Easy Access to Complex Analysis Tools for Climate Researchers and Climate Data End Users

Christian Pagé[1], Maarten Plieger[2], Wim Som de Cerff[2], Alessandro Spinuso[2], Rosa Filgueira[3], Malcolm Atkinson[3], Sissy Themeli[4], Iraklis Klampanos[4], and Vangelis Karkaletsis[4]

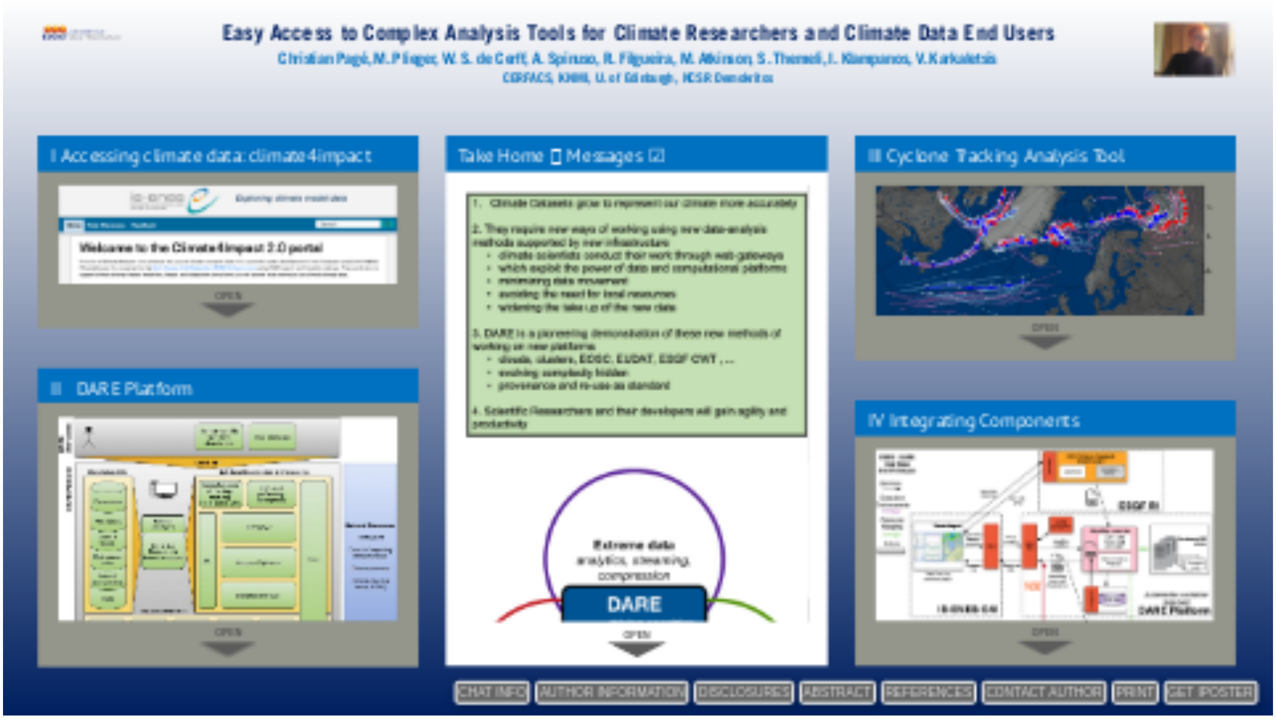
[1]CERFACS
[2]Royal Netherlands Meteorological Institute
[3]University of Edinburgh
[4]NCSR Demokritos

November 21, 2022

Abstract

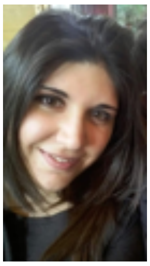
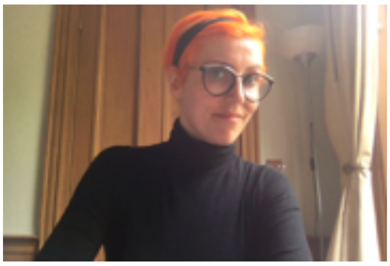
Researchers and end users using climate data face a challenge when they analyze the data they need. Data volumes are increasing very rapidly, and the ability to download all needed data is often no longer possible. Also, it can be complex to install, configure and use some advanced analysis tools on such large datasets. This is especially true when they are stored in a federated architecture like the ESGF. An example of a complex analysis tool used in climate research and adaptation studies is a tool to follow storm tracks. In the context of climate change, it is important to know if storm tracks will change in the future, in both their frequency and intensity. Storms can cause significant societal impacts, hence it is important to assess future patterns. Having access to this type of complex analysis tool is very useful, and integrating them with front-ends like the IS-ENES climate4impact (C4I) would enable the use of those tools by a larger number of researchers and end users. Integrating this type of complex tool is not an easy task. It requires significant development effort, especially if one of the objectives is also to adhere to FAIR principles. The DARE Platform enables research developers to faster develop the implementations of scientific workflows more rapidly. This work presents how such a complex analysis tool has been implemented to be easily integrated with the C4I platform. The DARE Platform also provides easy access to e-infrastructure services like EUDAT B2DROP, to store intermediate or final results and powerful provenance-powered tools to help researchers manage their work and data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements N°824084 and N°777413.

Easy Access to Complex Analysis Tools for Climate Researchers and Climate Data End Users



Christian Pagé, M. Plieger, W. S. de Cerff, A. Spinuso, R. Filgueira, M. Atkinson, S. Themeli, I. Klampanos, V. Karkaletsis

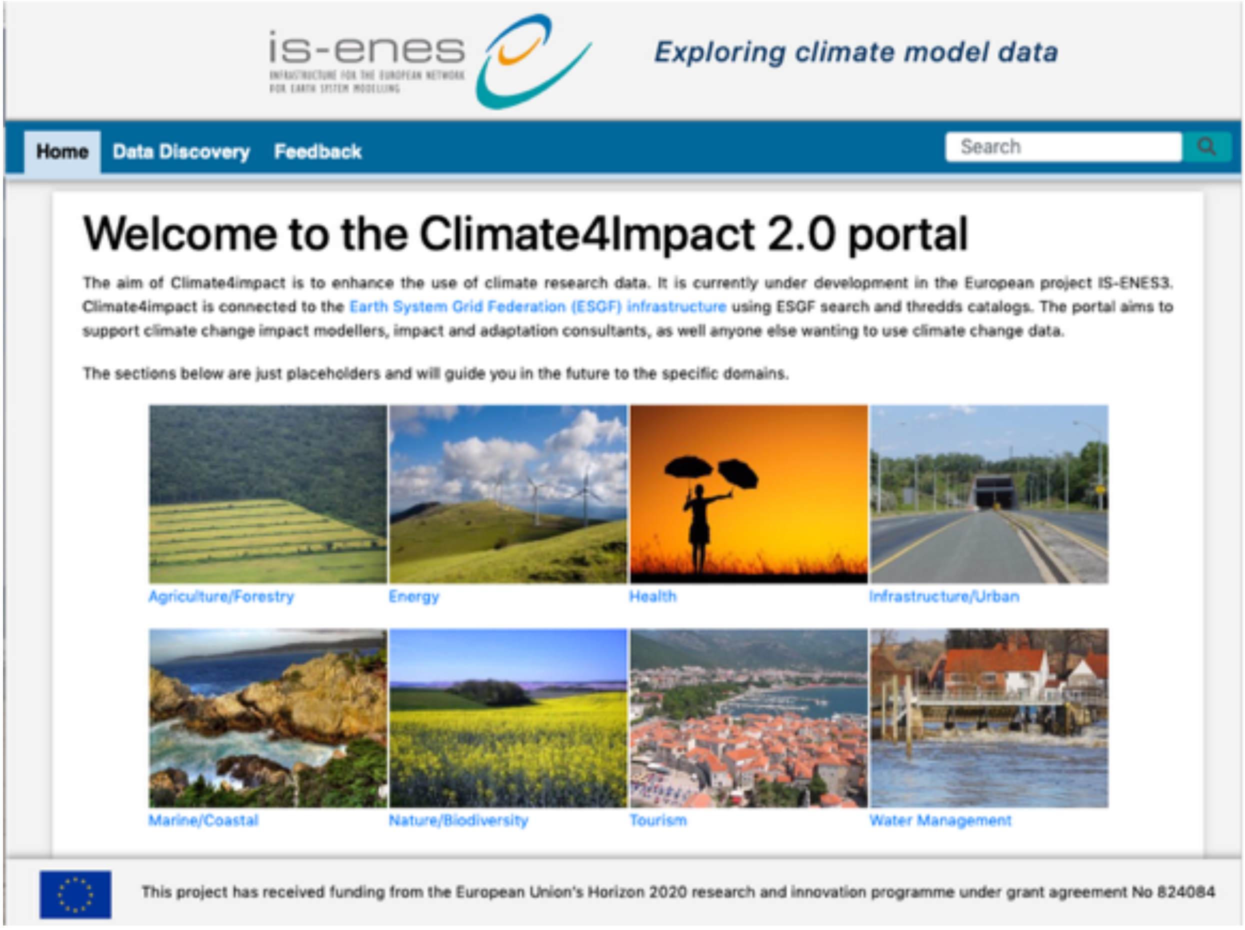
CERFACS, KNMI, U. of Edinburgh, NCSR Demokritos

PRESENTED AT:

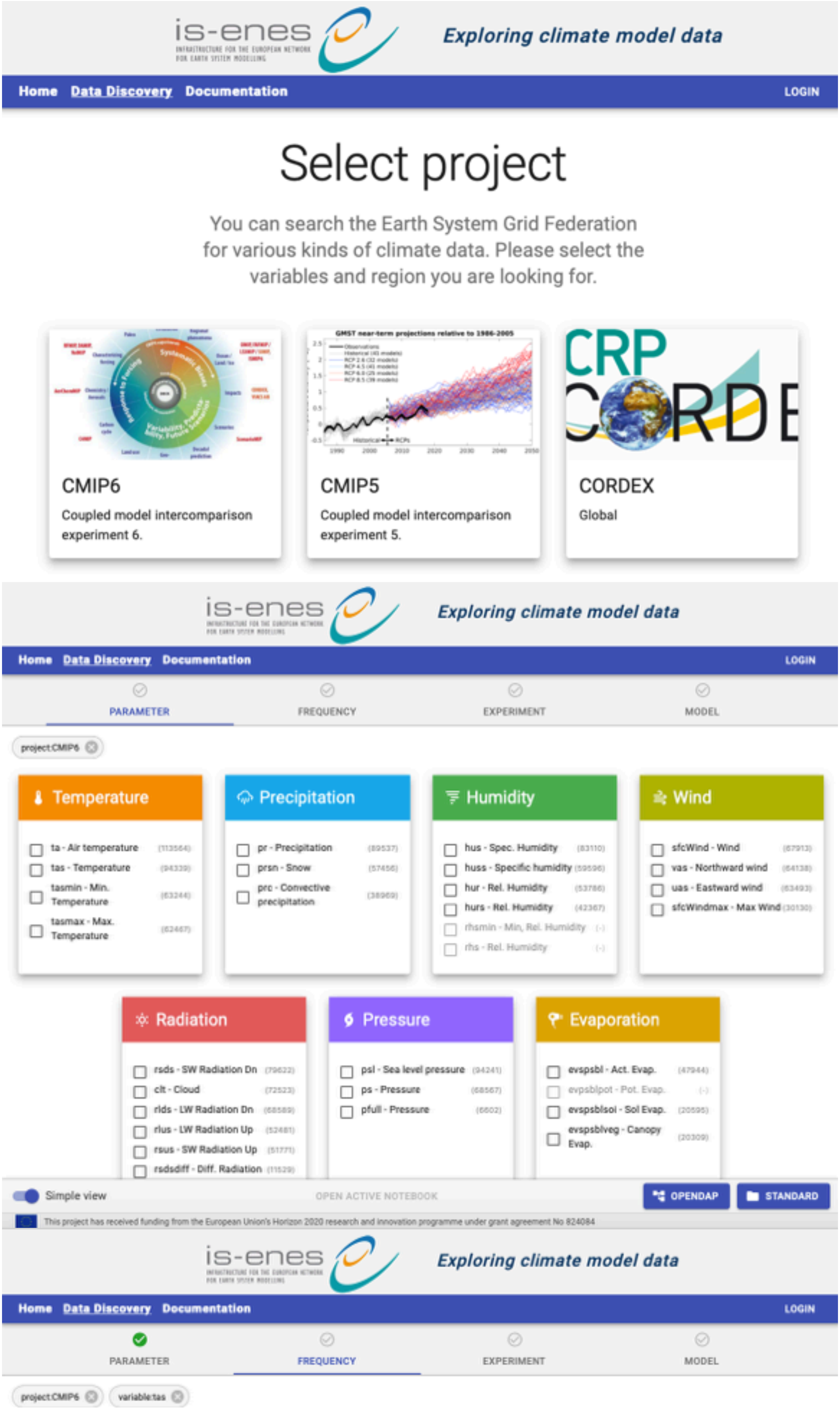
I ACCESSING CLIMATE DATA: CLIMATE4IMPACT

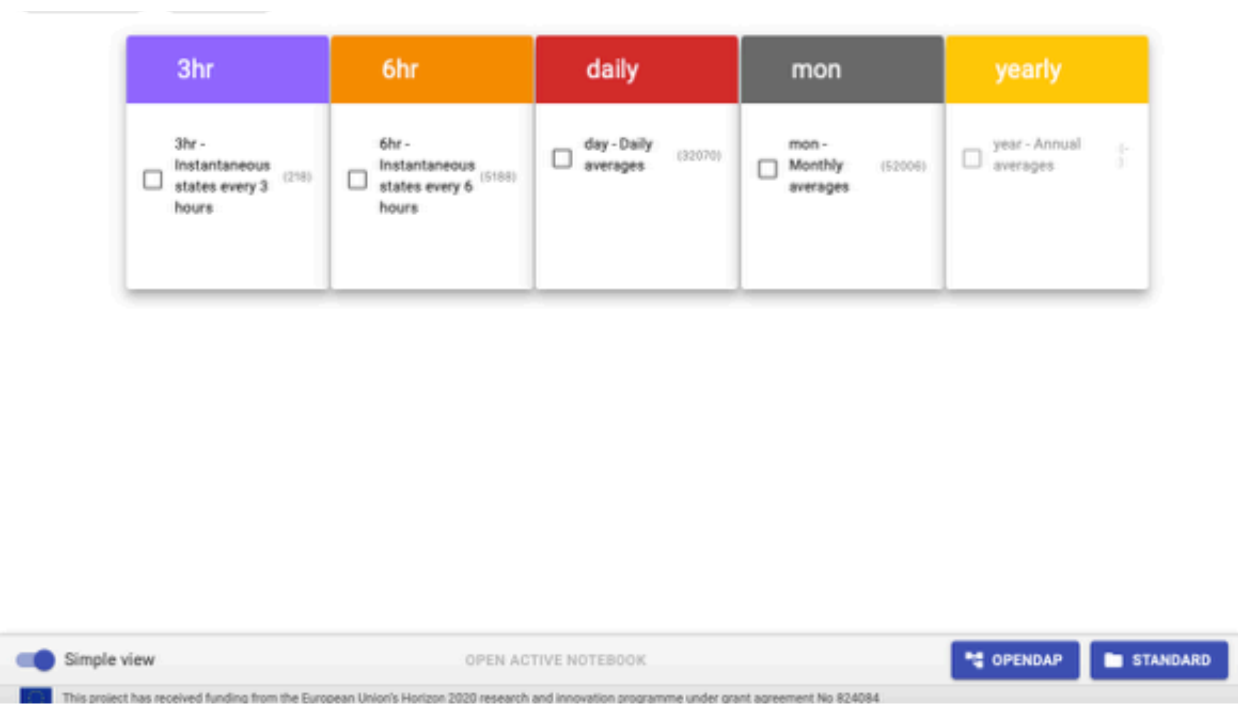


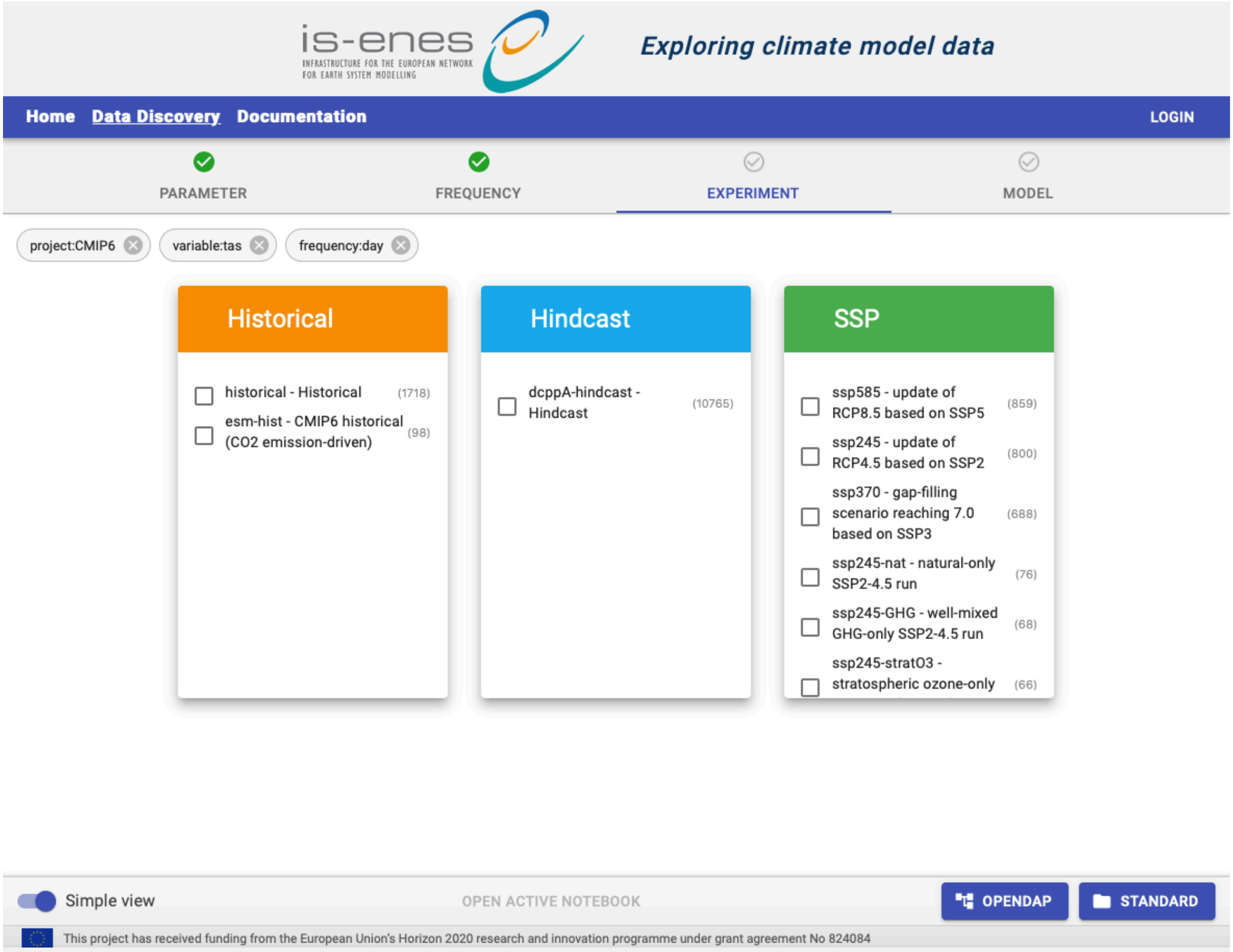
- https://climate4impact.eu (https://climate4impact.eu)
- Developed and managed by IS-ENES (IS-ENES3 EU Project (https://is.enes.org))
- Platform for researchers to explore climate data and perform analysis
- Connects to climate e-infrastructure services (ESGF (https://esgf-data.dkrz.de))
- Tailored for end-users
- Supports on-demand data processing and statistical downscaling

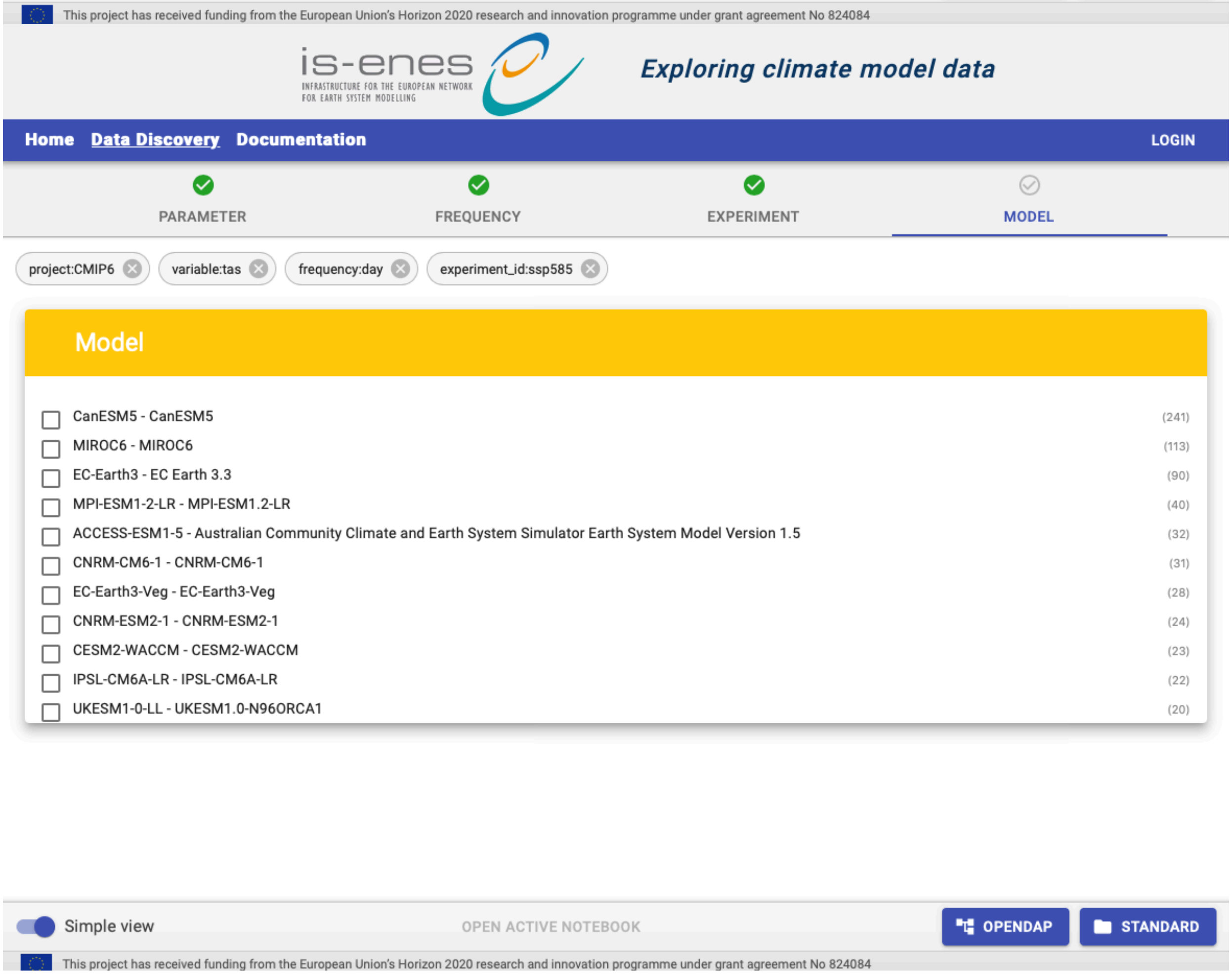
- Version 2.0 of the portal
  - Alpha stage currently
  - Jupyter Notebooks (JupyterLab) approach for data processing and analytics
  - Remote subsetting (spatial and temporal)
  - Revamped user workflow and data discovery interface
  - Going away from a file-based interaction
  - Full provenance and reproducibility, getting closer to FAIRness compliance

Below the user workflow in the Data Discovery of C4I.

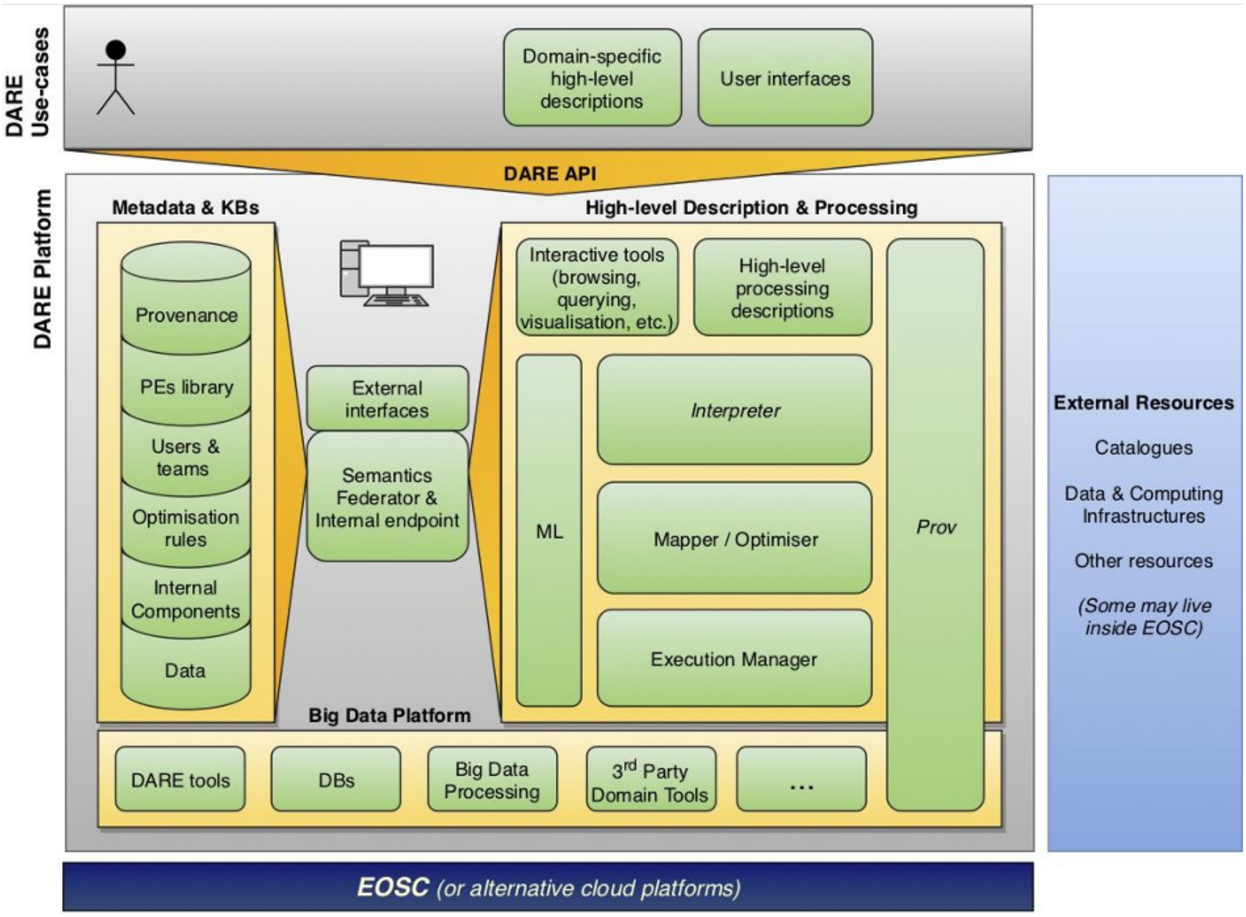

is-enes
Exploring climate model data
Home
Data Discovery
Documentation
LOGIN
Select project
You can search the Earth System Grid Federation for various kinds of climate data. Please select the variables and region you are looking for.
CMIP6
Coupled model intercomparison experiment 6.
CMIP5
Coupled model intercomparison experiment 5.
CORDEX
Global
is-enes
Exploring climate model data
Home
Data Discovery
Documentation
LOGIN
PARAMETER
FREQUENCY
EXPERIMENT
MODEL
project:CMIP6
Temperature
ta - Air temperature (113564)
tas - Temperature (94339)
tasmin - Min. Temperature (63244)
tasmax - Max. Temperature (62467)
Precipitation
pr - Precipitation (89537)
prsn - Snow (57456)
prc - Convective precipitation (38969)
Humidity
hus - Spec. Humidity (83110)
huss - Specific humidity (59596)
hur - Rel. Humidity (53786)
hurs - Rel. Humidity (42367)
rhsmin - Min. Rel. Humidity (-)
rhs - Rel. Humidity (-)
Wind
sfcWind - Wind (67913)
vas - Northward wind (64138)
uas - Eastward wind (63493)
sfcWindmax - Max Wind (30130)
Radiation
rsds - SW Radiation Dn (79622)
clt - Cloud (72523)
rlds - LW Radiation Dn (68589)
rlus - LW Radiation Up (52481)
rsus - SW Radiation Up (51771)
rsdsdiff - Diff. Radiation (11529)
Pressure
psl - Sea level pressure (94241)
ps - Pressure (68567)
pfull - Pressure (6602)
Evaporation
evspsbl - Act. Evap. (47944)
evpsblpot - Pot. Evap. (-)
evspsblsoi - Sol Evap. (20595)
evspsblveg - Canopy Evap. (20309)
Simple view
OPEN ACTIVE NOTEBOOK
OPENDAP
STANDARD
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084
is-enes
Exploring climate model data
Home
Data Discovery
Documentation
LOGIN
PARAMETER
FREQUENCY
EXPERIMENT
MODEL
project:CMIP6
variable:tas


3hr
6hr
daily
mon
yearly
3hr - Instantaneous states every 3 hours
6hr - Instantaneous states every 6 hours
day - Daily averages (32070)
mon - Monthly averages (52006)
year - Annual averages
Simple view
OPEN ACTIVE NOTEBOOK
OPENDAP
STANDARD
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084


is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING
Exploring climate model data
Home
Data Discovery
Documentation
LOGIN
PARAMETER
FREQUENCY
EXPERIMENT
MODEL
project:CMIP6
variable:tas
frequency:day
Historical
historical - Historical (1718)
esm-hist - CMIP6 historical (CO2 emission-driven) (98)
Hindcast
dcppA-hindcast - Hindcast (10765)
SSP
ssp585 - update of RCP8.5 based on SSP5 (859)
ssp245 - update of RCP4.5 based on SSP2 (800)
ssp370 - gap-filling scenario reaching 7.0 based on SSP3 (688)
ssp245-nat - natural-only SSP2-4.5 run (76)
ssp245-GHG - well-mixed GHG-only SSP2-4.5 run (68)
ssp245-stratO3 - stratospheric ozone-only (66)
Simple view
OPEN ACTIVE NOTEBOOK
OPENDAP
STANDARD
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084


is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING
Exploring climate model data
Home
Data Discovery
Documentation
LOGIN
PARAMETER
FREQUENCY
EXPERIMENT
MODEL
project:CMIP6
variable:tas
frequency:day
experiment_id:ssp585
Model
CanESM5 - CanESM5 (241)
MIROC6 - MIROC6 (113)
EC-Earth3 - EC Earth 3.3 (90)
MPI-ESM1-2-LR - MPI-ESM1.2-LR (40)
ACCESS-ESM1-5 - Australian Community Climate and Earth System Simulator Earth System Model Version 1.5 (32)
CNRM-CM6-1 - CNRM-CM6-1 (31)
EC-Earth3-Veg - EC-Earth3-Veg (28)
CNRM-ESM2-1 - CNRM-ESM2-1 (24)
CESM2-WACCM - CESM2-WACCM (23)
IPSL-CM6A-LR - IPSL-CM6A-LR (22)
UKESM1-0-LL - UKESM1.0-N96ORCA1 (20)
Simple view
OPEN ACTIVE NOTEBOOK
OPENDAP
STANDARD
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

II DARE PLATFORM



- **Composition of services** using **containers**
- Across service communication using exposed **REST APIs**
- **Scalable** and **flexible** due to **kubernetes** orchestration
- **Effortless cloud** infrastructure deployment
- Software isolation

Important aspects to consider

- Efficient Interfacing e-Infrastructures for Researchers is Challenging
- Technologies: fast-evolving but many are getting deprecated and obsolete
- Authentication and Security
- Scalability in Federated Environments
- DARE Platform
  - Hides complexity and heterogeneity
  - Provides automated Provenance & Lineage
  - Provides EUDAT & EOSC Compatibility

The *DARE Platform (https://project-dare.gitlab.io/dare-platform/)* is now available! With documentation and support. *Get more information here! (https://project-dare.gitlab.io/dare-platform/)*

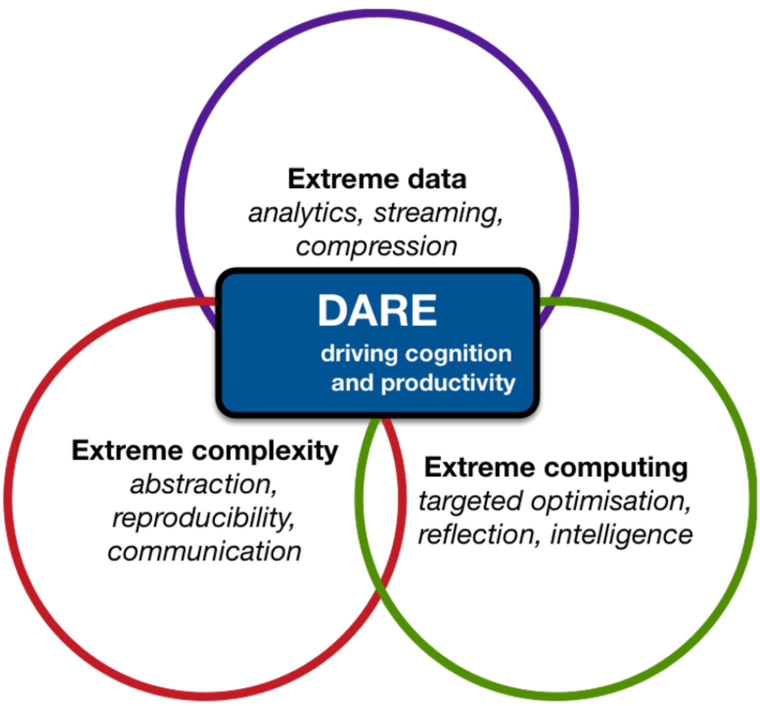
TAKE HOME 🏡 MESSAGES ☑️

1. Climate Datasets grow to represent our climate more accurately

2. They require new ways of working using new data-analysis methods supported by new infrastructure
   - climate scientists conduct their work through web gateways
   - which exploit the power of data and computational platforms
   - minimizing data movement
   - avoiding the need for local resources
   - widening the take up of the new data

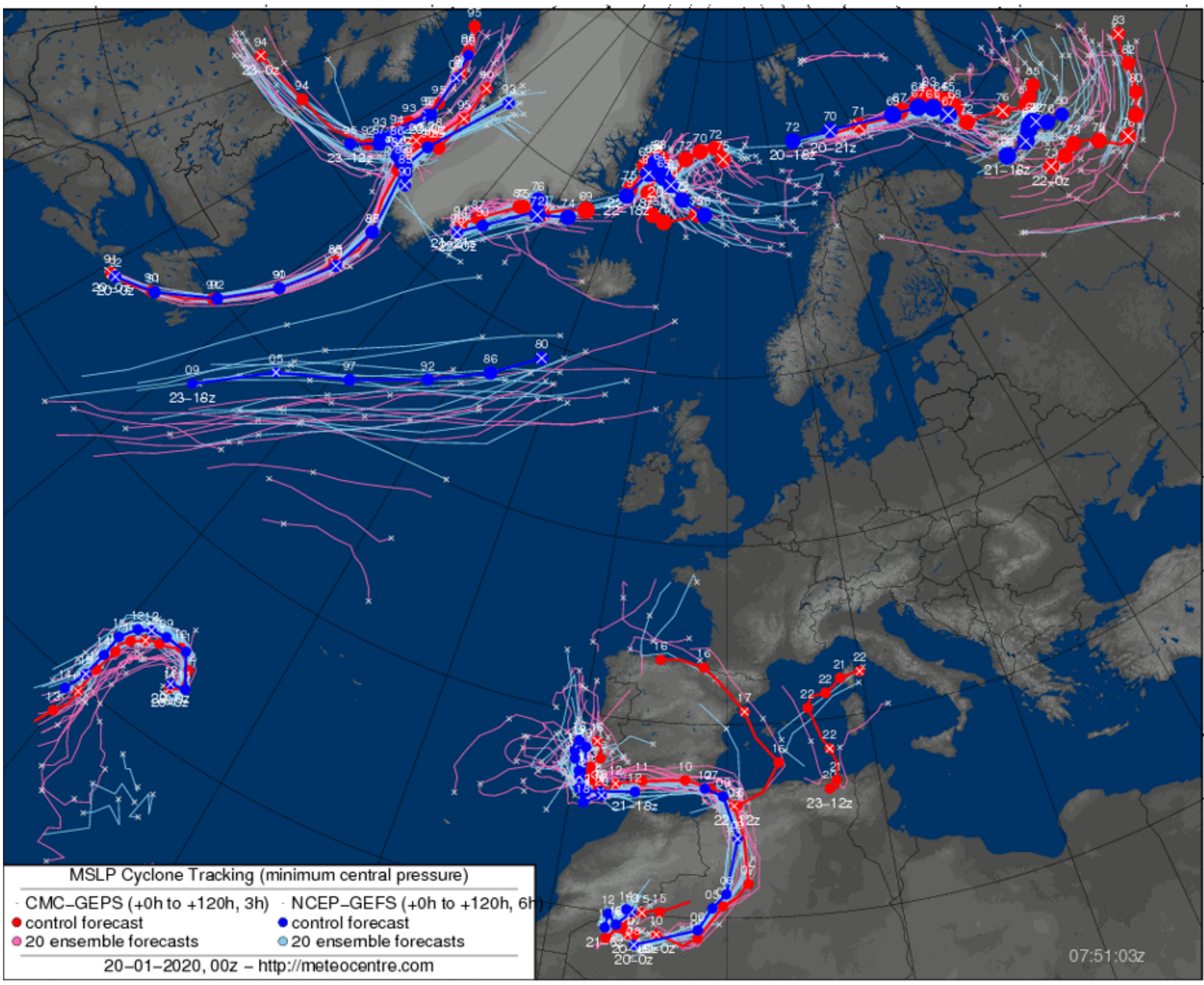
3. DARE is a pioneering demonstration of these new methods of working on new platforms
   - clouds, clusters, EOSC, EUDAT, ESGF CWT , ...
   - evolving complexity hidden
   - provenance and re-use as standard

4. Scientific Researchers and their developers will gain agility and productivity



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements N°824084 and N°777413.

III CYCLONE TRACKING ANALYSIS TOOL



Analysing climate scenarios requires several types of analysis tools

- Simple statistics (average, quantiles, anomalies, etc.)
- Climate indices (summer days, wet days, etc.)
- But also more complex analysis tools to analyze complex processes and features

Complex analysis tools are more difficult to be used by end users.

Example of a Complex Analysis Tool

Cyclone Tracking Software based on Sinclair (2004).

⇒Tracking of Extratropical and Tropical Cyclones

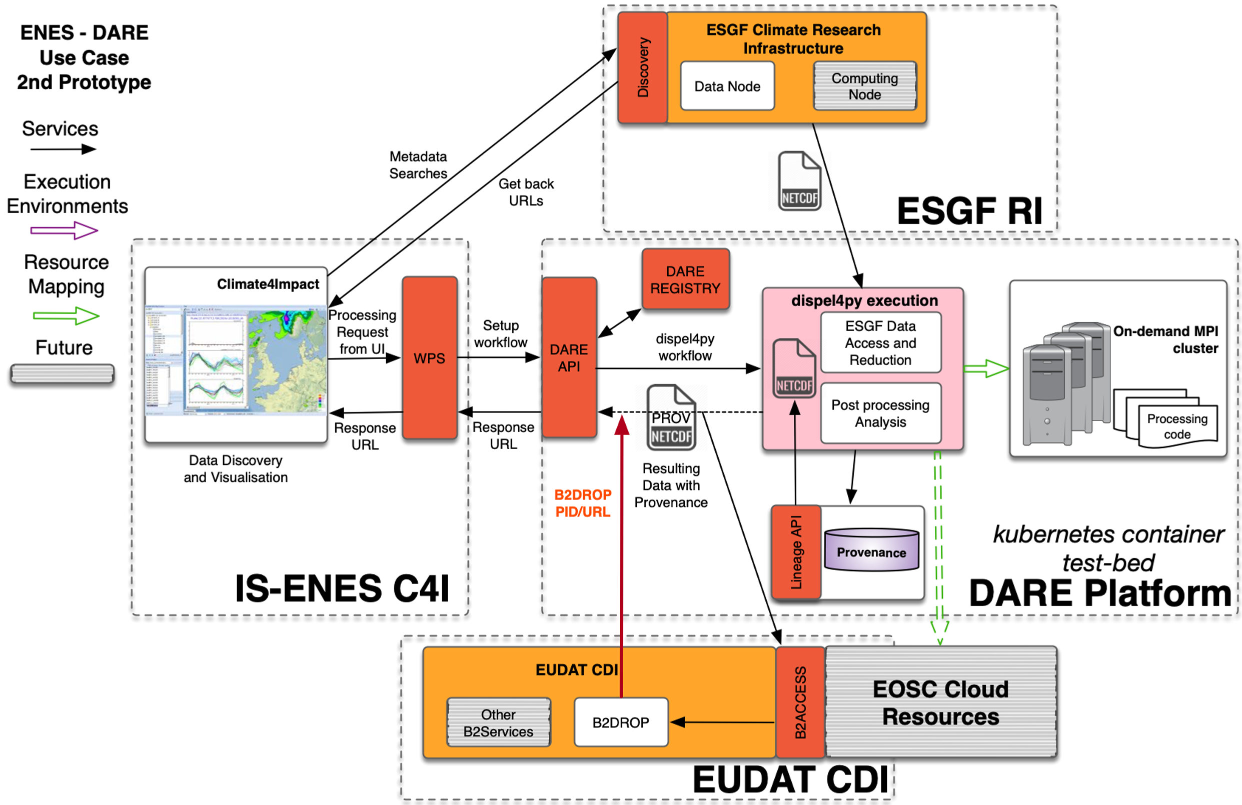
Installation and proper configuration is not trivial, and it requires large input data files.

Objective is to leverage this kind of tools to make them easily accessible to end users

Some background on this tracking algorithm and implementation

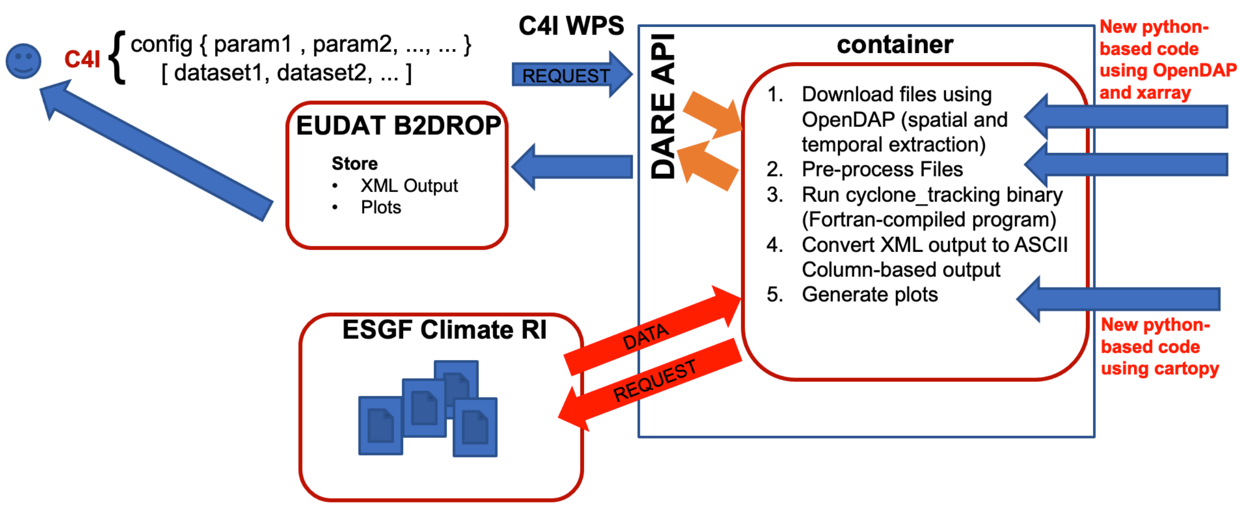
- Written in Fortran originally by Sinclair (USA)
- Modified extensively by UQAM and Ouranos (Canada)
- Tropical cyclone tracking added by the Meteorological Service of Canada
- Adapted to the NetCDF file format at CERFACS (France)

IV INTEGRATING COMPONENTS



The figure above presents the general overview of the integration

- Integration of these infrastructures: ESGF, DARE, EUDAT and ENES.
- Workflow has been implemented in a Jupyter Notebook, so it can also be used as a standalone, or in a Jupyter Lab environment.
- Connection to EOSC Resources and Services can be made through the use of EUDAT Integrated B2 Services.
- The DARE Platform is providing:
    - CWL workflow execution on kubernetes using a docker environment
    - Automated provenance using the PROV-O standard for both dispel4py and CWL workflows



Details about the workflow components

- Main component is the cyclone tracking program
  - Written originally in Fortran by Sinclair (USA)
  - Modified extensively by UQAM and Ouranos (Canada)
  - Tropical cyclone tracking added by the Meteorological Service of Canada
  - Adapted to the NetCDF file format at CERFACS (France)
- Workflow has been written using Python 3 processing elements
  - Separated in smaller standalone components
  - Encapsulated in bash shell scripts for integration into CWL
  - Integrated in a Common Workflow Language (CWL) workflow description
- Data access is using OpenDAP to access directly remote files, doing on-demand spatial and temporal subsetting that the OpenDAP protocol is providing
- A generic xarray-based python script has been written to pre-process input files and doing all needed subsetting steps using OpenDAP or local input files
- End results are stored into the user's EUDAT B2Drop Service
- The C4I Platform is providing, through its UI: user's algorithm configuration, input files' locations, and subsetting parameters

DISCLOSURES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements N°824084 and N°777413.

AUTHOR INFORMATION

Christian Pagé[1]

Maarten Plieger[2], Wim Som de Cerff[2], Alessandro Spinuso[2]

Rosa Filgueira[3], Malcolm Atkinson[3]

Sissy Themeli[4], Iraklis Klampanos[4], Vangelis Karkaletsis[4]

(1) CECI, Université de Toulouse, CNRS, Cerfacs, Toulouse, France

(2) Royal Netherlands Meteorological Institute (KNMI), R&D Observations and Data Technology, De Bilt, Netherlands

(3) University of Edinburgh, Scotland, United Kingdom

(4) National Centre for Scientific Research “Demokritos”, Agia Paraskevi, Greece

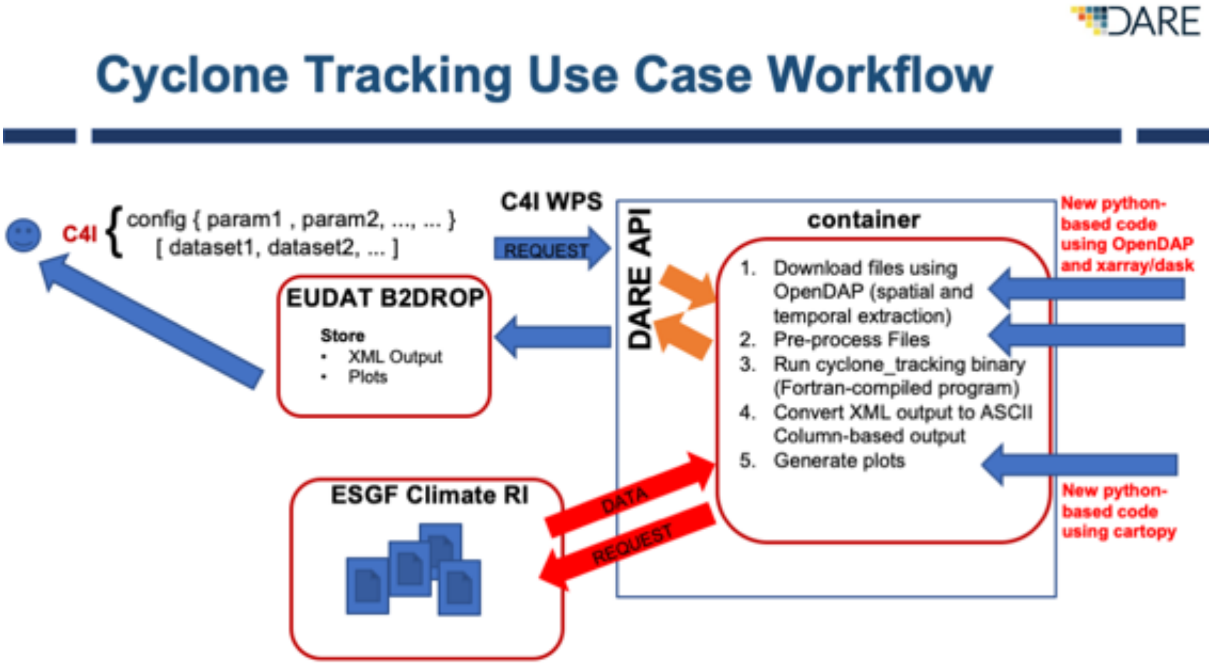
ABSTRACT

Researchers and end users using climate data face a challenge when they analyze the data they need. Data volumes are increasing very rapidly, and the ability to download all needed data is often no longer possible. Also, it can be complex to install, configure and use some advanced analysis tools on such large datasets. This is especially true when they are stored in a federated architecture like the ESGF.

An example of a complex analysis tool used in climate research and adaptation studies is a tool to follow storm tracks. In the context of climate change, it is important to know if storm tracks will change in the future, in both their frequency and intensity. Storms can cause significant societal impacts, hence it is important to assess future patterns. Having access to this type of complex analysis tool is very useful, and integrating them with front-ends like the IS-ENES climate4impact (C4I) would enable the use of those tools by a larger number of researchers and end users.

Integrating this type of complex tool is not an easy task. It requires significant development effort, especially if one of the objectives is also to adhere to FAIR principles. The DARE Platform enables research developers to develop the implementations of scientific workflows more rapidly. This work presents how such a complex analysis tool has been implemented to be easily integrated with the C4I platform. The DARE Platform also provides easy access to e-infrastructure services like EUDAT B2DROP, to store intermediate or final results and powerful provenance-powered tools to help researchers manage their work and data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements N°824084 and N°777413.



(https://agu.confex.com/data/abstract/agu/fm20/2/2/Paper_669222_abstract_649537_0.png)

REFERENCES

Sinclair, M. R., 2004: Extratropical Transition of Southwest Pacific Tropical Cyclones. Part II: Midlatitude Circulation Characteristics, Mon. Wea. Rev., 132, p. 2149.