

PHENET-EMPHASIS Data Management Training

Cloud infrastructure: deploy your own analysis workflow

Dr Yin Chen

EGI Foundation

04-06 December 2024, Paris

Get Slide at : <https://go.egi.eu/5EHdM>

Outline

Topic 1: What is Cloud computing?

Topic 2: What is EGI e-Infrastructure and how to access?

- *Hand-on:* get your EGI check-in account and join a VO

Topic 3: How to run your analysis workflow in EGI Notebook?

- *Hand-on:* get access to EGI Notebook

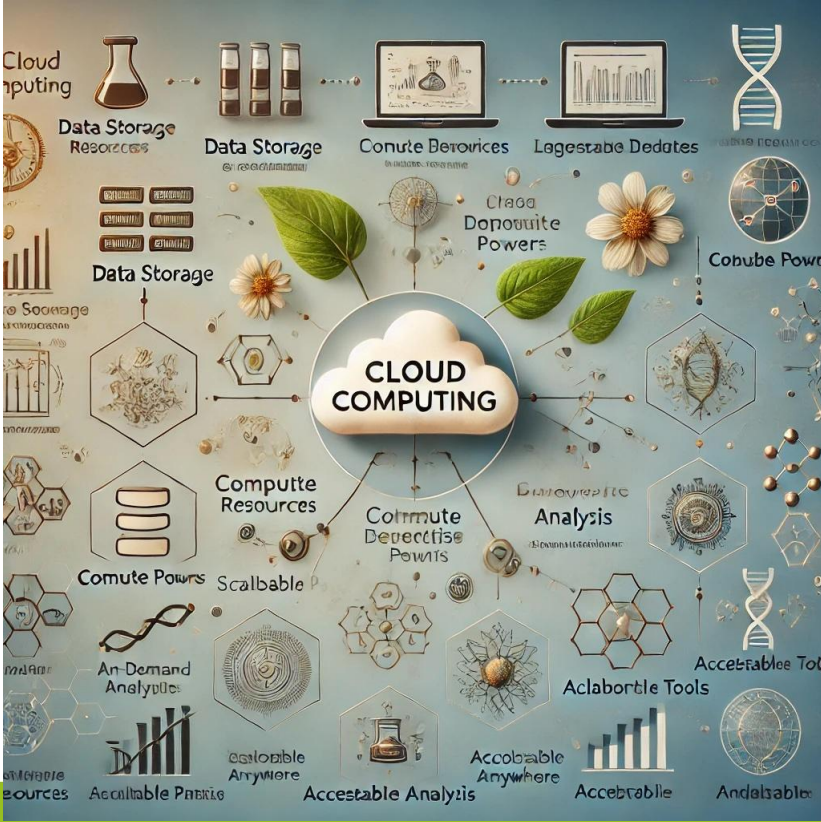
What is Cloud Computing?

Cloud computing provides **on-demand access to computing resources over the internet.**

- **No Need for Local Hardware** – Resources are hosted in remote data centers, eliminating the need for personal hardware or servers.
- **Three Main Services:**
 - **Compute Power** – Run applications, process data, or run simulations on remote servers.
 - **Storage** – Store large datasets securely and access them from anywhere.
 - **Software** – Access applications and tools through the internet without installation.
- **Accessible from Anywhere**
Just need an internet connection, no need to be on campus or in a specific lab.
- **Scalable and Flexible**
Use as much or as little as you need and scale up or down based on project demands.



The shift from traditional on-site infrastructure to Cloud computing has transformed the research landscape.



Outline

Topic 1: What is Cloud computing?

Topic 2: What is EGI e-Infrastructure and how to access?

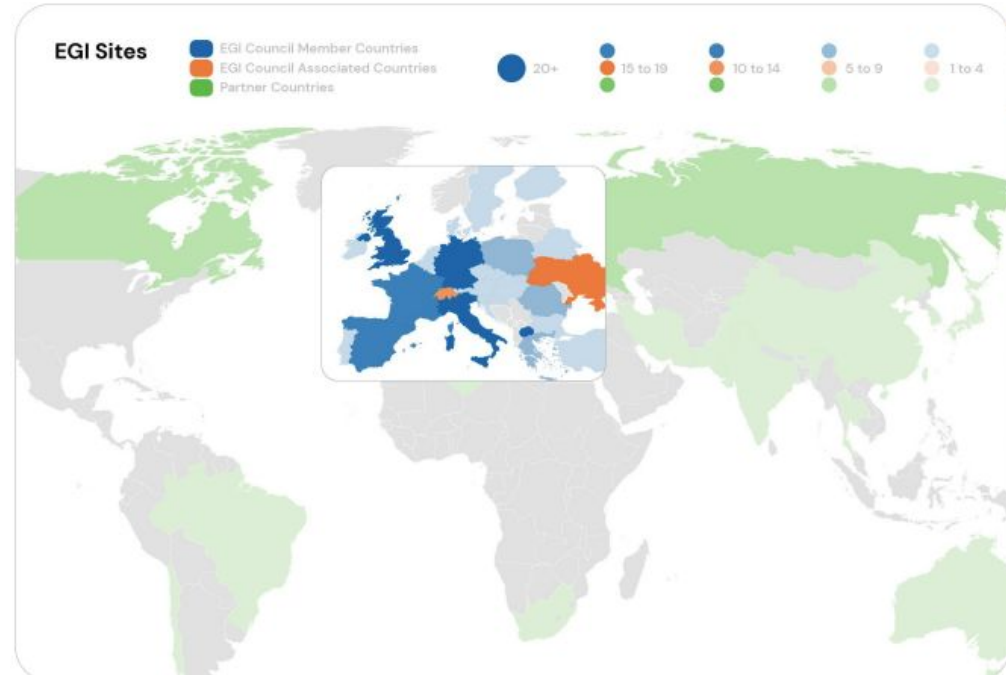
- *Hand-on:* get your EGI check-in account and join a VO

Topic 3: How to run your analysis workflow in EGI Notebook?

- *Hand-on:* get access to EGI Notebook

What is EGI e-Infrastructure?

- A **federated network of computational resources** spanning Europe and beyond
- **Established in 2002** to support the Large Hadron Collider (LHC) experiments.
- Transitioned from three EC-funded projects to form **EGI.eu**, a legal entity headquartered in Amsterdam.
- The **largest e-Infrastructure of its kind globally**, ensuring Europe remains at the forefront of innovation.
- Contributed to **two Nobel Prize-winning projects**: LHC (2013) and VIRGO (2017).
- Plays a pivotal role in the development of the **European Open Science Cloud (EOSC)**.





EGI in Numbers (2023)



2022

2023

+100,000 users in 2024!

Who is it for? (EGI community)



Physical Sciences & Engineering

Landmarks
CTA, ELI ERIC,
HL-LHC
SKAO
European XFEL

Projects
KM3NeT 2.0



Data Computing & Digital RIs

Projects
SoBigData++
EBRAINS
SLICES



Environment

Landmarks
ACTRIS ERIC
EPOS ERIC,
Euro-Argo ERIC
IAGOS
ICOS ERIC

Projects
DANUBIUS-RI
DISSCo
eLTER RI



Social & Cultural Innovation

Landmarks
CLARIN ERIC
DARIAH
CESSDA ERIC

Projects
E-RIHS
OPERAS



Health & Food

Landmarks
ELIXIR
INSTRUCT ERIC
BBMRI
EU-OPENSREEN ERIC

Projects
EMPHASIS
METROFOOD-RI

ESFRI research infrastructures supported by EGI

-  New RIs engaged in 2023
-  EGI Federation member

Key Numbers

95,000

Total number of users

+10.200

New users in 2023

Top 5 Cloud Communities

WeNMR	NBIS	Biomed	BioISI	ENVRI
41K	21K	1.5K	1K	967

By number of registered users

Top HTC Communities

Atlas, CMS, ALICE, LHCb,
Belle II, Virgo

Essential Partners and the Largest Adopters

Research infrastructures (RI) and research communities

13
new scientific communities

23
RIs from the ESFRI roadmap

49
RIs of pan-European scope using our services

1
new RIs engaged in 2023

Benefits for Researchers:

- **Scalable Infrastructure** – Handle complex or data-intensive tasks with ease by scaling resources up or down as needed.
- **Collaborative Platform** – Designed for cross-institution collaboration, allowing teams to work together effectively.
- **Access to Specialized Tools** – Pre-configured environments for scientific software and tools, making it easier to get started.

How to access EGI e-Infrastructure?

Two concepts

- Virtual Organisation (VO)
- Authentication & Authorization Infrastructure (AAI)

What is a Virtual Organization (VO)?

Imagine you and a group of researchers from different countries are working on the same scientific project, like studying climate change or biodiversity. Instead of each person setting up their own computers and storage, the VO helps the whole group use shared resources efficiently.

A **Virtual Organization (VO)** is a dynamic group of individuals or institutions that collaborate by sharing resources, data, and tools within a distributed computing environment to achieve a common goal securely and efficiently.

- *A place to share resources (computing tools, storage space, and data)*
- *Collaboration across borders*
- *Secure access*



What is Authentication & Authorization Infrastructure (AAI)?

Authentication is the process of verifying **who you are** before you can access a system, service, or resource. It ensures that only the right people can access the right resources.

*Imagine you are in a research facility: **Authentication** gets you into the building by proving who you are (e.g., showing your ID card).*

Authorization is the process of determining **what you are allowed to do** after your identity has been verified through authentication. It ensures that users only access the resources and perform actions they are permitted to.

*Imagine you are in a research facility: **Authorization** determines where you can go and what you can do inside (e.g., access certain labs, use specific equipment).*



What is the EGI Authentication & Authorization Service?

The **EGI Check-in** service is a **single sign-on system** that allows researchers to access multiple tools, resources, and services with **just one account**. It simplifies access management while ensuring secure collaboration across different organizations and platforms.

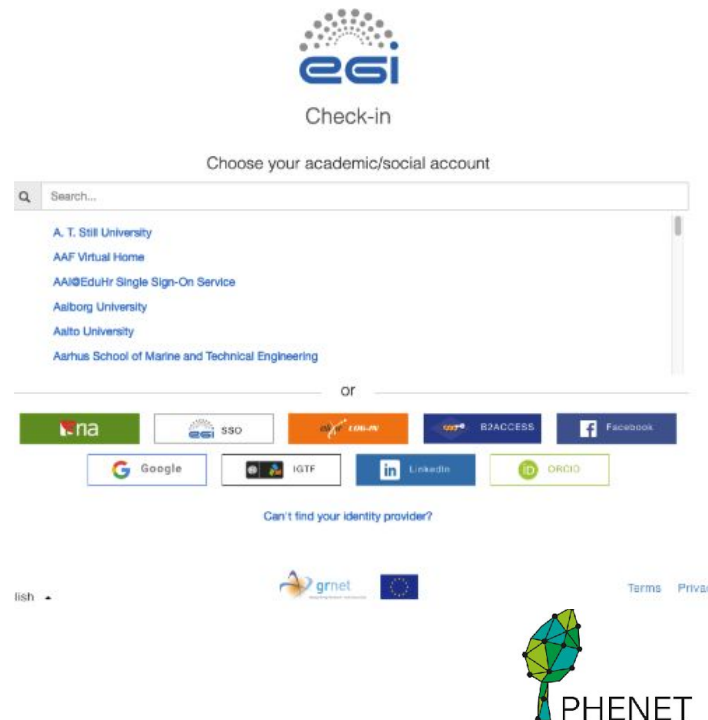
*Think of it like a **master key** for research tools:*

- *Instead of carrying different keys (logins) for each door (service), you use one key (your Check-in account) to open them all.*



Hand-on: Task 1: Get your first EGI Check-In account

- Go <https://aai.egi.eu/signup>
- Enter your login credentials to authenticate yourself with your Home Organisation
- After successful authentication, you may be prompted by your Home Organisation to consent to the release of personal information to the **EGI AAI Service Provider Proxy**.
- **User guide:**
<https://docs.egi.eu/users/aai/check-in/signup/>



Task 2: Join a VO

- Go: <https://aai.egi.eu/registry/>
- Login with your Check-in account.
- Expand the **People** drop down menu and click **Enrol**



- Join a VO: **vo.access.egi.eu**
- Click the **Begin** link of the Enrollment flow
- Join EMPHASIS VO: search for emphasis
- **User Guide:** <https://docs.egi.eu/users/aai/check-in/joining-virtual-organisation/>^{ET}



EGI VO for EMPHASIS & PHENET Community

● EGI Resources

- Provider: CESNET
- Cloud Compute: 108 vCPU cores, 402 GB of RAM, 1GPU with 4GB RAM
- Online Storage: 10.6TB

● EGI Services

- Check-In
- DataHub

● VO: vo.emphasis.eu

● Duration: 01/01/2022-30/06/2025

● SLA/OLA:

<https://documents.egi.eu/secure/ShowDocument?docid=3576>



SLA



OLA



This template (CC BY 4.0) is a



SE

Customer
Service Provider
First day of service
Last day of service
Status
Dissemination
Agreement identifier
Agreement identifier
Template version

The present Service Level Agreement ("the Agreement") is made between EGI Foundation (the Service Provider) and EMPHASIS/vo.emphasisproject.eu (the Customer) to define the provision and support of the provided services as described hereafter. Representatives and contact information are defined in Section 6.

Plant phenotyping refers to a quantitative description of the plant's anatomical, ontogenetical, physiological and biochemical properties. The PHS information system network was brought into the EMPHASIS ESFRRI from the PHENOME French project to ease data management and analysis within plant phenotyping communities. The system integrates various open source solutions, such as PostgreSQL, MongoDB, RDF4j databases, Apache HTTP, Apache Tomcat, iRODS. During the workshop we'd like to analyse and understand how the platform could be ported and hosted on EGI cloud resources as well as extended with a distributed file system and data archive.

The Customer is a consortium represented by the French National Institute for Agricultural Research (INRA).

In total, the Component Providers supporting this Agreement with the Customer offer:

Providers	Cloud Compute	Online Storage	DataHub	Check-in
IN2P3-IRIS (FR)	20 vCPU cores, 50 GB RAM, floating IPs available	1 TB		No. 1 COU group ¹ in COManage for the vo.emphasis.eu VO
CESNET-MCC (CZ)	88 vCPU cores, 1 GPU card with 4GB of RAM, 352 GB RAM, floating IPs available	11 TB		
Totals	108 vCPU cores, 402 GB of RAM, 1 GPU card with 4 GB of RAM, Floating IPs available	12 TB	10 TB available under the space: EMPHASIS	No. 1 COU group in COManage for the vo.emphasis.eu VO

The Component Providers are delivering a part of the Service(s) and are listed in Section 1.

This Agreement is valid from 01/01/2022 to 30/06/2023.

The Agreement was discussed and approved by the Customer and the Service Provider on 20/07/2022.

¹ <https://www.inrae.fr/en>

² https://vo.egi.eu/registry/vo_options/start/coef-65





Performance monitoring of the EMPHASIS VO

GOCDDB 5.10.2

Services
All Services in GOCDDB

Filter (clear)

Service Type: org.openstack.nova | NGI: (all)

Search for text in Hostname or Service Description: _____

Production Service: (all) | Monitored Service: (all) | Site Certification: (all)

Service Scopes: Service Scopes | Scope match: all (selected tags are AND'd)

Service Extension Name: (none) | Extension Value: _____

Include Closed Sites: Filter Services

40 Services (Showing 1 - 50)

Hostname	Service Type	Production	Monitored	Host Site	Scope(s)
api.cloud.ifta.es	org.openstack.nova	✓	✓	IFCA-LCG2	EGI, FedCloud, cms, _____

ARGO

Home | Contact | egi - Critical

Availabilities / Reliabilities

Availability/Reliability Table | Availability/Reliability Charts

Filter: **RedCloud**

Copy | Excel | CSV | PDF

Search: _____ Show entries

Month	2022-08		2022-09		2022-10		2022-11		2022-12	
	Av	Re	Av	Re	Av	Re	Av	Re	Av	Re
IOBT	100	100	100	100	99.97	99.97	99.99	99.99	100	100
BIF	99.93	99.93	99.99	99.99	100	100	99.99	99.99	100	100
CEIGA	100	100	100	100	100	100	100	100	100	100
CEIGA-CLOUD	99.93	99.93	99.99	99.99	99.99	99.99	99.97	99.97	99.99	99.99
CESNET-MCC	100	100	100	100	99.94	99.94	100	100	100	100
CETA-GRID	99.93	99.93	99.99	99.99	99.99	99.99	100	100	99.99	99.99
CLOUDIFM	99.93	99.93	99.99	99.99	100	100	100	100	100	100
CSTCLOUD-EGI	99.97	99.97	99.99	99.99	99.99	99.99	100	100	100	100

Performance of EMPHASIS VO monitored by EGI ARGO:

https://argo.egi.eu/egi/report-ar-group-details/SLA/SERVICEGROUPS/EGI_EMPHASIS_SLA/details

Services Performance Report | shows compliance with established SLA service targets

Audience: Isabelle Alic, David Rousseau
Report author: EGI SLA sla@mailman.egi.eu
Service: Cloud compute
Period: 2022-07 / 2022-12
Date of report: 18/01/2023
Date of next report: 2023-07
Documentation: <https://documentation.egi.eu/display/EGISLM/Service+Level+Targets+->Availability+Reliability>
Related agreements: <https://documents.egi.eu/document/3576>

Legend:
Underperforming
On Target

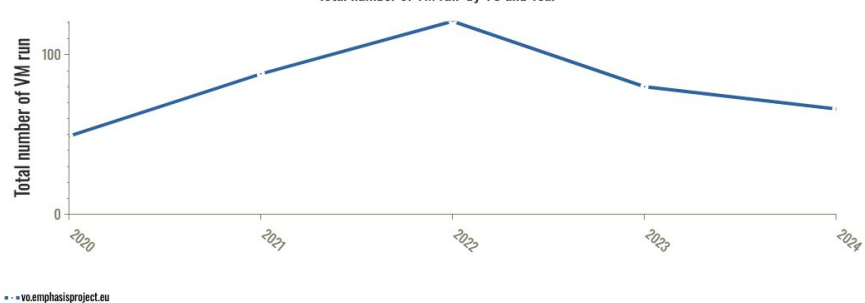
CESNET-MCC	Cloud Compute	
	Availability	Reliability
targets	95%	95%
previous reporting period	2022-04	98.16%
	2022-05	98.95%
	2022-06	100.00%
current reporting period	2022-07	95.61%
	2022-08	100.00%
	2022-09	100.00%
	2022-10	99.44%
	2022-11	100.00%
2022-12	100.00%	

CESNET-MCC	Cloud Compute	
	Availability	Reliability
targets	95%	95%
previous reporting period		
current reporting period	2022-01	99.79%
	2022-02	99.50%
	2022-03	99.38%
	2022-04	98.16%
	2022-05	98.95%
2022-06	100.00%	

IN2P3-IRES	Cloud Compute	
	Availability	Reliability
targets	95%	95%
2022-04	99.22%	99.22%

CYFRONET-CLOUD	Cloud Compute	
	Availability	Reliability
targets	95%	95%

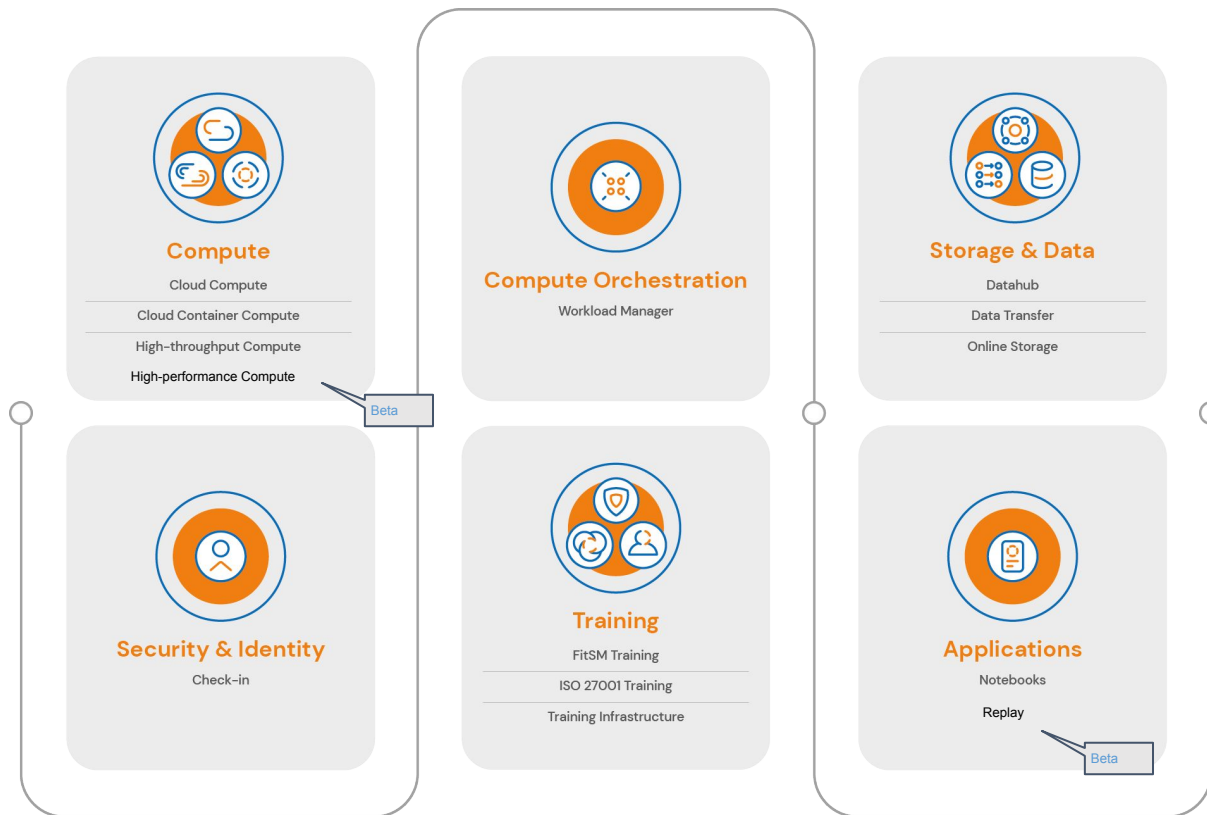
Total number of VM run by VO and Year



Usage of EMPHASIS VO stat monitored by the EGI Accounting Portal:

https://accounting.egi.eu/cloud/vm_num/VO/Year/2020/1/2024/11/custom-vo.emphasisproject.eu/only/nfrajobs/

EGI External Services (for Research)



Service catalogue:
<https://www.egi.eu/services/>

User documentation:
<https://docs.egi.eu>

Get Slide at : <https://go.egi.eu/5EHdM>





EGI Infrastructure Manager service for VM deployment

Go: <https://im.egi.eu/>

IM Dashboard Infrastructures Advanced External Links Enol Fernandez

Search...

Start a virtual machine with extra HD Launch a Kubernetes Virtual Launch an OSCAR Virtual

IM Dashboard Infrastructures Advanced External Links Enol Fernandez

SLURM virtual cluster



Launch a Storm Virtual Cluster



Launch a Kubernetes V

Description: TOSCA template for Dask, JupyterHub and Elasticity

Infrastructure Name:

HW Data **Kubernetes Data**

Number of WN in the cluster:

Number of CPUs for the front-end:

Amount of Memory for the front-end:

Flavor name of the front-end node:

Number of CPUs for the WNs:

My Infrastructures

Refresh + New deployment

Show 10 entries Search:

Name	Infrastructure uuid	Cloud Type	Cloud Info	Status	VMs	Actions
AiiDAlab	0b435d1a-7a32-11ed-aba7-4653ee2e5e57		Site: CESNET-MCC VO: vo.max-centre.eu	configured	0 1 2	Outputs
ops-console	9b585d1c-512a-11ed-a55f-32265dde0938		Site: IN2P3-IRES VO: vo.access.egi.eu	configured	0	+
jupyter	32e43eba-0e6b-11ed-9e84-22d37d76632e		Site: IISAS-FedCloud VO: vo.access.egi.eu	configured	0	+
managed-cluster	faff27f0-0da0-11ed-8d75-22d37d76632e		Site: TR-FC1-ULAKBIM VO: vo.access.egi.eu	configured	0 1	+
ranchera	9536ccb8-0d74-11ed-973f-f2b1dd77b9eb		Site: TR-FC1-ULAKBIM VO: vo.access.egi.eu	configured	0 1	+
jupysll	9a6c9c7e-fd3d-11ec-b04f-26934d900b86		Site: IISAS-FedCloud VO: vo.access.egi.eu	configured	0	Outputs
jup	cc638e44-f951-11ec-bb97-a2c3f9a44aae		Site: GSI-LCG2 VO: vo.access.egi.eu	configured	0	Outputs

- + Add nodes
- Remove nodes
- 🔍 Show template
- 📄 Log
- ⏹ Stop
- 🗑 Delete
- 🔄 Reconfigure
- 👤 Change Owner

Outline

Topic 1: What is Cloud computing?

Topic 2: What is EGI e-Infrastructure and how to access?

- *Hand-on:* get your EGI check-in account and join a VO

Topic 3: How to run your analysis workflow in EGI Notebook?

- *Hand-on:* get access to EGI Notebook

What is EGI Notebooks?



JupyterHub hosted in the EGI Cloud

- Offers Jupyter notebooks 'as Service'
- One-click solution: login and start using

Main Features:

- Easy access: Login with the EGI AAI Check-In service
- Persistent storage for notebooks
- Use EGI computing and storage resources from your notebooks



Notebooks

Notebooks is an environment based on Jupyter and the EGI cloud service that offers a browser-based, scalable tool for interactive data analysis. The Notebooks environment provides users with notebooks where they can combine text, mathematics, computations and rich media output.

The notebooks are limited to 2 CPU, 4GB RAM and 20GB of persistent storage per user. Access requires a valid EGI account and enrolling to the vo.notebooks.egi.eu VO.

[Start your notebooks!](#)

User communities/advanced users can have their customised EGI Notebooks service instance. EGI offers consultancy and support, as well as can operate the setup. Order a [community notebooks instance](#) via the Marketplace.



Notebooks is a service provided by CESNET, co-funded by EGI-ACE.

[Privacy policy](#) | [Terms of use](#)

How to access EGI Notebook?

- Go EGI Notebook: <https://notebooks.egi.eu/>
- Login with your check-in account
- Some examples:
https://notebooks.egi.eu/hub/user-redirect/lab/tree/PHENET_Training_examples





PHENET

PHENOTYPING & ENVIROTYPING
SOLUTIONS FOR AGROECOLOGY

 <https://www.phenet.eu/en>

 phenet_project

 PHENET