

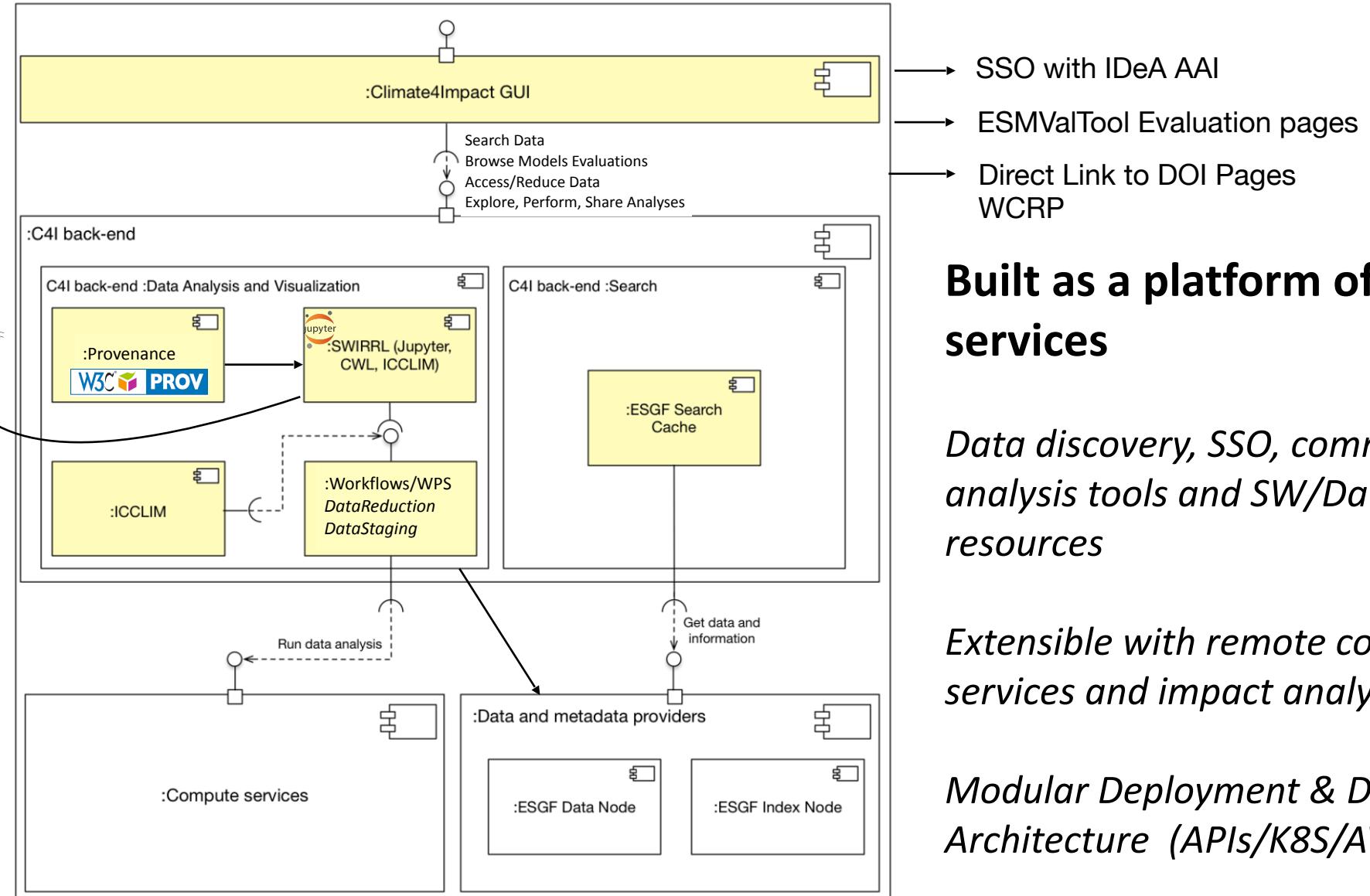
Climate4Impact v2

enhance the use of climate research data and methods

Alessandro Spinuso, Mats Veldhuizen,
Ian van der Neut,
Christian Pagé



Climate4Impact (v2) - Redesign



Built as a platform of integrated services

Data discovery, SSO, common workflows, analysis tools and SW/Data management resources

Extensible with remote computational services and impact analysis use cases

Modular Deployment & Decoupled Architecture (APIs/K8S/AWS)

Climate4Impact (v2) - Profiles and SSO

The figure illustrates the two user profiles available on Climate4Impact (v2) and the login process through the Earth System Grid Federation (ESGF).

Guest Profile: Limited features. Includes checkboxes for "Browse ESGF Data" and "Download file links".

Registered User Profile: Full feature access. Includes checkboxes for "Browse ESGF Data", "Download file links", "Access data from [Jupyter](#) in C4I workspace", "Access scientific [notebook presets](#)", "Subset large datasets", "Create [Binder](#) snapshots of your work", and "Trace and rollback operations".

A red arrow points from the "REGISTER" button in the Registered User profile box up towards the ESGF login interface, indicating the transition from registration to logging in.

BROWSE DATA button is located at the bottom left of the Guest profile box.

REGISTER button is located at the bottom right of the Registered User profile box.

ESGF Login Interface: Shows the "Login" screen with fields for "Select your institution", "I don't have an account", and "Sign in with one of these identity providers" (ID, GitHub).

Two User Profiles

Registration procedure

1 - Send us an email specifying affiliation and motivation to become a Registered User

2 - Once approved, login to C4I via CEDA/ESGF (SSO) using the same emailing account.

Climate4Impact (v2) - Nodes Selection by Service

Explicit Nodes Selection with Service Indication

The screenshot shows the is-enes Climate4Impact interface. On the left, there are filters for VARIABLE (tasmax selected), FREQUENCY (day selected), EXPERIMENT (ssp585 selected), MODEL, and MEMBER. The main search bar has filters: PROJECT: CMIP 6, NODES: ALL, variable:tasmax, frequency:day, experiment_id:ssp585, source:*. The results for Temperature show options like ta - Air temperature, tas - Temperature, tasmin - Min. Temperature, and tasmax - Max. Temperature. The results for Radiation show options like rsds - SW, rlus - LW, rsus - SW, rlds - LW, rsdsdiff - SW, and clt - Cloud. A modal window titled "Available ESGF Nodes" is open, showing a list of nodes and their subsetting modes. The modal title includes a checked checkbox for "Select & enable Rook WPS subsetting". The list includes:

Node	Subsetting Mode
esgf1.dkrz.de	Rook WPS
esgf3.dkrz.de	Rook WPS
aims3.llnl.gov	Opendap
cmip.dess.tsinghua.edu.cn	Opendap
cmip.fio.org.cn	Opendap
cordexesg.dmi.dk	Opendap
crd-esgf-drc.ec.gc.ca	Opendap
data.meteo.unican.es	Opendap
dataserver.nccs.nasa.gov	Opendap
dpesgf03.nccs.nasa.gov	Opendap
esg-cccr.tropmet.res.in	Opendap

On the right, there are sections for Wind and Evaporation, each with a list of variables. At the bottom of the modal is an "OK" button.

Nodes listed with subsetting support and mode.

Can scale to more services as they are implemented across the network

Climate4Impact (v2) WPS Integration

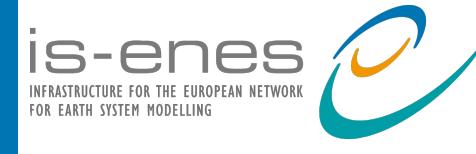
Search for CMIP5/6 data
(focussed & extended variable selection modes)

The screenshot shows the Climate4Impact search interface for CMIP6 data. On the left, a sidebar lists filters: VARIABLE, FREQUENCY, EXPERIMENT, MODEL, and MEMBER. The main area displays variables grouped by category: Temperature, Precipitation, Humidity, Wind, Radiation, Pressure, and Evaporation. Each category has a list of variables with counts in parentheses. A blue arrow points from this interface to the right-hand WPS parametrisation interface.

WPS parametrisation for data pre-processing and staging onto Computational Notebooks

The screenshot shows the WPS parametrisation interface. It includes a 'Select Parameters' dialog with sections for Operation Selection (set to Average), Spatial Parameters (bounding box: Min. Latitude -90.0, Max. Latitude 90.0, Min. Longitude -180.0, Max. Longitude 180.0), and Temporal Parameters (Start year 2015, End year 2100). To the right is a file browser listing datasets with their sizes and download options. A blue arrow points from the Climate4Impact search interface to this WPS interface.

Climate4Impact (v2) Workflows & Workspaces



Subsets are staged to a personal workspace

Select Parameters

Some ESGF nodes might not support subsetting. If you face issues with this functionality and no data becomes available, you can either try with a different node (mirror) or use the "Download" button to download the full data locally (via the "Link List") or to a C4I notebook.

Operation Selection

Operation to be used for rook subsetting

Average No Op

Spatial Parameters

Coordinates of the bounding box in lat / lon format

Min. Latitude * -90.0 Max. Latitude * 90.0
Min. Longitude * -180.0 Max. Longitude * 180.0

Temporal Parameters

Full years or specific dates in advanced mode

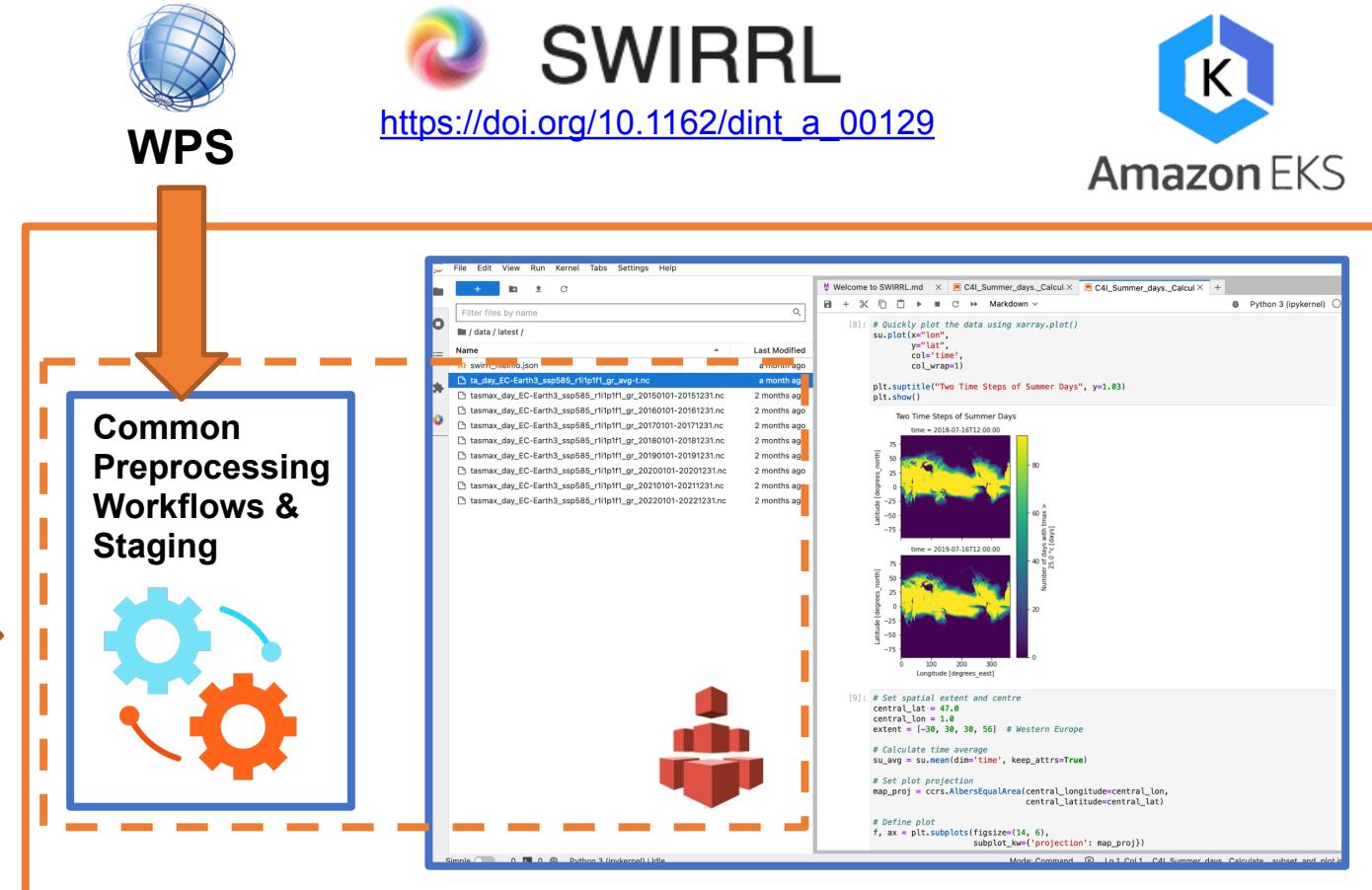
Advanced

Start year * 2015 End year * 2100

PROCESS

SET PARAMETERS

CLOSE



Only for registered accounts!
Cost-effective and secure

SWIRRL allows to setup shared data volumes (with versioning), enabling collaboration



Amazon EKS

Climate4Impact (v2) Workflows & Workspaces

The screenshot shows the SWIRRL interface. On the left, there's a sidebar with 'Notebook idle' status, a 'Github' button, a 'LOGIN' button, 'Create Snapshot' (with a 'SNAPSHOT' button), 'Data staging', 'ROLL BACK', 'Activities' (with a 'LOAD ACTIVITIES' button), and a table of recent actions:

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

A red circle highlights the first row in the table. A red arrow points from this circle to the top of the Jupyter Notebook window. The notebook cell [8] contains Python code to plot summer days, and the resulting heatmap is shown below. Cell [9] contains code to calculate a time average and plot it.

W3C PROV

```
> "prov:used": [...],
  "provone:hadPart": [
    {
      "prov:wasAssociatedWith": [...],
      "@type": [
        "Resource",
        "prov:Activity",
        "provone:Execution"
      ],
      "rdfs:label": "orchestrate",
      "prov:startedAtTime": "2022-11-29T14:13:02Z",
      "@id": "urn:roocs:orchestrate_16ca3e1f-fee6-4419-a264-65d273a801bf",
      "prov:endedAtTime": "2022-11-29T14:13:45Z"
    },
    {
      "prov:wasAssociatedWith": [
        ...
      ],
      "prov:wasActivityOfInfluence": [...],
      "@type": [
        "Resource",
        "prov:Activity",
        "provone:Execution"
      ],
      "rdfs:label": "average_ta_1",
      "roocs:apply_fixes": false,
      "@id": "urn:roocs:average_ta_1_6eabac6b-444e-4e9a-a71b-bcc1b16b9fb1",
      "roocs:dims": "time"
    },
    ...
  ],
  "@type": [...],
  "@context": {...},
  "prov:generated": [...],
  "swirrl:sessionId": "13abfb95-fd4a-463f-ba61-fc96d40a9b6c",
  "swirrl:message": "Succeeded",
  "prov:wasAssociatedWith": [...],
  "@id": "urn:uuid:49f0d7f5-5cc8-4893-8e17-2a18ed870540",
  "swirrl:jobId": "49f0d7f5-5cc8-4893-8e17-2a18ed870540",
  "prov:endedAtTime": "2022-11-29T13:13:54.220Z",
  "prov:atLocation": "POST /workflow/rookwps/run/",
  "prov:startedAtTime": "2022-11-29T13:09:03.676Z"
```

Thanks to the use of PROV by Rooks and SWIRRL, provenance is easily merged, stored and made available to users and machines (interoperable)

Tailored Workspace Environment (icclim)

The screenshot shows a Jupyter Notebook interface with a sidebar containing various workspace management tools like GitHub integration, login, snapshot creation, and activity logs.

ICCLIM C4I: Calculate the number of Summer Days

Example notebook that runs ICCLIM, which is pre-installed in the notebook.

The example calculates the number of summer days (SU indicator) for the dataset chosen by the user on C4I.

The data is read using xarray and a plot of the time series over a specific region is generated, as well as an average spatial map. Several output types examples are shown.

The dataset that is expected for this notebook are tasmax parameter (needed to calculate the SU indicator) for one specific climate model and experiment as well as one member. The time period should be continuous.

To keep this example fast to run, the following period is considered: 2015-01-01 to 2019-12-31, and plots are shown over European region.

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 | (main, Aug 22 2022, 20:35:26) [GCC 10.4.0]
numpy: 1.23.5
xarray: 2022.11.0
```

Tailored Workspace Environment (icclim)

The screenshot displays a Jupyter Notebook interface and a GitHub repository page side-by-side.

Jupyter Notebook Interface:

- Toolbar:** File, Edit, View, Run, Kernel, Tabs, Settings, Help.
- Header:** Notebook idle, Github, LOGIN, Please review your access using [this link](#) to revoke your access tokens.
- Left Sidebar:** Create Snapshot, Snapshot name, Data staging, ROLL BACK, Activities, LOAD ACTIVITIES, Type, Created at, Action (Rookwps WF, Download WF, Create).
- Code Cell:** [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
```

GitHub Repository View:

C4I Use Cases as Jupyter Notebooks

Project ID: 25761638 Request Access

13 Commits 1 Branch 0 Tags 1.5 MB Files 1.5 MB Storage

A collection of Jupyter Notebooks implementing some Use Cases.

Name	Last commit	Last update
C4I_Averaged_Temperature_An...	Some small fixes. Added deltaT_deltaP Not...	2 days ago
C4I_Summer_days_Calculate_....	Some small fixes. Added deltaT_deltaP Not...	2 days ago
C4I_deltaT_deltaP_Anomaly_20...	Some small fixes. Added deltaT_deltaP Not...	2 days ago
README.md	small readme and notebook edits	4 months ago

Reproducibility Controls

Workflow Monitoring



GitHub Authentication



Snapshot Controls



Data Staging Rollback

Activities History and Provenance



The screenshot shows a Jupyter Notebook interface with a GitHub integration sidebar on the left. The sidebar includes a user profile picture, a "SWITCH USER" button, and sections for "Create Snapshot" and "Data staging". In the "Data staging" section, there is a "ROLL BACK" button highlighted with a red circle. Below it is an "Activities" section with a "LOAD ACTIVITIES" button. A table lists activities: a "Library Update" for "pip install mpg" from 2023-01-12 13:48 with a "RESTORE" button; a "Rookwps WF" from 2022-11-29 14:09 with a green "Succeeded" status; and a "Download WF" from 2022-11-22 18:21 with a green "Succeeded" status. At the bottom, a "Create" activity from 2022-11-22 18:17 has a "RESTORE" button highlighted with a red circle.

The main notebook area shows a cell output with two plots titled "Two Time Steps of Summer Days". The top plot is for time 2018-07-16T12:00:00 and the bottom plot is for 2019-07-16T12:00:00. Both plots show a heatmap of "Number of days with tmax > 25.0 °C [days]" over a geographic area defined by Latitude [degrees_north] from -75 to 75 and Longitude [degrees_east] from 0 to 300. The color scale ranges from 0 to 80+ days. The notebook code includes:

```
[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

Users create intermediate shareable snapshots on GitHub

Update and rollback data and environments.

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

CanESM5 - CanESM5
MPI-ESM1-2-LR - MPI-ESM1.2-LR
UKESM1-0-LL - UKESM1.0-N96ORCA1
CNRM-CM6-1 - CNRM-CM6-1
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 concentrations
BCC-CSM2-MR - BCC-CSM 2 MR
CMCC-CM2-SR5 - CMCC-CM2-SR5
FGOAL_S-03 - FGOAL S-03

COMPARE MODEL PERFORMANCE

Climate impact result viewer

This application shows 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 calculated 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.

Temperature (K) Precipitation (mm/day)

Projected change (2050 versus 2000) Bias with respect to ERA5 (1981-2010)

Project_project CMIP6 Selected datasets:

- CNRM-CM6-1
- MIROC6
- CanESM5

Netherlands eScience Center

View static recipe output

Help Material and Feedback Form

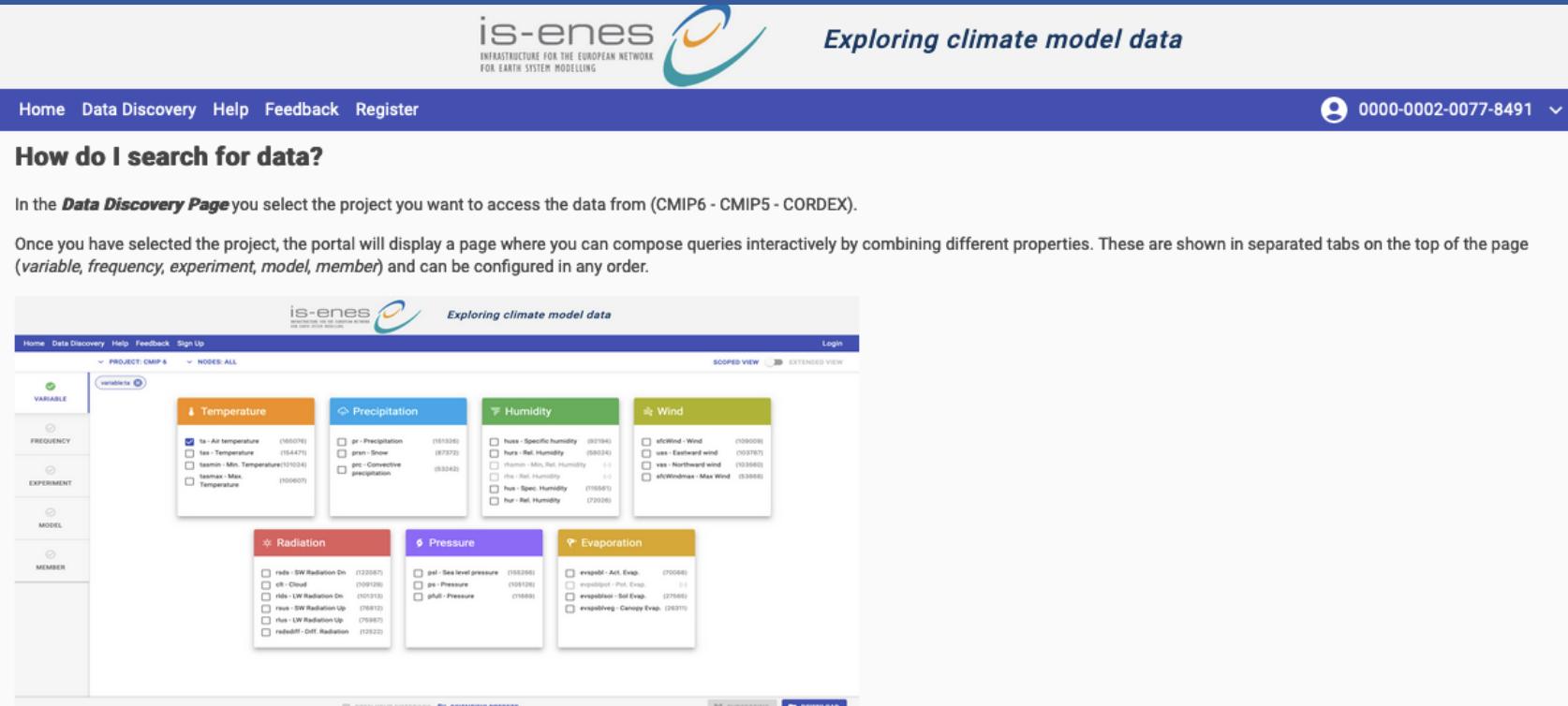
Exploring climate model data

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/Precipitation) compared to the past. The page is generated by using the [ESMValTool](#) and, in combination with the bias, gives some confidence on the performance.

Climate impact result viewer

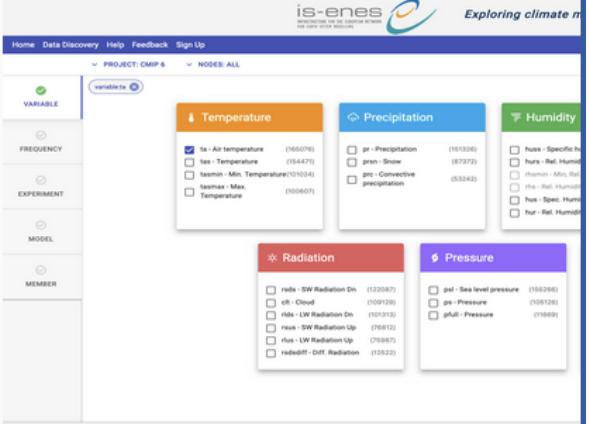
This application shows results from CMIP5 and CMIP6 models, calculated with ESMValTool. It is intended to provide some guidance for climate impact researchers, to select one or more datasets that adequately sample the spread of the CMIP ensembles.

Help Material and Feedback Form

How do I search for data?

In the **Data Discovery Page** you select the project you want to access the data from.

Once you have selected the project, the portal will display a page where you can search for data by variable, frequency, experiment, model, member and can be configured.



When configuring a specific property, to facilitate identifying the most relevant datasets, the knob at the top left of the screen, from which the current query is always displayed, can be changed by switching the knob at the top left of the screen, from which the current query is always displayed.

In the image above we show that the summary of the current query is always displayed, allowing for a live update of the available choices for each property.

When selecting the Model, by clicking on **Compare Model Performance**, it is possible to foresee a change on a variable (Temperature/Precipitation) compared to another one.



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

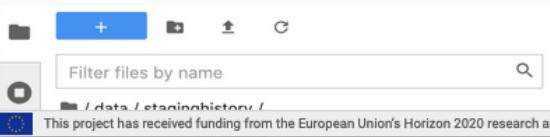


In this handy control panel you can check whether there are workflows running in the background. For instance, one might be copying data to your workspace just now.

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



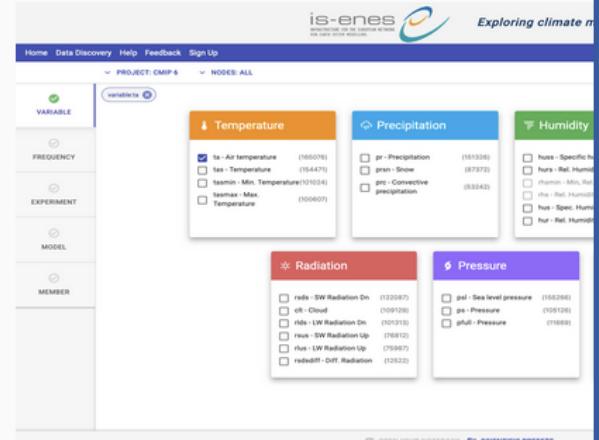
Exploring climate model data

Home Data Discovery Help Feedback Register

How do I search for data?

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

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



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

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

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



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 populat

SWIRRL manages the data in the **/data** folder, which is accessible from your *JupyterLab instance. To make sure that th
'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

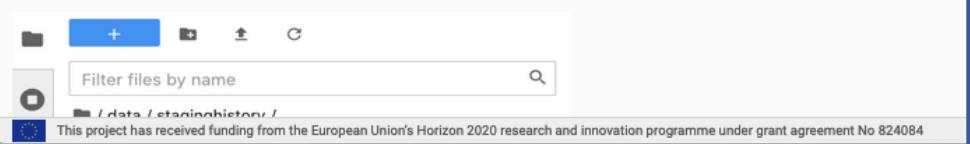
In this handy control panel you can check whether there are workflows running in the background. For instance, one might



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 wi



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)

CHANNEL TYPE	VISITS	ACTIONS	ACTIONS PER VISIT	AVG. TIME ON SITE
Direct Entry	260	1,638	6.3	6 min 45s
Websites	72	494	6.9	5 min 11s
Search Engines	5	33	6.6	10 min 19s

Frequency Overview

- 171 returning visits
- 8 min 11s average visit duration for returning visitors
- 6.8 actions per returning visit
- 30% returning visits have bounced (left the website after one page)
- 1,166 actions by the returning visits
- 166 new visits
- 5 min 19s average visit duration for new visitors

Actions: Event Categories

EVENT CATEGORY	EVENTS	EVENT VALUE
GenerateNotebook	1	-
NewNotebookSubsettingDialog	6	-
SubsetResultsWindow	7	-
ResumeNotebookDownloadDialog	10	-
DownloadMetalinkDialog	15	-
DownloadResultsWindow	45	-
OpenNotebook	45	-
TailoredSearch	131	-
FacetSidebar	492	-

Secondary dimension is Event Action.
Switch to Event Name

Visitor Map

337 visits

Countries Cities

World-Wide Visits

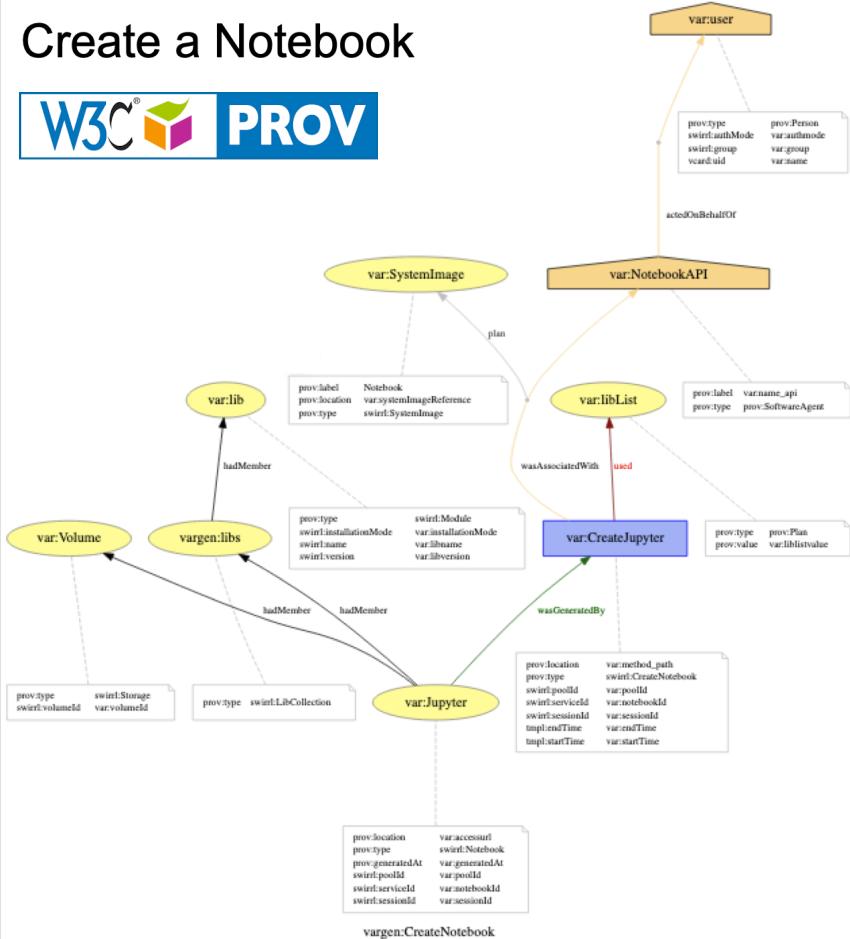
Tracking at Event Level allows to get useful insights at very high granularity.

Supported via C4I React front-end.

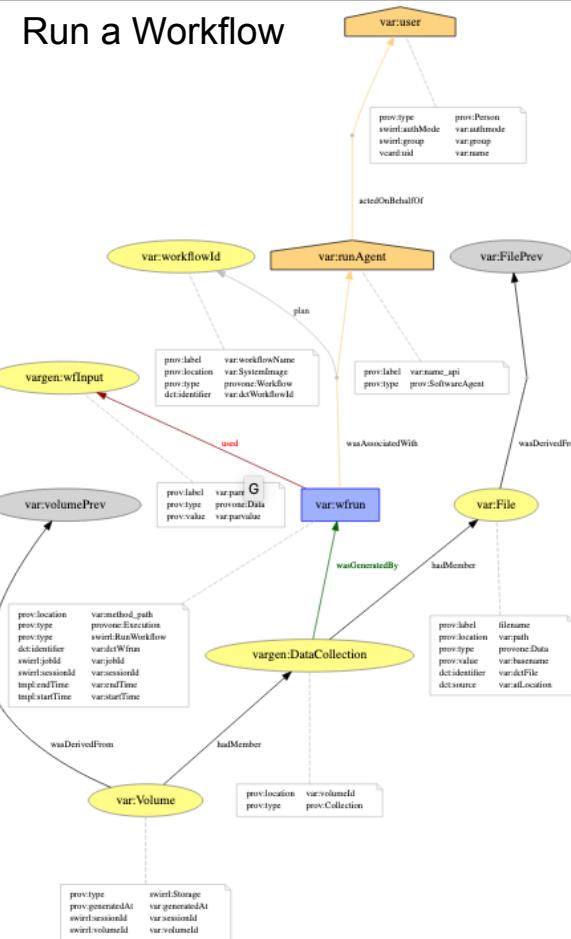
Web Analytics and Computational KPIs

Provenance used to Collect Computational KPIs

Create a Notebook



Run a Workflow



Notebook stats:

session_count	notebook_count
15	15

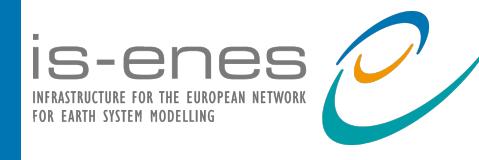
Successful workflow stats:

wf_types	session_count	run_count
download	13	16
rookswps	2	12

Failed workflow stats:

wf_types	session_count	run_count
download	2	2
rookswps	2	2

Climate4Impact (v2) - Conclusions



dev.climate4impact.eu

Ready to close C4I v1!?

Platform capabilities to be addressed collaboratively

- Access to CORDEX *requires the wider support of IdeA IDP by the CORDEX nodes*
- More Notebooks Use Cases *CERFACS/Icclim*
- Scientific Content and Descriptive Impact Use Cases to be extended by experts
- More ESGF nodes (at least in EU) to implement WPS *Beyond subsetting?*
- Roadmap to migrate to a new Search Interface STAC
- Use SWIRRL to add Visualisation Services and Collaborative Data Caches

Thank you!

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



Follow our channel
IS-ENES3 H2020