

Using the EGI FedCloud and the ESGF CWT-API in a WPS workflow to provide data analysis computations for the IS-ENES climate4impact platform

Christian Pagé, CERFACS

Xavier Pivan (CERFACS)

Asela Rajapakse (MPI-M)

Wim Som de Cerff, Maarten Plieger, Ernst de Vreede, Alessandro Spinuso (KNMI)

Lars Barring (SMHI)

Antonio Cofiño (UCantabria)

Alessandro d'Anca, Sandro Fiore (CMCC)

Current Situation: Climate Domain

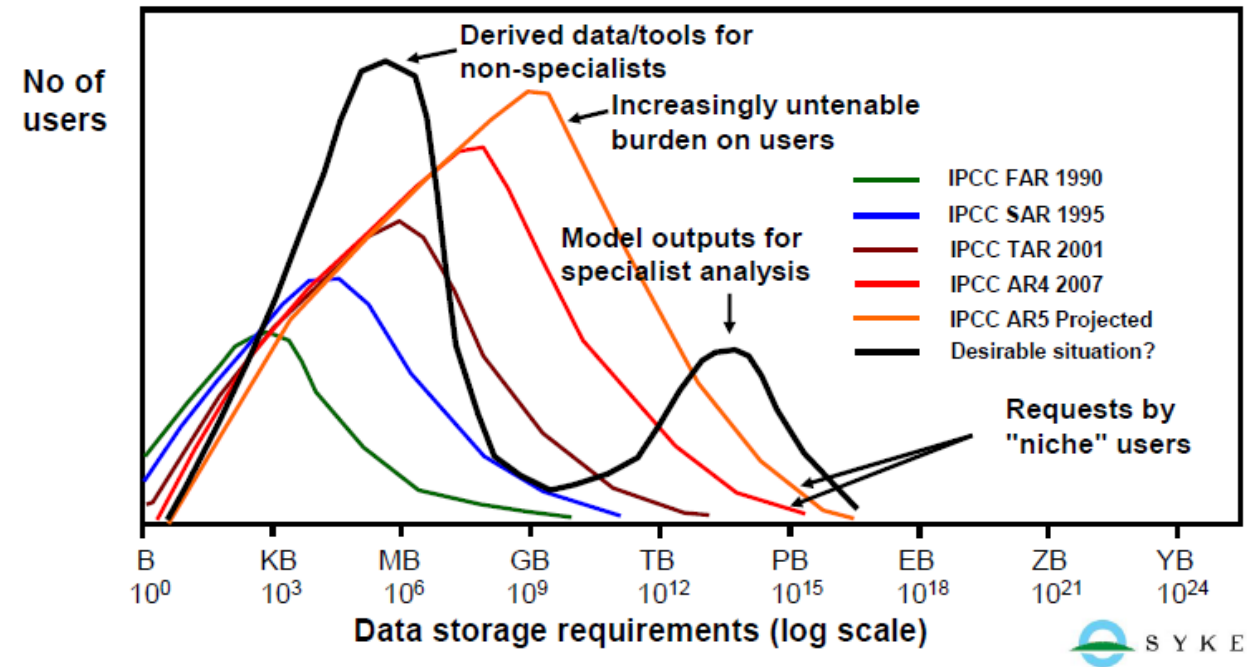
Data available for scientific analysis: a very large trend

- Limitations in data access means limitations in data analytics and scientific results

Download locally then Analyze:

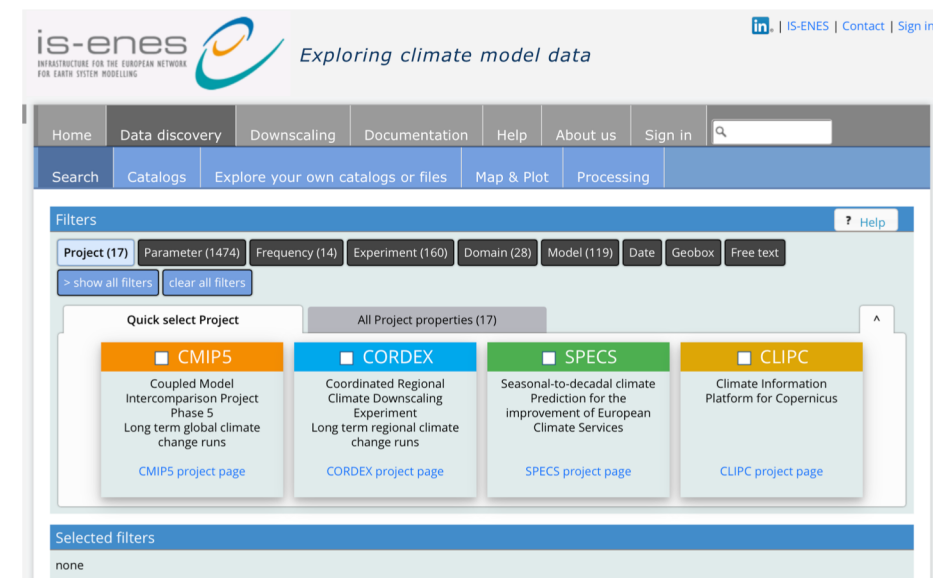
A workflow that cannot be sustained

- Climate researchers
- Impact researchers



Motivations

- Provide climate projections data to climate change impact researchers, facilitators, practitioners
 - Ease access with better intuitive interfaces
 - Provide more common data formats
 - Generate tailored products from data processing workflows

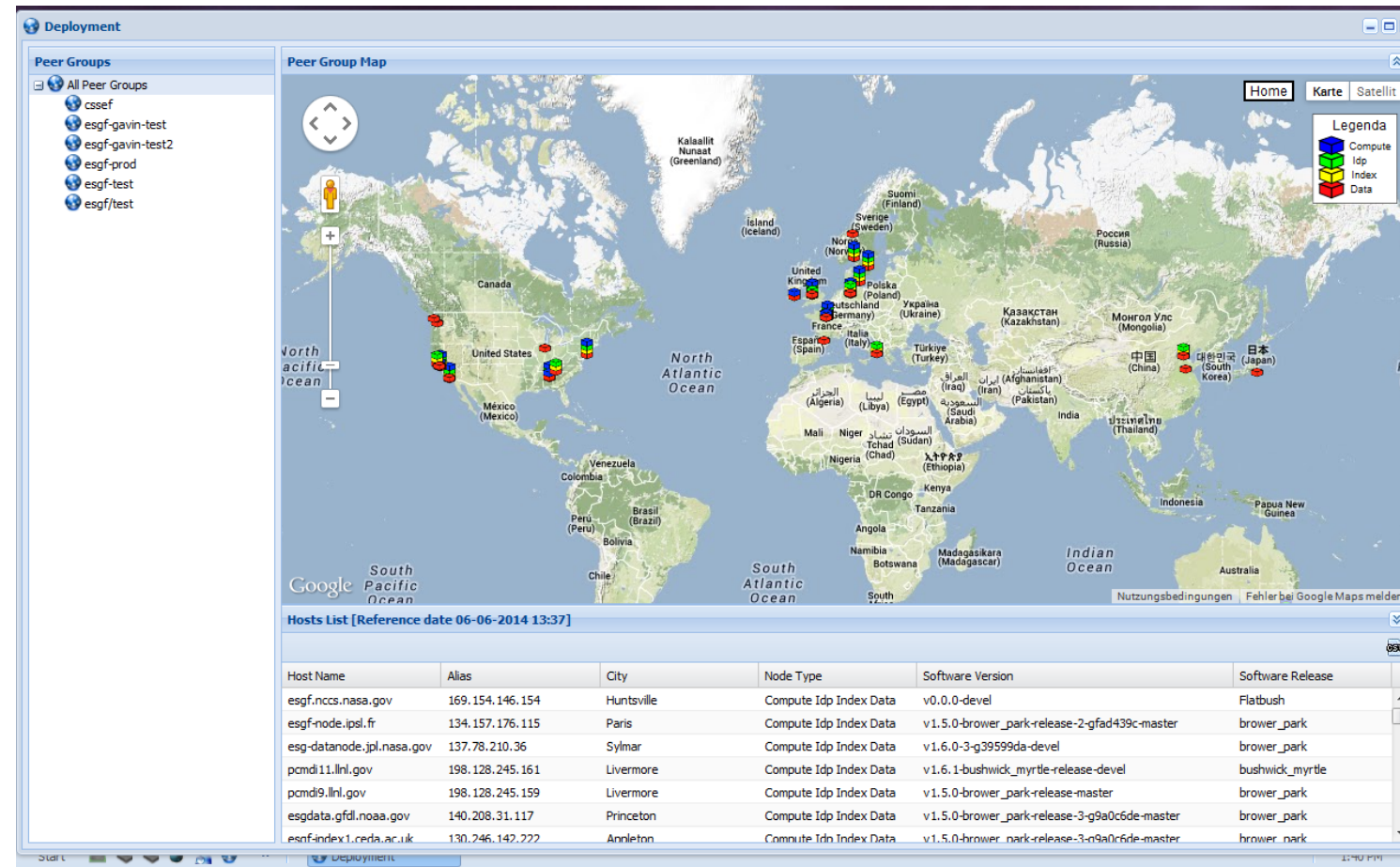


<http://climate4impact.eu>

Climate Data Distribution: ESGF

ESGF Data Nodes 2015:

- 40 worldwide
- 18 in Europe (coordinated in IS-ENES)



European Landscape & Components ESGF CWT

ESGF Computing Nodes: CWT API

- ◆ **Goal:** perform data analysis near the data storage
 - Better data access
 - Move away from the download/analyze workflow

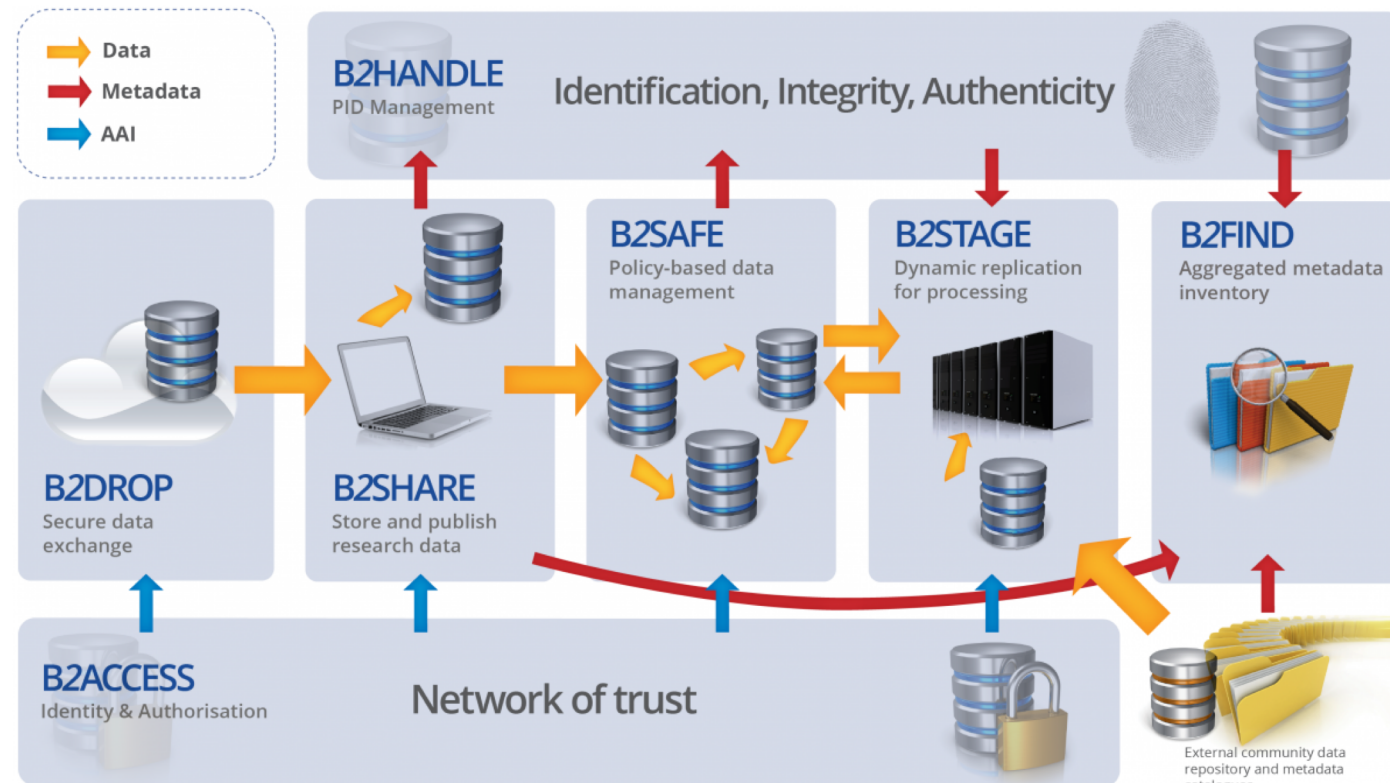
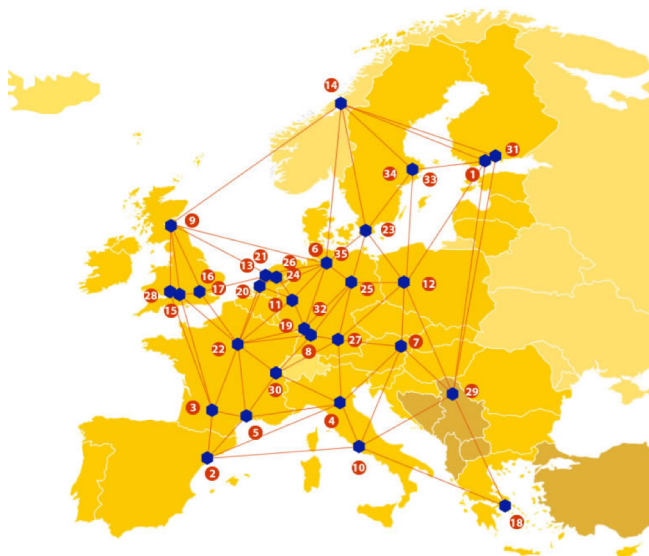


European Landscape & Components

EUDAT

EUDAT CDI B2 Service Suite

- ▶ Integrated B2 Services
- ▶ B2ACCESS: Common AAI
- ▶ Interface between EUDAT B2 Services and Communities infrastructures, such as Climate
- ▶ Prototype Workflow Service: GEF (Generic Execution Framework)

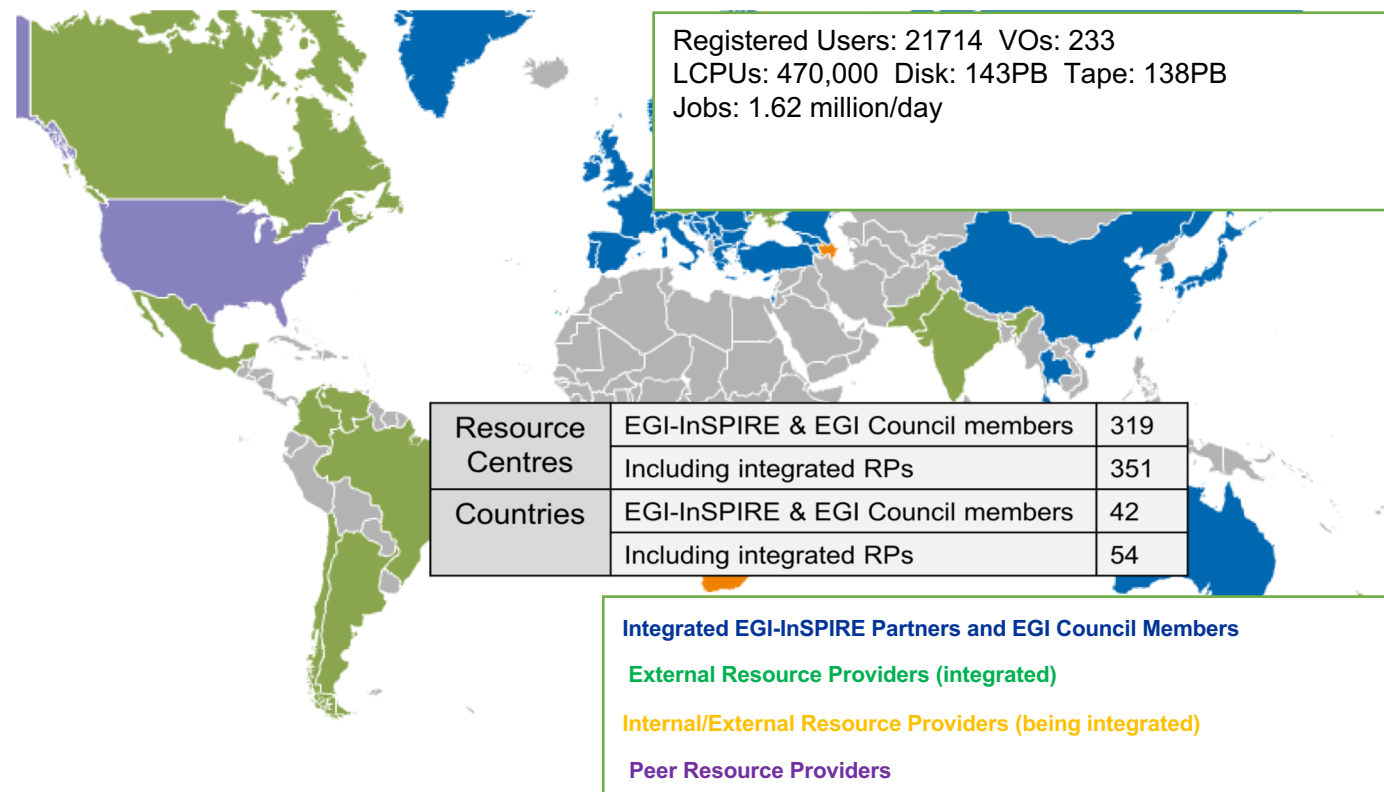


European Landscape & Components

EGI

EGI: European Grid Infrastructure

- ▶ Computing Power (FedCloud)
 - ▶ VM Based
 - ▶ Resource EndPoints: #Cores
#RAM #Linux
 - ▶ Storage
 - ▶ Disk & Tape

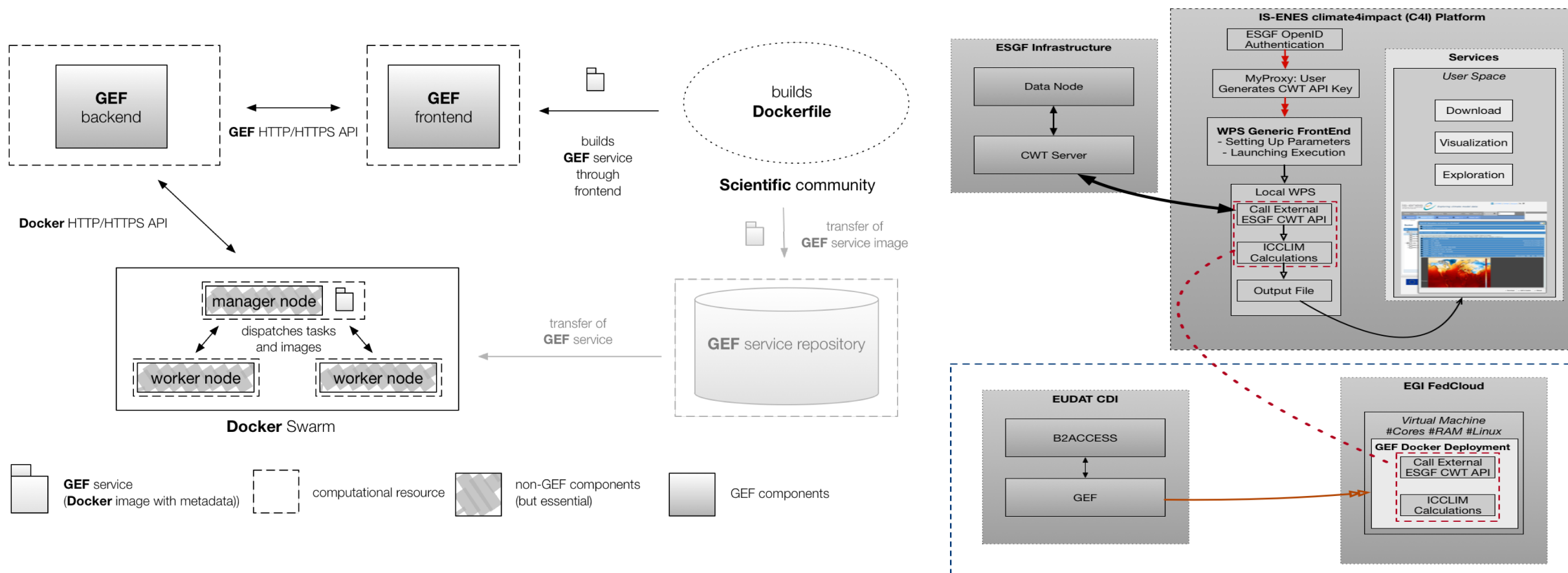


24 countries

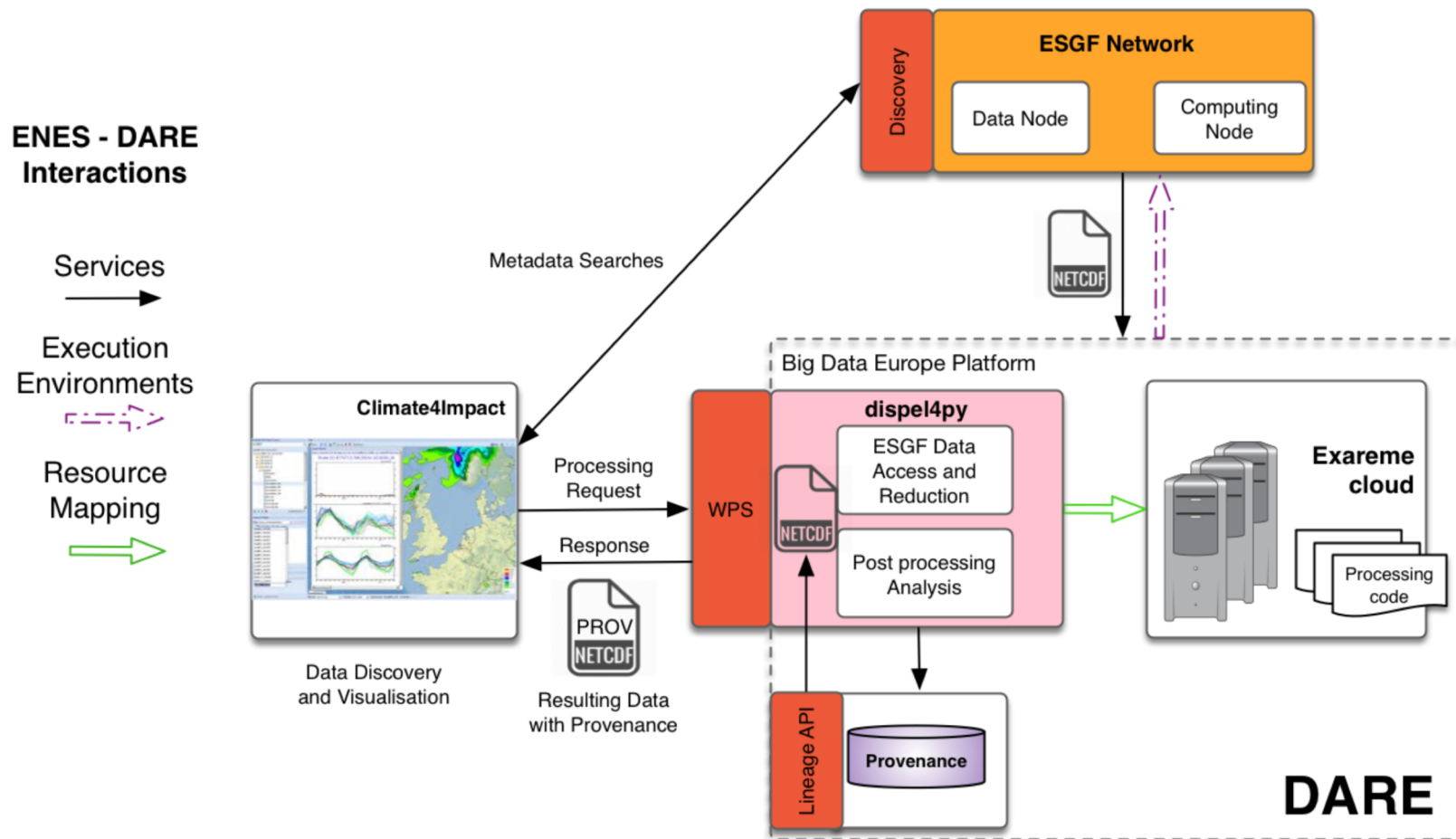
1 coordinating organization – EGI.eu

European Landscape & Components

EUDAT GEF & EGI



Using the DARE Platform as another backend



The DARE Platform

- Enable the **delegation of C4I Platform Data Analytics and Processing** to the **DARE** infrastructure (cloud-ready).
 - Typically, data reduction **on the order of 90%** can be achieved, depending on the users' analyses
- Streamline and **ease** the whole **data lifecycle**
- The DARE Platform will also:
 - **Be interoperable** with **EUDAT** by using its standards and services
 - **Interface** with **EOSC** and **Copernicus C3S-DIAS**

Merci!



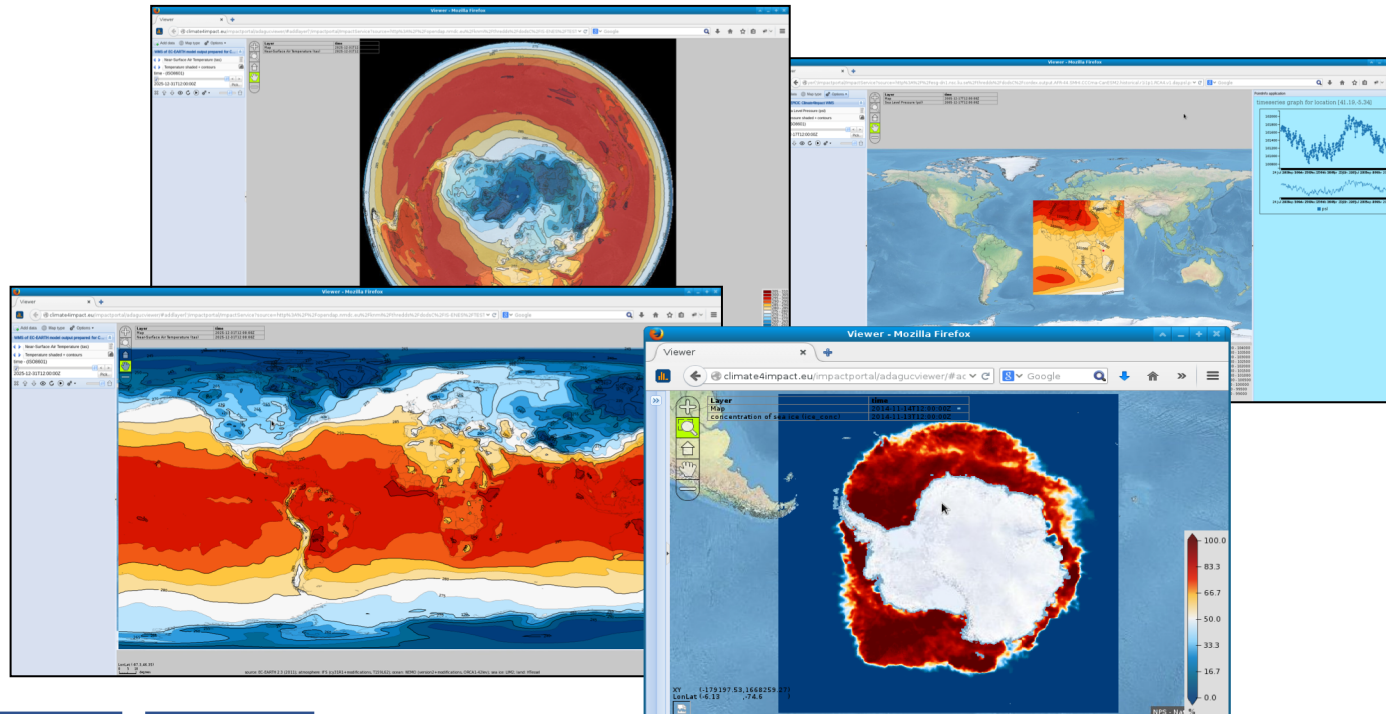
Supplemental Material



[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Real use of open standards and open source software:

- Data access over OPeNDAP → THREDDS
- Online analysis using Web Processing Services → PyWPS and ICCLIM
- Online visualization using Web Map Services → ADAGUC WMS
- Single Sign On using OpenId, delegation using MyProxy X509

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

Intro

Data

Search

Visualize

Process

Download

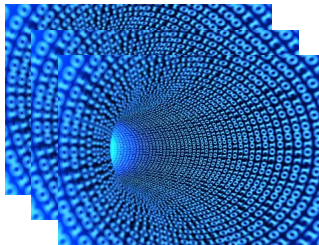
Summary

Data standards

Work done on client



Data files
HTTP / FTP



Array based data

OPeNDAP

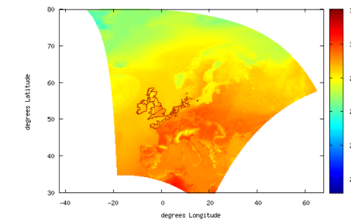
```
01001010010
01010010010
10010100100
10010011000
```

Referenced data

OGC® WCS

```
44 22 27 22 20 22 40 20 25 21 23 24 25 24
40 22 24 24 24 24 40 22 22 21 20 20 20 20 20
17 22 24 24 24 24 21 20 20 20 20 20 20 20 20
40 21 20 20 20 20 44 24 24 20 20 20 20 20 20
42 27 21 21 20 20 41 25 20 20 20 20 20 20 20
4 25 24 24 21 27 26 21 25 23 24 24 24 24
31 20 26 23 20 22 24 28 24 25 24
39 26 23 26 26 26 26 28 20 24 20 20
4 21 22 29 29 26 28 28 24 25 20
42 28 22 25 24 21 23 24 25
36 20 25 28 29 29 27 28 24
8 26 22 29 26 24 24 22
12 24 28 29 24 23 25
```

Graphics
OGC® WMS



Work done on server



(de Boer & Plieger, 2014)

Intro

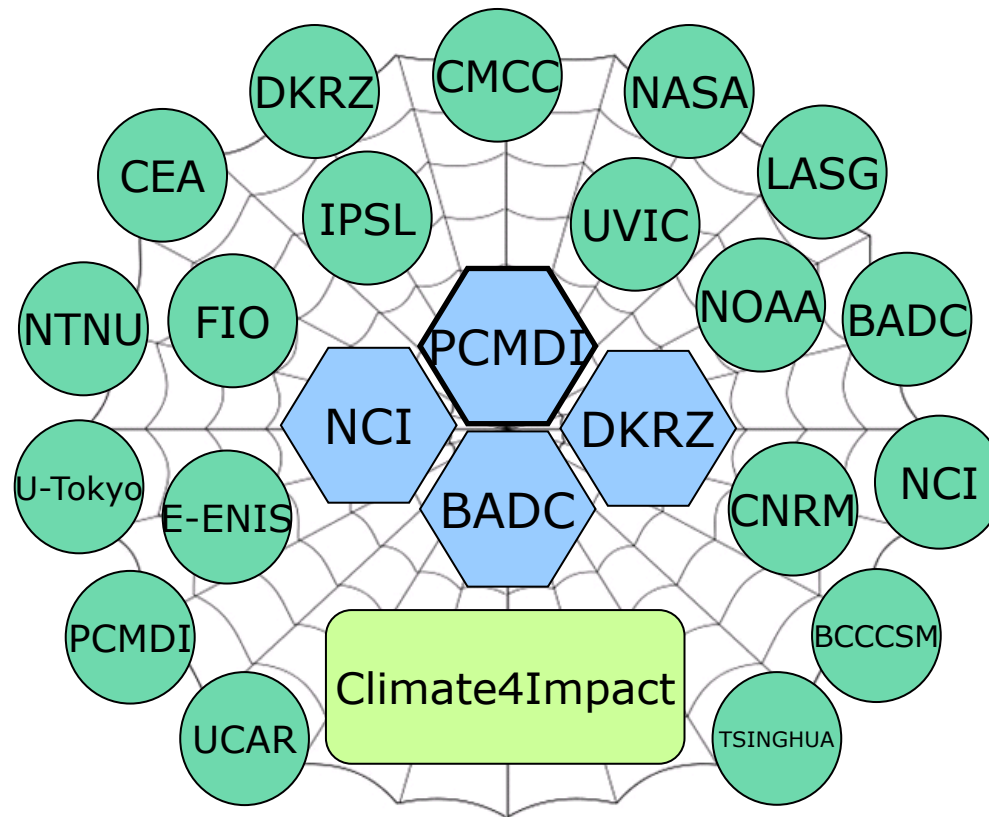
C4I

EUDAT

DARE

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Earth System Grid Federation



- Robust and distributed
- Index nodes (blue)
- Data nodes (green)
- Global Climate Model Data (CMIP5)
- Regional Climate Model Data (CORDEX)
- ~3 Petabyte of data
- Security: OpenId and X509
- **Search API offered**
- **OpenDAP data access offered**
- Climate4impact builds on and contributes to this global infrastructure

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

OPeNDAP – for data access and subsetting



- OPeNDAP is the name of the organization and the name of the protocol
 - Open-source Project for a Network Data Access Protocol
- Data is stored at remote server
- Data model is similar to NetCDF's data model (with differences)
 - N-dimensional array container, with variables, dimensions and attributes
- Only requested pieces of data are sent
 - Accessing small pieces of large files on a remote server can still be quick
 - Data is requested based on sub-setting along dimensions
- OPeNDAP resources can be opened locally on your computer as if it were local files using the NetCDF library
 - Local files versus remote files is transparent
- The concept of a file is gone, an OPeNDAP endpoint can represent thousands of files aggregated along a dimension
 - E.g. Usually concatenate a large time series observation to one endpoint using the time dimension

OPeNDAP within ESGF is served using the THREDDS data server

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

Intro

Data

Search

Visualize

Process

Download

Summary

The screenshot shows the IS-ENES Climate4Impact search interface. The browser address bar displays the URL: `climate4impact.eu/impactportal/data/basicsearch.jsp#project_CMIP5=1&variable_tasmin=1&time_frequency_day=1&experiment_rcp45=1`. The page header includes the IS-ENES logo and the tagline "Exploring climate model data". The navigation menu contains links for Home, Data discovery, Downscaling, Documentation, Help, About us, and Sign in. The search results section shows the following filters:

- Project:** ☒ CMIP5, ☐ CORDEX
- Variable:** ☐ Temperature, ☐ Precipitation, ☐ Windspeed, ☐ Shortwave radiation down, ☐ Surface specific humidity, ☒ Min temperature, ☐ Conv. precipitation, ☐ Max windspeed, ☐ Shortwave radiation up, ☐ Surface relative humidity, ☐ Max temperature, ☐ Snow, ☐ Eastward wind, ☐ Longwave radiation down, ☐ Specific humidity, ☐ Northward wind, ☐ Longwave radiation up, ☐ Relative humidity, ☐ Evaporation, ☐ Surface pressure, ☐ Diffuse radiation, ☐ Surface relative humidity, ☐ Potential evaporation, ☐ Pressure, ☐ Total cloud cover, ☐ Max relative humidity, ☐ Minimum relative humidity
- Frequency:** ☐ 3 hourly, ☒ daily, ☐ monthly
- Time frame:** ☐ Historical, ☐ RCP26, ☒ RCP45, ☐ RCP60, ☐ RCP85, ☐ Evaluation, ☐ 1pctCO2
- Domain:** Search domain (CORDEX)
- Models:** Found 182 model(s)

The search results section shows "Found 106 datasets. (see esgf query)".

Search based on feedback from impact researchers → quick preset of common queries

Intro

C4I

EUDAT

DARE

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

New faceted search

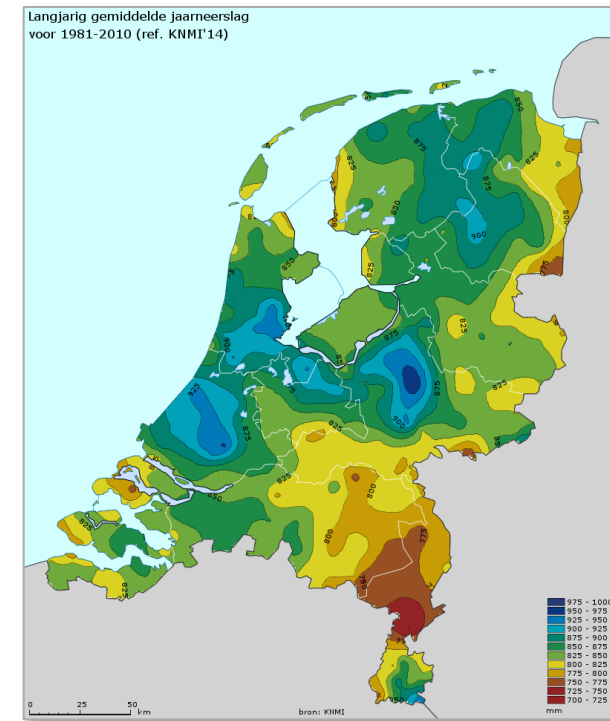
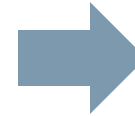
Faceted search allows to drill down search results using available filters in the federation
Results from a search query are treated as a new dataset

The screenshot displays the Climate4Impact web application in a Mozilla Firefox browser. The page title is "Climate4Impact" and the URL is "localhost/impactportal/data/esgfsearch.jsp". The application header includes the "is-enes" logo and the tagline "Exploring climate model data". A navigation bar contains links for Home, Data discovery, Downscaling, Documentation, Help, About us, and Account. Below this, a secondary navigation bar includes Search, Faceted Search (active), Catalogs, Explore your own catalogs or files, Map & Plot, and Processing. The main content area is titled "Faceted search" and features a "Filters" section with various filter categories and their counts: cf_standard_name (16), model (1), data_node (1), experiment_family (2), product (1), ensemble (10), project (1), institute (1), time_frequency (1), realm (1), cmor_table (1), experiment (1), variable_long_name (19), and variable (19). A "Selected filters" section shows the current selection: data_node : albedo2.dkrz.de, experiment : rcp45, project : CMIP5, time_frequency : day, variable : tas, and model : EC-EARTH. Below this, a "Datasets: Found 10, displaying 10 of 10 results." section lists the dataset identifiers. At the bottom of the results list is an "Add results to basket" button. The footer indicates the user is logged in as "https://pcmdi9.lln.gov/esgf-idp/openid/c4m".

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

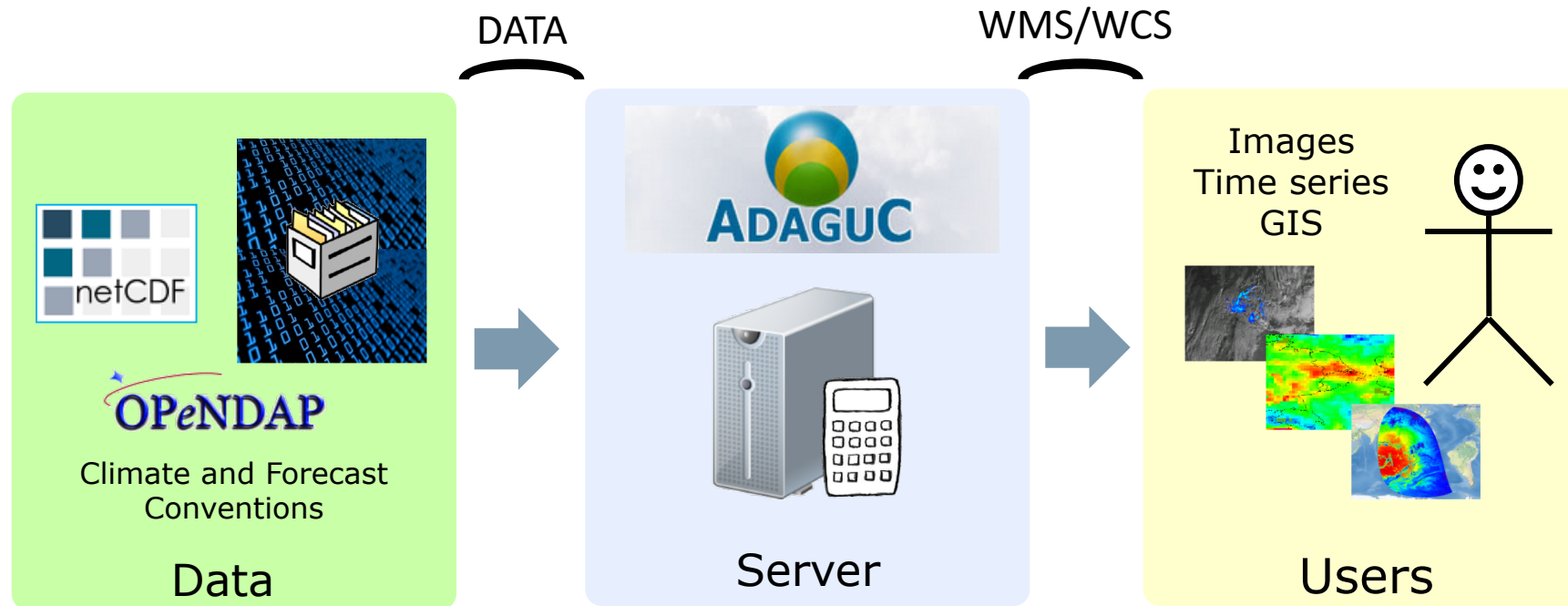
Web Map Service – for visualizations

- Generates visualizations of geospatial data in the form of 2D images, suitable for transfer over the internet (JPG/PNG/GIF)
 - REST based:
 - Compose an URL with key value pairs, and you will get an image!
 - Standard is developed and maintained by the Open Geospatial Consortium
 - Generated images are geo-referenced
 - Images from several sources can be easily combined
 - Images have dimensions
 - Time, elevation, member
 - WMS services can be viewed in many web based clients
 - OpenLayers, Leaflet, GoogleMaps, ADAGUC viewer, ...



[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

ADAGUC Web Map and Web Coverage server



e.g.:

OPeNDAP

OGC® WMS

Geographical visualization framework using open standards and formats: <http://adaguc.knmi.nl/>

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

Intro

Data

Search

Visualize

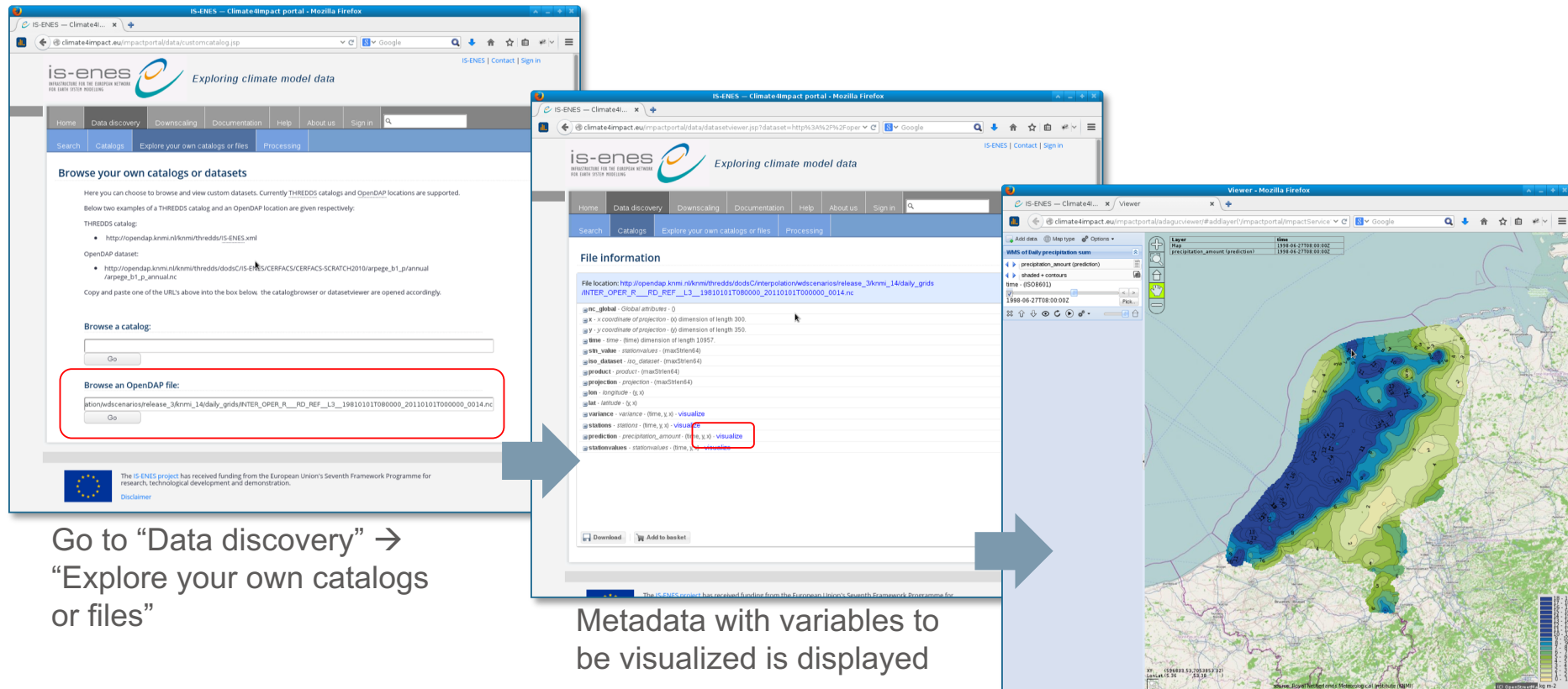
Process

Download

Summary

Web Map Services based on OPeNDAP resources

Climate4impact.eu allows for creation of WMS visualizations on OPeNDAP endpoints:



Intro

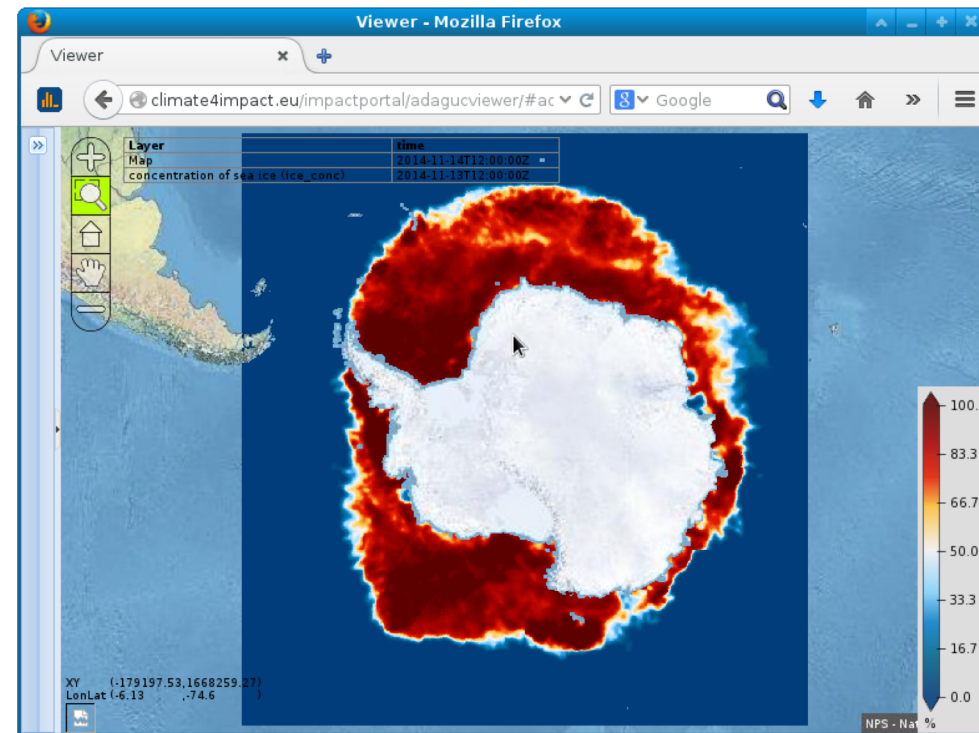
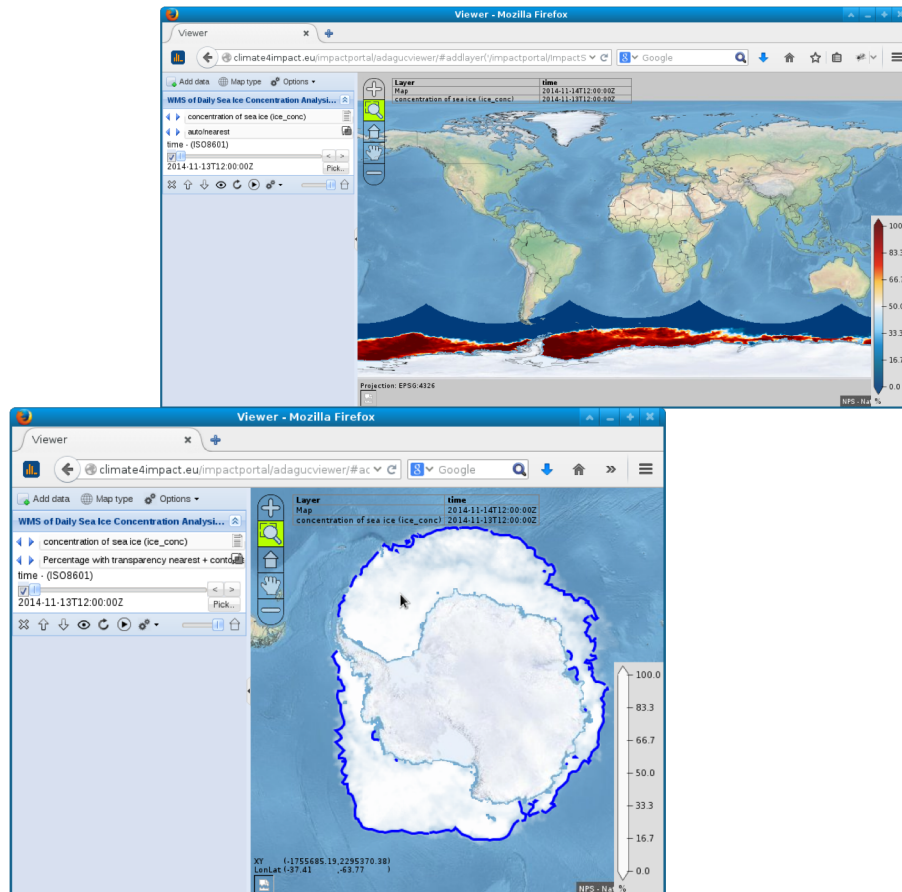
C4I

EUDAT

DARE

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Example: Many existing resources with OpenDAP enabled can already be visualized!



Concentration of sea ice for 2014-11-13

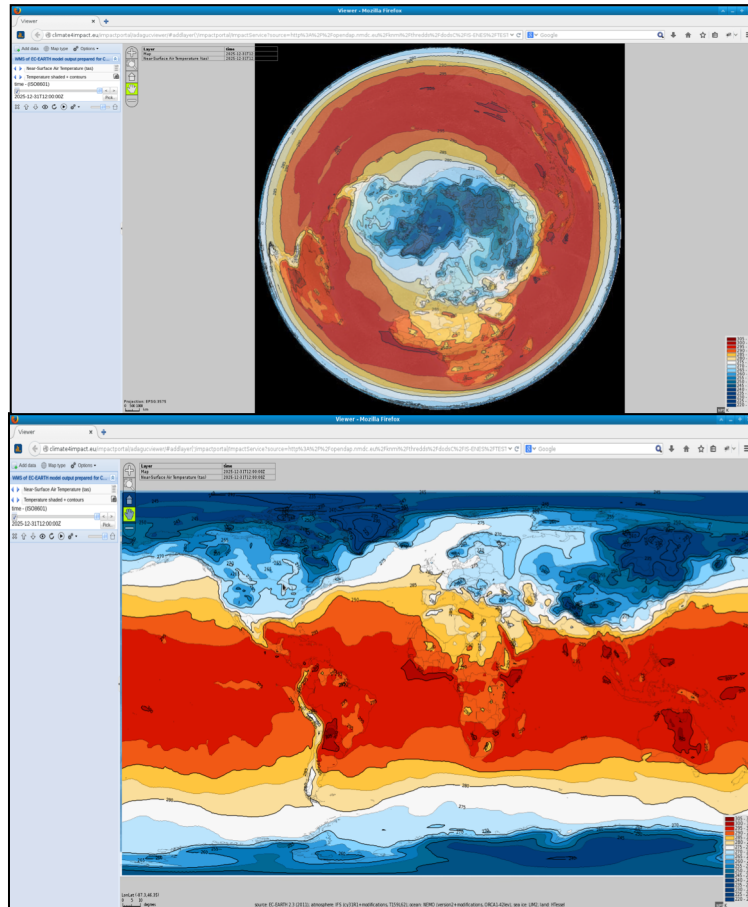
EUMETSAT Ocean and Sea ICE SAF from Norwegian Meteorological Institute

http://met.no/Hav_og_is/English/Access_to_data/

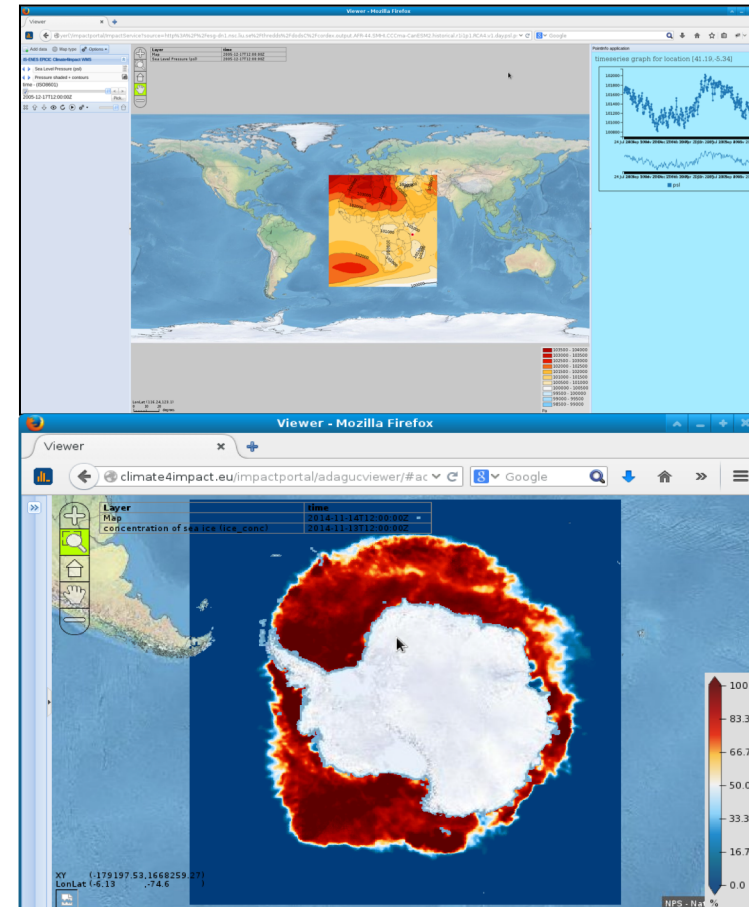
[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Example: Many existing resources with OpenDAP enabled can already be visualized!



CMIP5 - global climate models



CORDEX - regional climate models

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Web Processing Service for climate indices calculations

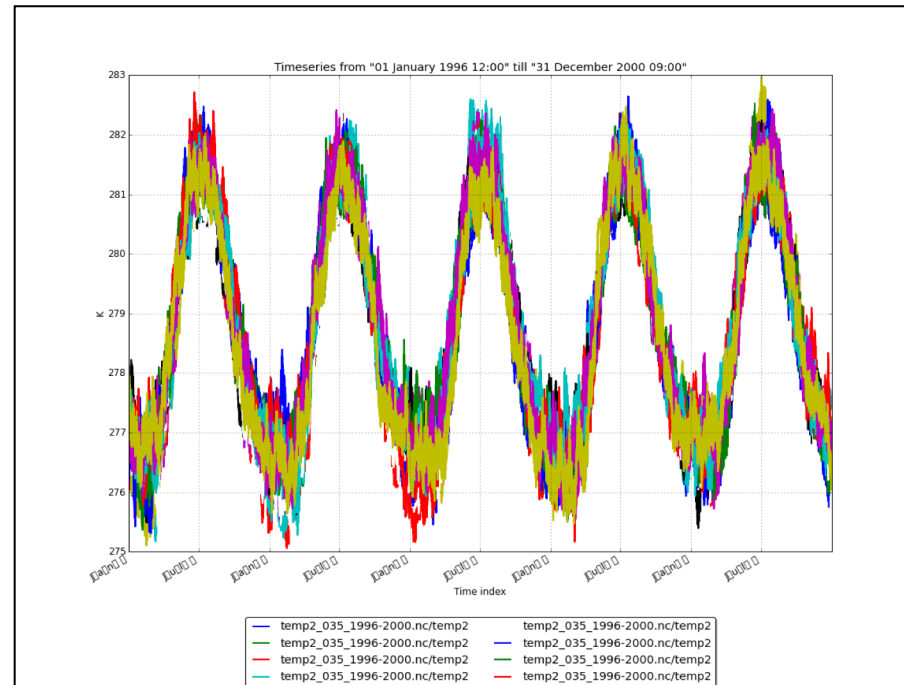
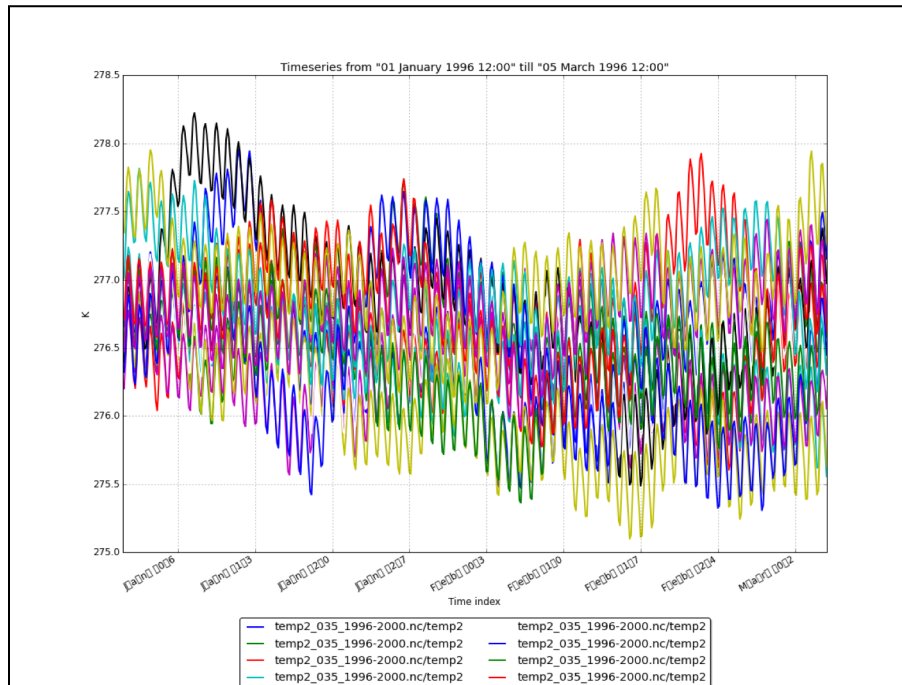
- Uses PyWPS and ICCLIM
- User interface is build automatically based on DescribeProcess XML file.
- Interface supports:
 - Link to basket
 - Comboboxes / select from list
 - Strings/text elements
- IS-ENES2 is working on an indices wizard for user friendly indices calculation
- Climate4impact WPS can be interfaced to other processing packages
- Climate indices calculations have added value!

The screenshot shows the IS-ENES Climate4impact portal in a Mozilla Firefox browser. The page title is 'IS-ENES — Climate4impact portal - Mozilla Firefox'. The URL bar shows 'climate4impact.eu/impactportal/account/wpsuseprocessor.jsp?processo'. The page features a navigation bar with links: Home, Data discovery, Downscaling, Documentation, Help, About us, and Account. Below the navigation bar, there are tabs for Account, Basket (12), Processing, and Monitor jobs (18). The main content area is titled 'Use a processor' and displays the 'Processing details and options' for the 'Calculate number of summer days' process. The details include the Title, Identifier (indice_ICCLIM_SU), Abstract, and Location. A 'Start processing' button is visible. The 'Options' section includes a 'Threshold' field set to 25, a 'Maximum temperature variable' field set to 'tasmax', a 'Slice mode (temporal grouping to apply for calculations)' dropdown set to 'year', an 'Input file(s)' section with a 'Please select a file from the basket' message, and an 'Output file name' field set to 'SU.nc'.

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

Multi member timeseries – created with WPS

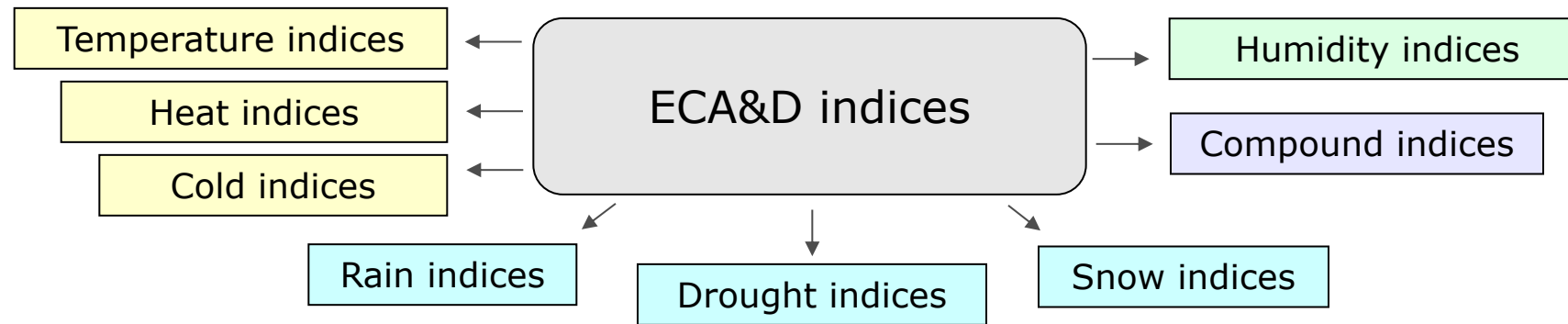


3 months
Temperature for several ensembles from the ESSENCE dataset

5 years

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

Indices calculation using ICCLIM – developed in IS-ENES



- Intra-period extreme temperature range [$^{\circ}$ C] - **ETR**
- Warm days (days with mean temperature > 90th percentile of daily mean temperature) - **TG90p**
- Summer days (days with max temperature > 25 $^{\circ}$ C) - **SU**



- Python code developed at CERFACS, started in September 2013
 - Generic and modular approach, can be reused in other environments
 - C functions called for optimization
- I/O interface is structured for optimal performance, with wrapper functions
- Some percentile-based indices (TG10p, TX10p, TN90p, etc): **OpenClimateGIS**

ICCLIM source code and documentation is available via <https://github.com/cerfacs-globc/icclim>

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)


Web Processing Service

Job progress can be viewed from anywhere



ENES Portal Interface for
the Climate Impact Communities

IS-ENES | [Contact](#) | [Login](#)

[Home](#) | [Data discovery](#) | [Map & Plot](#) | [Documentation](#) | [Help](#) | [About us](#) | [Account](#) |  (9) |

[Account](#) | [Basket \(9\)](#) | [Jobs \(7\)](#)

Processing jobs

Jobs for: <https://pcmdi9.llnl.gov/esgf-idp/openid/maartenplieger>

Started on:	WPS Identifier	Unique Id	Progress	View	X
2013-08-08 10:29:00Z	timeseries_avg2D	pywps-137595774038.xml	ready	view	X
2013-08-09 08:25:52Z	timeseries_avg2D	pywps-137603675248.xml	ready	view	X
2013-08-09 08:26:26Z	timeseries_avg2D	pywps-137603678625.xml	ready	view	X
2013-08-09 08:27:16Z	timeseries_avg2D	pywps-137603683692.xml	ready	view	X
2013-08-09 11:35:50Z	timeseries_avg2D	pywps-137604815013.xml	ready	view	X
2013-08-09 11:39:17Z	ensemble_dtdp	pywps-137604835705.xml	ready	view	X
2013-08-09 12:14:13Z	timeseries_avg2D	pywps-137605045340.xml	48 %	view	X

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

Intro

Data

Search

Visualize

Process

Download

Summary

- By default the basket contains:
 - “Remote data” for links
 - “My data” for your own data
- Script based download allows to select and download multiple files
- Client certificate (x509) is embedded in download script
 - No need for MyProxy login
 - No need for firewall changes
- The basket allows for uploading your own files
 - Can be used in processing or visualization

Climate4Impact

dev.climate4impact.eu/impactportal/account/basket.jsp

is-enes Exploring climate model data

Home Data discovery Downloading Documentation Help About us Account

Account Basket (11) Processing Monitor jobs (8)

Basket

File	DAP	HTTP	Filesize	Date
Remote data				
0.50 deg. regular grid			-	2015-01-22...
0.44 deg. rotated grid			-	2015-01-22...
tx_0.44deg_rot_v10.0.nc	true		691.9M	2015-01-22...
tn_0.44deg_rot_v10.0.nc	true		691.9M	2015-01-22...
tg_0.44deg_rot_v10.0.nc	true		691.9M	2015-01-22...
tasmax_day_IPSL-CM5A-LR_historical_r1i1p1_18500101-18991231.nc	true	true	673.2M	2015-03-19...
tasmax_day_IPSL-CM5A-LR_historical_r1i1p1_18500101-19491231.nc	true	true	1.346G	2015-03-19...
tasmax_day_IPSL-CM5A-LR_historical_r1i1p1_19000101-19491231.nc	true	true	673.2M	2015-03-19...
tasmax_day_IPSL-CM5A-LR_historical_r1i1p1_19500101-19991231.nc	true	true	673.2M	2015-03-19...
tasmax_day_IPSL-CM5A-LR_historical_r1i1p1_19500101-20051231.nc	true	true	754.0M	2015-03-19...
tasmax_AFR-44_CNRM-CERFACS-CNRM-CM5_rcp45_r1i1p1_CLMcom-CCLM4-8-17_v1_day_2096	true		-	2015-04-01...
My data				
polar_stereo_m.nc	true	true	906.824K	2015-01-23...
tas_WAS-44_ECMWF-ERAINT_evaluation_r1i1p1_IITM-RegCM4-1_v411_mon_198901-199012.nc	true	true	2.314M	2015-01-23...
tas_WAS-44_ECMWF-ERAINT_evaluation_r1i1p1_IITM-RegCM4-1_v411_day_19890101-19901231.nc	true	true	70.463M	2015-01-23...

View/Browse file Download file Script download Upload file Delete file(s) Reload basket

You are logged in as https://esg-dn1.nsc.liu.se/esgf-idp/openid/maartenpflieger

The IS-ENES project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration.

Disclaimer

Intro

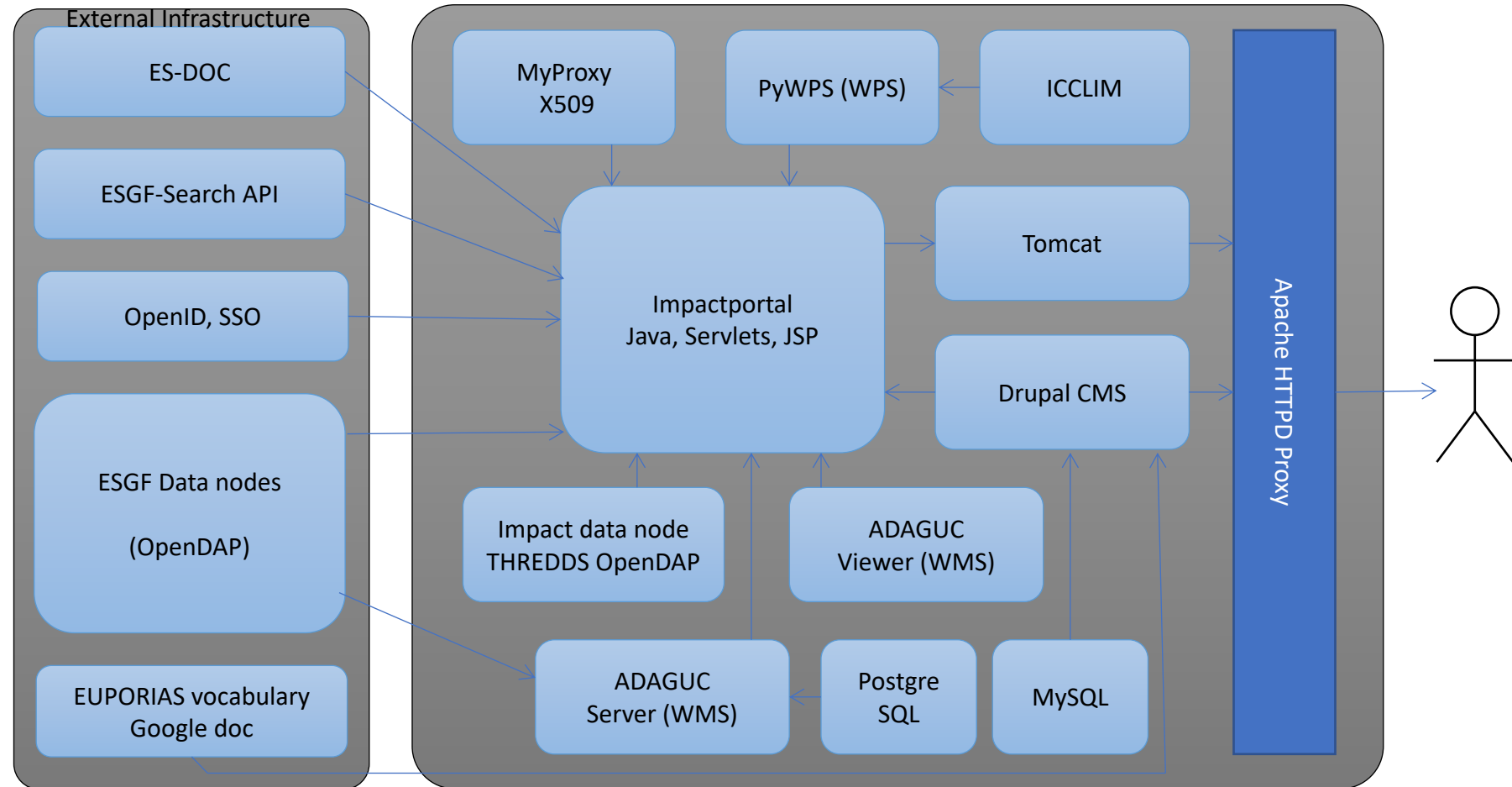
C4I

EUDAT

DARE

[Intro](#)[Data](#)[Search](#)[Visualize](#)[Process](#)[Download](#)[Summary](#)

How components fit together



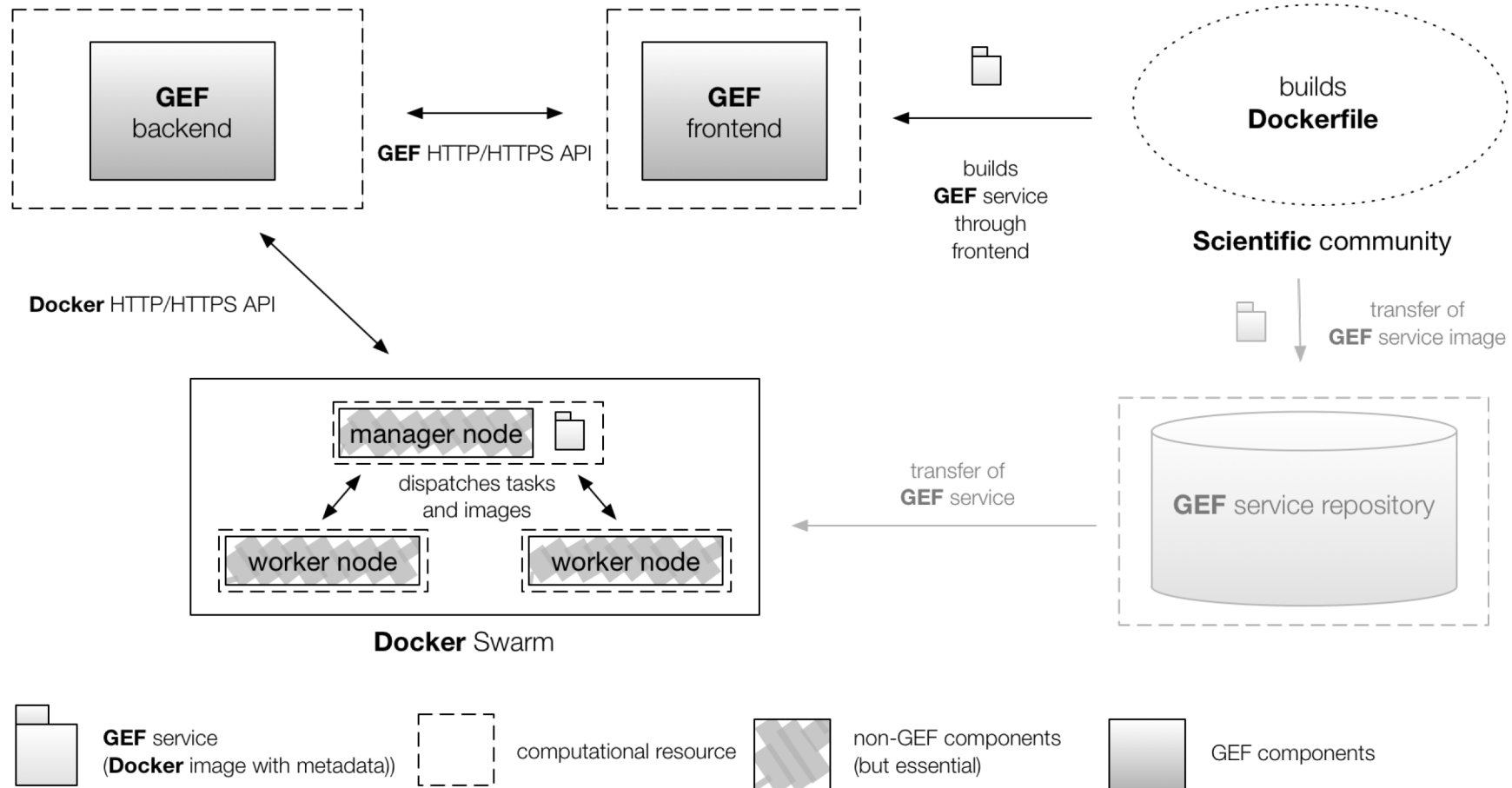
Climate4impact infrastructure

[Intro](#)[C4I](#)[EUDAT](#)[DARE](#)

About the EUDAT Generic Executive Framework

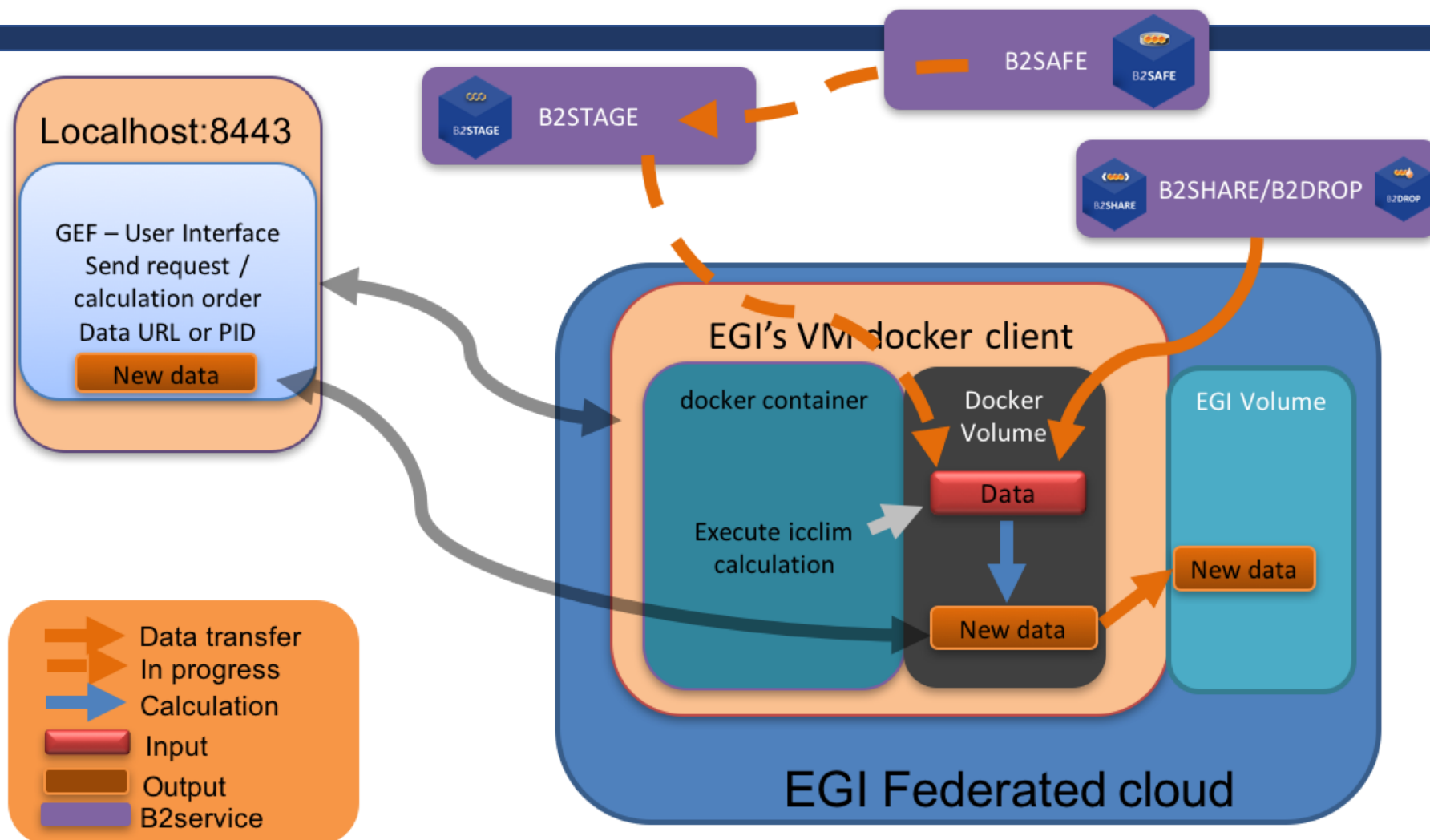
- Generic tool to encapsulate calculation using docker technology
- Generic -> convenient for all communities (Climate, Earth Science or Litterature)
- Community admin can create specific services

Generic Executive Framework (GEF)

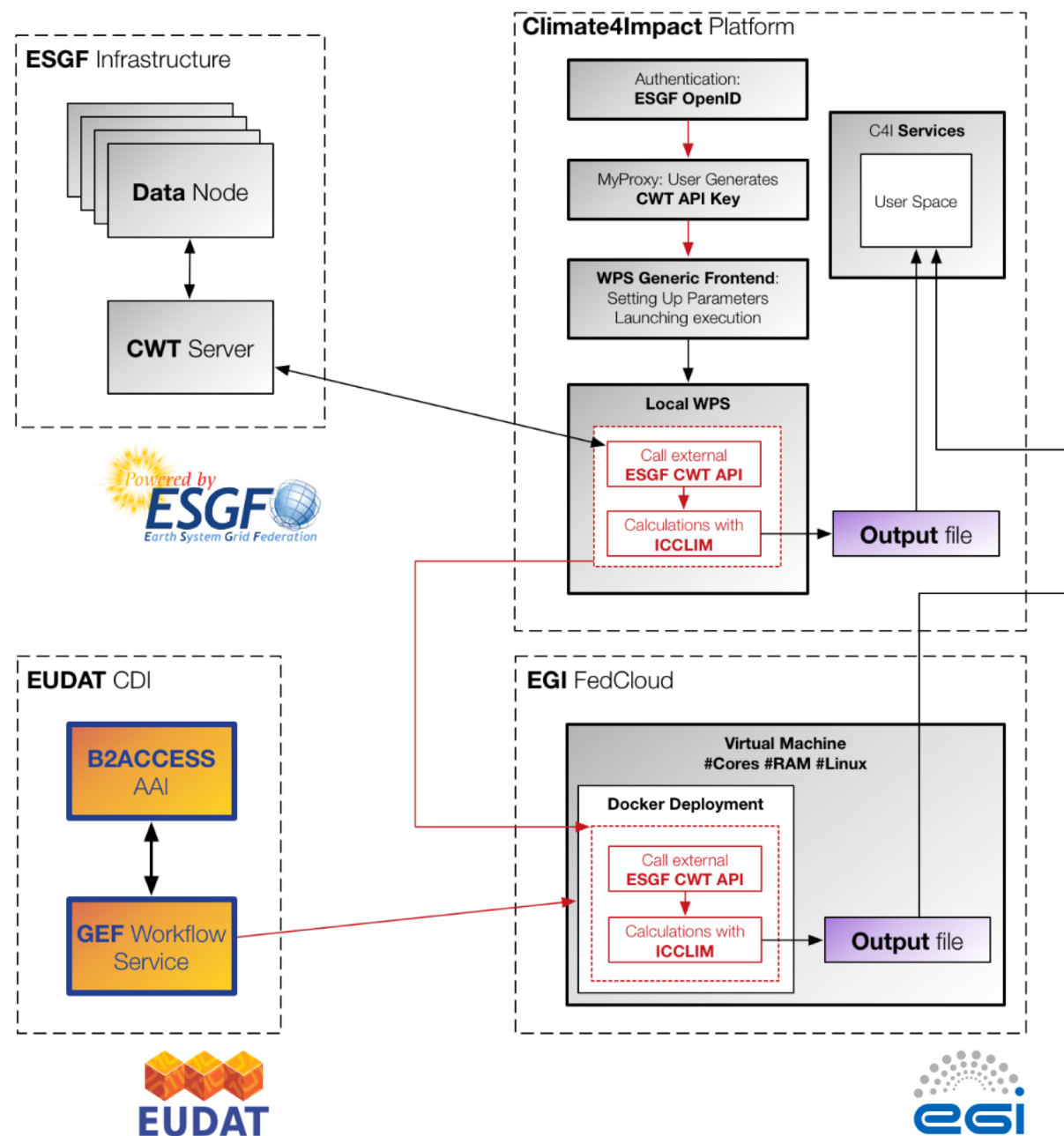


**Figure from GEF github

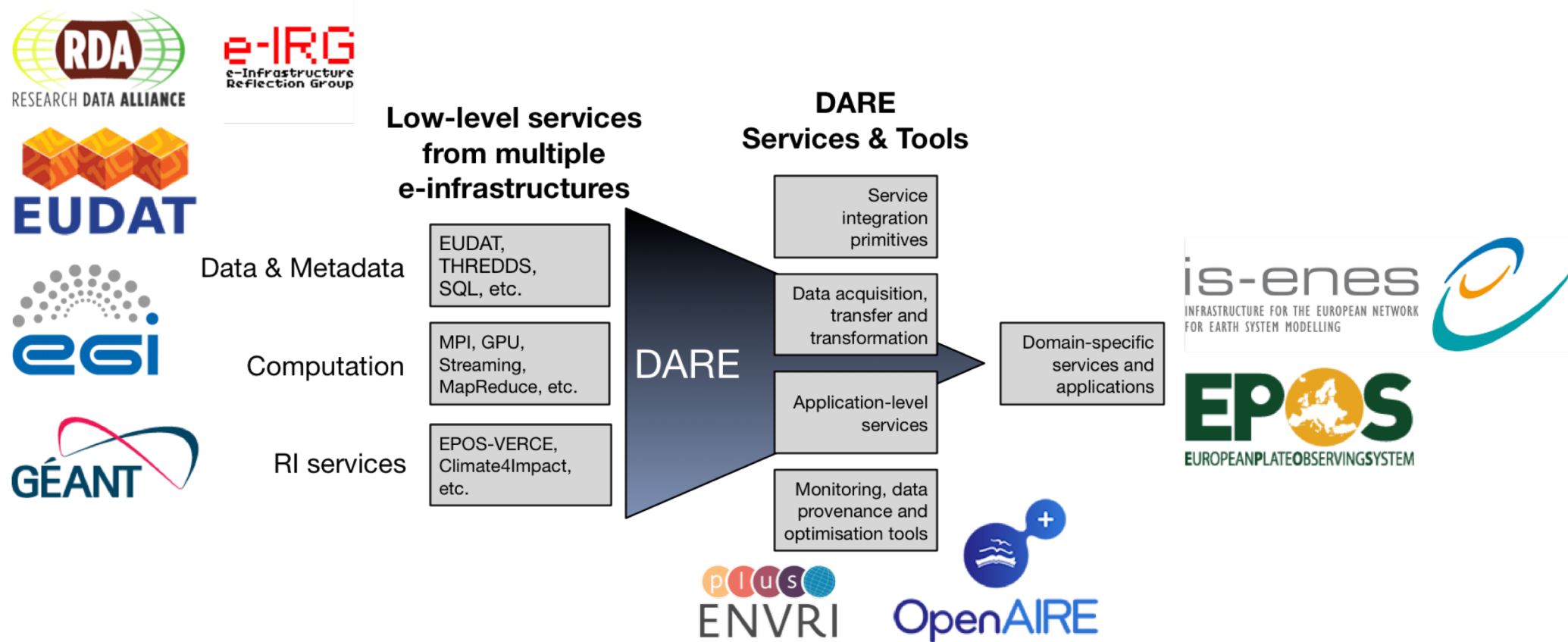
ENES Use Case Technical Aspects



ESGF – IS-ENES – EGI – GEF Integration Overview



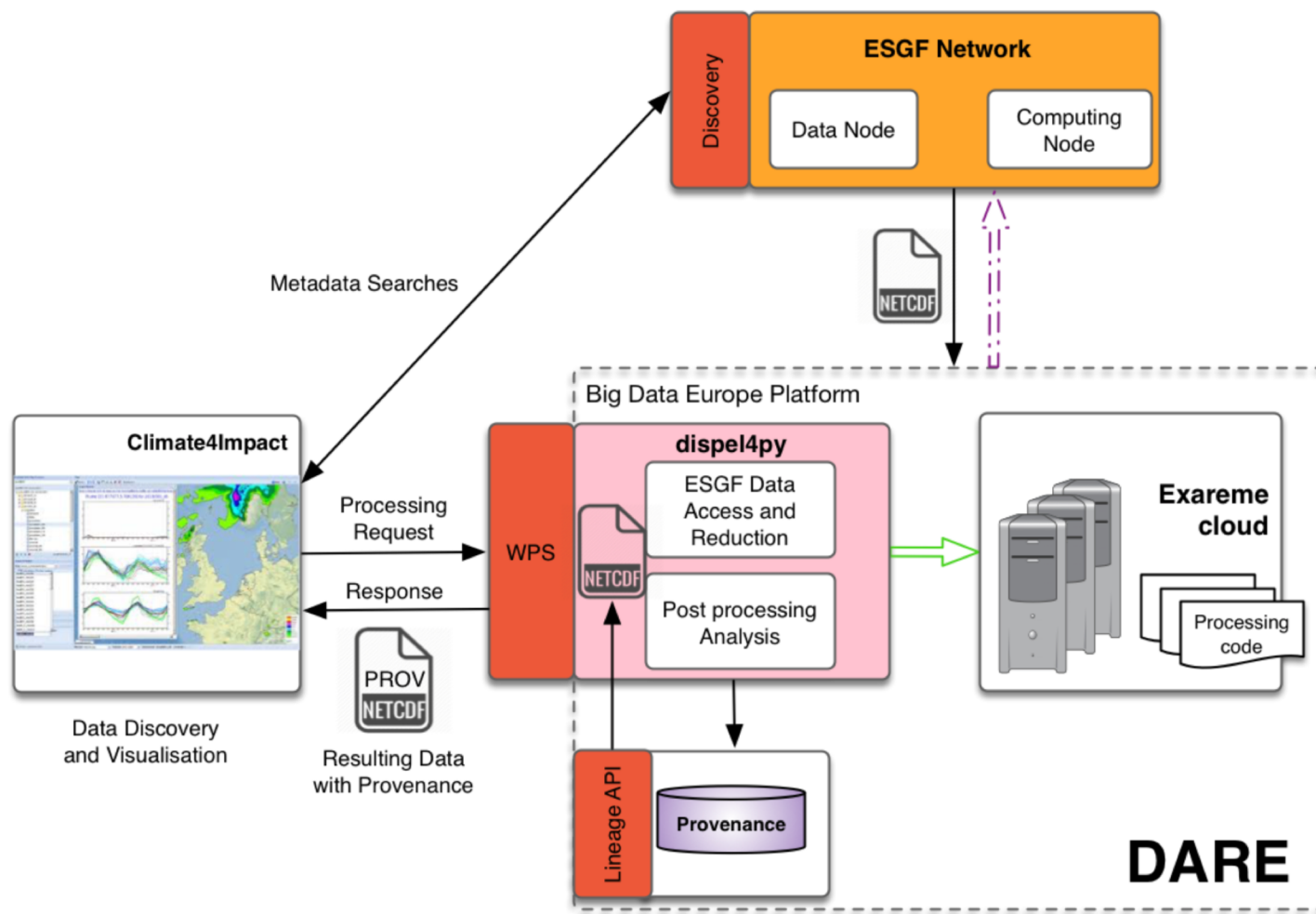
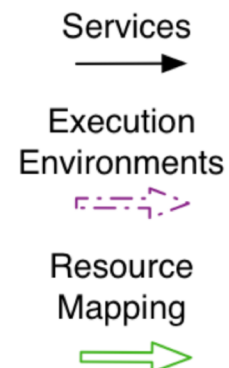
DARE and e-infrastructures



The positioning of DARE within the e-infrastructures context.

Using the DARE Platform as another backend

ENES - DARE Interactions



IS-ENES/C4I Pilot

Generic Use Case

Objective: Generate a multi-model multi-scenario time series average of the surface temperature using CMIP5 data

Scientific Workflow

- Spatially average over Western Europe (continents only)
- Time Period 1950-2100
- RCP 8.5 GES scenario
- All Global Climate Models available
- All members available
- Calculate the average time series
- Calculate the standard deviation
- Extract separate time series of every simulation
- Plot all those time series on a single graph

IS-ENES/C4I Pilot

Generic Use Case

Important Technical Details

- Input files are distributed on multiple ESGF data nodes
- Depending on the climate model, each input file can hold 5 to 10 years' worth of data
- Each file can have either Daily or Monthly data. Daily data will be used
- Typical input file size is on the order of 1 Gb per 5-year files

Estimation of input data volume

- 30 Gb per model per member, so an estimated size of about 300 Gb for 10 members

Data access

- Data files will be obtained through the C4I platform, which is already equipped with an efficient search service
- Authentication/Authorization is done through C4I