PHENET
PHENOTYPING & ENVIROTYPING
SOLUTIONS FOR AGROECOLOGY

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

PHENET

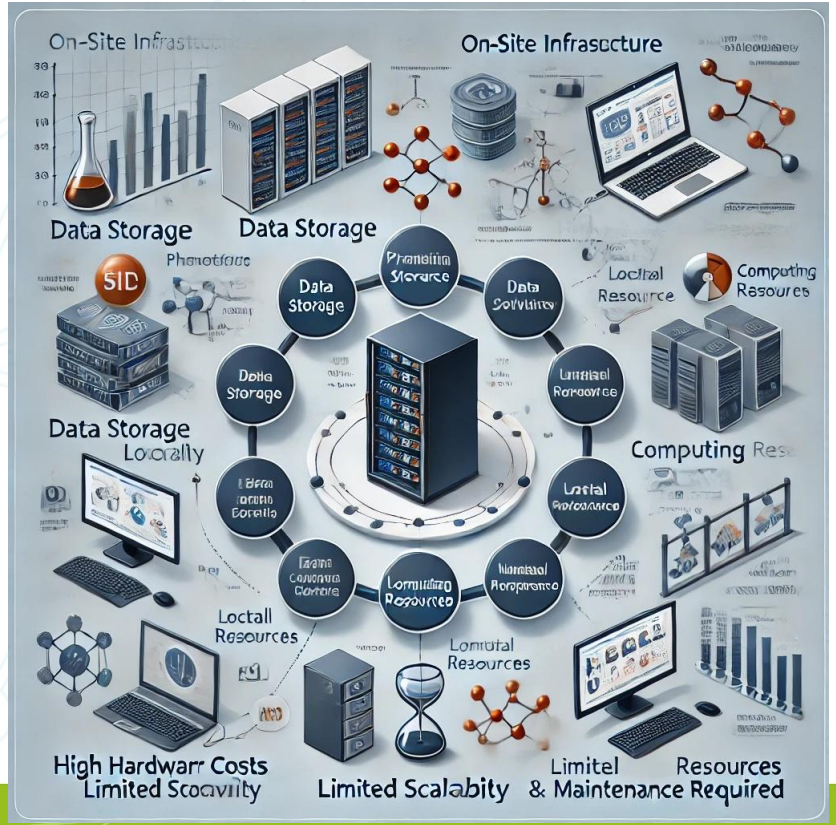
What is Cloud Computing?

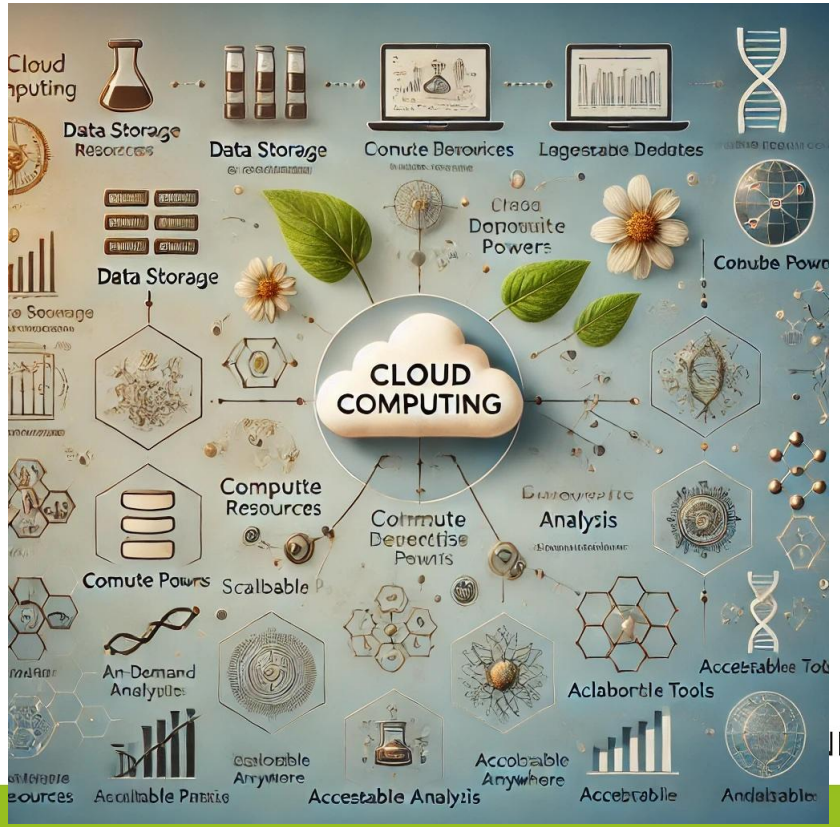
Cloud computing provides on-demand access to computing resources over the internet.

- **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.

PHENET

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

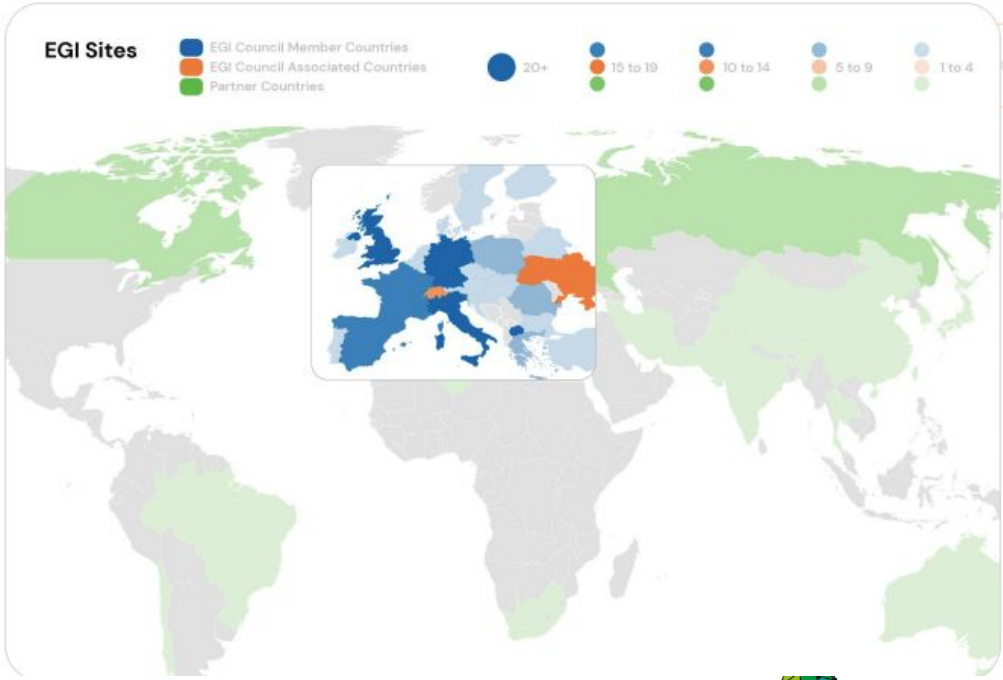
Topic 3: How to run your analysis workflow in EGI Notebook?

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

PHENET

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)**.



PHENET

EGI in Numbers (2023)

82M
Cloud CPU hours consumed

37
EU funded projects

94,200
Users

2,875
Enabled articles [1]

7.0B
HTC CPU hours consumed

2022 2023

+100,000 users in 2024!

Who is it for? (EGI community)

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)

PHENET

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*

PHENET

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).*

PHENET

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.

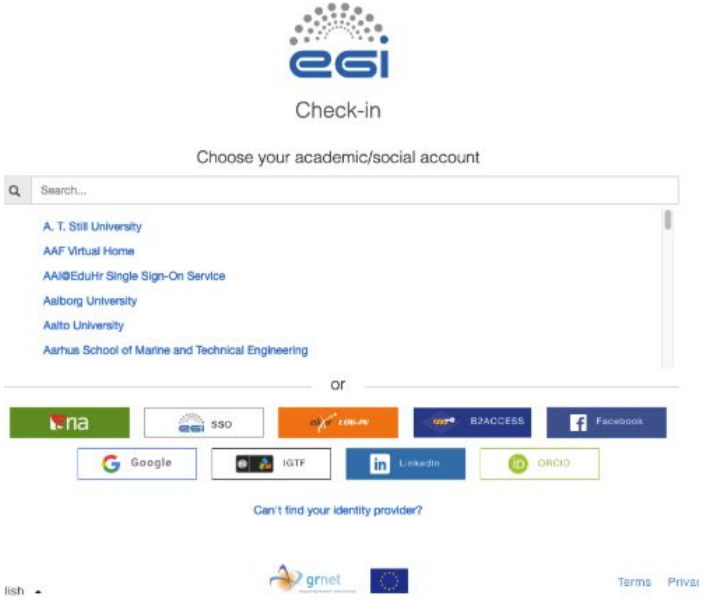
*Think of it like a **master key** for research tools:*

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

PHENET

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/



PHENET

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

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: v**o.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

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

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

EMPHASIS

SLA

OLA

cesnet

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 PHIS information system network was brought into the EMPHASIS ESFRI 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)**[1].

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

Providers	Cloud Compute	Online Storage	DataHub	Check-in
IN2P3-IRES (FR)	20 vCPU cores, 50 GB RAM, floating IPs available	1 TB		No. 1 COU group[2] 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.

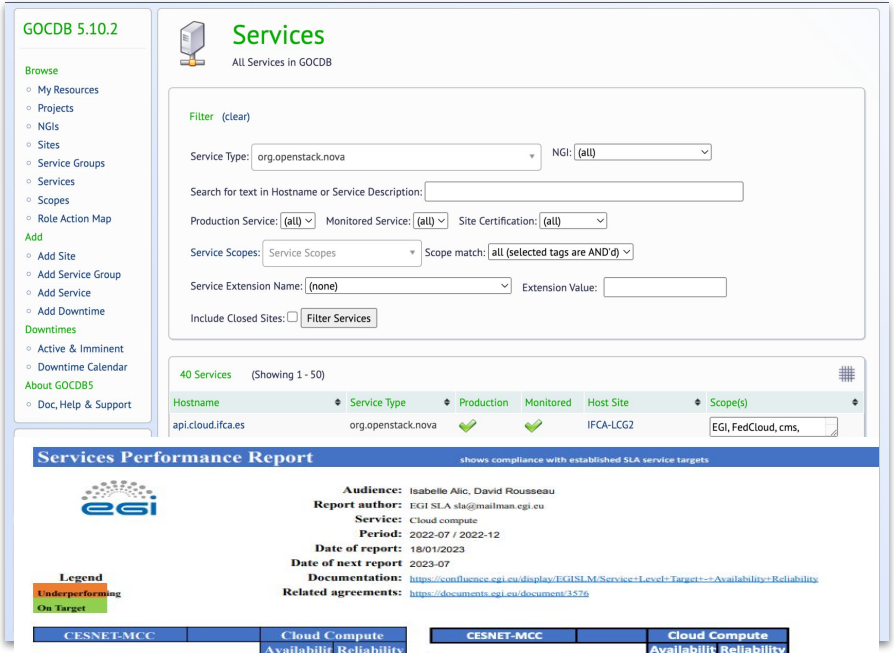
This Agreement is valid from **01/01/2022** to **30/06/2023**.

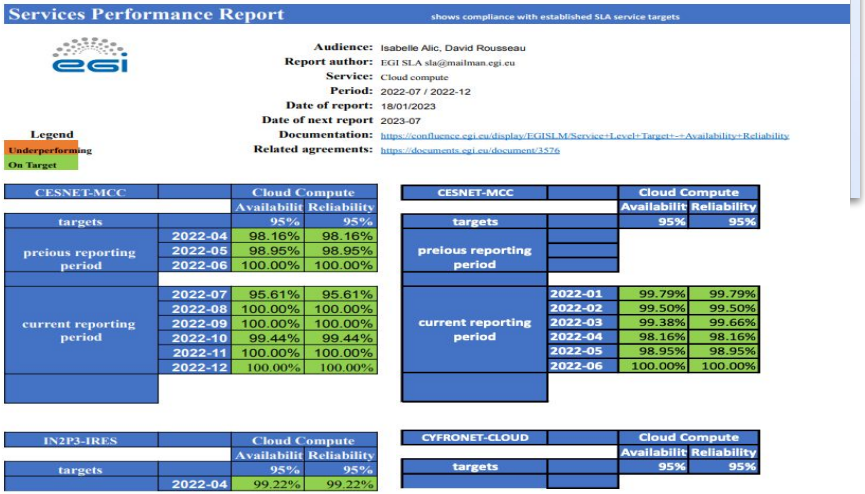
The Agreement was discussed and approved by the Customer and the Service Provider on **20/07/2022**.

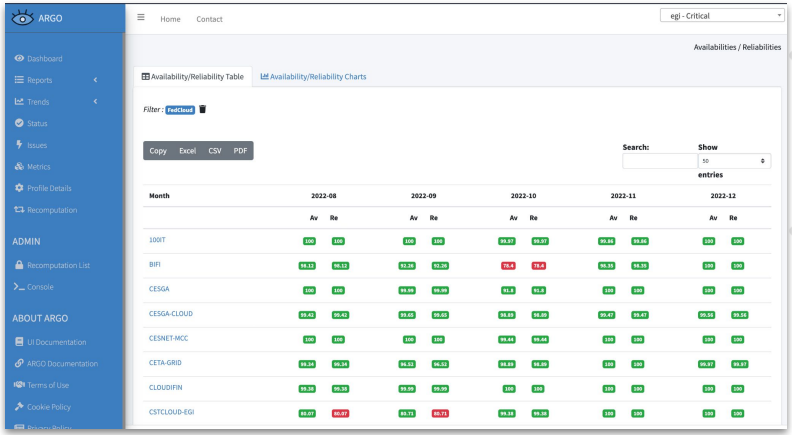
[1] https://www.inrae.fr/en
[2] https://aai.egi.eu/registry/co_petitions/start/coef:45

4

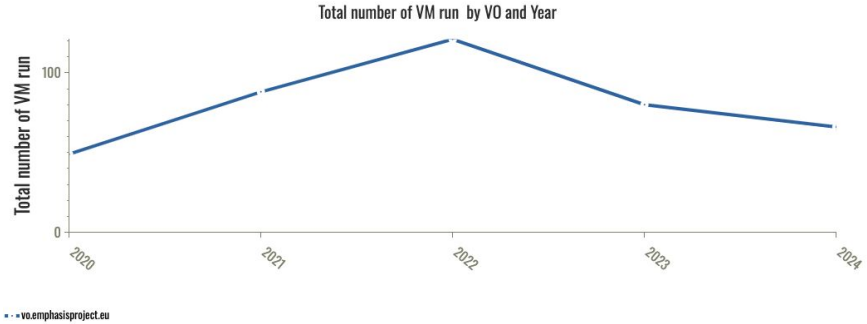
Performance monitoring of the EMPHASIS VO





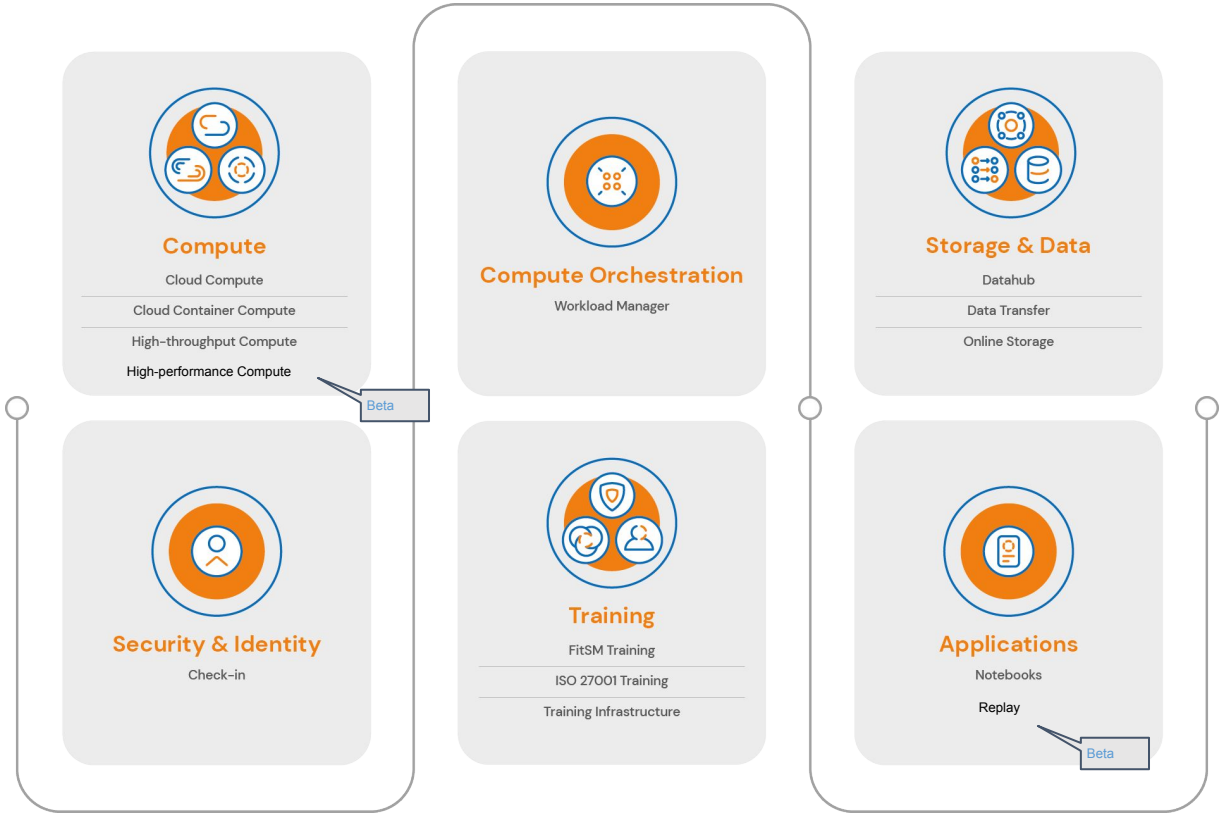


Performance of EMPHASIS VO monitored by EGI ARGO:
https://argo.egi.eu/egi/report-ar-group-details/SLA/SERVICEGROUPS/EGI_EMPHASIS_SLA/details



Usage of EMPHASIS VO stat monitered by the EGI Accounting Portal:
https://accounting.egi.eu/cloud/vm_num/VO/Year/2020/1/2024/11/custom-vo.emphasisproject.eu/onlyinfrajobs/

EGI External Services (for Research)



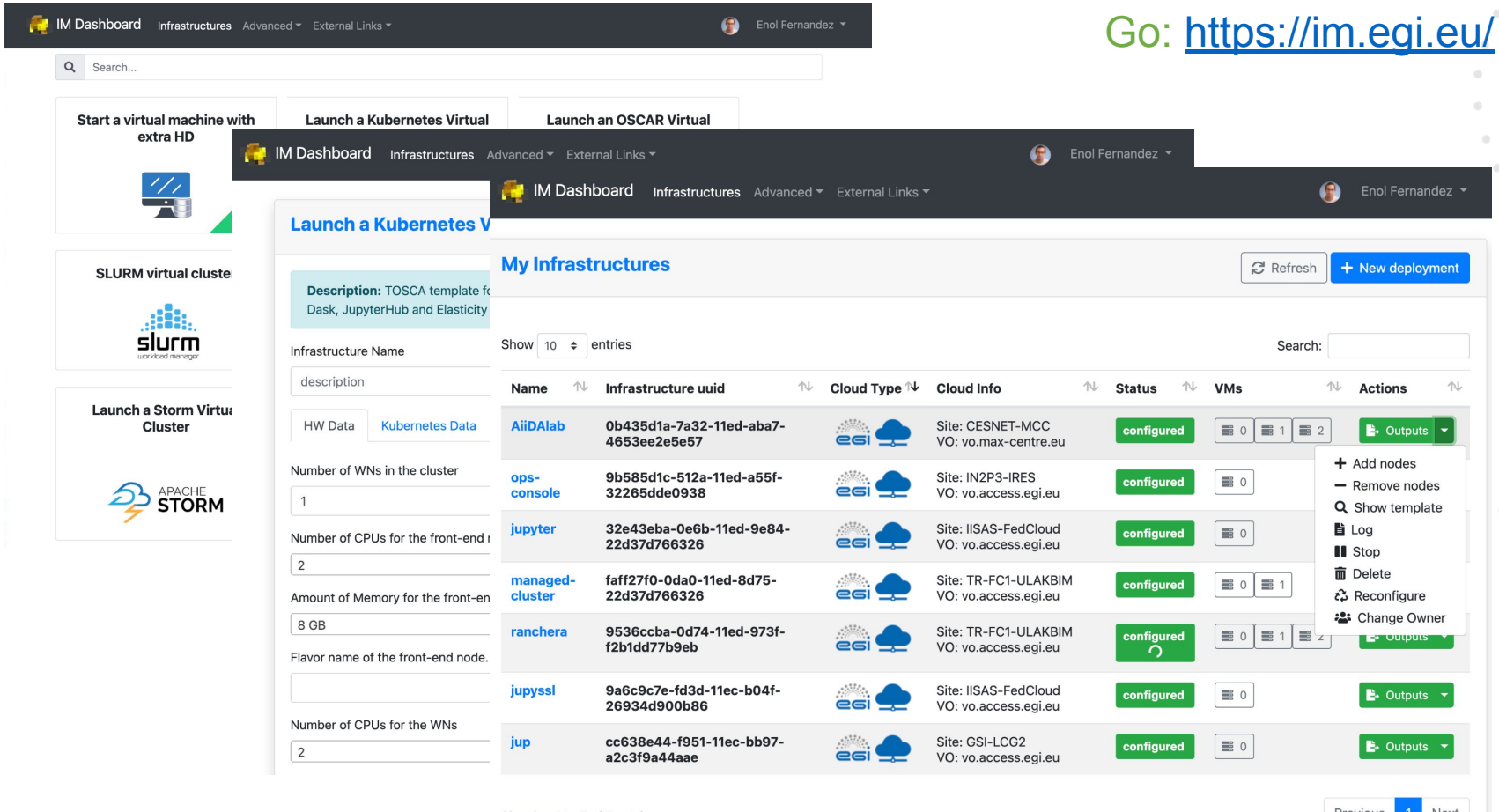
Service catalogue:
https://www.egi.eu/services/

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/



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

PHENET

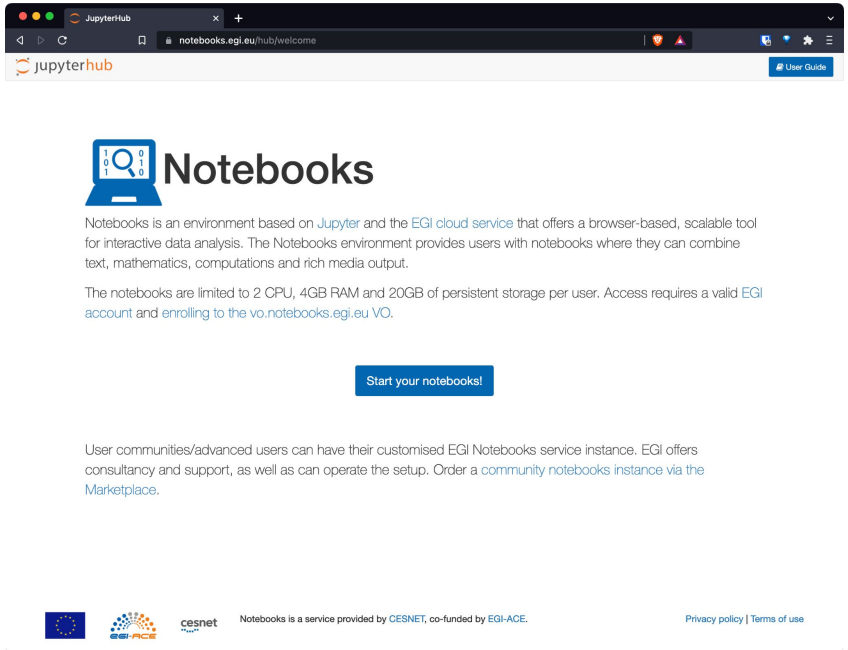
What is EGI Notebooks?

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



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

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

PHENET

PHENOTYPING & ENVIROTYPING SOLUTIONS FOR AGROECOLOGY

https://www.phenet.eu/en

phenet_project

PHENET