



## Virtual workshop on Compute and Analytics

# Compute Services requirements for the climate impact community using C4I

CERFACS: Christian Pagé

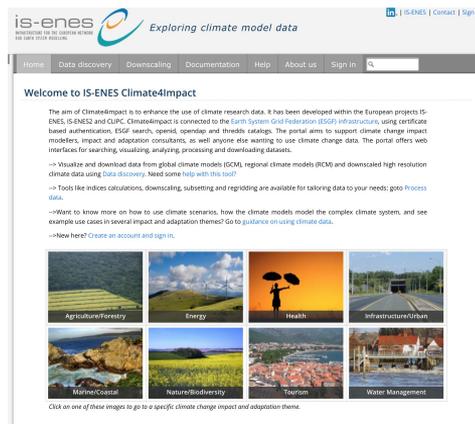
KNMI: Wim Som de Cerff, Maarten Plieger, Janette Bessembinder, Bernadet  
Overbeek

DKRZ: Stephan Kindermann

SMHI: Lars Barring



- **C4I Platform**
  - Current Status
  - DARE coordination
  - IS-ENES3 C4I developments
- **User engagement**
  - User requirements
  - Use case example
- **Improvements and needs from compute and analytics**



<https://climate4impact.eu>



## C4I Status

- CMIP6 data findable and can be processed (ICCLIM, subsetting)
- Hardware failure in October 2018 (downtime)
- Migrated to AWS, will be moved to SurfSara
- MyProxy replaced by OAUTH2 with certificates
- Average of 2.100 unique users each month (AWStats)

The screenshot shows the 'is-enes Exploring climate model data' website. The search results page displays a list of project properties for CMIP6. The 'Selected filters' section shows 'Project: CMIP6'. Below this, it indicates 'Found 43213 datasets. Displaying page 1 of 1729'. A list of datasets is visible, including entries like 'CMIP6.CMIP.EC-Earth-Consortium.EC-Earth3-LR.pi.Control.r11p1f1.Oman.thetan.gv.v20190103'.

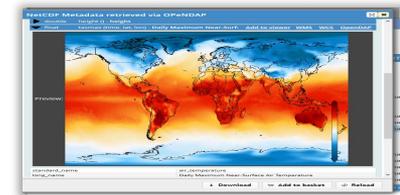
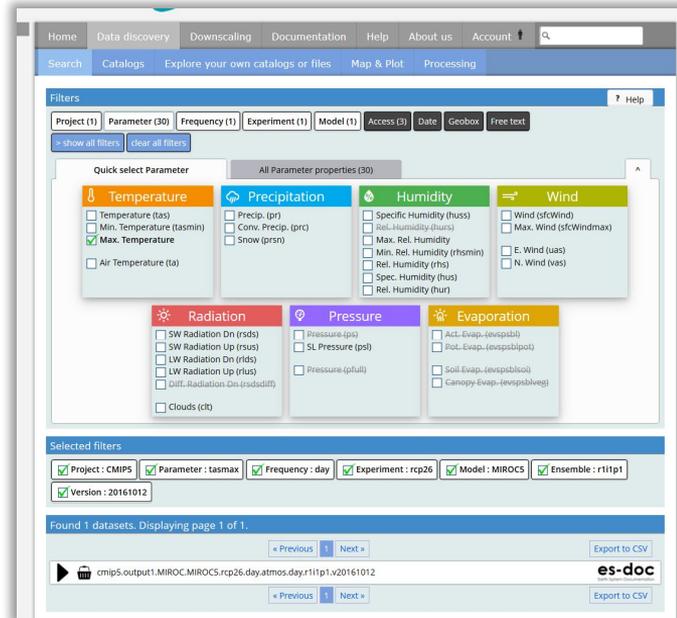
Search showing CMIP6 data

The screenshot shows a detailed view of a dataset in the is-enes interface. The 'Selected filters' section shows 'Project: CMIP6' and 'Parameter: tas'. It indicates 'Found 309 datasets. Displaying page 1 of 13'. A 'NetCDF Metadata retrieved via OPENDAP' window is open, showing metadata for 'tas\_3hr\_IPSL-CM6'. Below the metadata, a 'Text filter' section is visible. The main visualization is a world map showing temperature anomalies, with a color scale ranging from blue (cooler) to red (warmer). The map shows significant warming over land and in the mid-to-high latitudes.

Visualization of CMIP6 data

# Example use case: Calculating Summer Days (SU) - step 1

- Calculate number of days where maximum temperature is above 25 degrees per European country, based on experiment RCP 2.6 and model MIROC5
- Go to C4I and sign in
- Go to Search and select:
  1. Model: CMIP5
  2. Parameter: tasmax
  3. Time frequency: daily
  4. Experiment: rcp26
  5. Model: MIROC5,
  6. Ensemble: r11p1
  7. Select the latest version
- Select a file from the dataset and add it to your basket





- Go to Processing and select ICCLIM simple indicator calculations
- Select SU, Summer days. Leave the threshold to 25 degrees Celsius
- Select the file from your basket and click “Start processing”
- Inspect the output

**Processor CLIPIC ICCLIM simple indicator calculator Execute**

**Title** CLIPIC ICCLIM simple indicator calculator Execute

**Identifier** clip\_simpleindicator\_execute

**Abstract** Using ICCLIM single input indices of temperature TG, TX, TN, TXx, TNx, TNn, SU, TR, CSU, G04, FD, CF, ID, HD17; of rainfall: CDD, CWD, RR, RR1, SDII, R10mm, R20mm, RX1day, RX5day; and of snowfall: SD, SD1, SDS, SDS0 can be computed.

**Location** [https://climate4impact.eu/impactportal/WPS?service=WPS&version=1.0.0&request=describeProcess&identifier=clip\\_simpleindicator\\_execute](https://climate4impact.eu/impactportal/WPS?service=WPS&version=1.0.0&request=describeProcess&identifier=clip_simpleindicator_execute)

---

**Processing succeeded! Showing report:**

Identifier	Title	Type	MimeType	Value
opendapURL	opendapURL	integer		<a href="https://climate4impact.eu/impactportal/DAP?eds.ac.uk.openid.Maarten.Plijer/WPS_Scratch/WPS_clip_simpleindicator_execute_20190404T134222/ou_ictlim.nc">https://climate4impact.eu/impactportal/DAP?eds.ac.uk.openid.Maarten.Plijer/WPS_Scratch/WPS_clip_simpleindicator_execute_20190404T134222/ou_ictlim.nc</a>

---

**Processing inputs**

**Indicator name** (*indicatorName*)  
The indicator to calculate  
SU - Number of Summer Days (Tmax > 25°C)

**Time slice mode** (*sliceMode*)  
Selects temporal grouping to apply for calculation  
year - Yearly time period

**Indicator threshold** (*threshold*)  
Threshold for temperature (degrees Celsius)

**NetCDF Metadata retrieved via OPeNDAP**

- Go to Processing and select Polygon overlay
- For “Input File B - Gridded data”, choose the latest result with SU from your basket. This is the most recent folder under WPS\_Scratch
- As variable select “SU”, as time range select “\*”
- Click “Start processing”

Processor Polygon overlay

Basket

File	DAP	HTTP	Filesize	Date
Remote data				
cmp5_output1_NCC_NorESM1-ME_historical_mon_land_Lmon_r11p1_v2			0	2018-11-09 15:02:28Z
cmp5_output1_NOAA-GFDL_GFDL-ESM2G_piControl_day_atmos_day_r			0	2018-11-09 15:06:00Z
tx_0_25deg_reg_v17_0.nc			4.632G	2018-12-05 11:03:23Z
tas_Amon_MIROC6_historical_r11p1f1_gn_195001-201412.nc	true		-	2019-03-29 09:16:10Z
tasmax_Amon_MIROC6_historical_r11p1f1_gn_195001-201412.nc	true	true	56.81M	2019-03-29 09:18:23Z
tasmax_day_GISS-E2-1-G_historical_r11p1f1_gn_19500101-20141	true		-	2019-03-29 09:23:37Z
cmp5_output1_MIROC6_MIROC6_historical_day_atmos_day_r5r1p1_v201			0	2019-03-29 13:52:25Z
tasmax_day_MIROC6_historical_r5r1p1_20000101-20091231.nc	true	true	478.5M	2019-03-29 13:52:46Z
tasmax_day_ACCESS1-3_historical_r3r1p1_18500101-18741231.nc	true	true	1.017G	2019-04-01 08:57:12Z
tasmax_day_MIROC5_rcp26_r11p1_20060101-20091231.nc	true	true	191.4M	2019-04-04 12:33:28Z
My data				
WPS_Settings				2019-04-04T12:34:22Z
WPS_Scratch				2019-04-04T12:34:22Z
WPS_clipc_simpleindicator_execute_20190404T123422Z				2019-04-04T12:34:24Z
out_kclm.nc	true	true	572.5K	2019-04-04 12:36:27Z
WPS_clipc_simpleindicator_execute_20190401T090210Z				2019-04-01T09:02:10Z
WPS_is_enes_wps_polygonoverlay_20190329T140922Z				2019-03-29T14:09:33Z
WPS_is_enes_wps_polygonoverlay_20190329T140621Z				2019-03-29T14:06:26Z

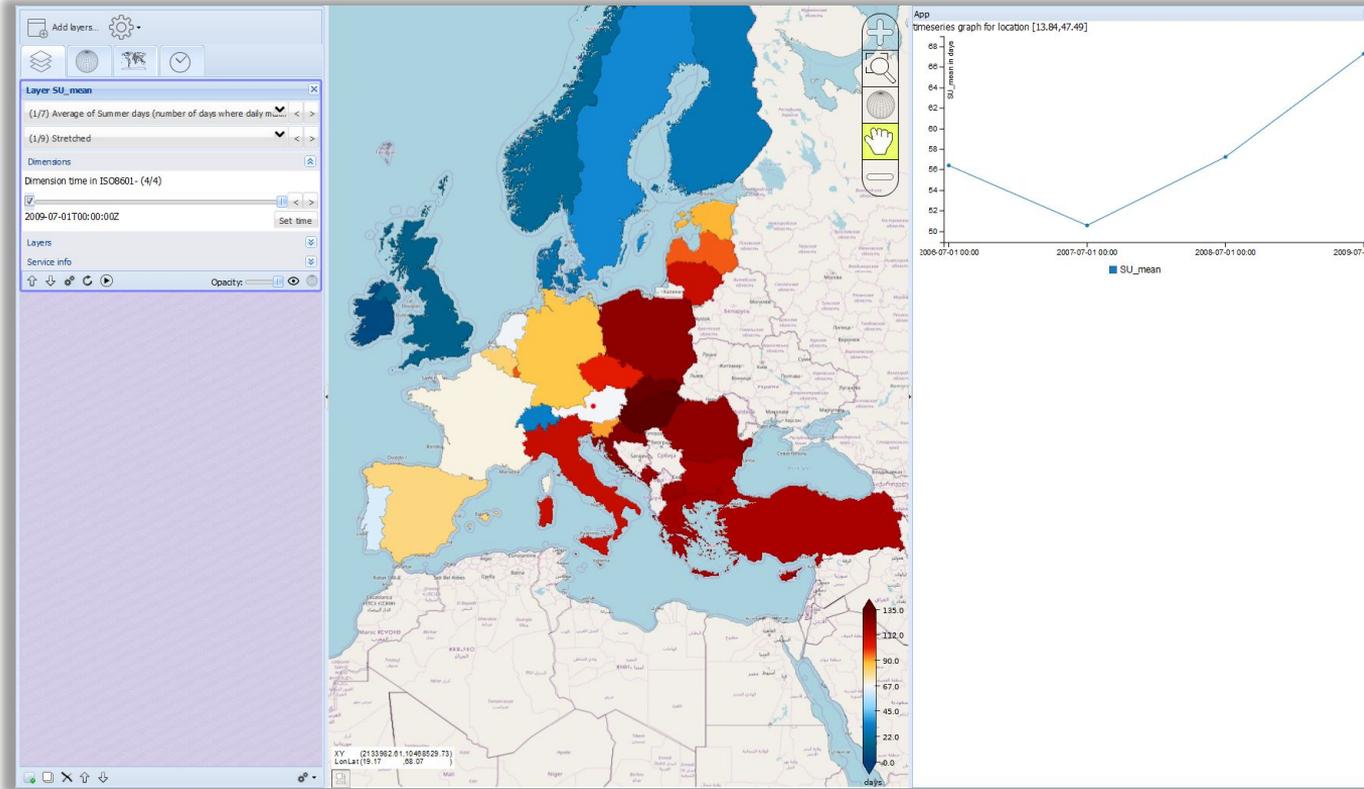
Output CSV filename (csvnutstafilename)

min/d / max: 1

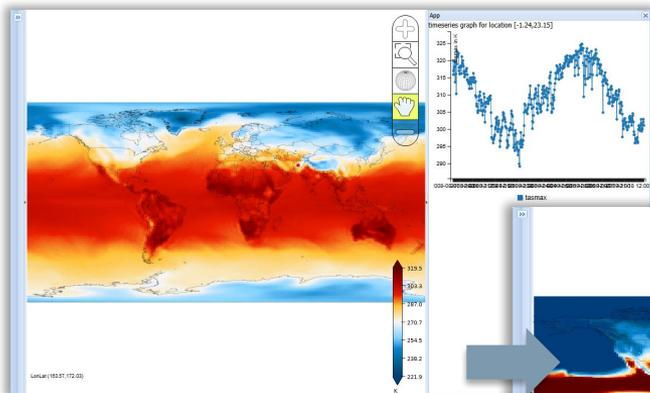
Buttons: Link file, Search, View, Download, Script download, Upload, Delete, Reload, Use file(s)

Footer: Name of the comma separated value (CSV) output file. Will contain statistics in table form. Can be opened in a spreadsheet program.

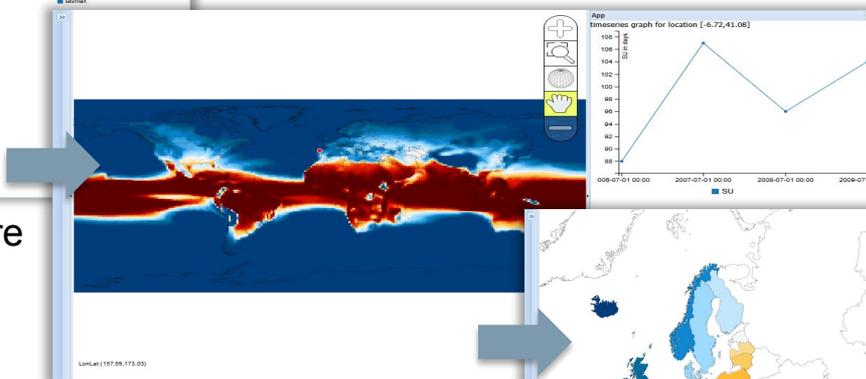
# Summer days per European country for MIROC5 / RCP26.



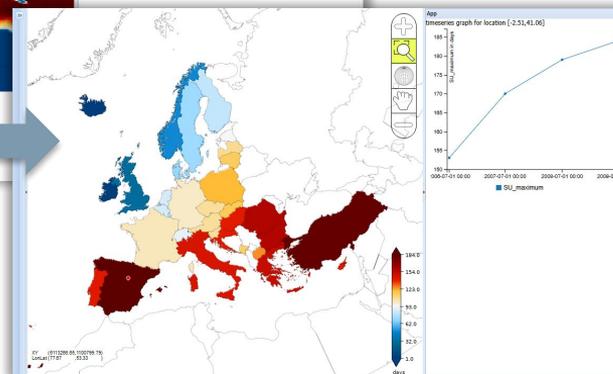
# To summarize this use case:



Daily Maximum temperature



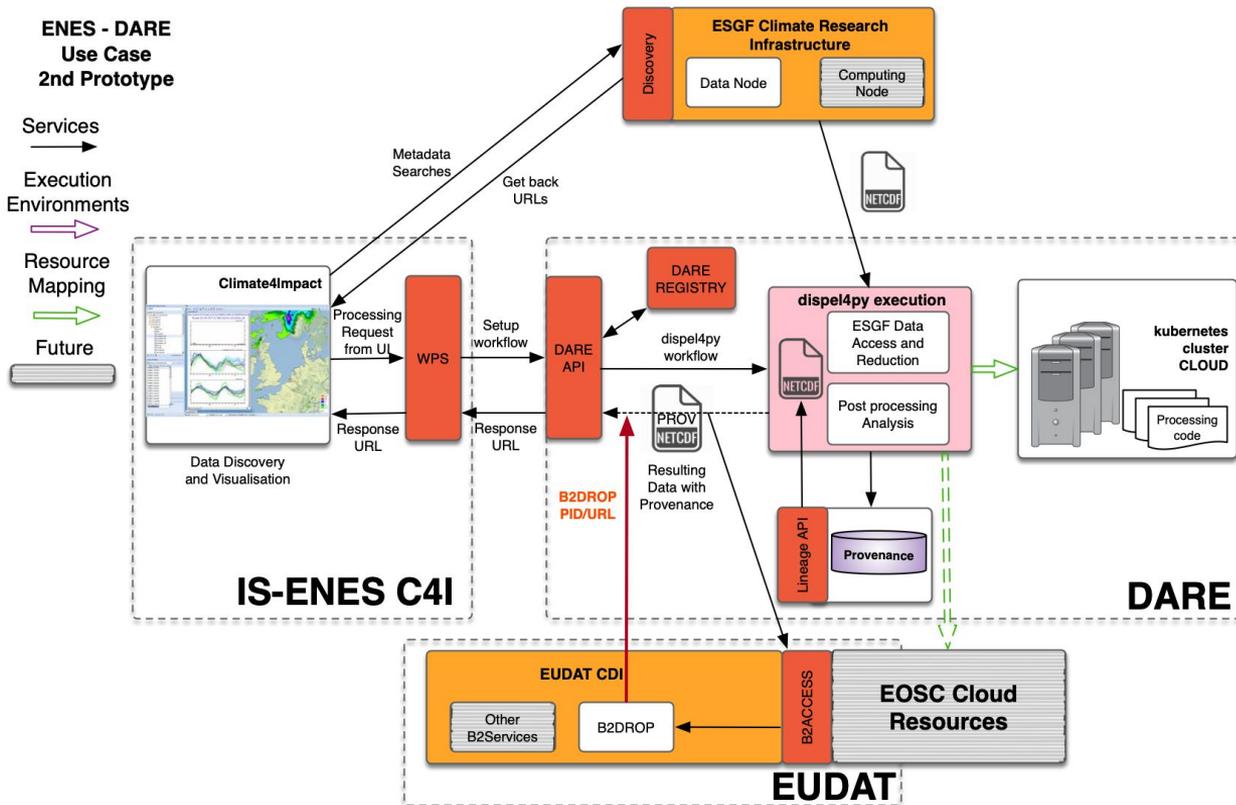
Summer days ( $T > 25$ )



Maximum summer days per country



## H2020-DARE: 01/18-12/20





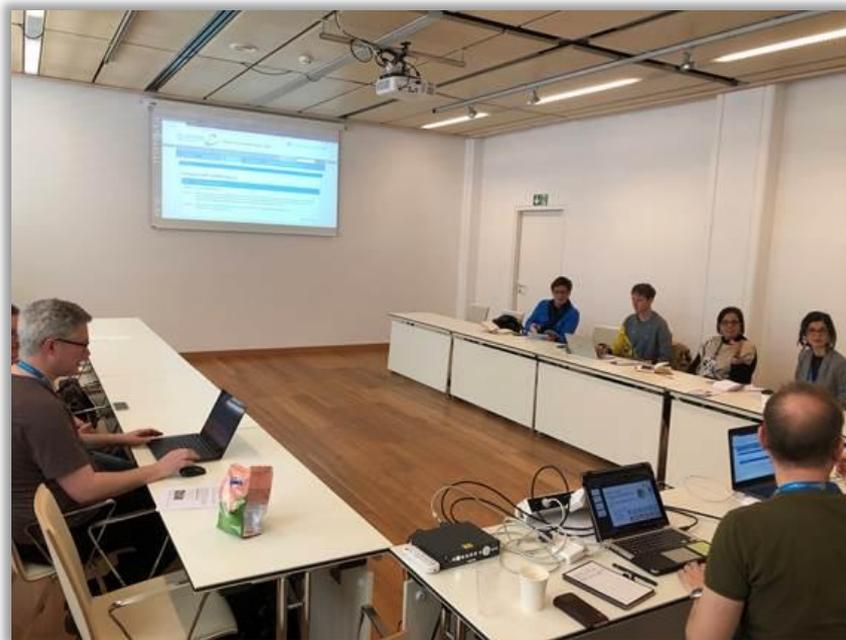
## Planned C4I Tasks in IS-ENES3

- Task in WP3, WP7 and WP10 (virtual WP....)
  - WP3: Community building, standards (metadata), requirements
  - WP7: Operational C4I service
  - WP10: Developments of C4I

## Goals:

- Users' engagement: climate research community, climate impact community as well as interdisciplinary research community
- Make all ESGF climate model data and services (CMIPs, CORDEX) accessible
- Separate computing and portal infrastructure
- Provide advanced data processing and new user-friendly interfaces for data analytics

- Receive feedback from users
- Keep users in the loop, show them new features





- For impact studies often relatively high **spatial resolutions** are required:
  - EUROCORDEX is therefore very interesting for them
- Impact research needs: **several climate scenarios (ensemble) is needed**
  - Information on biases
  - Changes under the various RCPs
  - => for a large range of climate variables (not just averages, also extremes, variability)
- **Terminology** in climate science is often a problem for impact researchers
  - Provide more information and guidance
  - Expert / Non-expert mode could be needed (variable names, etc.)
  - Provide guidance/default values for scenario parameters
  - Avoid having impact researchers deal with files
- For use in impact models generally **bias-corrected data** are needed
  - Impact researchers are often not able to do a good bias correction themselves



- Provide better tools for impact researchers to overcome:
  - Large **amounts of data**
  - **NetCDF format.**
- Tools to “cut out” certain regions, time periods
- Tools to process the data without the need to download the data
  
- A very robust platform is needed
  - Avoid cryptic **error messages**
  - Optimize to overcome the **extremely long times for processing**
  - Need to better guide and inform the user: tailored error messages, guidance on how to proceed
  
- Very complete **guidance** material is needed
  
- Training must be provided for impact researchers

## Improvements and needs from compute and analytics

- Currently C4I handles ESGF data on file level
  - Files are pieces of data with approximately 5 years of climate data
  - **We want to make it easier to process long sequences of data**
  - Compare experiments
- We want to bring the processing to the data
  - Collaborate with ESGF CWT and results from H2020 DARE?
  - **Calculations should run faster**
- Currently provenance tracking is too limited
  - Enhance usage of W3C PROV-DM standard and WPS\_PROV toolkit
  - **Track actions, record how a product was generated, make process repeatable.**

# Questions



- For impact studies often relatively high **spatial resolutions** are required. EUROCORDEX is therefore very interesting for them
- For impact research often not enough resources are available to take into account a whole ensemble of climate models, but two is regularly possible. **How to select 2 model runs** (or a few more) that show the range of impacts? Information on biases, changes under the various RCPs for a large range of climate variables (not just averages, also extremes, variability) would be very useful to make a good selection (is it possible to generate this info with ESMval and make this available in an easy way?)
- **Terminology** in climate science is often a problem for impact researchers. In the C4I portal there are many options/filters. Some could be formulated in a different way, such that it is clearer for a broader range of users. Further more, more guidance/explanation can also help and default options. E.g. most impact researcher do not understand the names of the files or they do not know what "Tasmax" means. For the spatial resolution e.g. daily data could be used as default
- For use in impact models generally **bias-corrected data** are needed. Impact researchers are often not able to do a good bias correction themselves



- Many impact researcher have problems with large **amounts of data and the NetCDF format**. Therefore the possibilities to “cut out” certain regions, time periods with in the C4I is very useful for them. Also the various options to process the data without the need to download the data is very useful
- At the moment the C4I portal regularly gives **error messages**, takes **extremely long times for processing** (with regularly error messages at the end). Would it be possible to give some more feedback on the time required for e.g. processing and providing indications on what is the problem in the case of error messages?
- If we want a broader group to use the C4I portal, **more guidance** material is needed and more training. People from the broader group often do not have much background knowledge on climate data and to use the portal, one needs a considerable level of knowledge on climate data