

- Scriptable (WGET/cURL)

C4I v2 Workflows and Workspaces

Only for registered and validated accounts!
Cost-effective and secure

Full Data Staging…

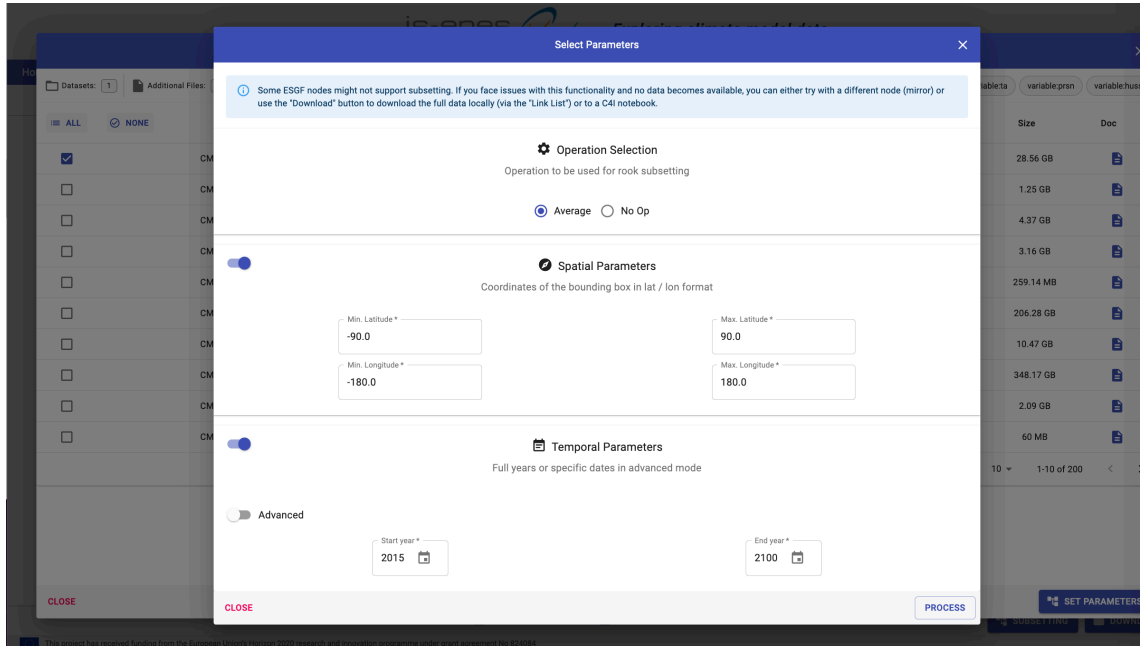


Workspaces offer very large storage

Full Data staging allows 3GB per request
(prevents timeout, given the occasional high load on the nodes)

&

…Subsetting onto personal workspaces

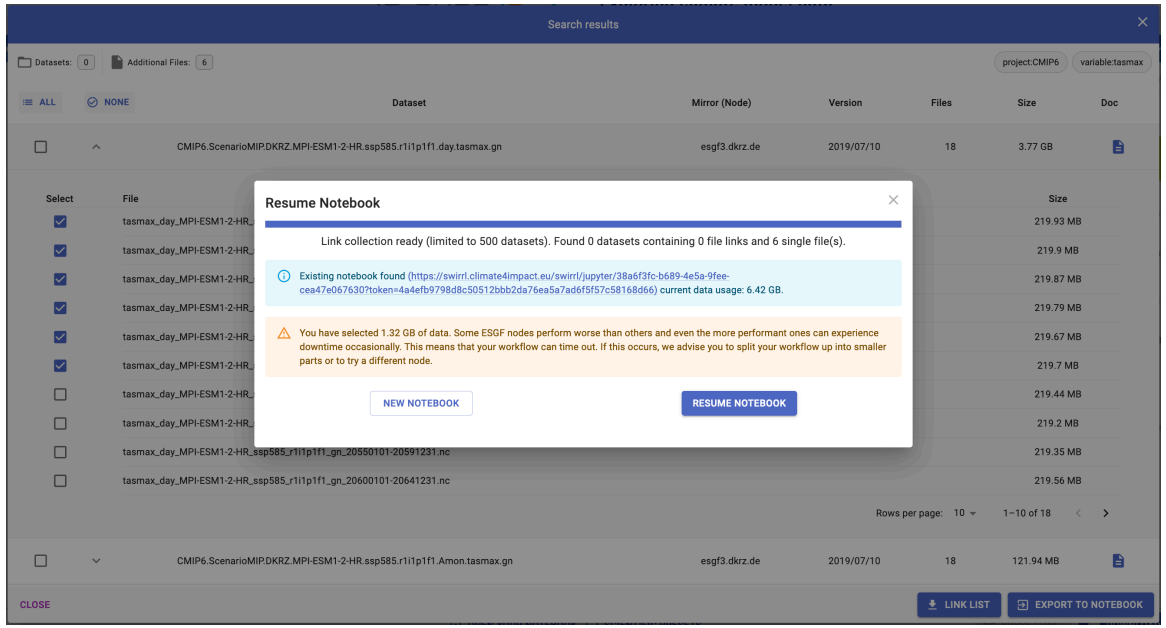


The remote subsetting option, where available, helps reducing the requested data based on user's needs

C4I v2 Workflows and Workspaces

Only for registered and validated accounts!
Cost-effective and secure

Full Data Staging…

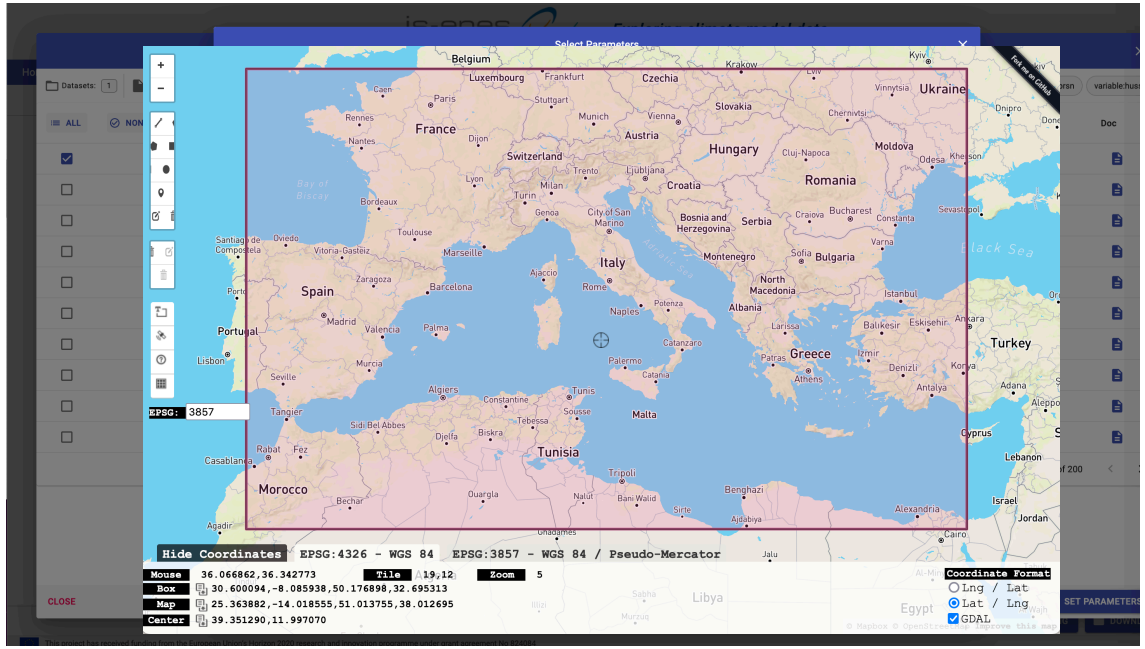


Workspaces offer very large storage

Full Data staging allows 3GB per request
(prevents timeout, given the occasional high load on the nodes)

&

…Subsetting onto personal workspaces

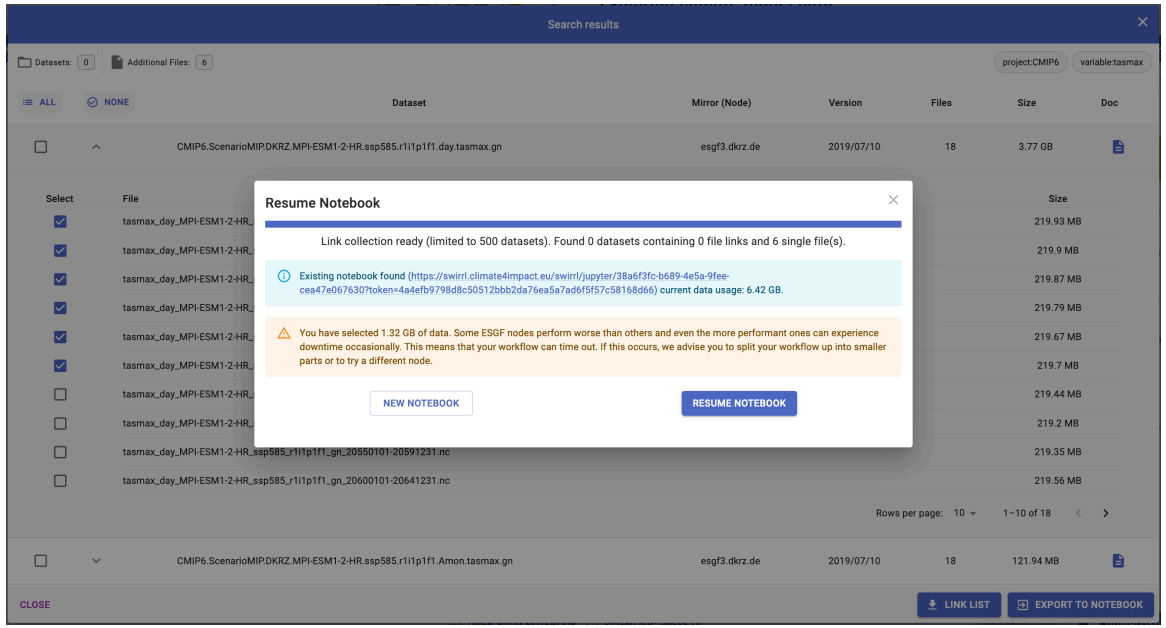


The remote subsetting option, where available, helps reducing the requested data based on user's needs

C4I v2 Workflows and Workspaces

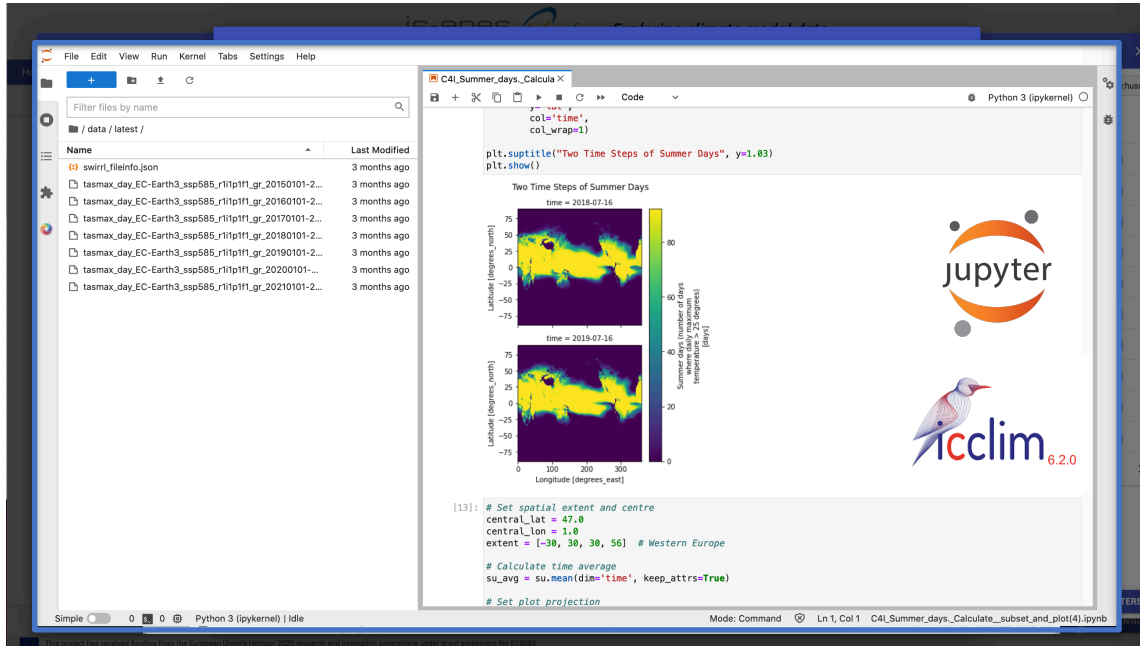
Only for registered and validated accounts!
Cost-effective and secure

Full Data Staging…



&

…Subsetting onto personal workspaces



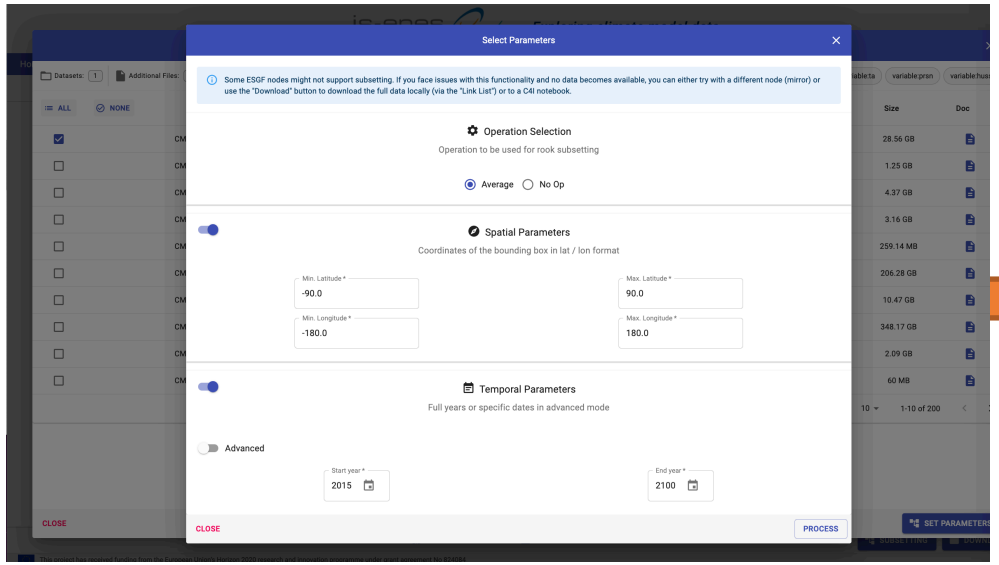
Workspaces offer very large storage

Full Data staging allows 3GB per request
(prevents timeout, given the occasional high load on the nodes)

The remote subsetting option, where available, helps reducing the requested data based on user's needs

SWIRRL - VRE API

Manage **data-driven** collaborative **workspaces**, combining reproducible **Notebooks, Workflows and Visualisation** tools on the Cloud



SWIRRL

https://doi.org/10.1162/dint_a_00129
https://gitlab.com/KNMI-OSS/swirrl

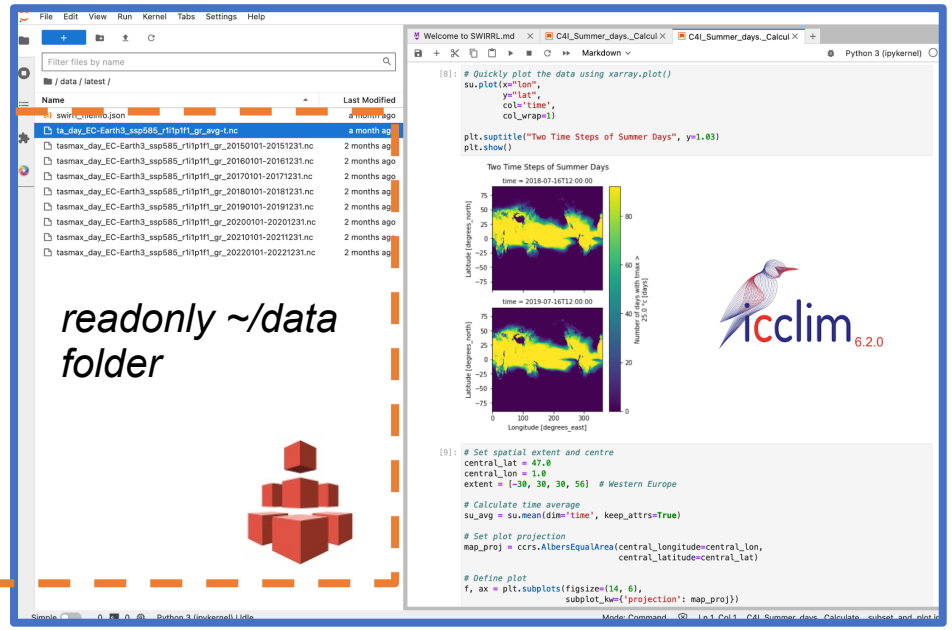
ESGF/WPS

Amazon EKS

Kubernetes

Common Preprocessing Workflows & Staging

readonly ~/data folder



icclim 6.2.0

AWS Fargate

icclim sample notebooks

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

File Edit View Run Kernel Tabs Settings Help

Notebook idle

Github

LOGIN

Please review your access using this link to revoke your access tokens.

Create Snapshot

Snapshot name SNAPSHOT

Data staging

ROLL BACK

Activities

LOAD ACTIVITIES

Type	Created at	Action
Rookwps WF	2022-11-29 14:09	
Succeeded		
Download WF	2022-11-22 18:21	
Succeeded		
Create	2022-11-22 18:17	RESTORE

Welcome to SWIRRL.md | C4I_Summer_days._Calcul | C4I_Summer_days._Calcul

Code

ICCLIM C4I: Calculate the num

Example notebook that runs ICCLIM, which is pre-

The example calculates the number of summer da

The data is read using xarray and a plot of the tim

output types examples are shown.

The dataset that is expected for this notebook are

and experiment as well as one member. The time p

To keep this example fast to run, the following per

Preparation of the needed modules

[1]:
```
import icclim

import sys
import glob
import os
import datetime
import cftime

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import xarray as xr
import cartopy.crs as ccrs

print("python: ",sys.version)
print("numpy: ", np.__version__)
print("xarray: ", xr.__version__)
print("pandas: ", pd.__version__)
print("icclim: ", icclim.__version__)
print("cftime: ", cftime.__version__)
```

```
python:  3.10.6 | packaged by conda-forge
numpy:  1.23.5
xarray:  2022.11.0
```

IS-ENES Climate Data Infrastructure for Climate 4 Impact › C4I Use Cases as Jupyter Notebooks

C

C4I Use Cases as Jupyter Notebooks

Project ID: 25761638

Star 0

27 Commits 1 Branch 0 Tags 6.7 MB Project Storage

A collection of Jupyter Notebooks implementing some Use Cases.

Updated to match icclim 5.0.2.
Christian Pagé authored 1 year ago 5825906e

master notebooks Find file Clone

README

Name	Last commit	Last update
C4I_Averaged_Temperature_Anomaly_2...	Updated with icclim 5.0.2.	1 year ago
C4I_Summer_days._Calculate__subset_a...	Updated with icclim 5.0.2.	1 year ago
C4I_TX90p._Calculate__subset_and_plot...	Updated with icclim 5.0.2.	1 year ago
C4I_deltaT_deltaP_Anomaly_2081-2100_...	Updated with icclim 5.0.2.	1 year ago
C4I_userindex_maxtemp_freezing.ipynb	Updated with icclim 5.0.2.	1 year ago
README.md	Updated to match icclim 5.0.2.	1 year ago
cordex_download.ipynb	Adjustments to dask config.	1 year ago

README.md

Reproducibility Controls

Workflow Monitoring

GitHub Authentication

Snapshot Controls

Data Staging Rollback

Activities History and Provenance

File Edit View Run Kernel Tabs Settings Help

Notebook idle

Github

aspinuso

SWITCH USER

Please review your access using this link to revoke your access tokens.

Create Snapshot

Snapshot name

SNAPSHOT

Data staging

ROLL BACK

Activities

LOAD ACTIVITIES

Type	Created at	Action
Library Update	2023-01-12 13:48	RESTORE
`pip install mpg`		
Rookwps WF	2022-11-29 14:09	
Succeeded		
Download WF	2022-11-22 18:21	
Succeeded		
Create	2022-11-22 18:17	RESTORE

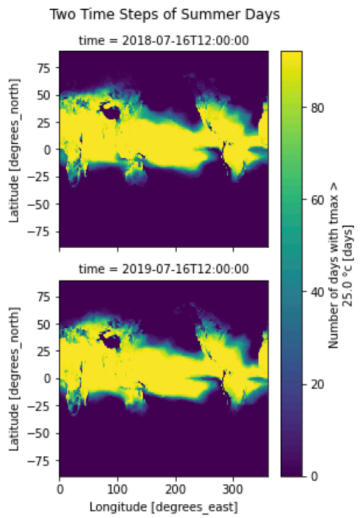
jovyan@jupyter-38a6f3fc-l | C4I_Summer_days._Calcul

Markdown | Python 3 (ipykernel)

2019-07-16 12:00:00	43.859545	3.515625	2.0	48.0

```
[7]: # Time subsetting: this is just an example on how to do it
start_date = "2018-01-01"
end_date = "2019-12-31"

su = su_xr["SU"].sel(time=slice(start_date, end_date))
```

```
[8]: # Quickly plot the data using xarray.plot()
su.plot(x="lon",
        y="lat",
        col='time',
        col_wrap=1)

plt.suptitle("Two Time Steps of Summer Days", y=1.03)
plt.show()
```



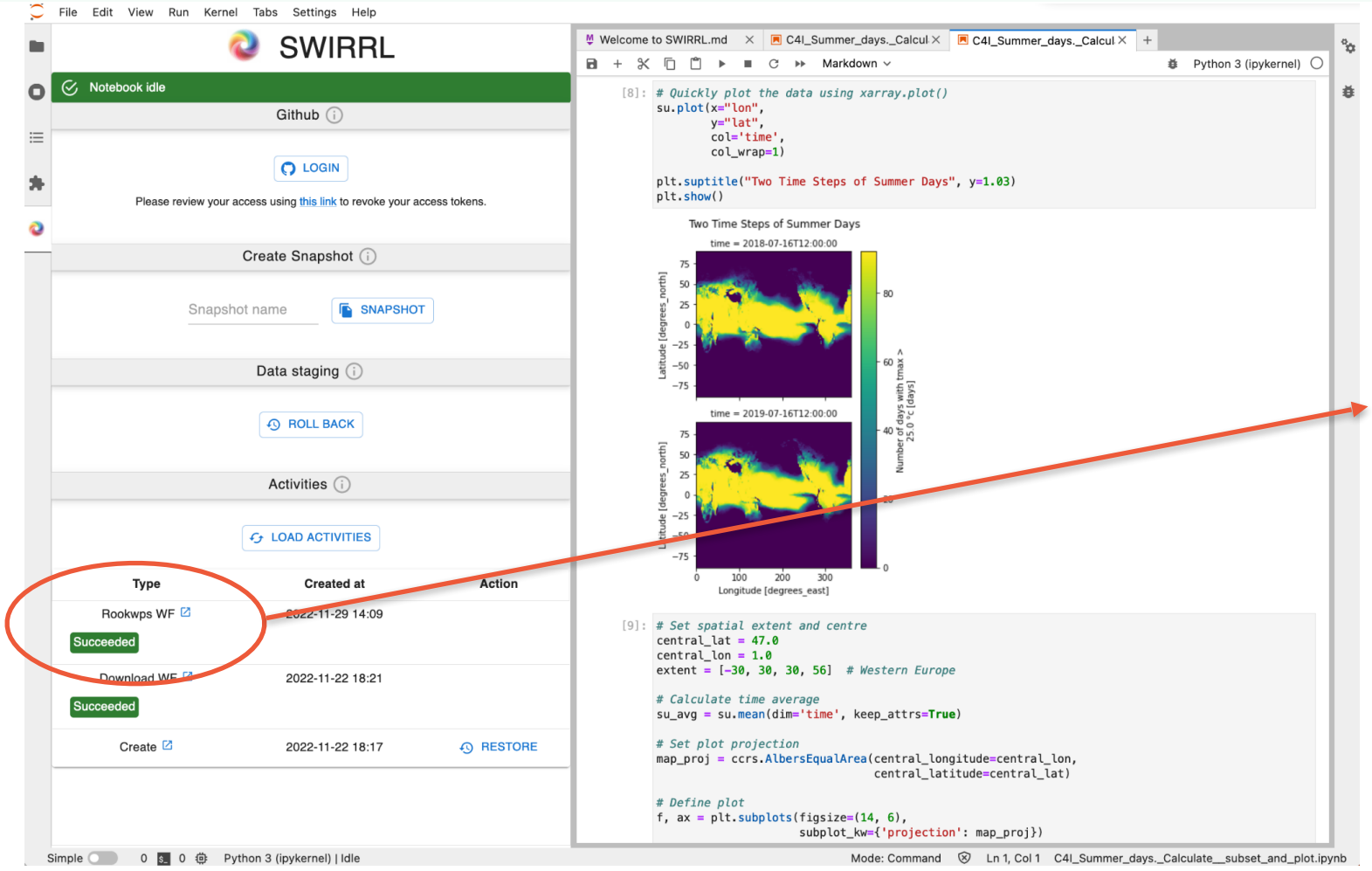

```
[9]: # Set spatial extent and centre
central_lat = 47.0
central_lon = 1.0
extent = [-30, 30, 30, 56]  # Western Europe
```

Simple | 1 | 0 | Python 3 (ipykernel) | Idle | Mode: Command | Ln 1, Col 1 | C4I_Summer_days._Calculate__subset_and_plot.i

Users create intermediate and shareable snapshots on GitHub,
Binders, compatible with https://mybinder.org/

Rollback workflows and Restore environments to a previous state

Reproducibility Controls



automated provenance management and data-versioning, fostering traceability and reproducibility.

Help Material and Feedback Form

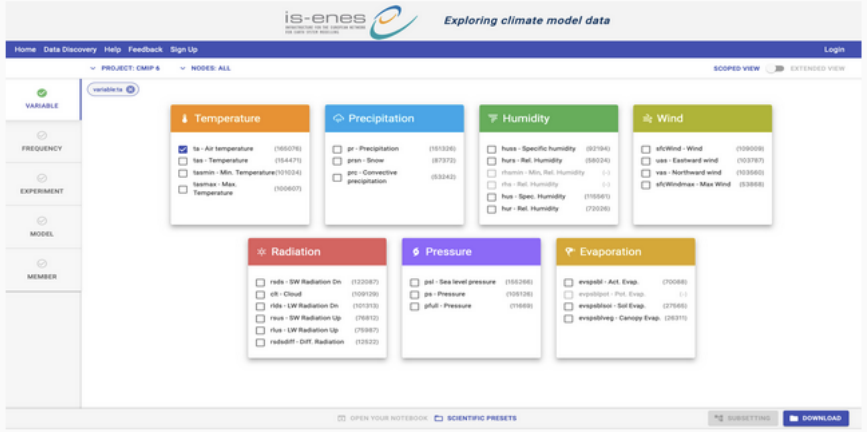
is-enes — Exploring climate model data

Home Data Discovery Help Feedback Register 0000-0002-0077-8491

How do I search for data?

In the ***Data Discovery Page*** you select the project you want to access the data from (CMIP6 - CMIP5 - CORDEX).

Once you have selected the project, the portal will display a page where you can compose queries interactively by combining different properties. These are shown in separated tabs on the top of the page (*variable*, *frequency*, *experiment*, *model*, *member*) and can be configured in any order.



When configuring a specific property, to facilitate identifying the most relevant options, a subset of the possible values is presented, grouped by different topics. This is the *Scoped* view. This presentation mode can be changed by switching the knob at the top left of the screen, from *Scoped* to *Extended*. The *Extended* presentation mode will display, for each property, all the possible values available for the current query configuration.

In the image above we show that the summary of the current query is always displayed on the top of the screen. Each value can be interactively removed. Adding or removing a property value will always result in the live update of the available choices for each property.

When selecting the Model, by clicking on ***Compare Model Performance*** at the bottom of the Model panel, users will be prompted with a page showing how the selected models (and the others available) foresee a change on a variable (Temperature/Precipitaton) compared to the past. The page is generated by using the ESMValTool and, in combination with the bias, gives some confidence on the performance.

Help Material and Feedback Form

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

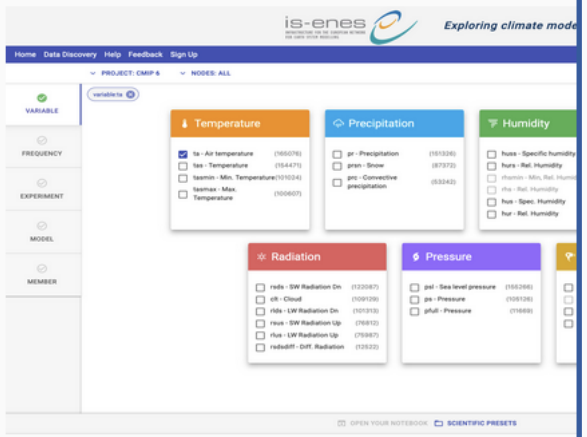
Exploring climate model data

Home Data Discovery Help Feedback Register 0000-0002-0077-8491

How do I search for data?

In the ***Data Discovery Page*** you select the project you want to access the d

Once you have selected the project, the portal will display a page where you
(*variable, frequency, experiment, model, member*) and can be configured in



When configuring a specific property, to facilitate identifying the most releva
can be changed by switching the knob at the top left of the screen, from *Sc*
configuration.

In the image above we show that the summary of the current query is alway
the live update of the available choices for each property.

When selecting the Model, by clicking on ***Compare Model Performance*** at
foresee a change on a variable (Temperature/Precipitaton) compared to the

Home Data Discovery Help Feedback Register 0000-0002-0077-8491

Data Management: the /data folder

Thanks to the support for workflows that perform data-staging operations from remote providers, workspace are populated with the data you are interested in.

SWIRRL manages the data in the ***/data*** folder, which is accessible from your ***JupyterLab** instance. To make sure that this folder is always consistent and your analysis reproducible, this is set as *'read only'*. However, you can always request to add more data.

/data/latest

Is the data that you have requested already available?

Check your ***/data/latest*** folder! Here you will find the most updated version the data available for you to use.

Not there yet? Click on the SWIRRL Explorer

In this handy control panel you can check whether there are workflows running in the background. For instance, one might be copying data to your workspace just now.

You will read more about the functionalities of SWIRRL Explorer in this short guide!

/data/staginghistory

Each time data is staged to your workspace, SWIRRL keeps track of the changes by maintaining a staging history. You will find this in the ***/data/staginghistory*** folder.

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

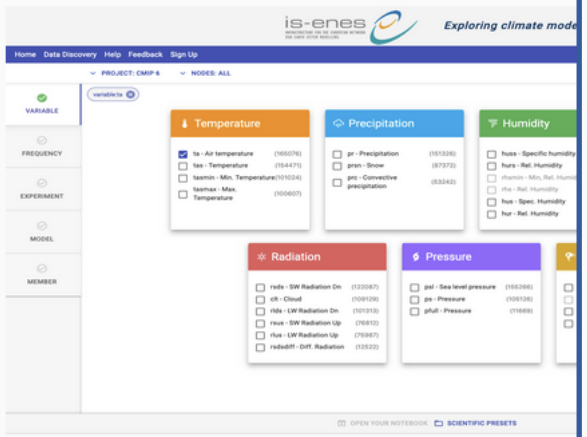
Help Material and Feedback Form

is-enes
Exploring climate model data
Home Data Discovery Help Feedback Register 0000-0002-0077-8491

How do I search for data?

In the ***Data Discovery Page*** you select the project you want to access the d

Once you have selected the project, the portal will display a page where you
(*variable, frequency, experiment, model, member*) and can be configured in



When configuring a specific property, to facilitate identifying the most releva
can be changed by switching the knob at the top left of the screen, from *Sc*
configuration.

In the image above we show that the summary of the current query is alway
the live update of the available choices for each property.

When selecting the Model, by clicking on ***Compare Model Performance*** at
foresee a change on a variable (Temperature/Precipitaton) compared to the

Home Data Discovery Help Feedback Register 0000-0002-0077-8491

Data Management: the /data folder

Thanks to the support for workflows that perform data-staging operations from remote providers, workspace are populated with the

SWIRRL manages the data in the ***/data*** folder, which is accessible from your ***JupyterLab*** instance. To make sure that this folder is a
'read only'. However, you can always request to add more data.

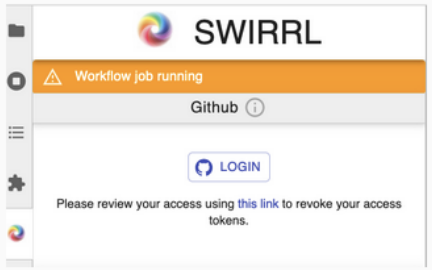
/data/latest

Is the data that you have requested already available?

Check your ***/data/latest*** folder! Here you will find the most updated version the data available for you to use.

Not there yet? Click on the SWIRRL Explorer

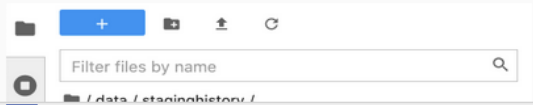
In this handy control panel you can check whether there are workflows running in the background. For instance, one might be copying



You will read more about the functionalities of SWIRRL Explorer in this short guide!

/data/staginghistory

Each time data is staged to your workspace, SWIRRL keeps track of the changes by maintaining a staging history. You will find this in



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

Email *

Il tuo indirizzo email

What is your function? *

This helps us get a better understanding of whom our end users are.

- [ ] Climatologist
- [ ] Impact Researcher
- [ ] Hydrologist
- [ ] Geologist
- [ ] Altro:

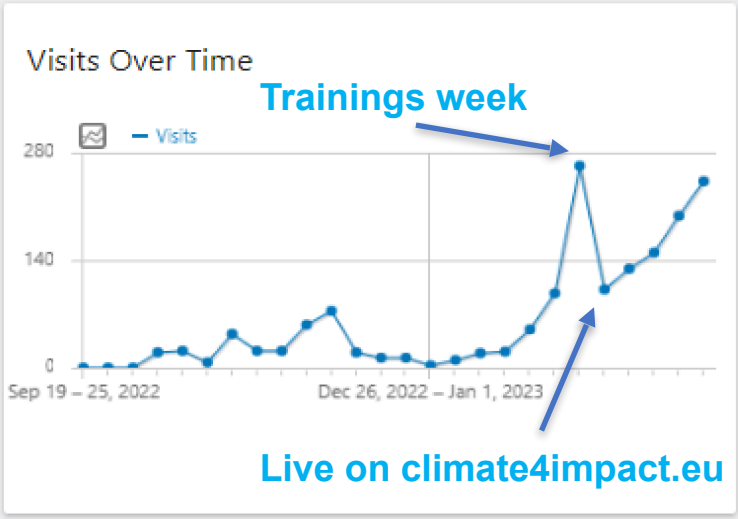
What kind of feedback to you have? *

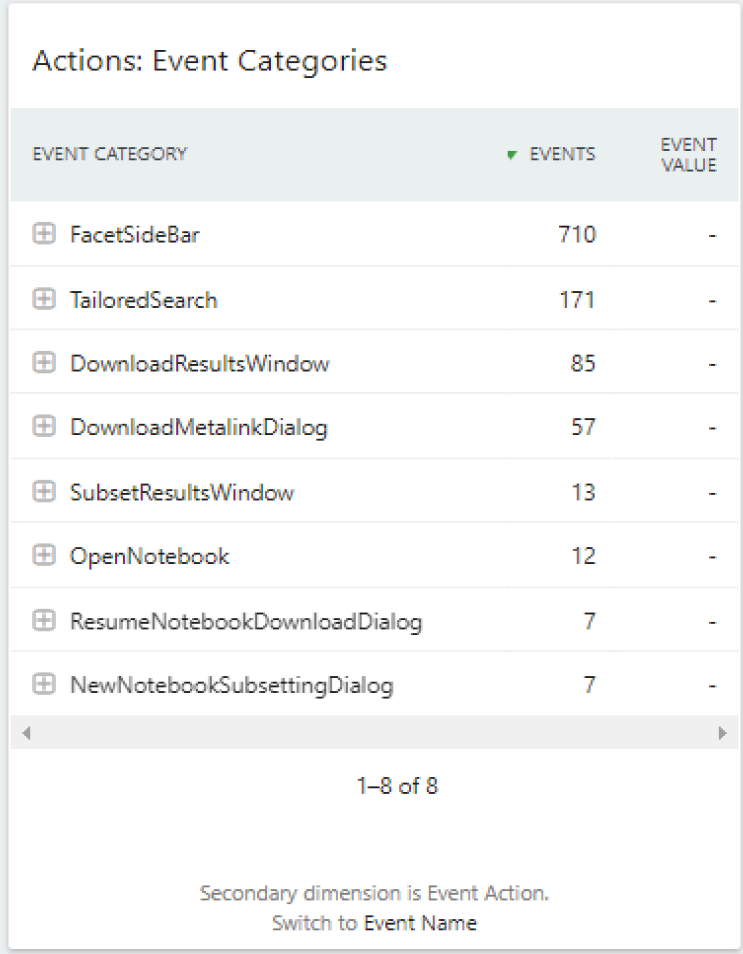
This helps us organize the feedback we receive.

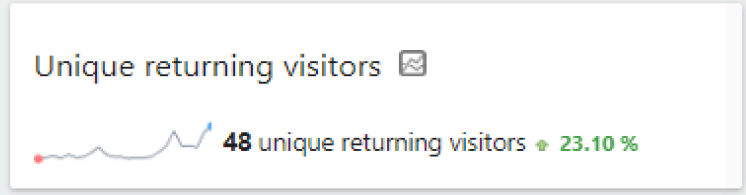
- ( ) Feature Request
- ( ) Bug Report

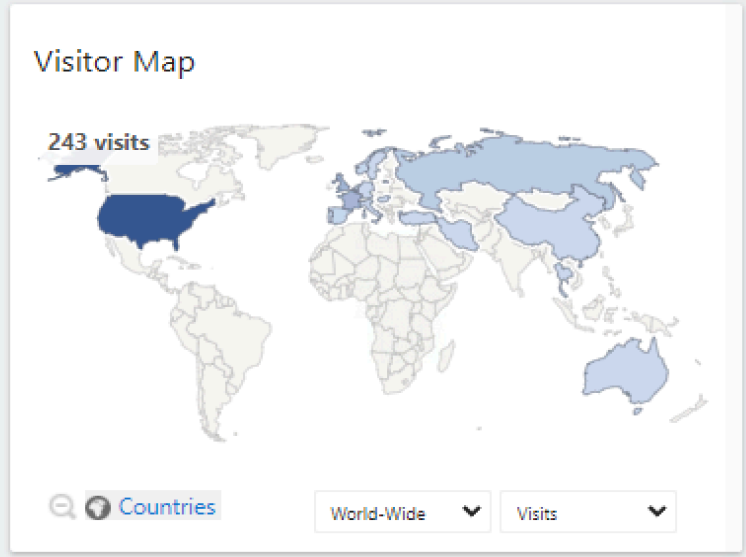
Web Analytics and Computational KPIs

GDPR Compliant Web analytics service (Matomo)





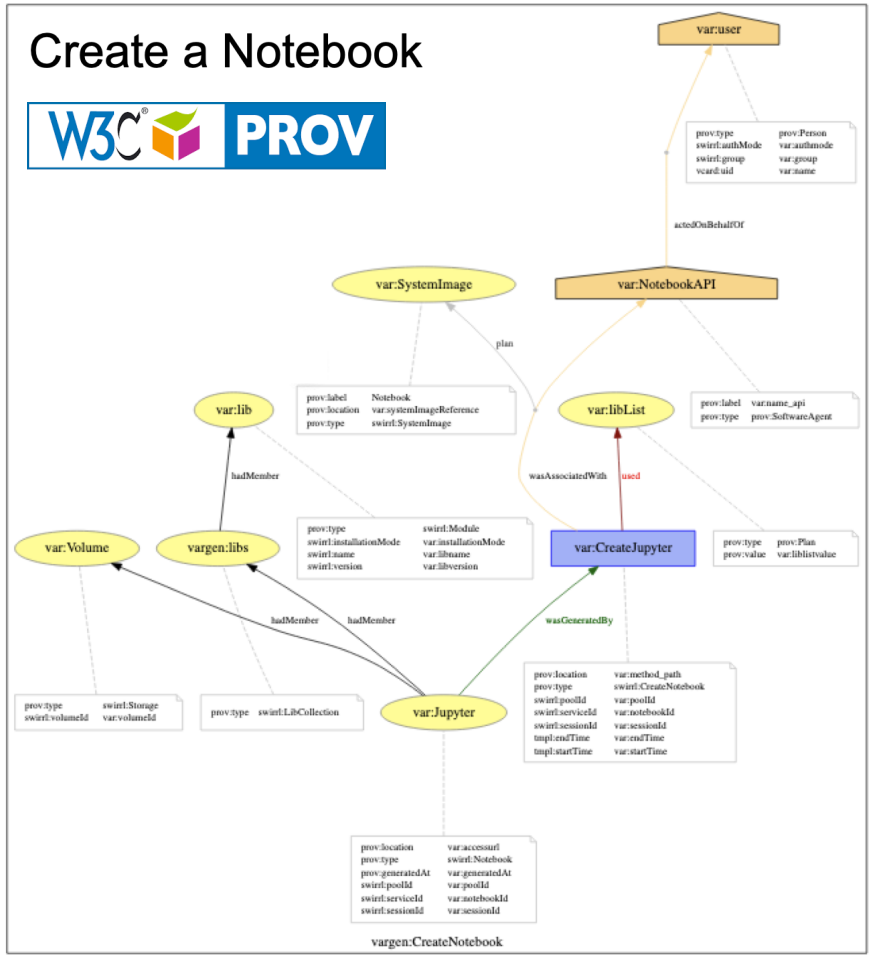


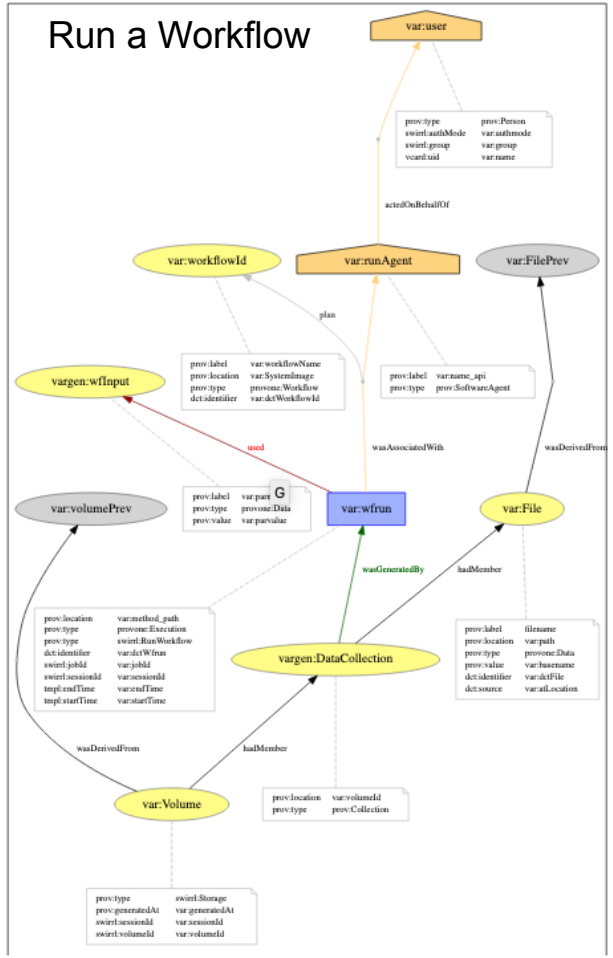


Tracking at Event Level allows to get useful insights at a very high granularity.

Web Analytics and Computational KPIs

Provenance used to Collect Computational KPIs (examples)





Notebook stats:

session_count	notebook_count
34	34

Workflow stats:

wf_types	run_count	success %
download	98	82%
rookswps	18	72%

Staged-Data

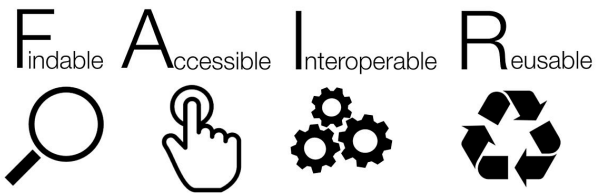
Total Data Files	Unique Data Fies
1104	949

>90 Users Registered via the new ESGF SSO system at CEDA

Climate4Impact v2

Improvements from v1

- **Search Usability**
- **Models Performance Comparison** (ESMValTool)
- **Single Sign On** with ESGF
- **Flexible analysis features** (Workspaces with ICCLIM Notebooks, Data Staging & Subsetting Workflows)
- **Automated reproducibility mechanisms fostering FAIR research**



- **Decoupled Architecture** (SWIRRL/SSO/WPS)

Ongoing and Future work.....

- **Access to CORDEX** *with ESGF SSO supported at more CORDEX nodes* (ongoing)

- **More ESGF nodes to implement Subsetting WPS** (e.g. CMCC coming soon)

- **More Notebooks Use Cases** using ***icclim***
  https://gitlab.com/is-enes-cdi-c4i/notebooks

- **Scientific Guidance produced by experts**

- **Use SWIRRL to create Collaborative Data Caches**
(frequently requested by lecturers for educational purposes)