



# ECAS: a data science environment for climate change in the EGI federated infrastructure

**Fabrizio Antonio, Donatello Elia, Sandro Fiore**

On behalf of the ECAS Team

---

 [eosc-hub.eu](https://eosc-hub.eu)

 [@EOSC\\_eu](https://twitter.com/EOSC_eu)

**EOSC-hub week 2020**  
Online | 18-20 May 2020

**Impact of EOSC-hub on science communities**



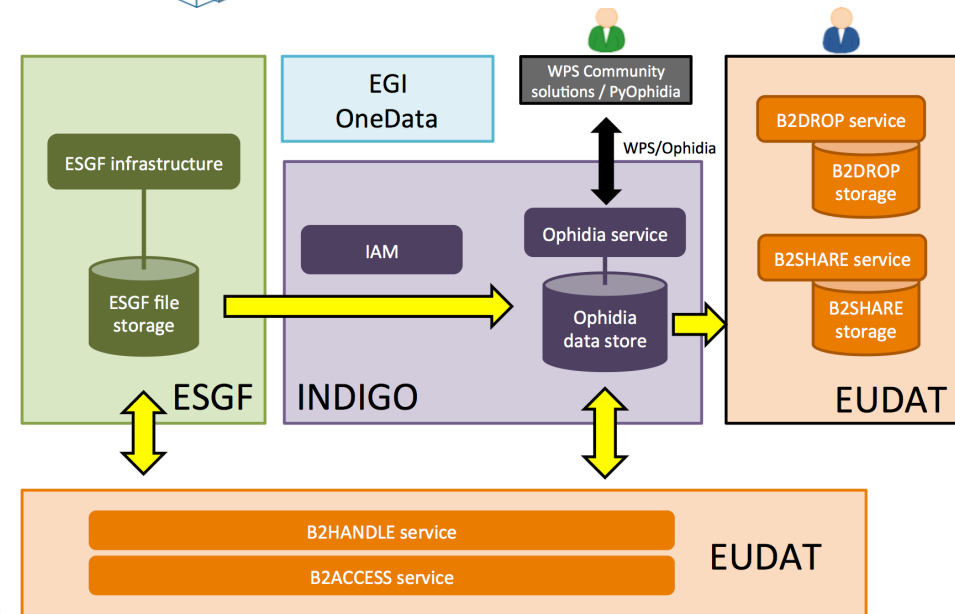
EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536.



The *ENES Climate Analytics Service (ECAS)*, proposed by CMCC & DKRZ in the *EU H2020 EOSC-Hub project*, supports **climate data analysis experiments** with a strong focus on **data intensive analysis**, **provenance management**, and **server-side approaches**

It is one of the ***EOSC-Hub Thematic Services*** as well as a ***Compute Service*** in the ***IS-ENES3 project***

ECAS consists of multiple integrated components from INDIGO-DataCloud, EUDAT, ESGF and EGI, centered around the ***Ophidia HPDA framework***



**ECASLab** provides a user-friendly environment for scientific analysis based on:

The *ECAS* integrated service

A **JupyterHub** instance providing a graphical environment for users experiments

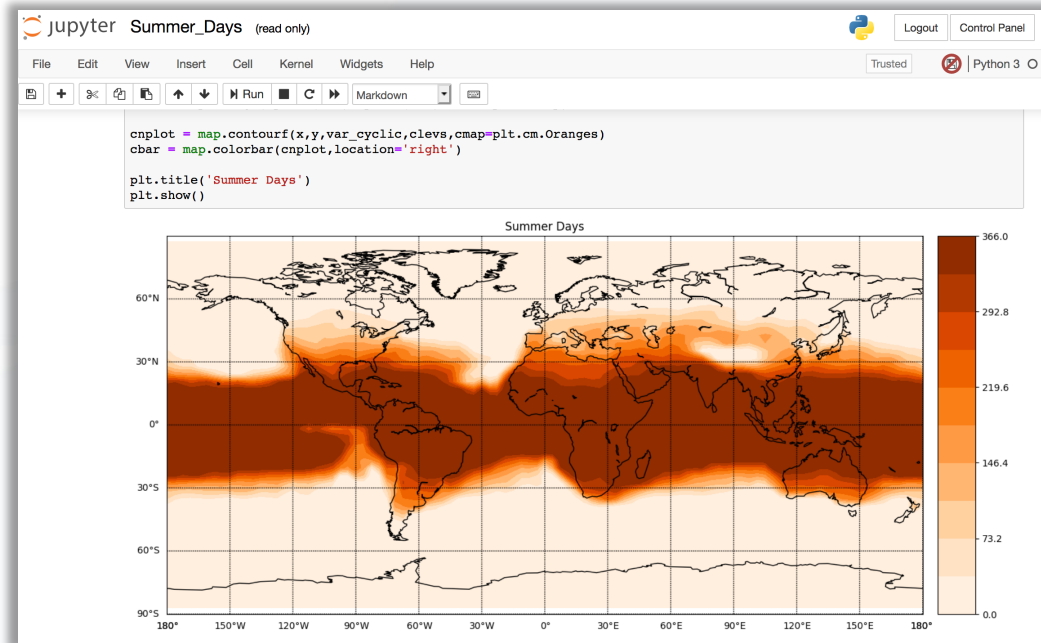
A wide set of **Python scientific modules** for data manipulation, analysis and visualization (PyOphidia, NumPy, Pandas, Dask, Matplotlib, basemap, Cartopy)

A set of ECAS usage example **notebooks** (<https://github.com/ECAS-Lab/ecas-notebooks>)

Two major instances hosted by:

✓ **CMCC** <https://ecaslab.cmcc.it>

✓ **DKRZ** <https://ecaslab.dkrz.de>



*Ophidia* (<http://ophidia.cmcc.it>) is a CMCC Foundation research project addressing data challenges for eScience<sup>1</sup>

It provides:

a *High Performance Data Analytics (HPDA)* framework for multi-dimensional scientific data joining HPC paradigms with scientific data analytics approaches

**in-memory** and **server-side** data analysis exploiting parallel computing techniques and database approaches

a **multi-dimensional, array-based**, storage model and partitioning schema for scientific data leveraging the datacube abstraction

end-to-end mechanisms to support **complex experiments and large workflows** on scientific datacubes, primarily in **climate domain**



## Ophidia



1. S. Fiore, A. D'Anca, C. Palazzo, I. T. Foster, D. N. Williams, G. Aloisio, "Ophidia: toward big data analytics for escience", *ICCS 2013*

**PyOphidia** provides a Python interface to submit commands to the Ophidia Server and to easily retrieve the results (e.g. in Jupyter Notebooks)

Two modules available:

**Client class:** supports the submissions of Ophidia commands and workflows as well as the management of sessions

**Cube class:** provides the datacube type abstraction and the methods to manipulate, process and get information from cubes objects

```
from PyOphidia import cube, client
cube.Cube.setclient(read_env=True)

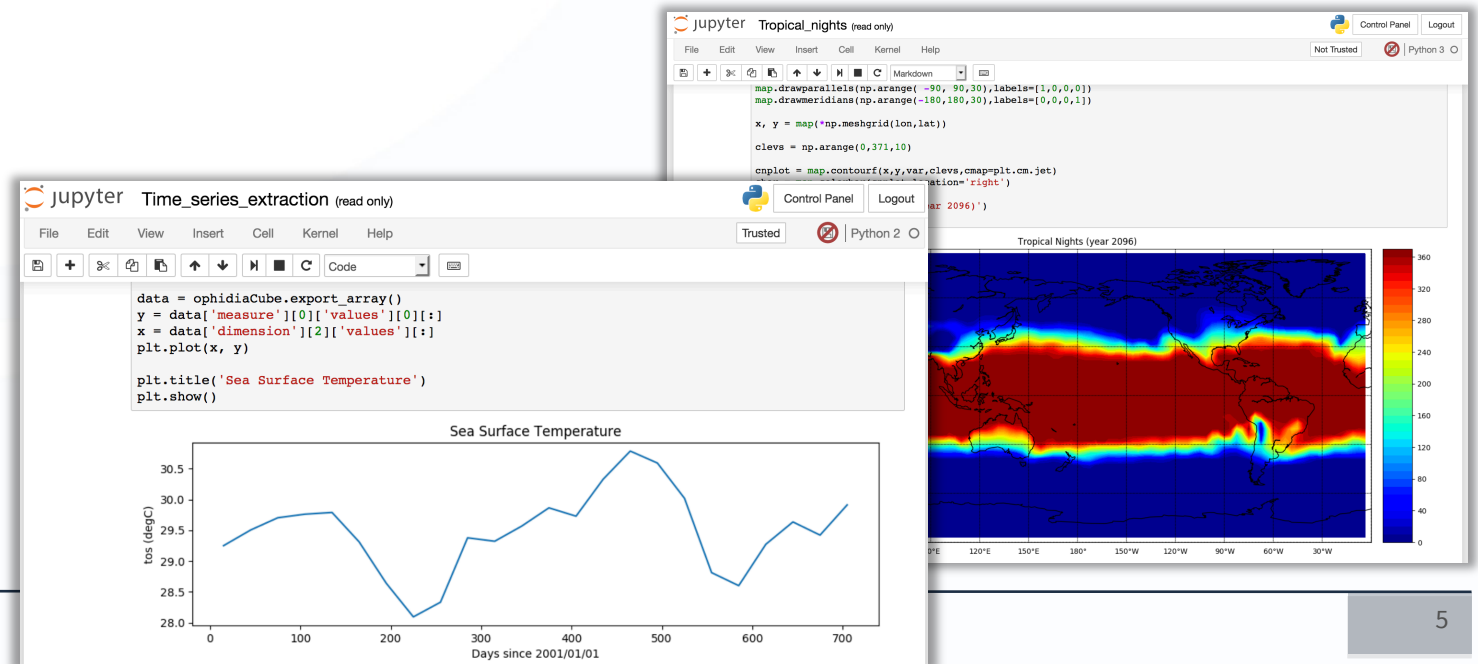
mycube =
cube.Cube.importnc(src_path='/public/data/ecas_training
/file.nc', measure='tos', imp_dim='time',
import_metadata='yes', ncores=5)
mycube2 = mycube.reduce(operation='max',ncores=5)
mycube3 = mycube2.rollup(ncores=5)
data = mycube3.export_array()

mycube3.exportnc2(output_path='/home/test',
export_metadata='yes')
```

<https://github.com/OphidiaBigData/PyOphidia>

<https://pypi.org/project/PyOphidia/>

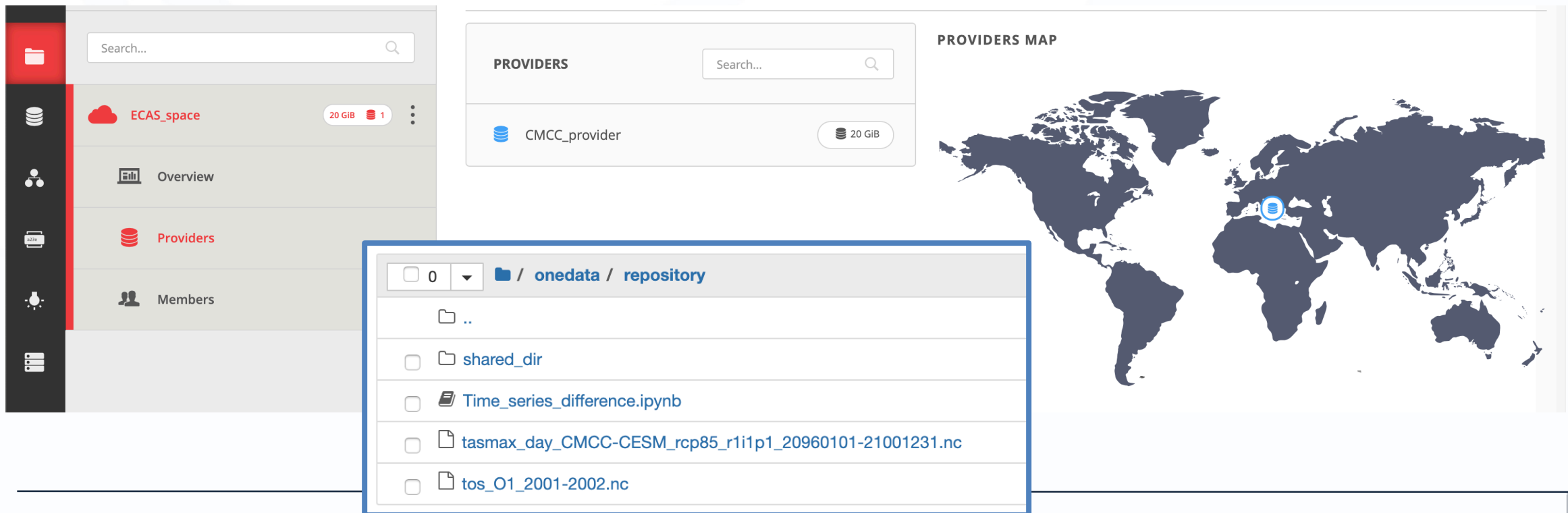
<https://anaconda.org/conda-forge/pyophidia>



*Onedata Provider deployed at CMCC SuperComputing Center to support a shared read-only repository (**ECAS\_space**)*

*Single **OneClient** instance set up to interact with the provider*

*Data folders mounted on the ECAS users home through **NFS** in read-only mode*



The screenshot displays the Onedata web interface. On the left is a navigation sidebar with icons for home, storage, providers, and members. The main content area is divided into three sections:

- ECAS\_space**: A provider card showing 20 GiB of storage and 1 file. Below it are links for Overview, Providers, and Members.
- PROVIDERS**: A searchable list of providers, currently showing **CMCC\_provider** with 20 GiB of storage.
- PROVIDERS MAP**: A world map with a blue location pin over Europe, indicating the provider's location.

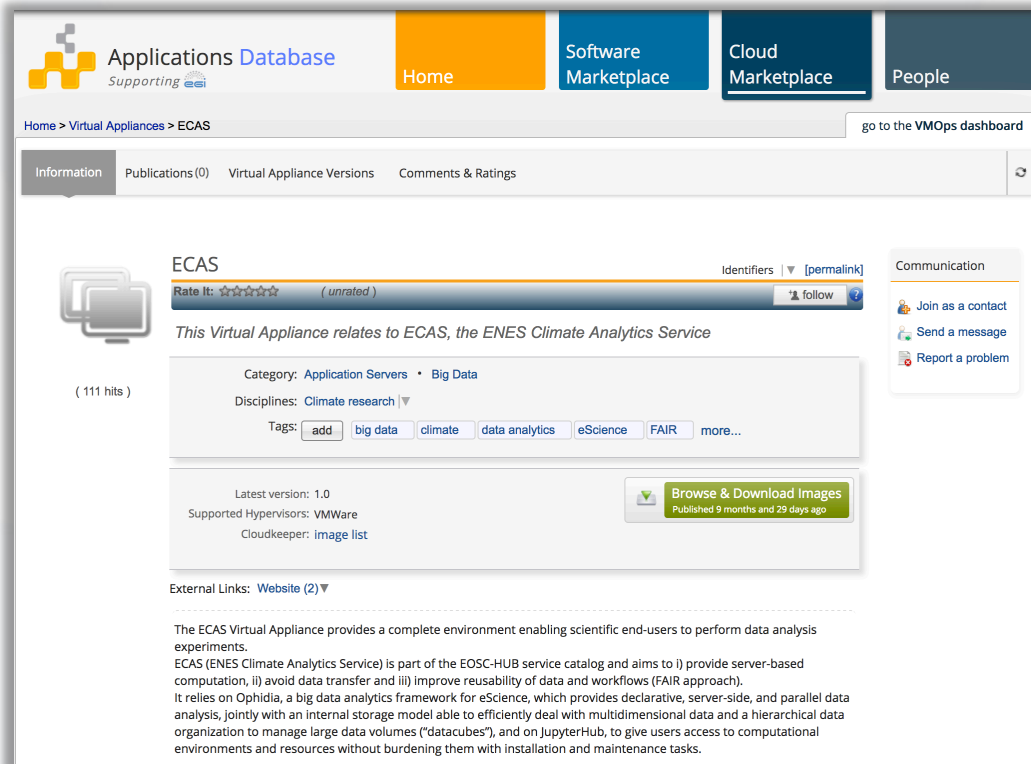
A file browser window is overlaid on the interface, showing the path `/ onedata / repository`. The file list includes:

- `..` (parent directory)
- `shared_dir` (directory)
- `Time_series_difference.ipynb` (file)
- `tasmax_day_CMCC-CESM_rcp85_r1i1p1_20960101-21001231.nc` (file)
- `tos_O1_2001-2002.nc` (file)

ECAS has been integrated into the EGI FedCloud, considering two scenarios<sup>1</sup>:

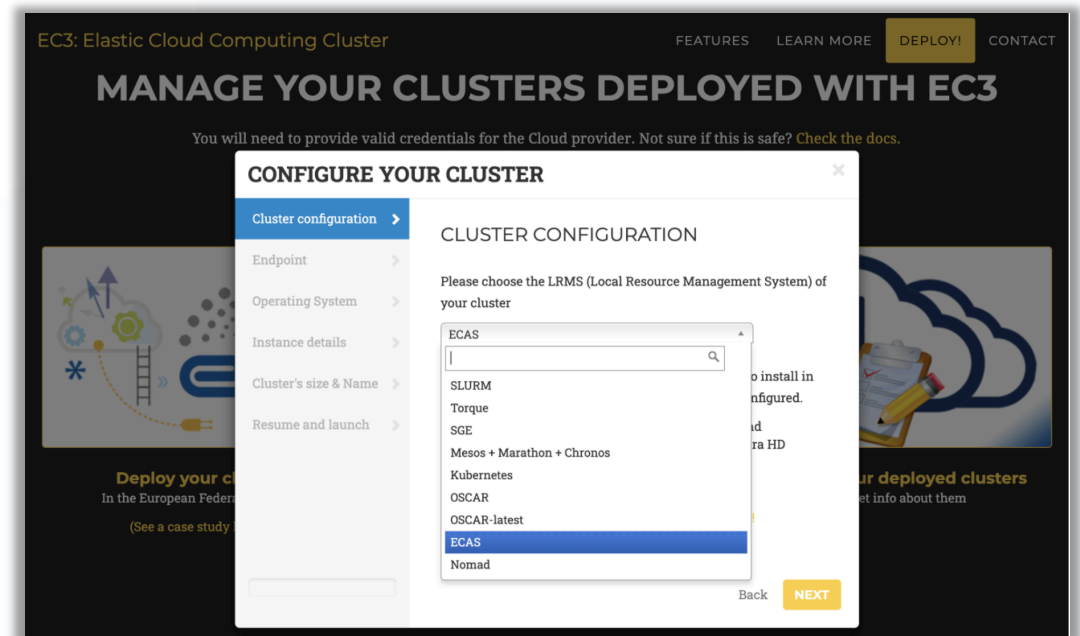
A ready-to-use ECAS single-node VMI  
available from the EGI AppDB

<https://appdb.egi.eu/store/vappliance/ecas>



The screenshot shows the EGI Applications Database interface. The top navigation bar includes 'Applications Database', 'Home', 'Software Marketplace', 'Cloud Marketplace', and 'People'. The breadcrumb trail is 'Home > Virtual Appliances > ECAS'. The main content area displays the 'ECAS' virtual appliance page, which includes a rating bar, a description ('This Virtual Appliance relates to ECAS, the ENES Climate Analytics Service'), category ('Application Servers'), disciplines ('Climate research'), and tags ('big data', 'climate', 'data analytics', 'eScience', 'FAIR'). A 'Browse & Download Images' button is visible, along with a 'Communication' sidebar with options like 'Join as a contact' and 'Send a message'.

A multi-node ECAS cluster dynamically  
provisioned on the EGI FedCloud through  
the Elastic Cloud Computing Cluster (EC3)



The screenshot shows the EC3 Elastic Cloud Computing Cluster interface. The main heading is 'MANAGE YOUR CLUSTERS DEPLOYED WITH EC3'. A modal window titled 'CONFIGURE YOUR CLUSTER' is open, displaying a 'CLUSTER CONFIGURATION' form. The 'Cluster configuration' sidebar on the left lists options: Endpoint, Operating System, Instance details, Cluster's size & Name, and Resume and launch. The main form area shows a dropdown menu for selecting the LRMS (Local Resource Management System) of the cluster, with 'ECAS' selected. Other options in the dropdown include SLURM, Torque, SGE, Mesos + Marathon + Chronos, Kubernetes, OSCAR, OSCAR-latest, and Nomad. A 'NEXT' button is visible at the bottom right of the modal.

1. Elastic deployment of ECAS on EGI: <https://www.egi.eu/about/newsletters/elastic-deployment-of-ecas-on-egi/>

*ECAS single-instance VMI uploaded to the EGI AppDB*

*VMI assigned to a set of trusted VOs*

EGI AppDB Dashboard VM Operations 100% Cloud availability

**Topology Builder** Create a new topology and deploy it to the infrastructure

Usage of vo.access.egi.eu VO resources

0% 4 CPUs available    0% 8 GB available    0% 100 GB available

Select a resource template (flavour) for your VM instance

✓ ECAS Virtual Appliance   
 ✓ vo.access.egi.eu Virtual Organization   
 ✓ INFN-CATANIA-STACK Site Endpoint   
 </> VM Template VM Image resource t

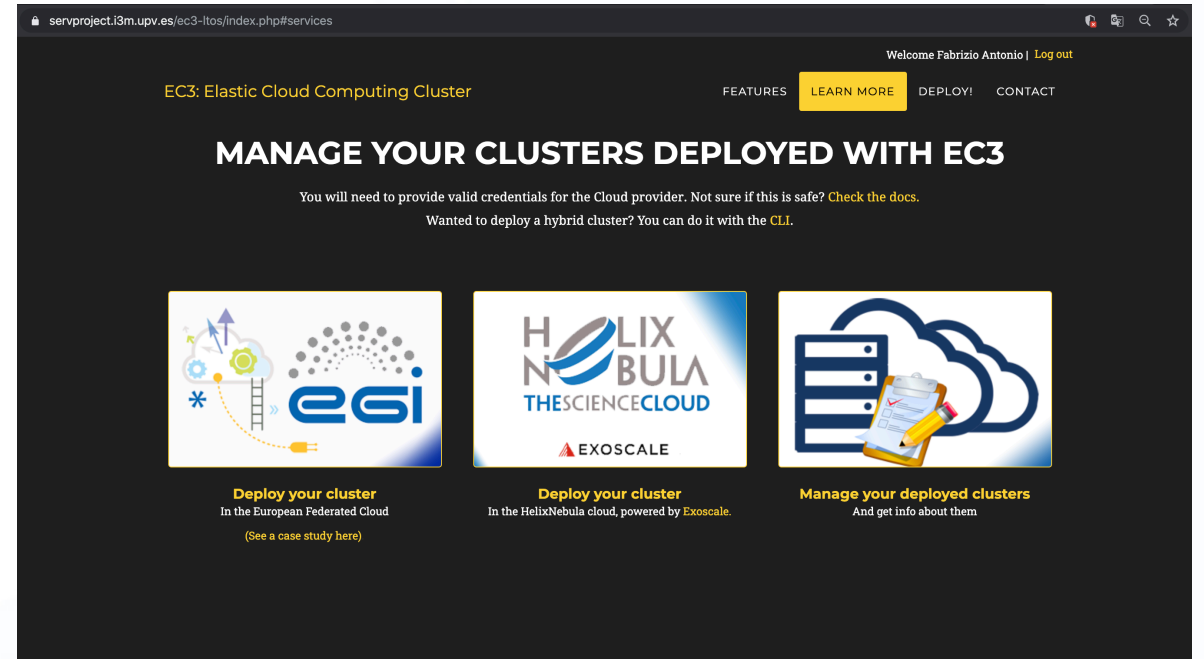
Name	CPUs		Memory	Disk	Connectivity		OS
	Logical	Physical			In	Out	
⋮ b8f64e9d-d5bc-40c3-bdc6-d55fe3625c72	1	1	64 MB	1 GB	✓	✓	linux
⋮ f1356df5-fc28-43ba-bba0-2ad0dcd01b1b	1	1	512 MB	1 GB	✓	✓	linux
⋮ 2	1	1	2048 MB	20 GB	✓	✓	linux
⋮ 8c3bded5-f640-4870-8ed0-f73d3f8eca8e	1	1	2048 MB	20 GB	✓	✓	linux
⋮ fea52bb1-6a69-4d6c-99fb-531c9cf73e67	2	2	4096 MB	40 GB	✓	✓	linux
⋮ ce1d7a0e-32cf-44c8-a8ba-6a387f516b69	4	4	8192 MB	80 GB	✓	✓	linux
⋮ 831e1465-3551-4ce3-a663-cebee7b04710	8	8	16384 MB	160 GB	✓	✓	linux



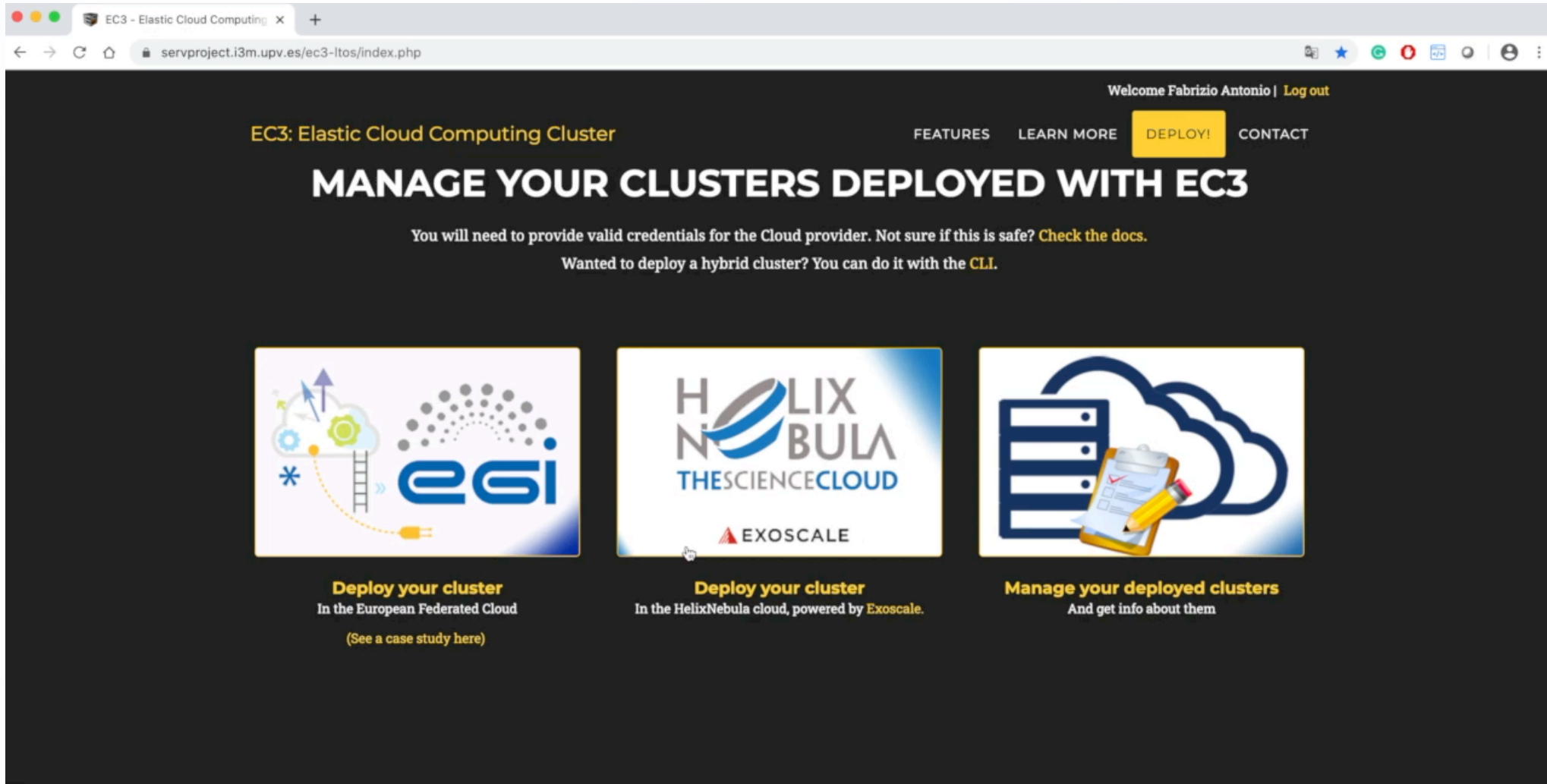
*Multi-node ECAS environment dynamically provisioned on the FedCloud through the EC3 LToS service according to the user requirements*

*RADL file for the Infrastructure Manager (IM) to define the cluster setup in terms of resources, infrastructure and software configuration, and contextualization*

**DEMO**



The screenshot shows the EC3 web interface. At the top, it says "servproject.i3m.upv.es/ec3-itos/index.php#services" and "Welcome Fabrizio Antonio | Log out". Below that, "EC3: Elastic Cloud Computing Cluster" is displayed with navigation links for "FEATURES", "LEARN MORE", "DEPLOY!", and "CONTACT". The main heading is "MANAGE YOUR CLUSTERS DEPLOYED WITH EC3". A message states: "You will need to provide valid credentials for the Cloud provider. Not sure if this is safe? Check the docs." Below this, it says "Wanted to deploy a hybrid cluster? You can do it with the CLI." There are three main sections: 1. "Deploy your cluster In the European Federated Cloud (See a case study here)" with an EGI logo. 2. "Deploy your cluster In the HelixNebula cloud, powered by Exoscale." with the HelixNebula logo. 3. "Manage your deployed clusters And get info about them" with a server and cloud icon.



The screenshot shows a web browser window with the URL `servproject.i3m.upv.es/ec3-ltos/index.php`. The page content includes a navigation bar with "Welcome Fabrizio Antonio | Log out", "FEATURES", "LEARN MORE", "DEPLOY!" (highlighted in yellow), and "CONTACT". The main heading is "MANAGE YOUR CLUSTERS DEPLOYED WITH EC3". Below this, there are three instructions: "You will need to provide valid credentials for the Cloud provider. Not sure if this is safe? Check the docs." and "Wanted to deploy a hybrid cluster? You can do it with the CLI." The main content area features three cards: 1) "Deploy your cluster In the European Federated Cloud (See a case study here)" with an "ESI" logo; 2) "Deploy your cluster In the HelixNebula cloud, powered by Exoscale." with "HELIX NEBULA THE SCIENCE CLOUD" and "EXOSCALE" logos; 3) "Manage your deployed clusters And get info about them" with a server rack and clipboard icon.

timestamp	idworkflow	name	username	ip_address	client_address	#tasks	#success_tasks	duration
2018-03-29 01:08:48	1171802	oph_list	*4E6FA15F82F3...	192.168.88.150	193.204.199.213	1	1	0.194882
2018-03-29 01:09:05	1171804	oph_randcube	*4E6FA15F82F3...	192.168.88.150	193.204.199.213	1	1	0.613289
2018-03-29 08:29:54	1171806	oph_list	*0C0EBF0FBAAF...	192.168.88.150	193.204.199.213	1	1	0.185737
2018-03-29 09:16:38	1171808	oph_list	*70410					
2018-03-29 09:36:34	1171810	oph_list	*70410					

timestamp	idtask	idworkflow	operator	#cores	success_flag	duration
2018-03-29 01:08:48	1171803	1171802	oph_list	1	1	0.153245
2018-03-29 01:09:05	1171805	1171804	oph_randcube	1	1	0.578767
2018-03-29 08:29:54	1171807	1171806	oph_list	1	1	0.152150
2018-03-29 09:16:38	1171809	1171808	oph_list	1	1	0.223756
2018-03-29 09:36:34	1171811	1171810	oph_list	1	0	0.153697

Ophidia Server logs

**Registered users**

**180**

**Number of jobs**

**1750K**

**Number of core-hours**

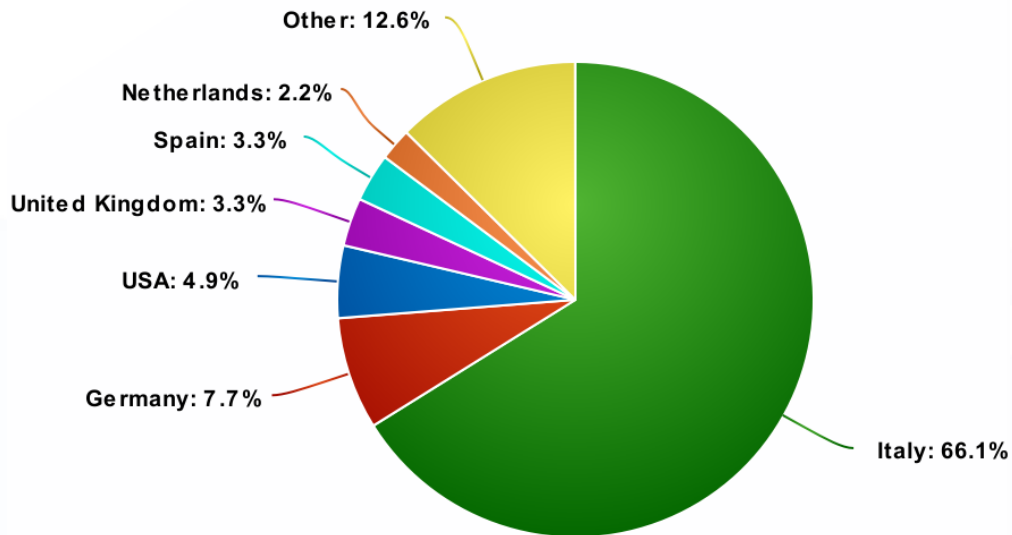
**~10K**

Other metrics about users and computing resources usage

<https://github.com/ECAS-Lab/ecas-accounting>

**Training events**

**8**



## Conclusions

**ECAS** enables scientific end-users to perform data analysis experiments on large volumes of multidimensional data by exploiting a PID-enabled, server-side, and parallel approach

The integration of **ECAS** in the **EGI cloud-based resources** allows researchers to deploy on demand a full ECAS elastic cluster on the **EGI Infrastructure** thanks to the **EC3 platform**.

## Next steps

Integration of **EGI Check-in** in ECAS

Stronger integration with **Onedata**

CMCC ECASLab instance: <https://ecaslab.cmcc.it/>

DKRZ ECASLab instance: <https://ecaslab.dkrz.de/>

ECASLab repository: <https://github.com/ECAS-Lab>

Ophidia Website: <http://ophidia.cmcc.it>

Ophidia Doc: <http://ophidia.cmcc.it/documentation>

PyOphidia repository: <https://github.com/OphidiaBigData/PyOphidia>



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



These activities are supported in part by EOC-Hub and IS-ENES3 projects:

*EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536*

*IS-ENES3 is a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824084*

# Thanks

