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

IS-ENES3 1st General Assembly
25-27th March 2020

Perspectives on compute services for ESGF

Sandro Fiore (CMCC)

This presentation includes material from the IS-ENES3/ESGF Virtual Workshop on Compute & Analytics (Dec 2, 2019).

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

Compute Services & IS-ENES3

- **A compute service aims to provide remote access to analysis and processing capabilities**

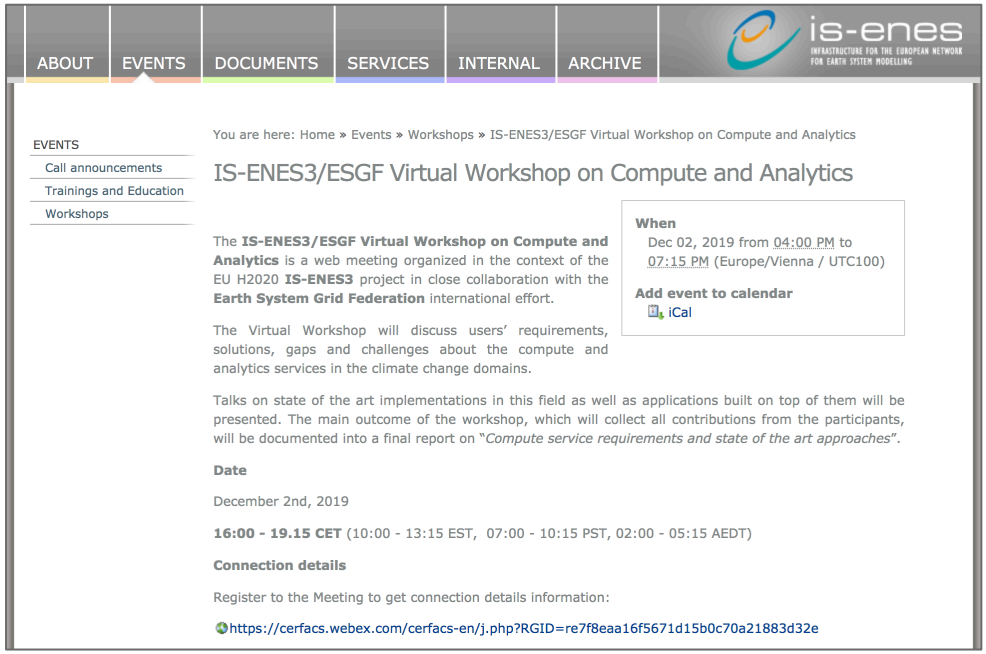
- In this respect, IS-ENES3 aims to:
  - develop Compute Service solutions (WP10)
  - gather community requirements about analysis/processing (WP3)
  - define the future compute service roadmap as part of long-term plan (WP5)
  - provide operational compute services for end-users (WP7)
    - Virtual Access
    - Trans-National Access
  - define sustainability plan (WP2)

- **Outcome: the Compute Services developed/supported in IS-ENES3 will enrich and complement the existing ENES CDI data & metadata service offering**

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

IS-ENES3/ESGF Virtual Workshop on Compute & Analytics

IS-ENES3 1st General Assembly
25-27th March 2020



ABOUT | EVENTS | DOCUMENTS | SERVICES | INTERNAL | ARCHIVE

EVENTS
- Call announcements
- Trainings and Education
- Workshops

You are here: Home » Events » Workshops » IS-ENES3/ESGF Virtual Workshop on Compute and Analytics

IS-ENES3/ESGF Virtual Workshop on Compute and Analytics

When
Dec 02, 2019 from 04:00 PM to 07:15 PM (Europe/Vienna / UTC100)

Add event to calendar
iCal

The **IS-ENES3/ESGF Virtual Workshop on Compute and Analytics** is a web meeting organized in the context of the EU H2020 **IS-ENES3** project in close collaboration with the **Earth System Grid Federation** international effort.

The Virtual Workshop will discuss users' requirements, solutions, gaps and challenges about the compute and analytics services in the climate change domains.

Talks on state of the art implementations in this field as well as applications built on top of them will be presented. The main outcome of the workshop, which will collect all contributions from the participants, will be documented into a final report on "*Compute service requirements and state of the art approaches*".

Date

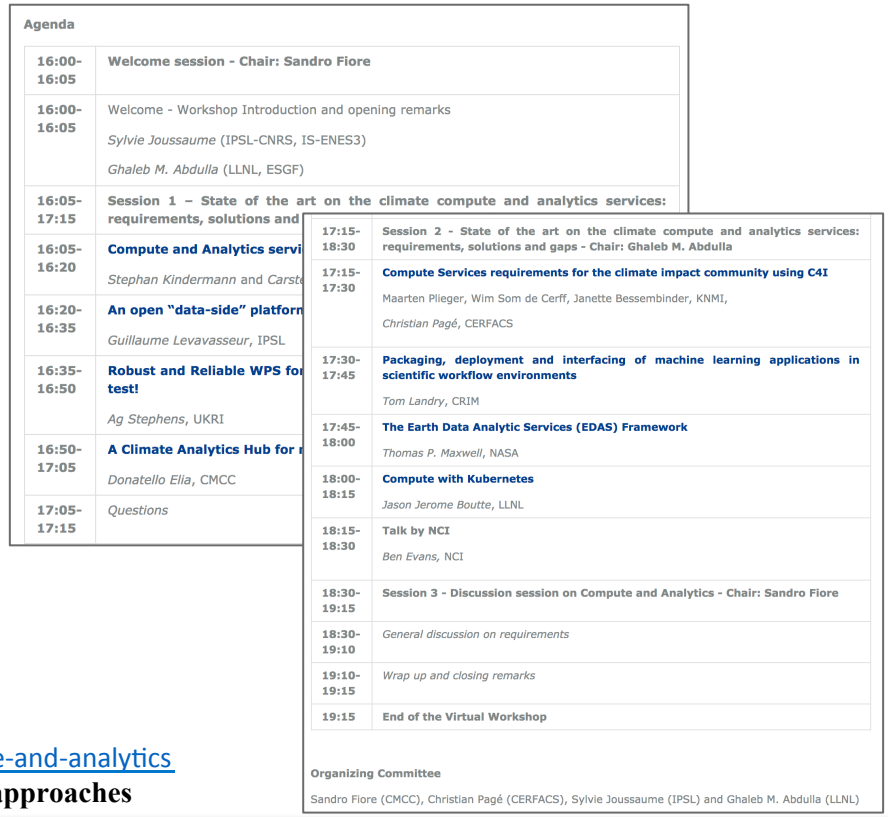
December 2nd, 2019

16:00 - 19.15 CET (10:00 - 13:15 EST, 07:00 - 10:15 PST, 02:00 - 05:15 AEDT)

Connection details

Register to the Meeting to get connection details information:

https://cerfacs.webex.com/cerfacs-en/j.php?RGID=re7f8eaa16f5671d15b0c70a21883d32e



Agenda

16:00-16:05	**Welcome session - Chair: Sandro Fiore**
16:00-16:05	Welcome - Workshop Introduction and opening remarks *Sylvie Joussaume* (IPSL-CNRS, IS-ENES3) *Ghaleb M. Abdulla* (LLNL, ESGF)
16:05-17:15	**Session 1 – State of the art on the climate compute and analytics services: requirements, solutions and**
16:05-16:20	**Compute and Analytics servi** *Stephan Kindermann* and *Carst*
16:20-16:35	**An open "data-side" platforn** *Guillaume Levavasseur*, IPSL
16:35-16:50	**Robust and Reliable WPS fo test!** *Ag Stephens*, UKRI
16:50-17:05	**A Climate Analytics Hub for** *Donatello Elia*, CMCC
17:05-17:15	*Questions*
17:15-18:30	**Session 2 - State of the art on the climate compute and analytics services: requirements, solutions and gaps - Chair: Ghaleb M. Abdulla**
17:15-17:30	**Compute Services requirements for the climate impact community using C4I** Maarten Plieger, Wim Som de Cerff, Janette Bessembinder, KNMI, *Christian Pagé*, CERFACS
17:30-17:45	**Packaging, deployment and interfacing of machine learning applications in scientific workflow environments** *Tom Landry*, CRIM
17:45-18:00	**The Earth Data Analytic Services (EDAS) Framework** *Thomas P. Maxwell*, NASA
18:00-18:15	**Compute with Kubernetes** *Jason Jerome Boutte*, LLNL
18:15-18:30	**Talk by NCI** *Ben Evans*, NCI
18:30-19:15	**Session 3 - Discussion session on Compute and Analytics - Chair: Sandro Fiore**
18:30-19:10	*General discussion on requirements*
19:10-19:15	*Wrap up and closing remarks*
19:15	**End of the Virtual Workshop**

Organizing Committee

Sandro Fiore (CMCC), Christian Pagé (CERFACS), Sylvie Joussaume (IPSL) and Ghaleb M. Abdulla (LLNL)

Date: December 2, 2019 - Around 25 participants from EU and US
Agenda of the workshop (2 sessions):
https://is.enes.org/events/workshops/is-enes3-esgf-virtual-workshop-on-compute-and-analytics
Outcome reported in D5.1: Compute service requirements and state of the art approaches

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

IS-ENES3/ESGF Virtual Workshop on Compute & Analytics

Compute Service solutions/deployments:

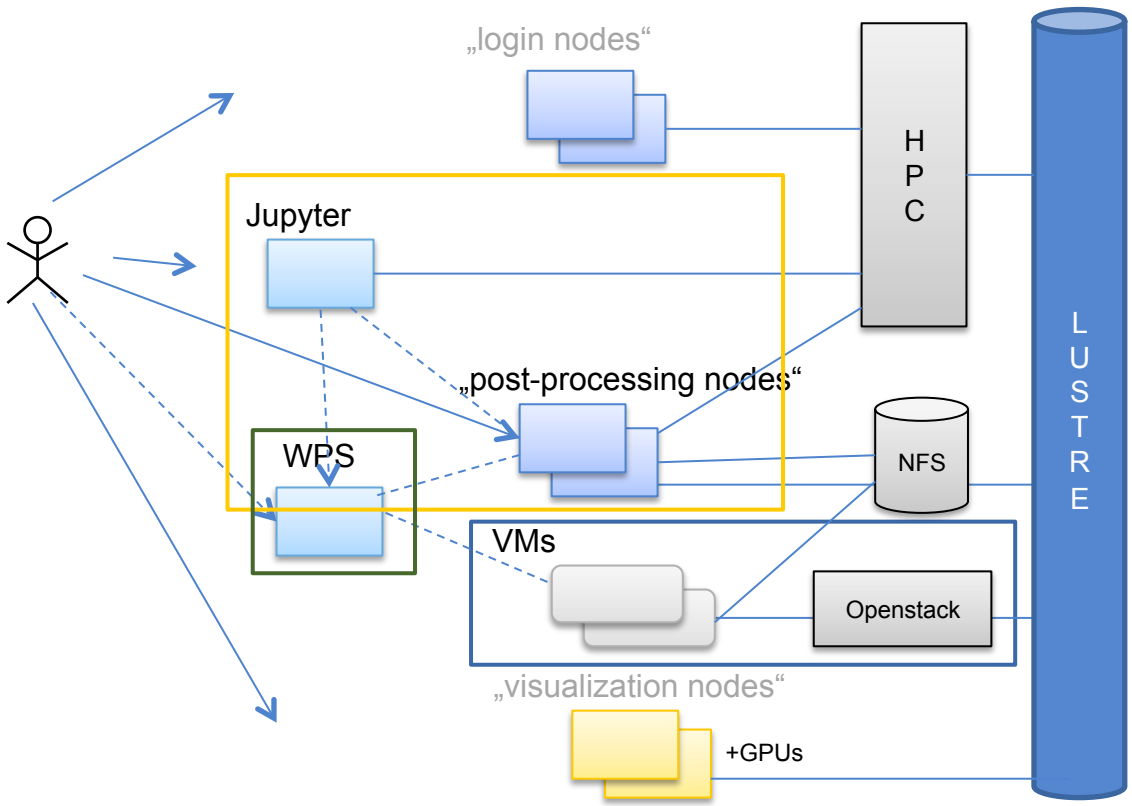
- CMCC
- DKRZ
- IPSL
- UKRI
- NASA
- LLNL

Application perspective:

- CERFACS/KNMI
- CRIM

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

Compute Service @DKRZ



Major usage scenarios

Batch processing

Interactive processing
- Jupyter hub service
- Birdhouse: used for Copernicus WPS deployment
- ECAS deployment for EOSC

Service provisioning
- ex: CMIP6 model evaluation result generation
- uses ESMValTool
- continuous processing of CMIP6 evaluation results.

Visualization

Credits: Stephan Kindermann (DKRZ)

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

Compute Service @IPSL

ESPRI = "Common Services for Research at IPSL"
ESPRI is a **mutualized** data analysis **platform** providing **optimal** access to climate observations **and** model results, together **close** to the computing facility used by IPSL community(ies).

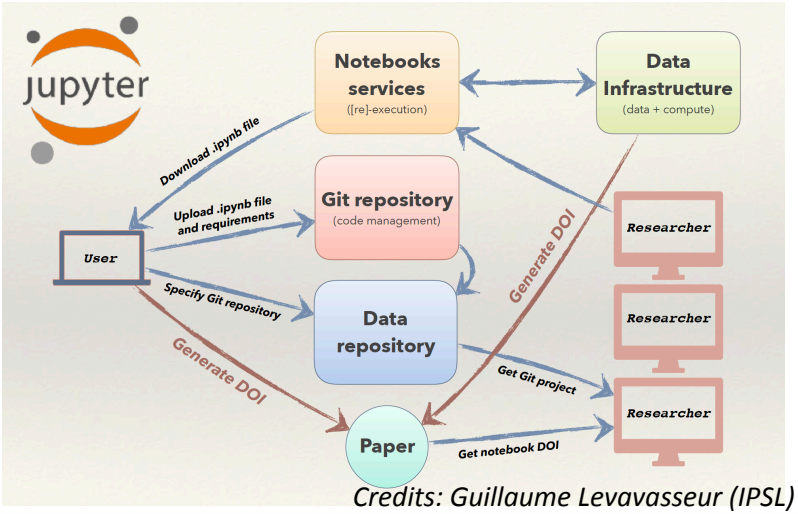
ESPRI, the "local" level:

- facilitates the **distribution**, **access** and **analysis** of international **climate data,**

CLIMERI-France, the national level, relies on **ESPRI**

Proposed solution:

- To Improve/complete our analytic environment
- Kubernetes instance soon in production on ClimServ cluster
- Jupyter Notebooks for training purposes and analysis traceability,
- ESGF data node hosting (vesg.polytechnique.ipsl.fr ?).
- WPS deployed for climate services (Copernicus)



Credits: Guillaume Levavasseur (IPSL)

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

Compute Service @UKRI

Proposed solution

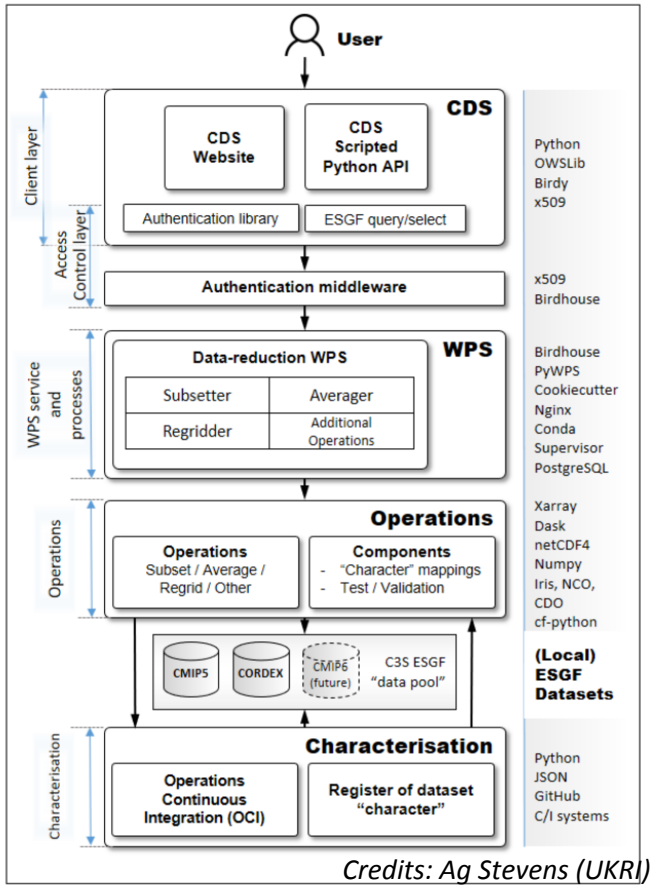
- Layered approach
- Strong focus on *robustness* rather than *functionality*
- Python codebase, build on PyWPS, Birdhouse, Xarray and any existing libraries/tools.
- Begin with basic operations:
  - Subsetting, Averaging and (some) regridding

Characterization of data

In spite of CMOR, data is heterogeneous; this can affect analysis code

For each project/activity (e.g. CMIP5, CORDEX):

- Test all operations against a large, representative sample of the available data.
- Validate the outputs in a comprehensive way.
- Most importantly, “characterise” the data sets in a public register.



Credits: Ag Stevens (UKRI)

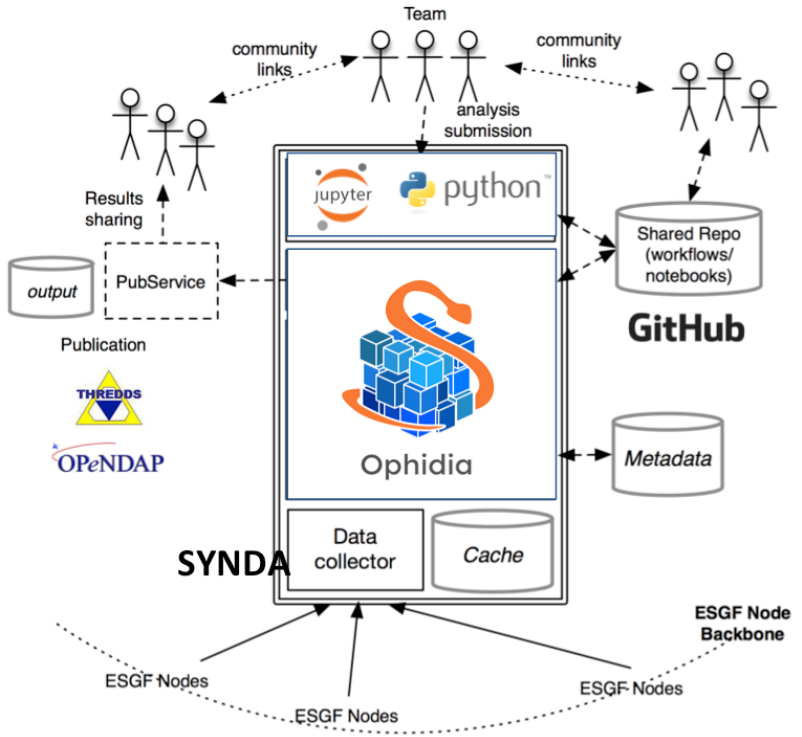
The IS-ENES3 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

Compute Service @CMCC

ECAS represents the **Analytics-Hub** solution adopted at **CMCC**. Its main components are:

- A data science environment based on JupyterHub and a set of high-level **scientific libraries** for analysis/plotting
- A **WPS** interface
- **Ophidia** as internal **analytics** engine
- The **data collector** (Synda) and to gather relevant datasets from ESGF
- A local **storage pool**

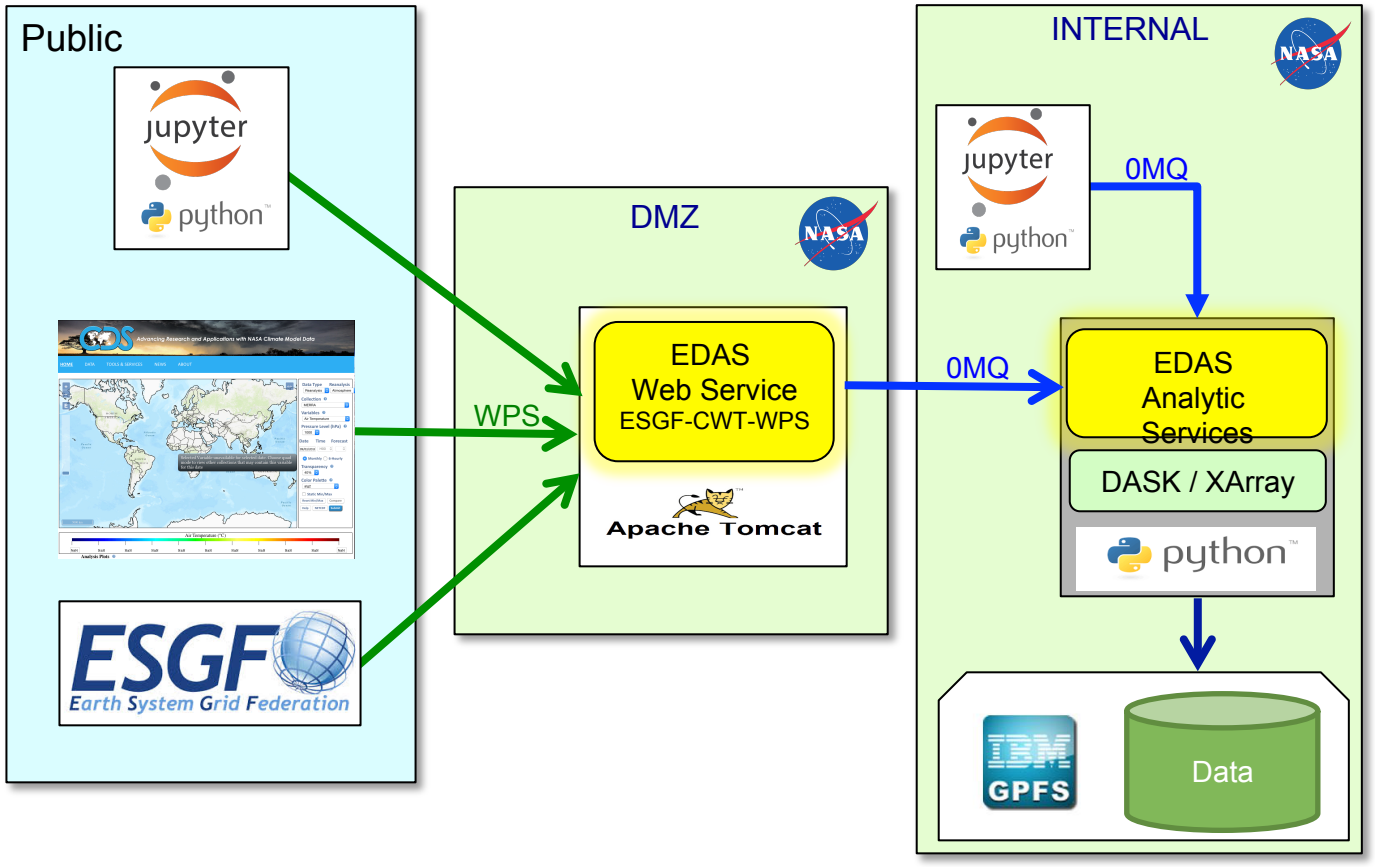
ECAS is also one of the **EOSC-hub Thematic Services** (in collaboration with DKRZ).



Credits: Donatello Elia (CMCC)

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

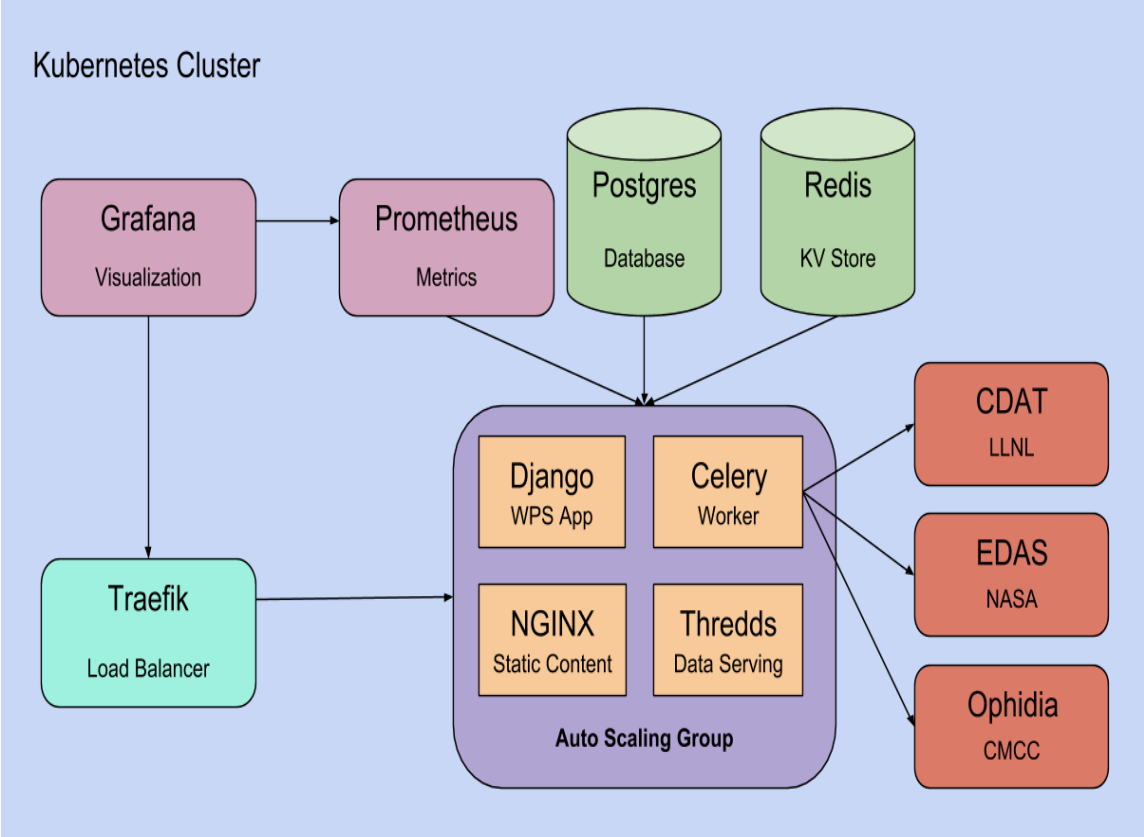
Compute Service @NASA



Credits: T. Maxwell (NASA)

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

Compute Service @LLNL



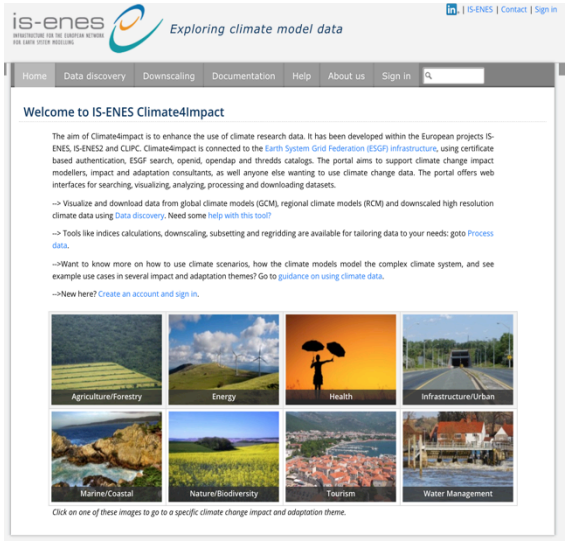
Proposed solution

- Microservices/containers approach
- Cloud-based solution running in a Kubernetes cluster
- Interface with different back-end as a result of the Compute Working Team activity
- Access via JupyterHub
- Back-end (selected set of processes: Aggregate, Subset, Max, Min, Subtract, Sum)

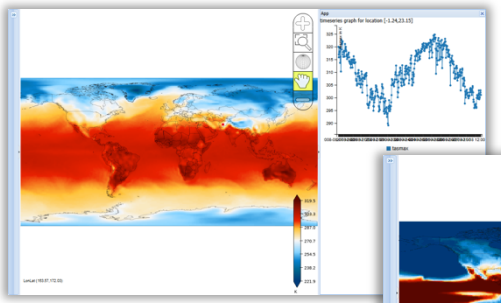
Credits: Jason Boutte (LLNL)

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

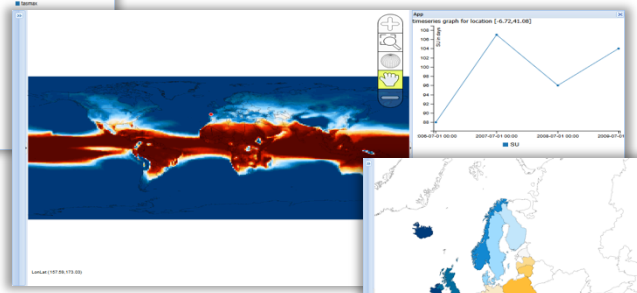
Application needs: C4I “use case”



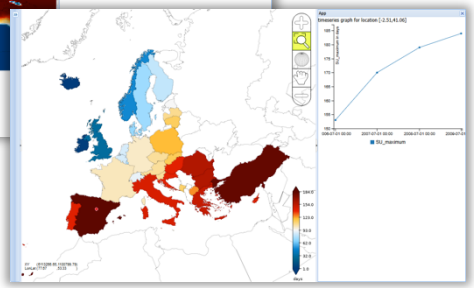
https://climate4impact.eu



Daily Maximum temperature



Summer days (T > 25)



Maximum summer days per country

Improvements and needs from compute and analytics services

- Currently C4I handles ESGF data on file level
  - *Make easier to process long sequences of data*
- We want to bring the processing to the data
  - *Calculations should run faster*

Credits: C. Page’ (CERFACS), W. Som de Cerff (KNMI) M. Plieger (KNMI) , J. Bessembinder (KNMI), B. Overbeek (KNMI)

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

Conclusions (I): Different solutions

Unlike other parts of ESGF, several solutions are under development or are supported at different sites.

Differences are driven by :

- contexts (e.g. national/institutional, Copernicus, EOSC, etc.)
- target users and requirements
- priorities

Diversity can help tackling a wider spectrum of scenarios

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

Conclusions (II): Common aspects

However there are common (relevant) aspects/directions that must be highlighted as well.

Key examples relate to:

- **Service interface** (WPS)
- **Security** (ESGF context, IdEA working group)
- **Programmatic access** interface (Python-based) and **data science software** eco-system
- **Virtualization**, **containerization**, **orchestration solutions and cloud tech.**

They can help providing a good interoperability basis for more complex scenarios!

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

Conclusions (III): End-users perspective

Scientists are cautious about using a processing service for several reasons:

- Awareness
- Transparency
- Robustness
- Trust

Not a big surprise...

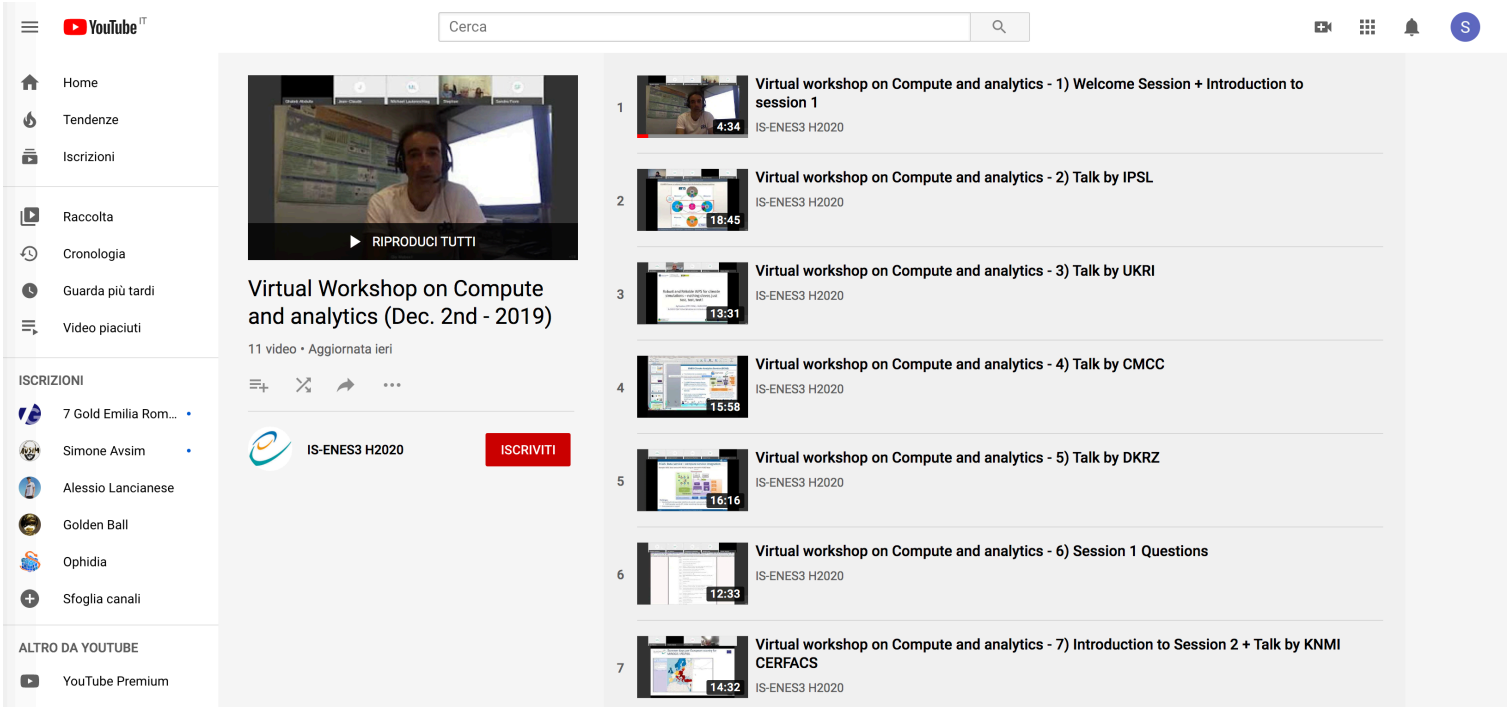
- it can take time for the community to move towards a new approach
- It should be clearly explained what each solution is extremely good at
- Specific needs require a “tailored compute service” (heterogeneous data)
- Some actions can be taken to increase usability, awareness and adoption (webinars, demos, training, access calls, etc.).

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

IS-ENES3/ESGF Virtual Workshop on Compute & Analytics

In case you missed it!

IS-ENES3 1st General Assembly
25-27th March 2020



Videos are available on the IS-ENES3 YouTube Channel: https://www.youtube.com/playlist?list=PLFvev1W5vG7N69d4mD0Aa6FgNiuisQDcl

Thanks Sophie!

The IS-ENES3 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

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

Join the community on ZENODO !