



ESGI Tutorial at ISGC 2023

Giuseppe La Rocca

Community Support Team Lead.

TLP: GREEN Limited disclosure



Community Support Team Lead. @ EGI Foundation
Member of the EGI Foundation team since Dec. 2015
MSc in Computer Science Engineering from the
University of Catania (Italy).
EGI-ACE Community Manager
Based on Catania, Italy

giuseppe.larocca@egi.eu
<https://www.egi.eu/people/giuseppe-la-rocca/>

14:00	Intro about the EGI and the EGI infrastructure <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 14:00 - 14:20
	EGI VO for AP <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 14:20 - 14:50
15:00	DEMO - How to use the EGI Cloud <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 14:50 - 15:10
	DEMO - Getting started with the EGI Notebooks <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 15:10 - 15:30
16:00	Hands-on with the EGI Notebooks and Replay services <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 16:00 - 16:30
	Approach to reproducible data science with EGI and EOSC <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 16:30 - 16:45
17:00	Overview of the hands-on - Full data lifecycle <i>Auditorium, BHSS, Academia Sinica</i>	<i>Giuseppe La Rocca</i> 16:45 - 17:30



Introduction about the EGI and the EGI Infrastructure

TLP: GREEN Limited disclosure

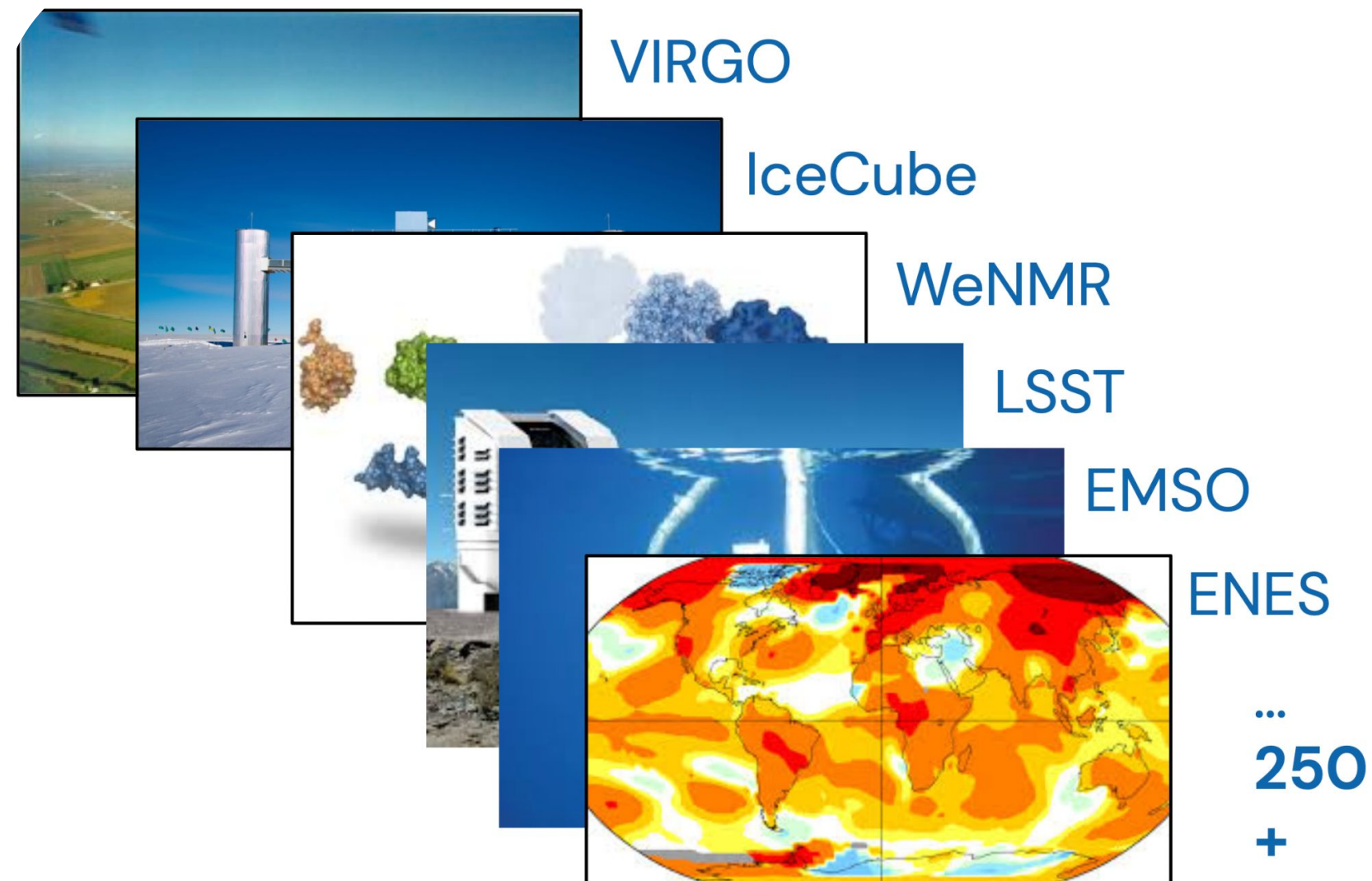
The EGI Federation is an international e-infrastructure

We provide advanced computing and data analytics for research and innovation



2010

From the high-energy physics compute grid (WLCG)



2022

To a multi-disciplinary, multi-technology infrastructure



Vision

All researchers have seamless access to services, resources and expertise to collaborate and conduct world-class research and innovation

Mission of the EGI Federation

Deliver open solutions for advanced computing and data analytics in research and innovation

Mission of the EGI Foundation

Enable the EGI Federation to serve international research and innovation together

- Participant Country
- Associated Participant Country
- Participant
- Associated Participant
- Headquarter of international research organisation or institutional representative

EGI Council Participants:

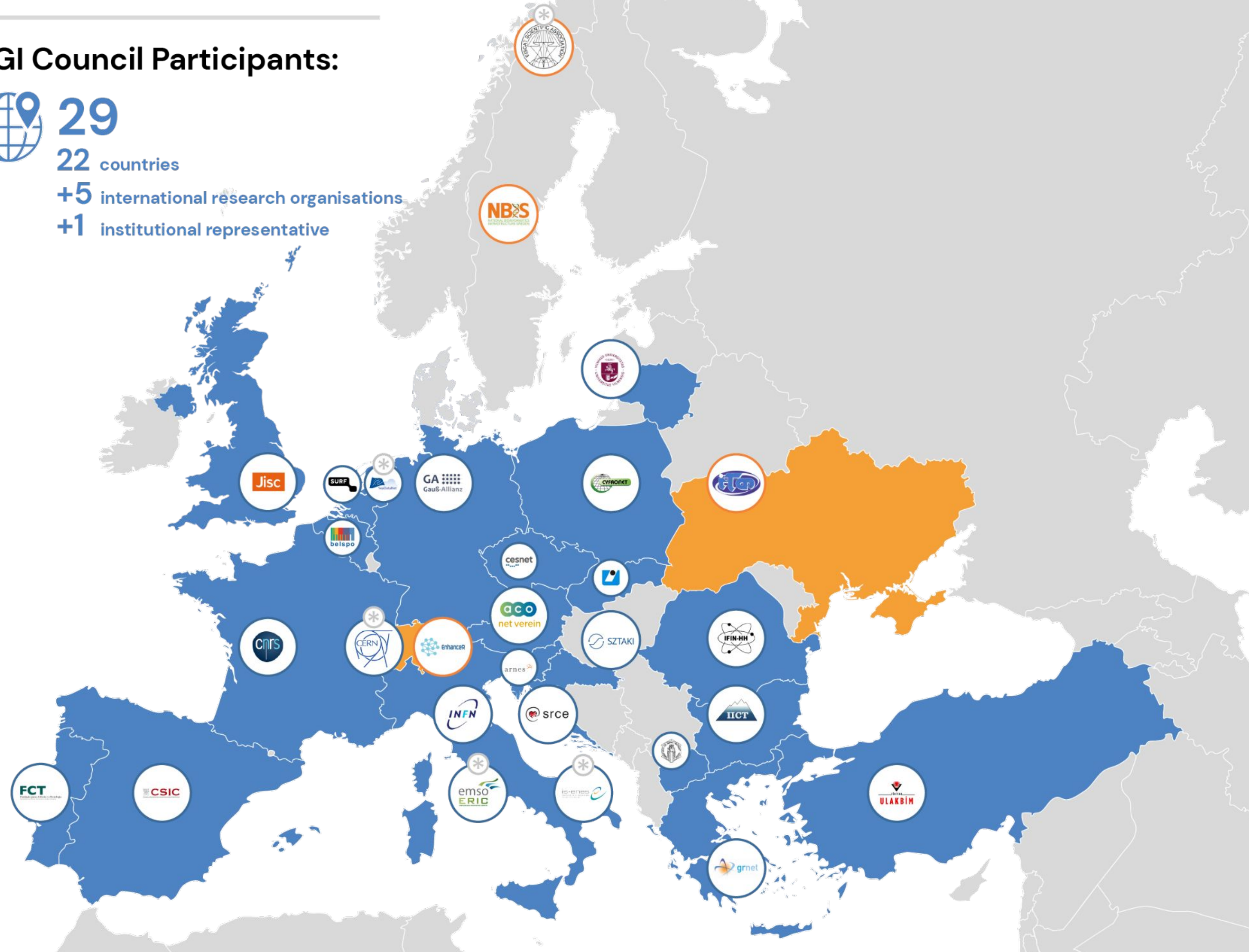


29

22 countries

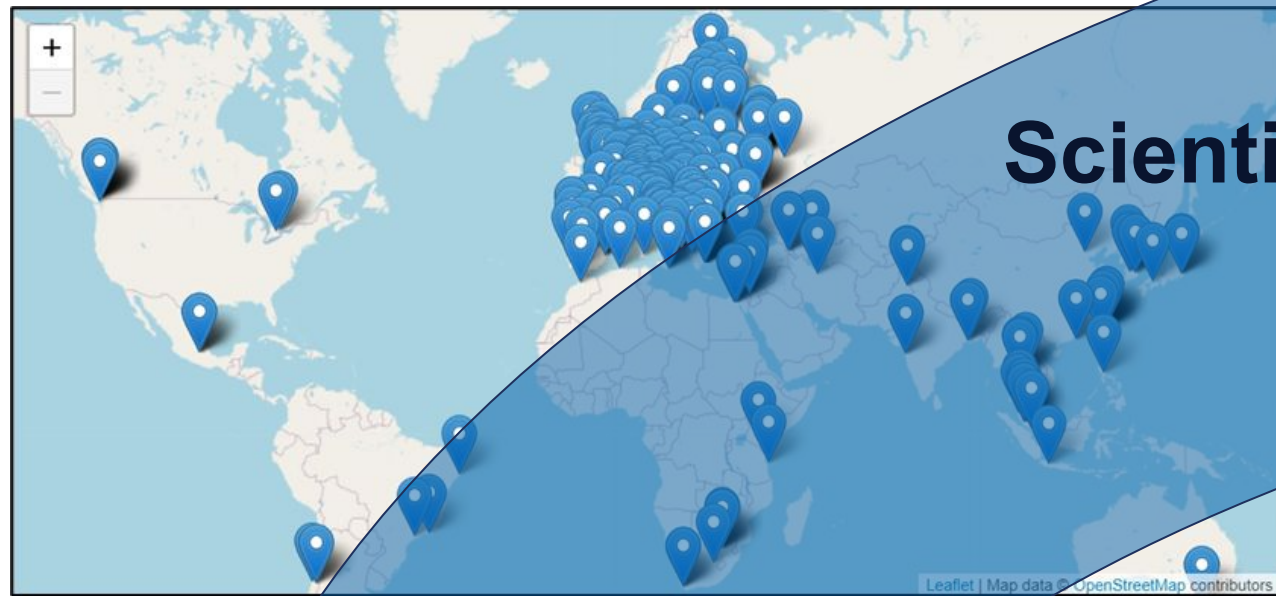
+5 international research organisations

+1 institutional representative



International Partnerships



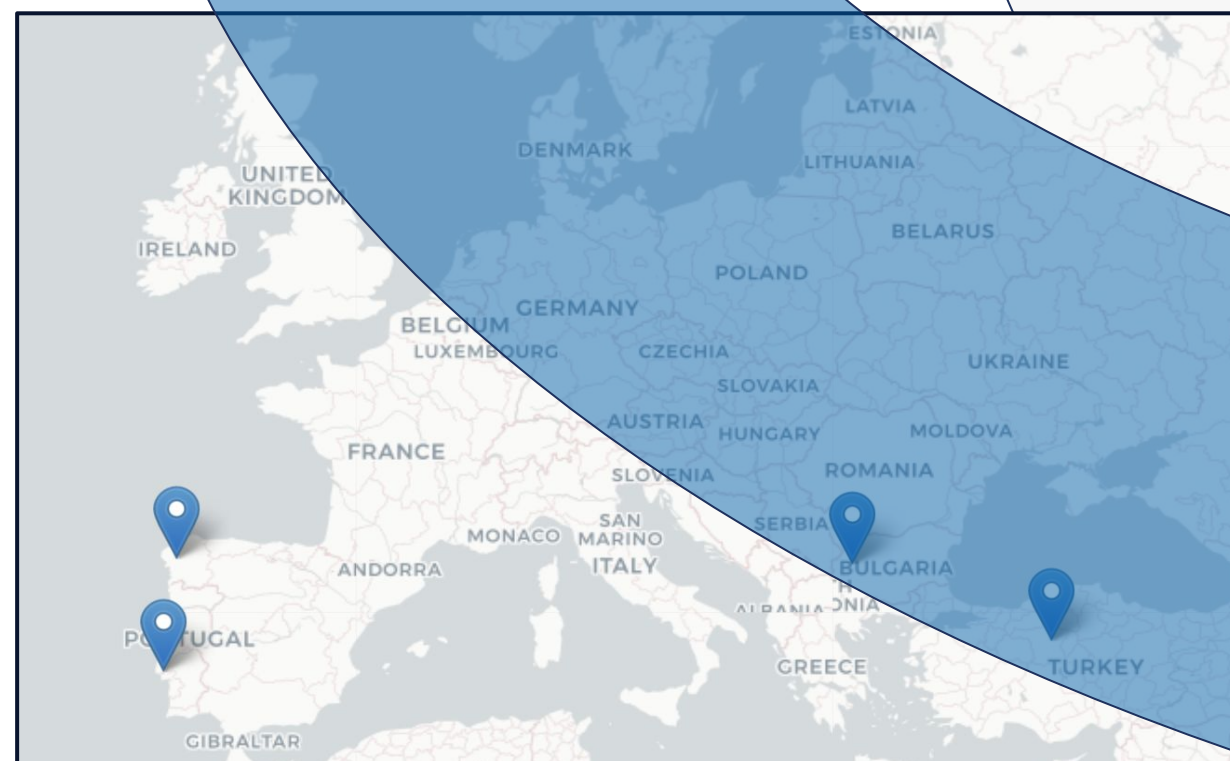
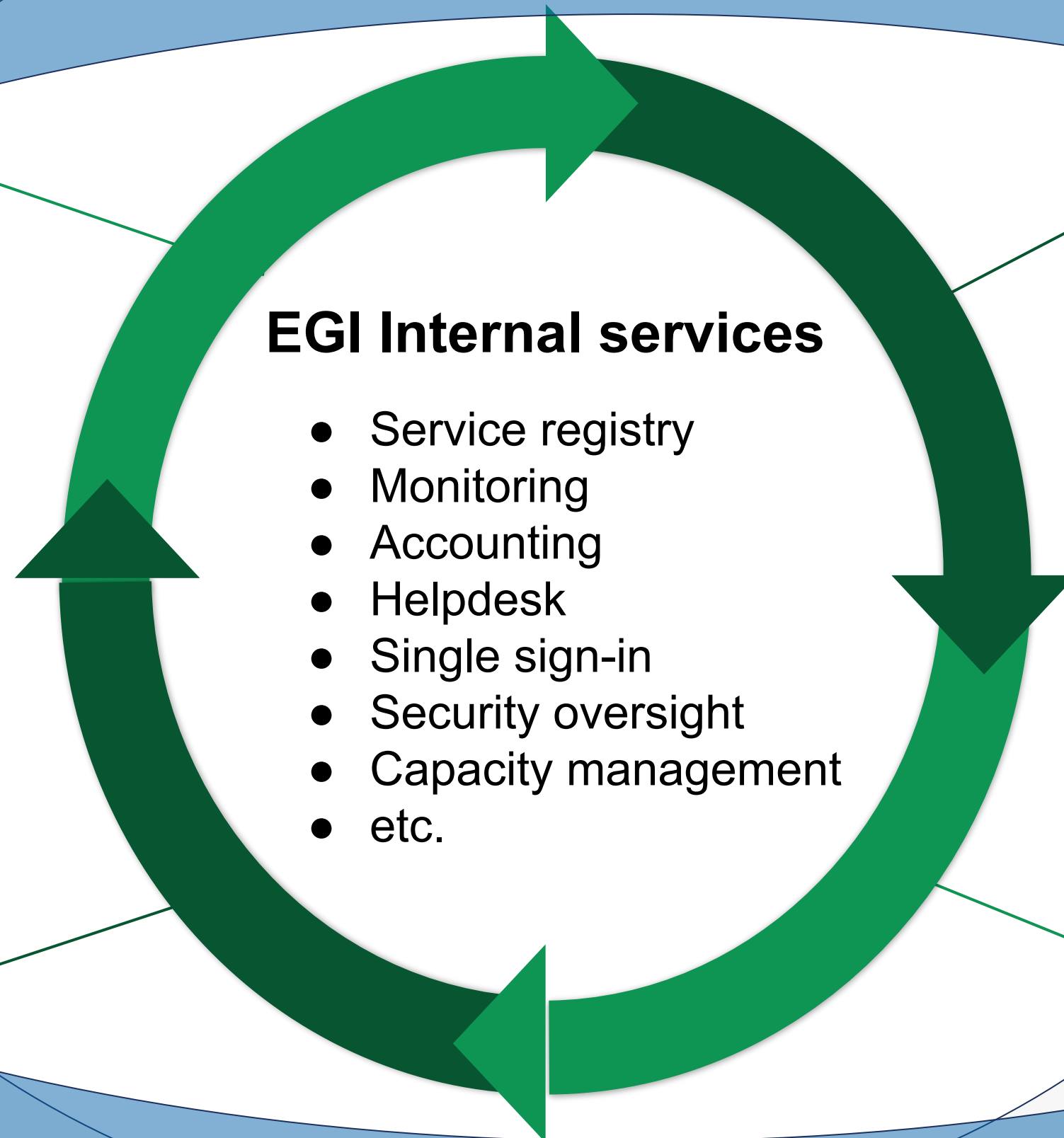


High-Throughput Compute
(ARC-CE, HTCondorCE, SRM, webdav, XrootD)

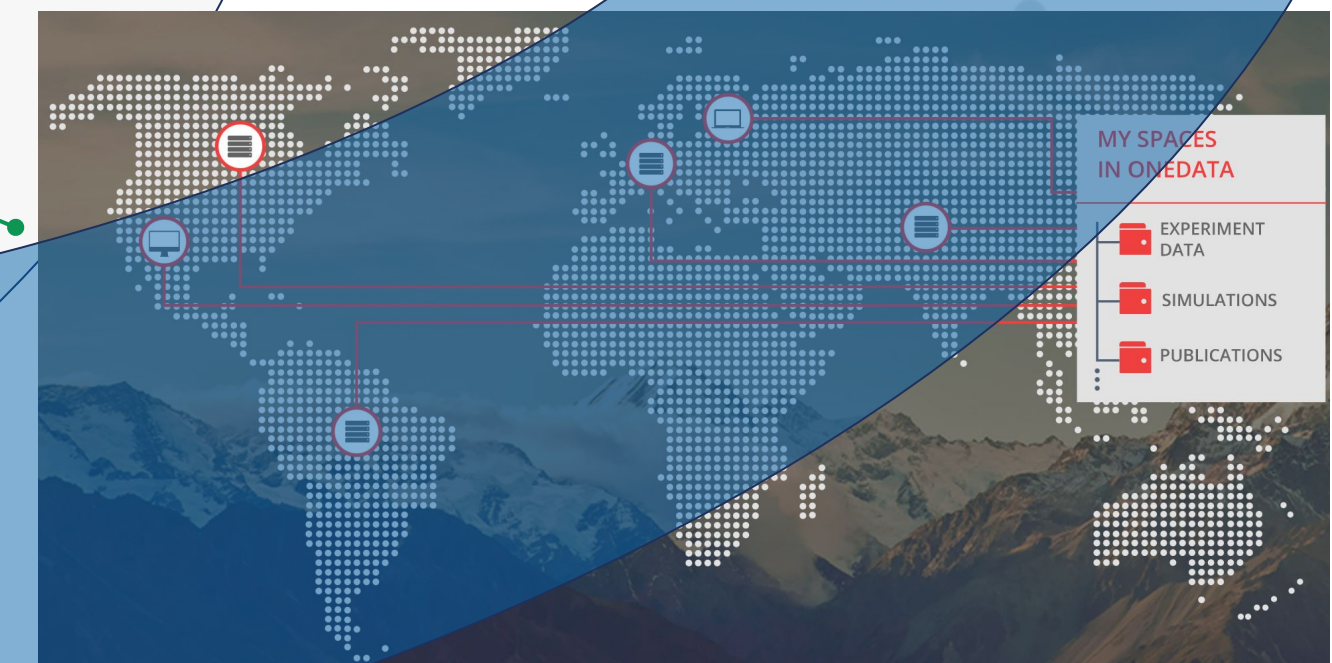
Platform and community specific services built on top:
Scientific gateways, Virtual Research Environments, Application portals, etc.



Cloud Compute
(based on OpenStack. Since 2013)



High Performance Compute

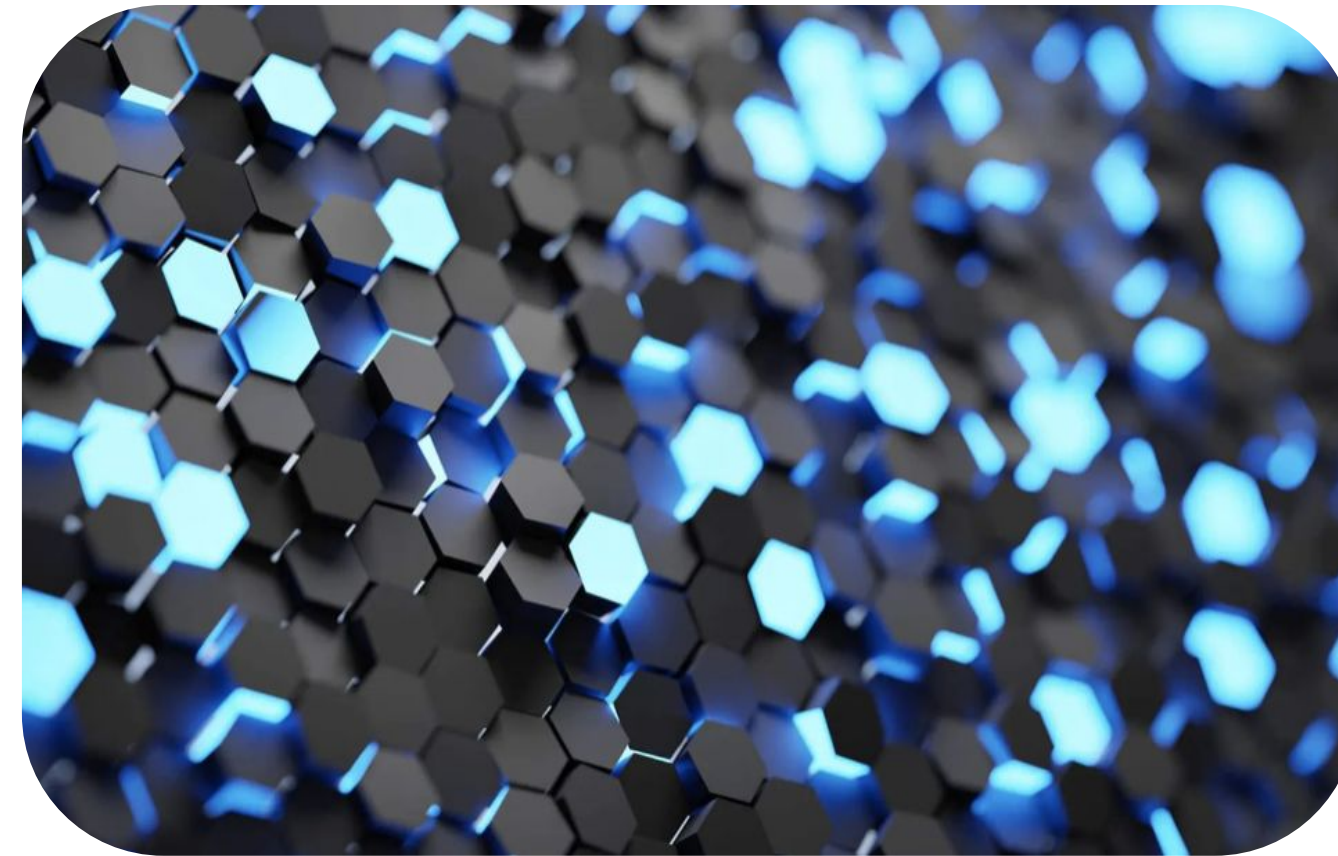


Data
(based on OneData)



Services for Research

Our large-scale computing and data analytics services are helping scientists to accelerate their research.



Services for Federation

Our internal services provide tools for coordination of the EGI Federation, improving how we work together.



Services for Business

We help companies to exploit and provide services and resources for compute- and data-intensive research and innovation.

Access policies

✓ Wide access

Users can freely access scientific data and digital services provided by EGI resource providers

Example:

- EGI Notebooks open instance
- EGI Cloud pool for application piloting

✓ Policy-based

Users are granted access based on policies defined by the EGI resource providers or by the EGI Foundation

Example:

- Biomed HTC VO pool for life sciences
- INFN-Bari (IT), IN2P3-IRES (FR), ULAKBIM (TR) provide Clouds for the National Bioinformatics Infrastructure of Sweden

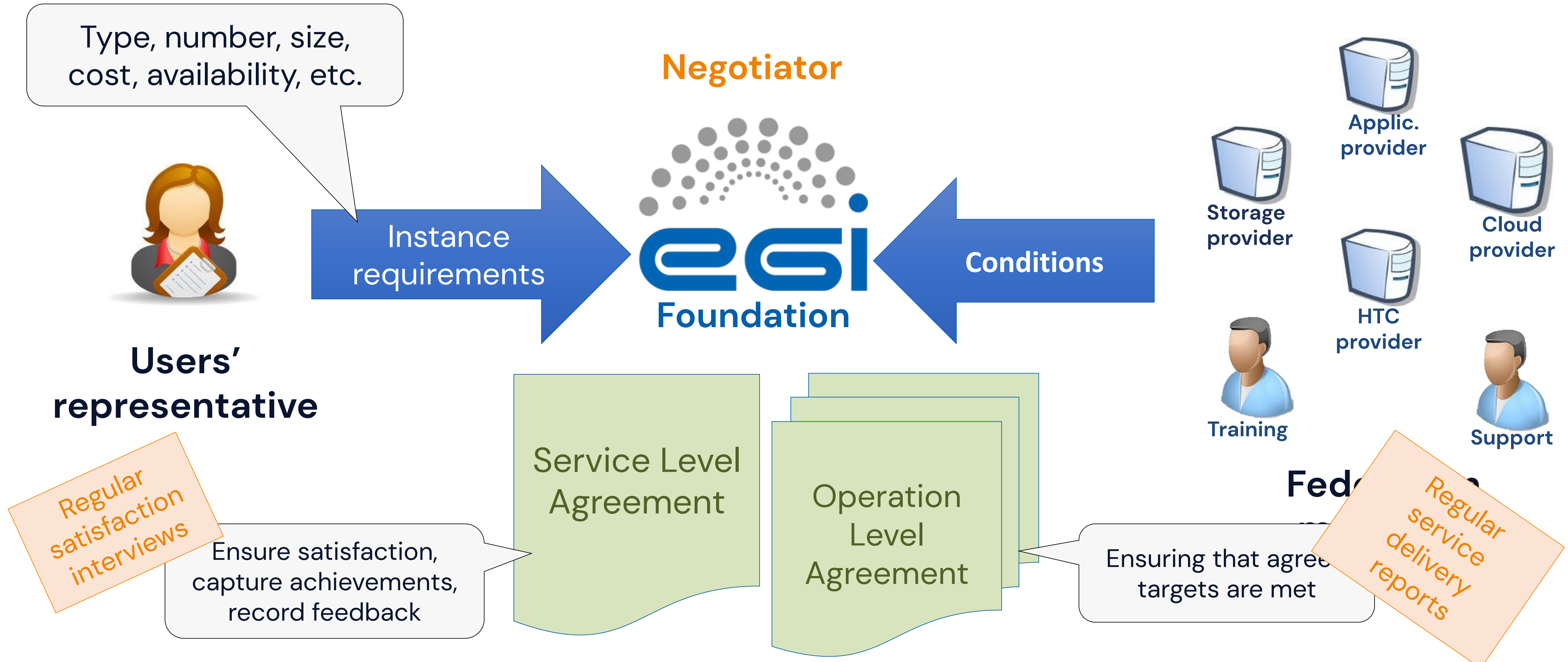
✓ Market-driven

Users can negotiate a fee to access services either directly with EGI resource providers or indirectly with the EGI Foundation

Example:

- FitSM training by EGI Foundation
- 4 EGI Cloud providers deliver to Exprivia

Allocating services and resources





An SLA example



EGI VO SERVICE LEVEL AGREEMENT

The Services are defined by the following properties:

Cloud Compute (category: Compute)

Description: <https://www.egi.eu/services/cloud-compute/>

- Resource Centre: **CESGA (Country: Spain)**
 - Cloud Compute
 - Number of virtual CPU cores: 192
 - Memory per core (GB): 2.5GB. A total of 512GB is provided
 - Local disk (GB): 50
 - Public IP addresses:
 - Allocation type: Pledged
 - Payment mode offer: Sponsored
 - Other technical requirements:
 - Duration: 01/12/2019 – 30/06/2023
 - Supported VOs: vo.emso-eric.eu
 - VO ID card: <https://operations-portal.in2p3.fr/vo/view/voname/vo.emso-eric.eu>
 - VO-wide list: <https://vmcaster.appdb.egi.eu/store/vo/vo.emso-eric.eu/image.list>
- Resource Centre: **RECAS-BARI (Country: Italy)**
 - Cloud Compute
 - Number of virtual CPU cores: 300
 - Memory per core (GB): A total of 1.2TB is provided
 - Local disk (GB):
 - Public IP addresses:
 - Allocation type: Pledged
 - Payment mode offer: Sponsored
 - Other technical requirements:
 - Duration: 01/12/2019 – 30/06/2023
 - Supported VOs: vo.emso-eric.eu
 - VO ID card: <https://operations-portal.in2p3.fr/vo/view/voname/vo.emso-eric.eu>
 - VO-wide list: <https://vmcaster.appdb.egi.eu/store/vo/vo.emso-eric.eu/image.list>

Online Storage (category: Storage)

- Resource Centre: **CESGA (Country: Spain)**
 - Online Storage
 - Guaranteed storage capacity [TB]: 0.6
 - Opportunistic storage capacity [TB]:

Online Storage (category: Storage)

Description: <https://www.egi.eu/services/online-storage/>

- Resource Centre: **CESGA (Country: Spain)**
 - Online Storage
 - Guaranteed storage capacity [TB]: 0.6
 - Opportunistic storage capacity [TB]:



2 Service hours and exceptions

The Services operate during the following hours: twenty-four (24) hours a day, seven (7) days a week, three hundred sixty-five (365) days a year.

The following exceptions apply:

- ² CDMI, POSIX, SWIFT, etc.
- ³ DPM, dCache, STORM, etc.
- ⁴ CDMI, POSIX, SWIFT, etc.
- ⁵ DPM, dCache, STORM, etc.
- ⁶ <http://accounting.egi.eu/>
- ⁷ <http://argo.egi.eu/>



- Planned maintenance windows or service interruptions ("scheduled downtimes") will be notified via email in a timely manner i.e. 24 hours before the start of the outage⁹.
- Downtime periods exceeding 24 hours need justification.

3 Support

Support is provided via EGI Service Desk¹⁰. Access requires a valid X.509 or the login via a EGI SSO account¹¹. Support is available between:

- Monday to Friday.
- From 9:00 to 17:00 in the time zone of the relevant Resource Centres.

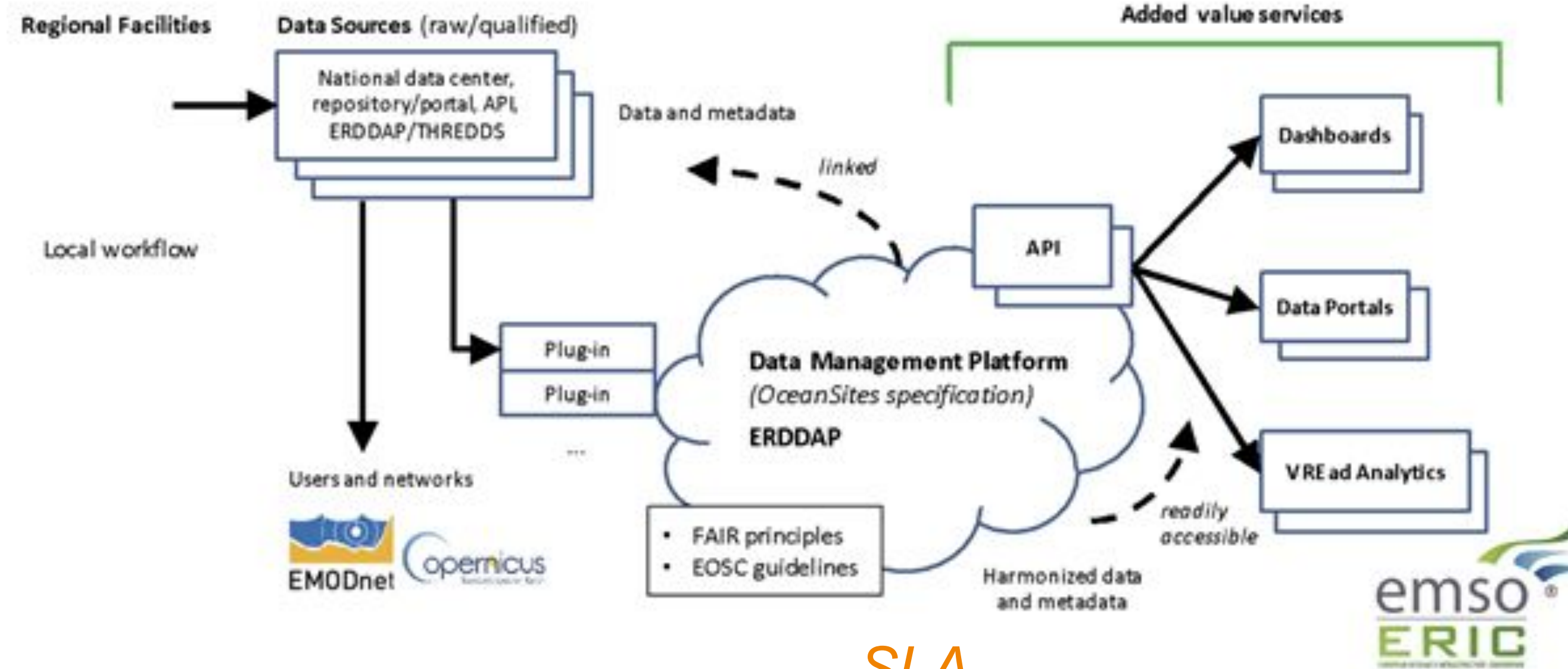
Service times always apply with the exception of public holidays in the country of the supporting Resource Centres.

3.1 Incident handling

Incidents will be handled according to the Quality of Support level that is estimated according to the impact of the outage or service quality degradation.

The Quality of Support in this Agreement has level: **Medium**¹²

Incident priority	Response time
Less urgent	5 working days



SLA



OLA-1

OLA-2



Read the full story:
<https://www.egi.eu/cas-e-study/emso-eric/>

Check the SLA-OLA documents and reports:
<https://documents.egi.eu/document/353>



ESG allocations for the Asian Pacific (AP)

TLP: GREEN Limited disclosure

- A Virtual Organisation (alias resource pool) in the EGI cloud federation
- With a community specific Notebooks installation
- With a community specific Replay installation

- Purposes:
 - o To facilitate open science in the region → IN THE SCOPE OF THE TUTORIAL TODAY
 - o To enable sharing of cloud resources from the region
 - o To facilitate application sharing from the region
 - o To facilitate data sharing related to environmental sciences from the region

For Users

- ▼ Getting Started
 - Architecture
 - OpenStack
 - Providers
 - Command Line
- ▼ Communities
 - Disaster Mitigation and Agriculture**
 - ▶ Tutorials
 - ▶ Authentication & Authorization
 - ▶ Compute
 - ▶ Data
 - ▶ Security
 - ▶ Development Environments
 - ▶ Training Infrastructure

[For Users](#) / [Getting Started](#) / [Communities](#) / [Disaster Mitigation and Agriculture](#)

Disaster Mitigation and Agriculture

EGI infrastructure for the Disaster Mitigation and Agriculture community

This is the documentation to support the **Disaster Mitigation and Agriculture** community in the Asian Pacific region.

About the community

Hazard risk estimation and prediction by numerical simulation is crucial to disaster mitigation studies and applications. The Disaster Mitigation and Agriculture community investigates in-depth the mechanisms of the selected disaster events and develops the appropriate simulation models to reproduce the processes by case studies. The collaboration framework aims at becoming an open science platform of disaster mitigation so that all the tools, data, resources and simulation facilities are sharable, and the simulations are reproducible.

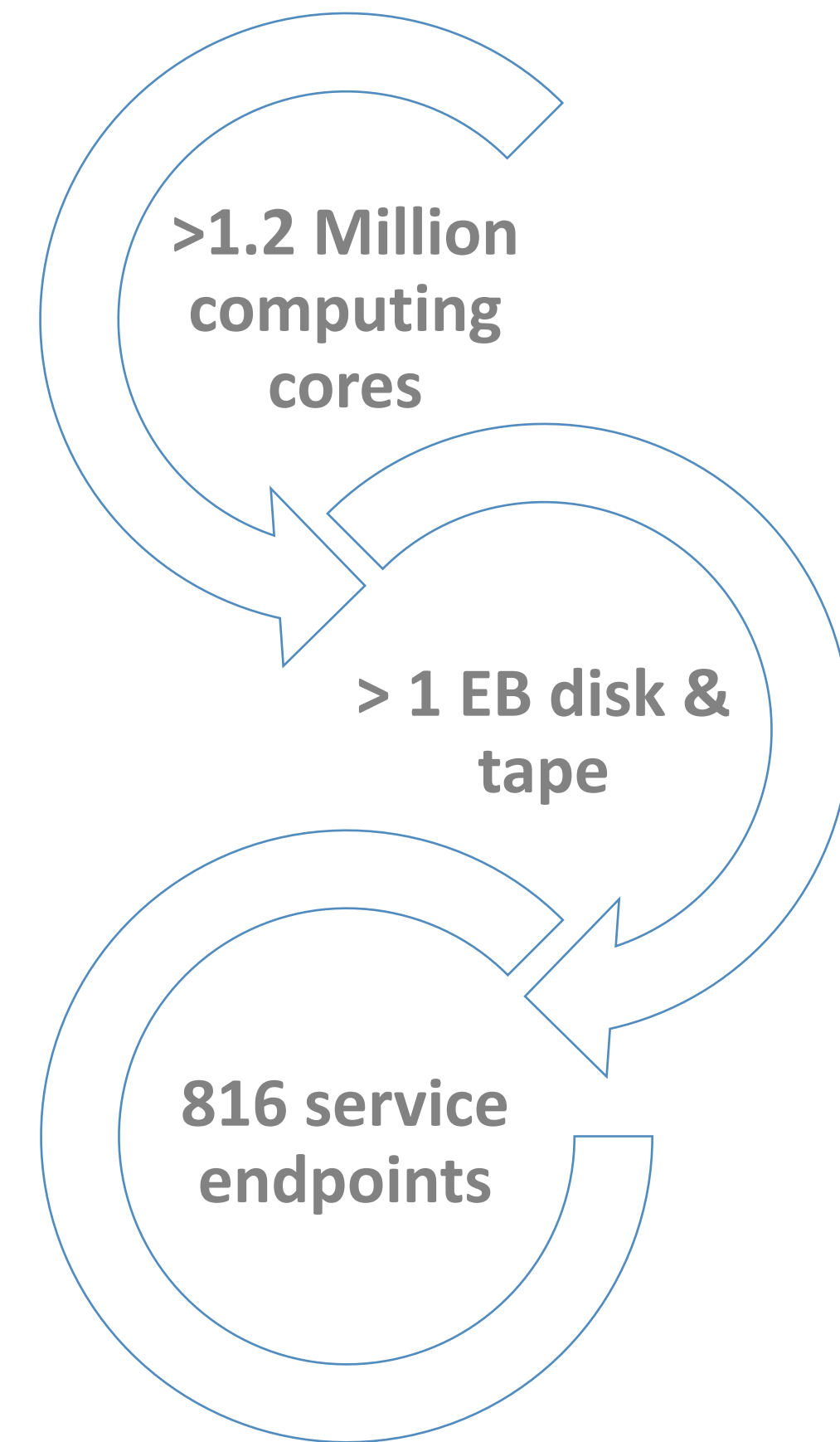
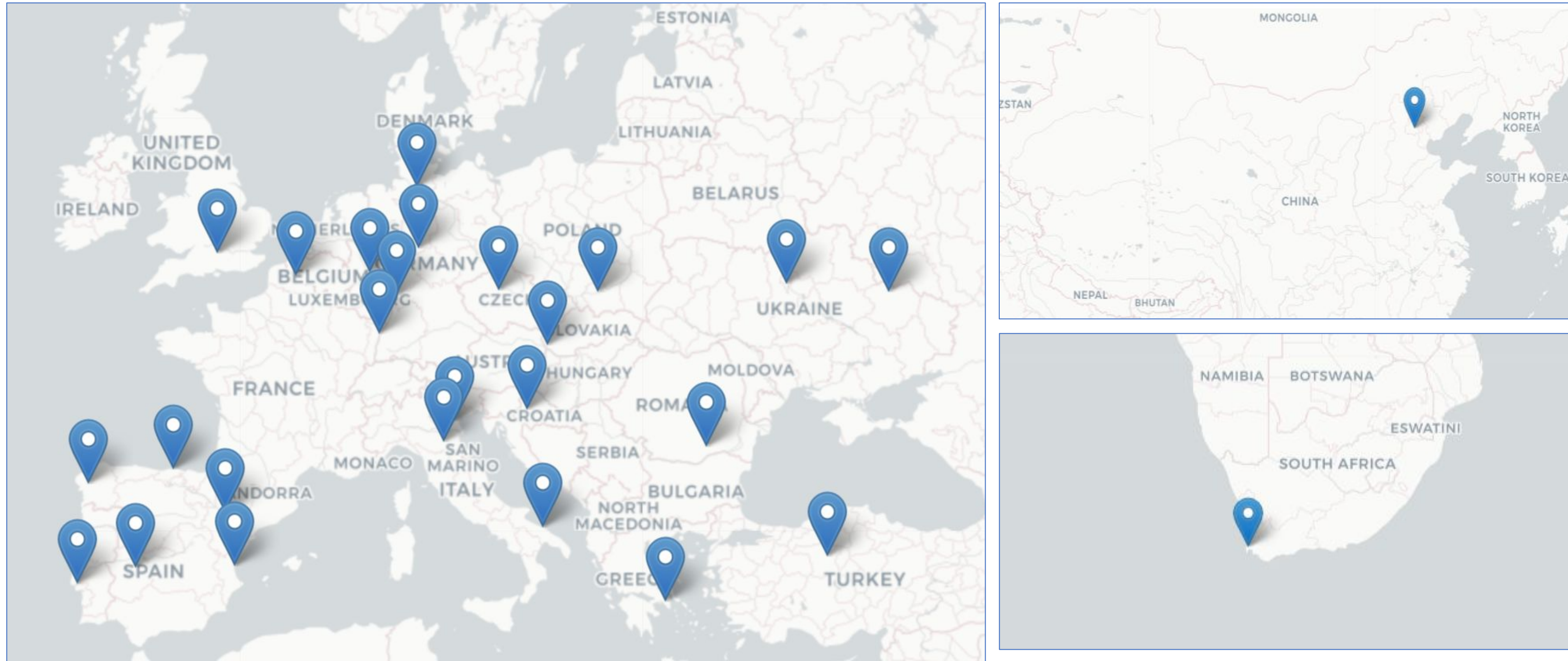
This knowledge base is enriched by the simulation models, portals, data and visualisation facilities that are contributed to by the members:

- Academia Sinica, Taiwan (Leading Partner, represented by [Academia Sinica Grid Computing Centre \(ASGC\)](#))
- [Institute of Earth Science, Academia Sinica, Taiwan](#)
- [Research Centre of Environmental Changes, Academia Sinica, Taiwan](#)



Demo: How to use the ESGI Cloud resources pool for Asia Pacific

TLP: GREEN Limited disclosure



- Operational since 2013
- **28** Cloud Providers (OpenStack)
- Providing different functionalities and more dynamic platforms for scientists
- **69,863,034** Cloud CPU/h consumed in 2022
- **105** Virtual Organisations supported (Virtual Organisation = Resource pool)
- Access via GUI, CLI, OpenStack native and libcloud APIs, Terraform

Self-deployed by users

- Spin up some VMs and deploy kubernetes (works with existing tooling targeting OpenStack)

Automated deployment with Infrastructure Manager (IM)

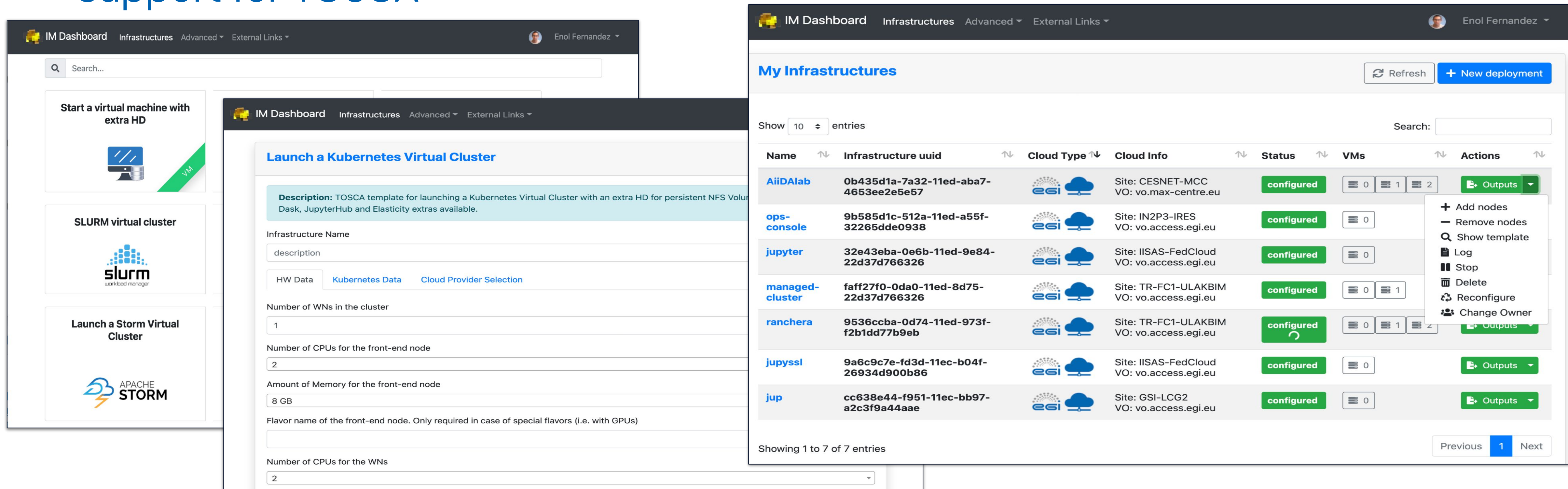
- Select site, size of VMs and go

Rancher-based access to managed clusters (under development)

- Rancher integrated with EGI Check-in
- Access to pre-configured clusters managed by expert operators

Infrastructure Manager orchestrates deployment of Virtual Infrastructures

- Multi-cloud following Infrastructure as a Code (IaC) paradigm
- Rely on Ansible and Helm to manage definition of complex infrastructure
- Manages whole life-cycle: deployment, monitoring, reconfiguration (such as scaling-up or down) and removal.
- Support for TOSCA

The screenshot displays the IM Dashboard interface, which is used for managing virtual infrastructures. It features a navigation bar with 'IM Dashboard', 'Infrastructures', 'Advanced', and 'External Links'. The main content area is divided into several sections:

- Start a virtual machine with extra HD:** A card with a VM icon and a green checkmark.
- SLURM virtual cluster:** A card with the SLURM logo and 'workload manager' text.
- Launch a Storm Virtual Cluster:** A card with the Apache Storm logo.
- Launch a Kubernetes Virtual Cluster:** A detailed form for launching a Kubernetes cluster. It includes a description: 'TOSCA template for launching a Kubernetes Virtual Cluster with an extra HD for persistent NFS Volume, JupyterHub and Elasticity extras available.' The form has tabs for 'HW Data', 'Kubernetes Data', and 'Cloud Provider Selection'. Fields include 'Infrastructure Name' (description), 'Number of WNs in the cluster' (1), 'Number of CPUs for the front-end node' (2), 'Amount of Memory for the front-end node' (8 GB), 'Flavor name of the front-end node' (optional), and 'Number of CPUs for the WNs' (2).
- My Infrastructures:** A table listing deployed infrastructures. The table has columns for Name, Infrastructure uuid, Cloud Type, Cloud Info, Status, VMs, and Actions. A dropdown menu is open over the 'Actions' column, showing options like 'Add nodes', 'Remove nodes', 'Show template', 'Log', 'Stop', 'Delete', 'Reconfigure', and 'Change Owner'.

Name	Infrastructure uuid	Cloud Type	Cloud Info	Status	VMs	Actions
AiiDAlab	0b435d1a-7a32-11ed-aba7-4653ee2e5e57	EGI	Site: CESNET-MCC VO: vo.max-centre.eu	configured	0 1 2	Outputs
ops-console	9b585d1c-512a-11ed-a55f-32265dde0938	EGI	Site: IN2P3-IRES VO: vo.access.egi.eu	configured	0	
jupyter	32e43eba-0e6b-11ed-9e84-22d37d766326	EGI	Site: IISAS-FedCloud VO: vo.access.egi.eu	configured	0	
managed-cluster	faff27f0-0da0-11ed-8d75-22d37d766326	EGI	Site: TR-FC1-ULAKBIM VO: vo.access.egi.eu	configured	0 1	
ranchera	9536ccba-0d74-11ed-973f-f2b1dd77b9eb	EGI	Site: TR-FC1-ULAKBIM VO: vo.access.egi.eu	configured	0 1 2	Outputs
jupysl	9a6c9c7e-fd3d-11ec-b04f-26934d900b86	EGI	Site: IISAS-FedCloud VO: vo.access.egi.eu	configured	0	Outputs
jup	cc638e44-f951-11ec-bb97-a2c3f9a44aae	EGI	Site: GSI-LCG2 VO: vo.access.egi.eu	configured	0	Outputs

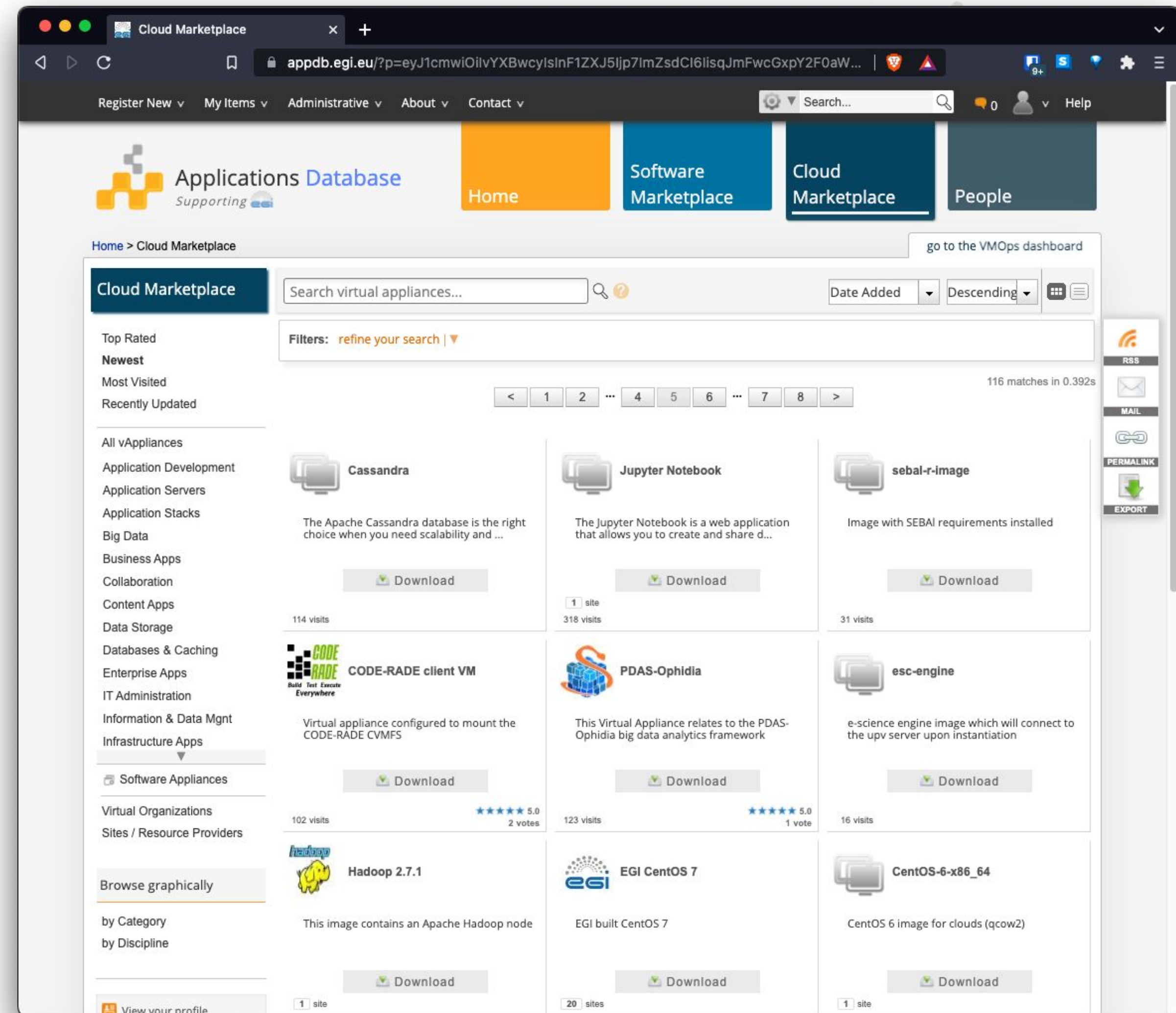
Common registry for Virtual Appliances (VA)

- VM image + metadata

Available for running at the EGI cloud or on any hypervisor

Community-level management of VAs

- Automatic distribution to providers

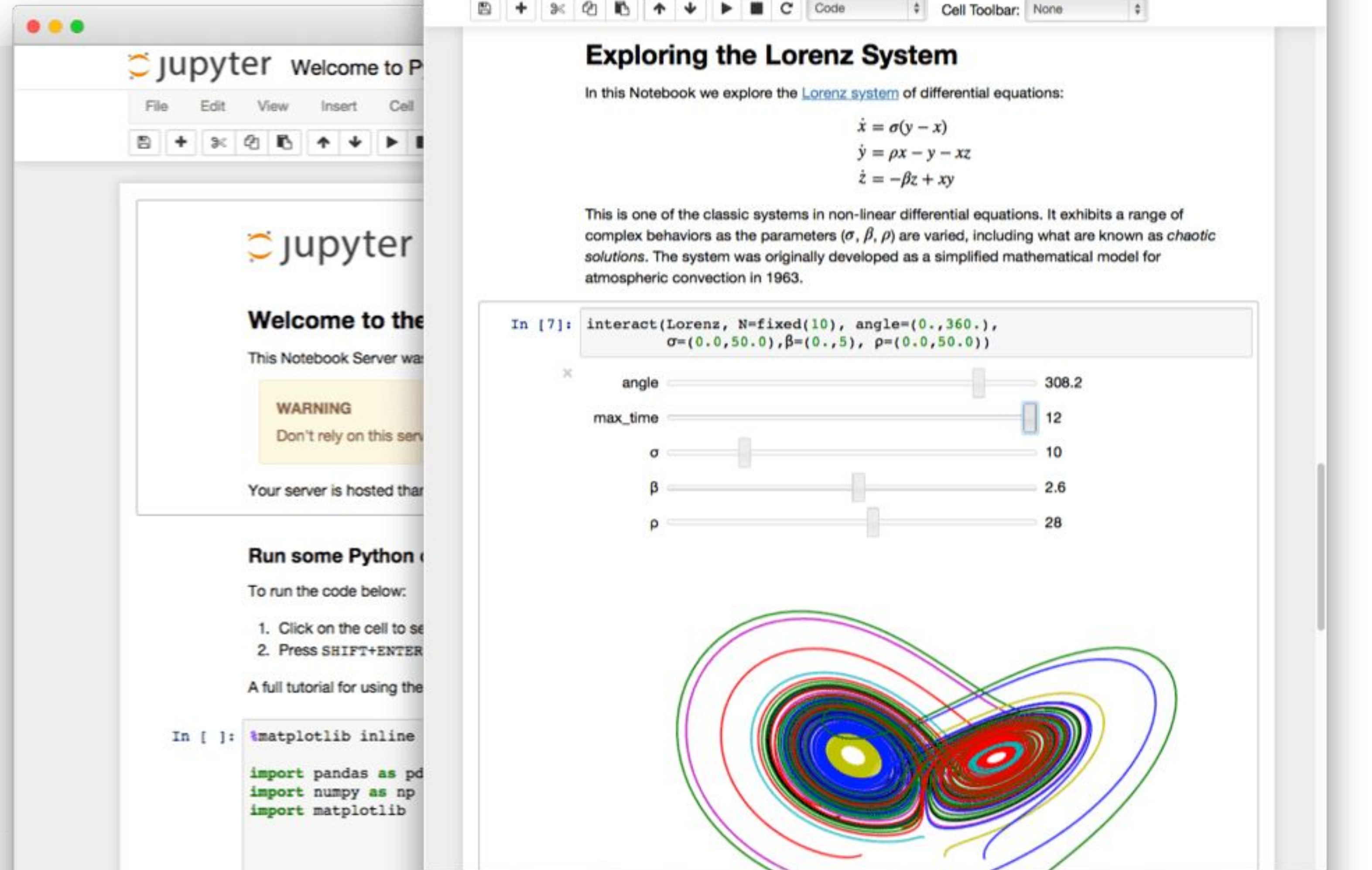




EGi Notebooks in the Asia Pacific VO

TLP: GREEN Limited disclosure

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.





Documentation
Text formatted using
Markdown/LaTeX

Interactive
browser based
environment



The screenshot shows a JupyterLab notebook titled 'FirstNotebook.ipynb' running on a Python 3 kernel. The interface includes a menu bar (File, Edit, View, Run, Kernel, Hub, Tabs, Settings, Help) and a toolbar with icons for file operations and execution. The notebook content is as follows:

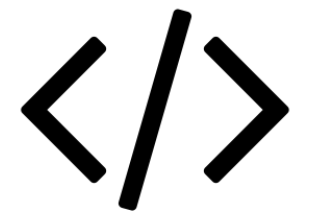
```
[4]: print("hello")
hello

[5]: import seaborn as sns
sns.set(style="darkgrid")

# Load an example dataset with long-form data
fmri = sns.load_dataset("fmri")

# Plot the responses for different events and regions
sns.lineplot(x="timepoint", y="signal",
             hue="region", style="event",
             data=fmri)
```

The output of the second cell is a line plot showing the signal over time for different brain regions (parietal and frontal) and events (stim and cue). The y-axis is labeled 'signal' and ranges from -0.1 to 0.3. The x-axis represents time. The plot shows two main peaks: a larger one around timepoint 10 and a smaller one around timepoint 20. The parietal region (blue) shows a higher peak than the frontal region (orange). The 'stim' event (solid lines) shows a higher signal than the 'cue' event (dashed lines). Shaded areas around the lines represent confidence intervals.



Code
Use your favourite
language



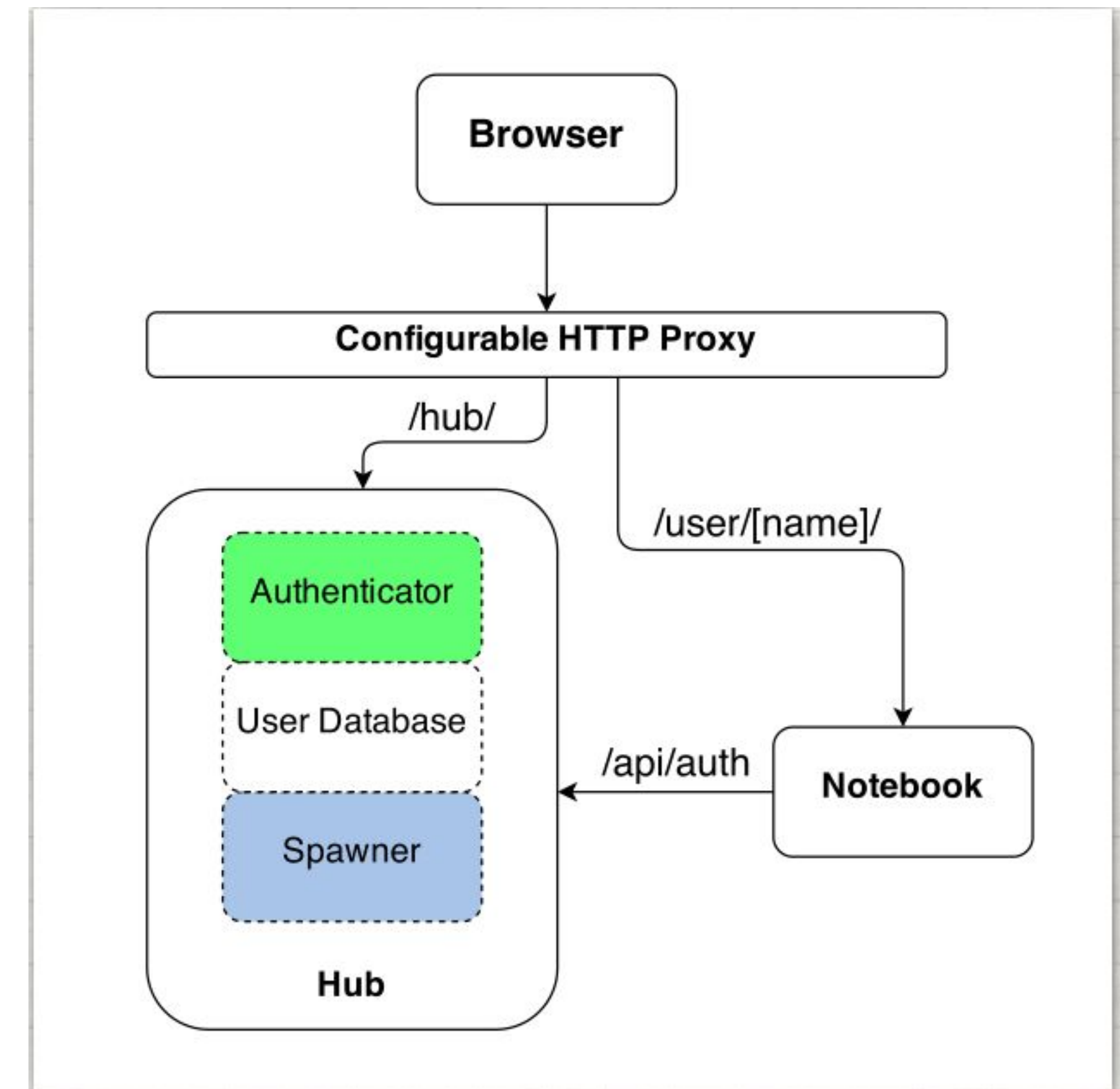
Output
Results of the code
execution(e.g. plots)



Jupyter is single user by design

JupyterHub is a multi-user version designed for companies, classrooms and research labs:

- Manages Authentication
- Spawns single-users notebooks servers on-demand
- Gives each user a complete Jupyter server





JupyterHub hosted in the EGI Cloud

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

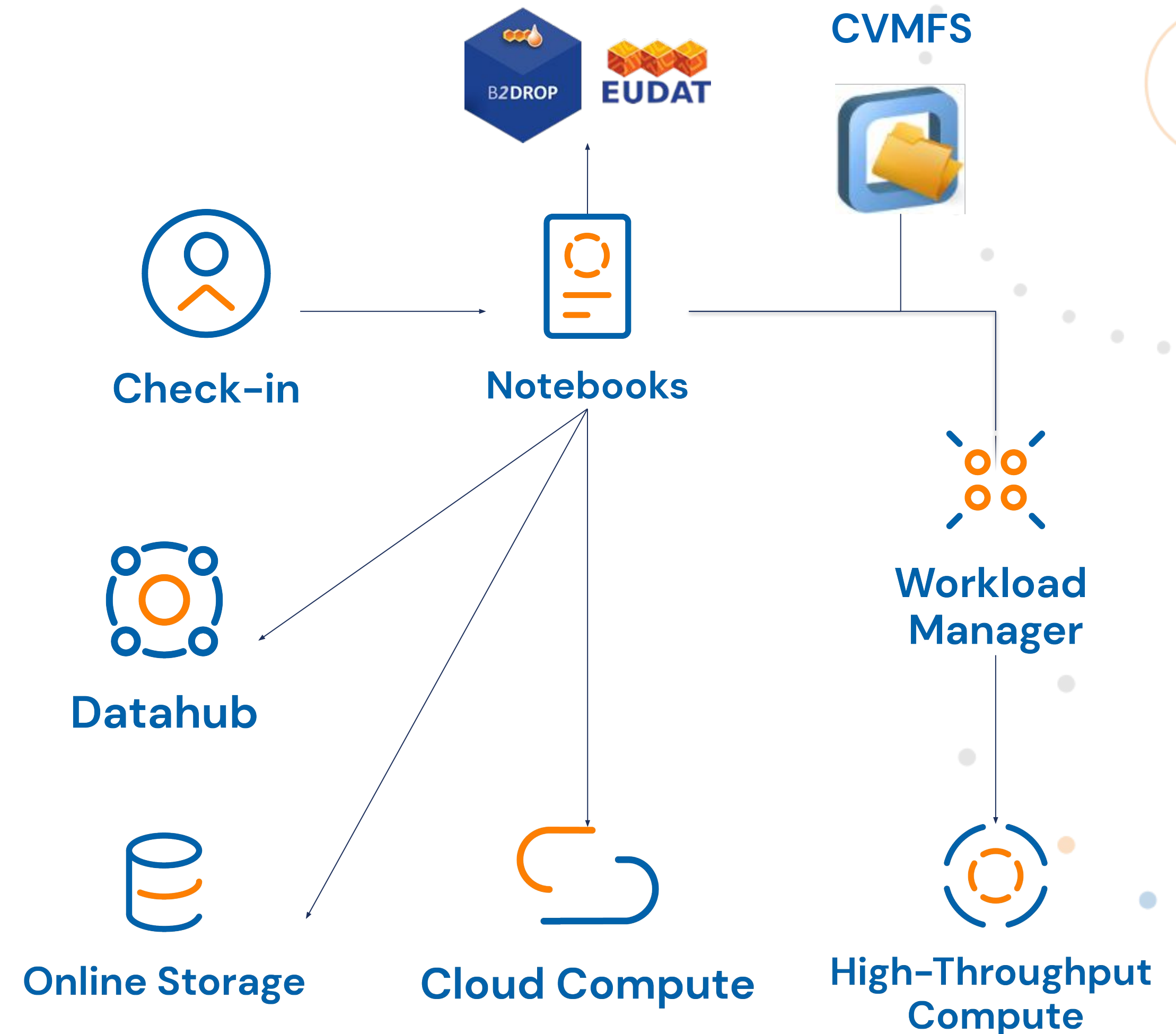
Main Features:

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

<https://notebooks.egi.eu>

The screenshot shows the EGI Notebooks website in a browser window. The browser tab is titled "EGI Notebooks" and the address bar shows "notebooks.egi.eu/hub/welcome". The page features the EGI logo in the top left and a "User Guide" link in the top right. The main heading is "Notebooks" with a notebook icon. Below the heading, there is a paragraph describing the service as a browser-based, scalable tool for interactive data analysis. A prominent blue button labeled "Continue with EGI Check-in" is centered on the page. At the bottom, there are logos for the European Union, EGI-ACE, and CESNET, along with text stating "Notebooks is a service provided by CESNET, co-funded by EGI-ACE." and links for "Privacy policy" and "Terms of use".

1. Runs on EGI Cloud providers (Cloud Compute / Online Storage)
2. Uses Check-in for authentication + access to other services
3. Makes user-level software available from CVMFS
 - DIRAC client for submission of jobs to Workload Manager / High Throughput Compute
 - Fedcloudclient for interaction with Cloud Compute
4. Transparent access to Datahub spaces
5. Access to 3rd party services: B2DROP





Starting your server

EGI Notebooks offers different computing environments:

- **Default:** Python, R, Julia, Octave and a wide range of data science libraries
- **MATLAB (Basic/Full):** Run MATLAB on EGI resources (requires a license!)

Server Options

- Default EGI environment – 6 GB RAM / 2 core

The Default notebook environment includes Python, R, Julia and Octave kernels

- MATLAB Environment (Basic) – 4GB RAM / 4 cores

The MATLAB environment (requires a [valid license](#)), includes Python and MATLAB kernels

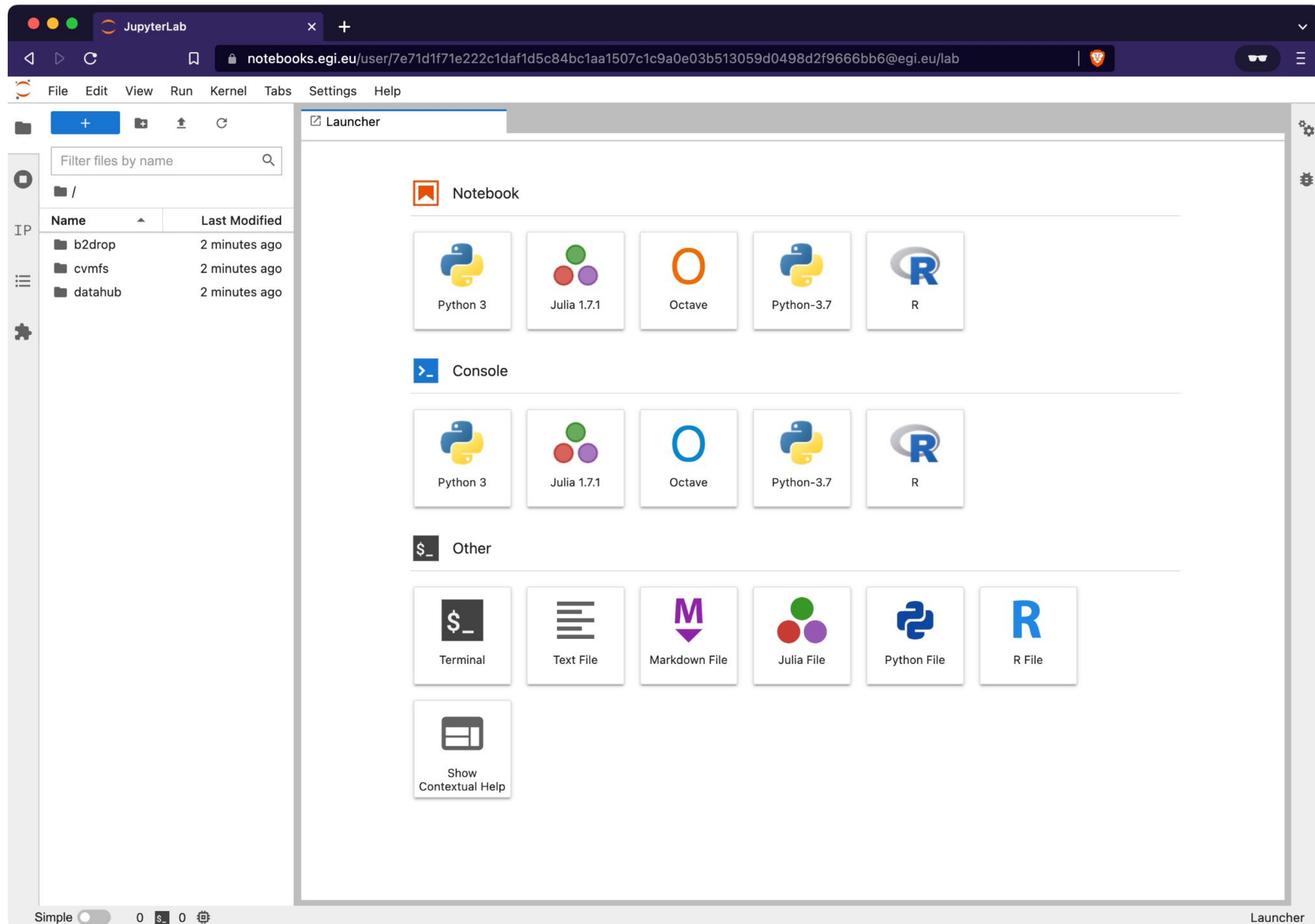
- MATLAB Environment (Full) – 4GB RAM / 4 cores

The MATLAB environment (requires a [valid license](#)), includes Python, MATLAB kernels and additional MATLAB packages

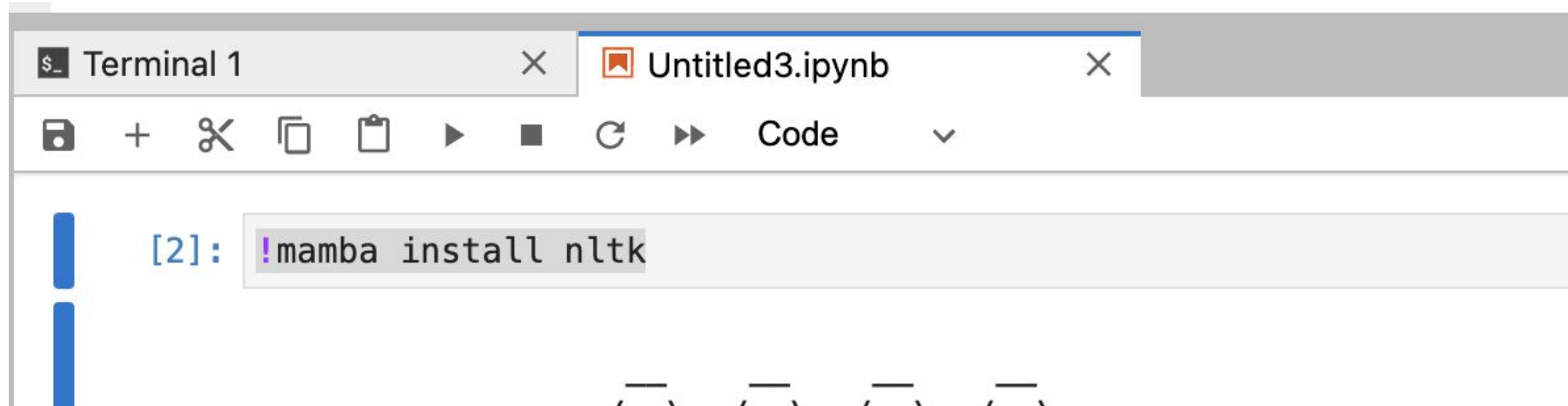
B2DROP connection

Start

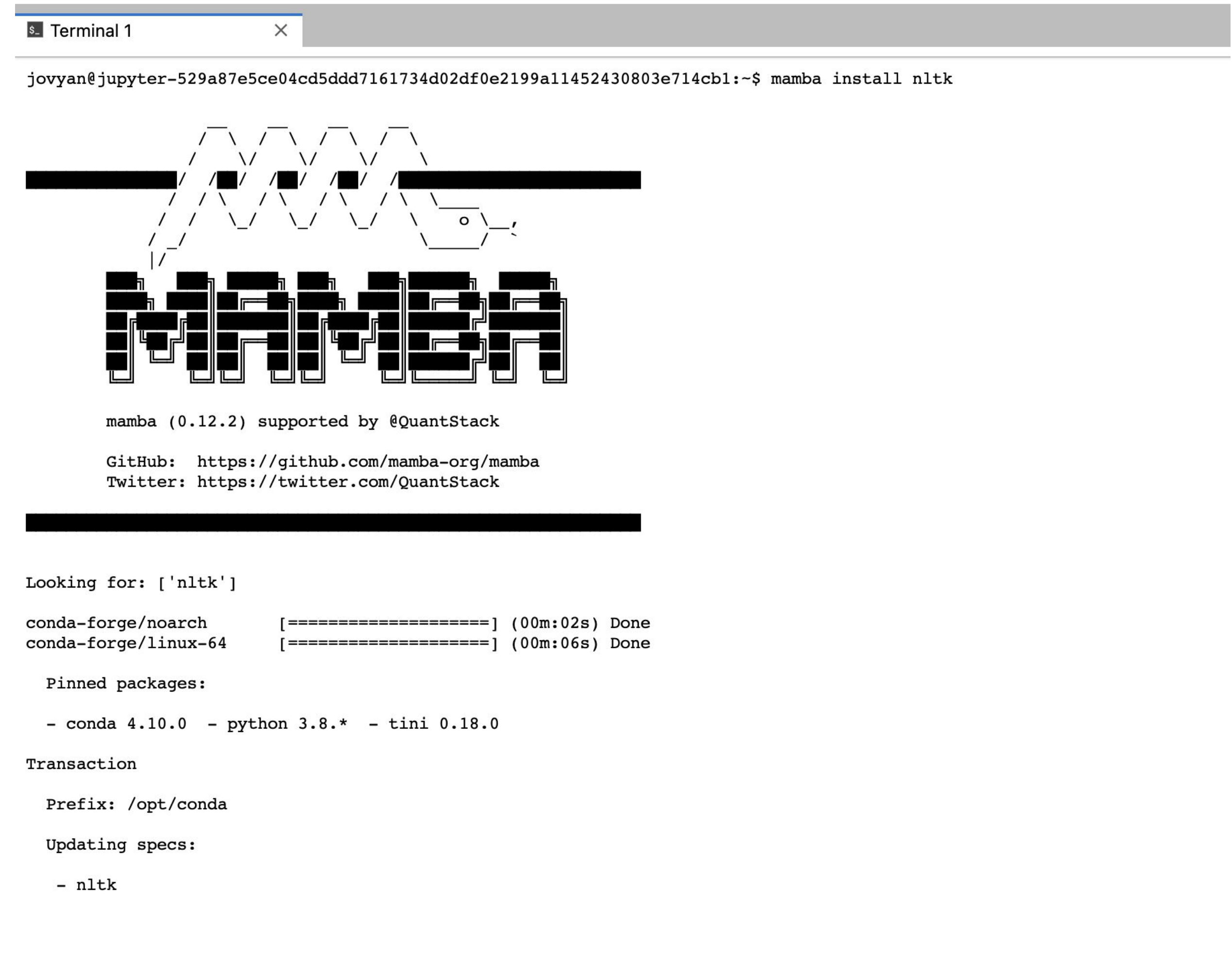
JupyterLab interface



- (Most) software in the environment is managed with conda (**mamba**)
- Adding packages can be done at any time with **mamba** or **pip**
- This is installed in /opt/conda which is wiped out for every new server session!
 - Include a installation cell on your notebooks to avoid not having the needed libraries
 - Ask us for adding them to the environment!



```
[2]: !mamba install nltk
```



```
Terminal 1
jovyan@jupyter-529a87e5ce04cd5ddd7161734d02df0e2199a11452430803e714cbl:~$ mamba install nltk

mamba (0.12.2) supported by @QuantStack

GitHub: https://github.com/mamba-org/mamba
Twitter: https://twitter.com/QuantStack

Looking for: ['nltk']

conda-forge/noarch      [=====] (00m:02s) Done
conda-forge/linux-64   [=====] (00m:06s) Done

Pinned packages:

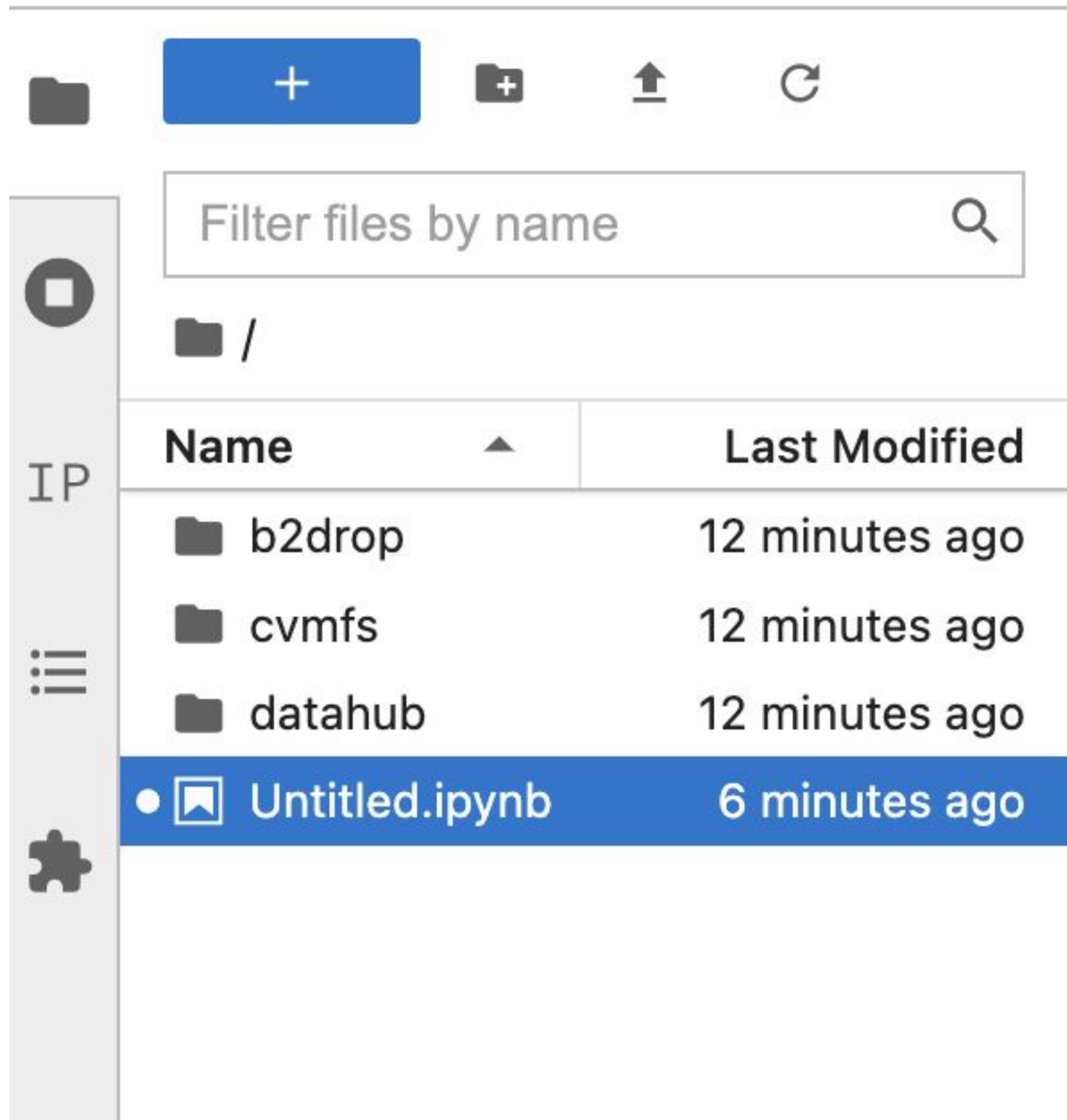
- conda 4.10.0 - python 3.8.* - tini 0.18.0

Transaction

Prefix: /opt/conda

Updating specs:

- nltk
```



- **Persistent home**
 - Can be used to store data (10GB limit)
 - Files will be kept even if the notebook server dies
- **nbgitpuller**
 - Get code from any git repository from a single URL
 - <https://hub.jupyter.org/nbgitpuller/link.html>
- **CVMFS**
 - Selected CVMFS repositories available
 - Easy to add community specific ones
- **DataHub**
 - Access to your accessible spaces in datahub
 - Share data and assign PIDs to shared spaces
- **B2DROP**
 - Access data stored in EUDAT's B2DROP service
- **External web server**
 -



Easily clone GitHub repositories into your home with a shareable link.

Try it: <https://go.egi.eu/x8ouj>

[JupyterHub](#) [Launch from Canvas](#) [Binder](#)

`user-redirect/git-pull?repo=https%3A%2F%2Fgithub.com%2FPetrP-training%2FSentinel&urlpath=lab%2Ftree%2FSentinel%2F&branch=main`

JupyterHub URL ✓
The JupyterHub to send users to. [nbgitpuller](#) must be installed in this hub.

Git Repository URL ✓ ✓
Use `main` instead of `master` for new GitHub repositories

File to open ✓
This file or directory from within the repo will open when user clicks the link.

Application to Open

- Classic Jupyter Notebook
- RetroLab
- JupyterLab
- RStudio
- Shiny
- Custom URL

- CernVM-FS (CVMFS) is implemented as a POSIX read-only file system in user space (a FUSE module).
- Files and directories are hosted on standard web servers and mounted in the universal namespace **/cvmfs**.

Main features

- Allows centrally managed software distribution across federated environments.
- Allows to make content available as a read-only file system that efficiently downloads and caches files on demand

Documentation

- <https://docs.egi.eu/users/compute/software-distribution/>

Service based on **Onedata** technology

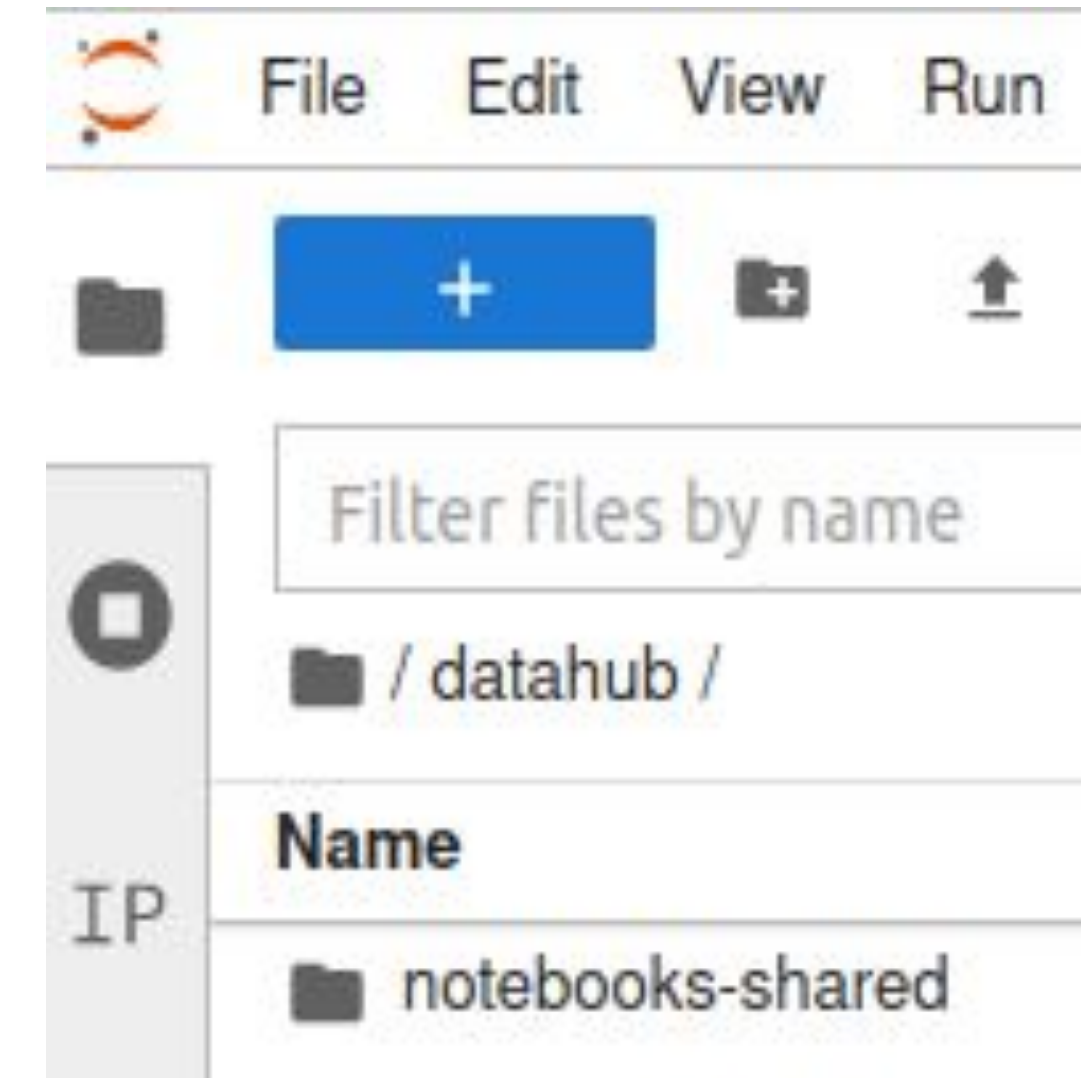
It allows **transparent data access** under a common namespace regardless of the location open access

Access restricted to members of a Virtual Organization (VO)

Data can be accessed via a **GUI** or **APIs**

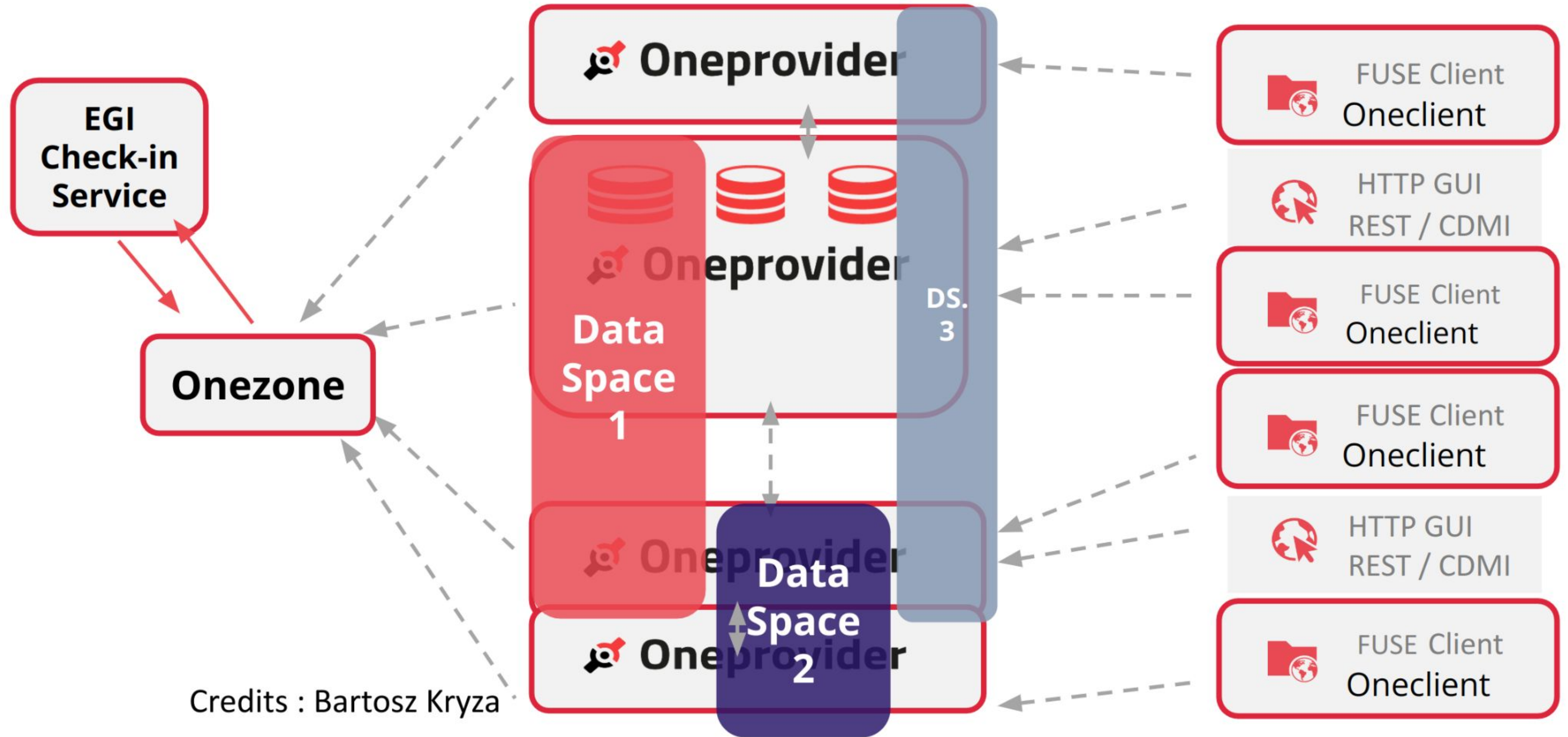
Allows **replication** of data from data providers for **resiliency** and **availability** purposes. Replication may take place either **on demand** or **automatically**.

Easy integration with other EGI components thanks to integration with **EGI Check-in** service, the EGI Authentication and Authorization Infrastructure (AAI)



ONEDATA

EGI DataHub – Architecture





B2DROP is EUDAT Sync & Share solution

- Based on NextCloud
- WebDav interface that can be accessed remotely

If you want to access your space from the EGI Notebooks:

- Go to [B2DROP security configuration](#)
- Create an app password
- Copy credentials into the EGI Notebooks
- Tick the “Remember B2DROP credentials” box to not keep these credentials in your settings

Server Options

- Default EGI environment – 6 GB RAM / 2 core
The Default notebook environment includes Python, R, Julia and Octave kernels
- MATLAB Environment (Basic) – 4GB RAM / 4 cores
The MATLAB environment (requires a [valid license](#)), includes Python and MATLAB kernels
- MATLAB Environment (Full) – 4GB RAM / 4 cores
The MATLAB environment (requires a [valid license](#)), includes Python, MATLAB kernels and additional MATLAB packages

B2DROP connection

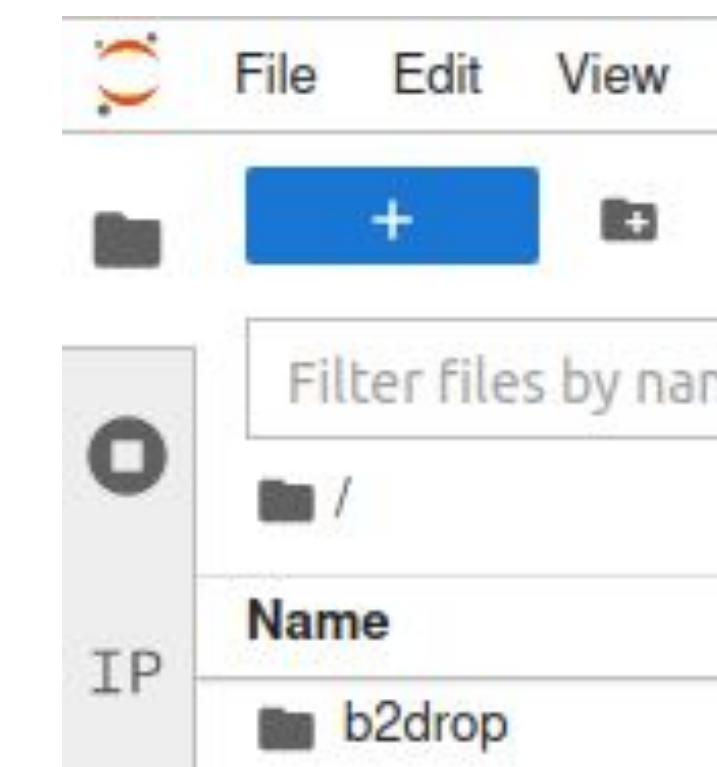
B2DROP app Username

Create new app password at [B2DROP security configuration](#)

B2DROP app Password

Remember B2DROP credentials

Start





Getting access with the EGI Notebooks in the Asia Pacific VO

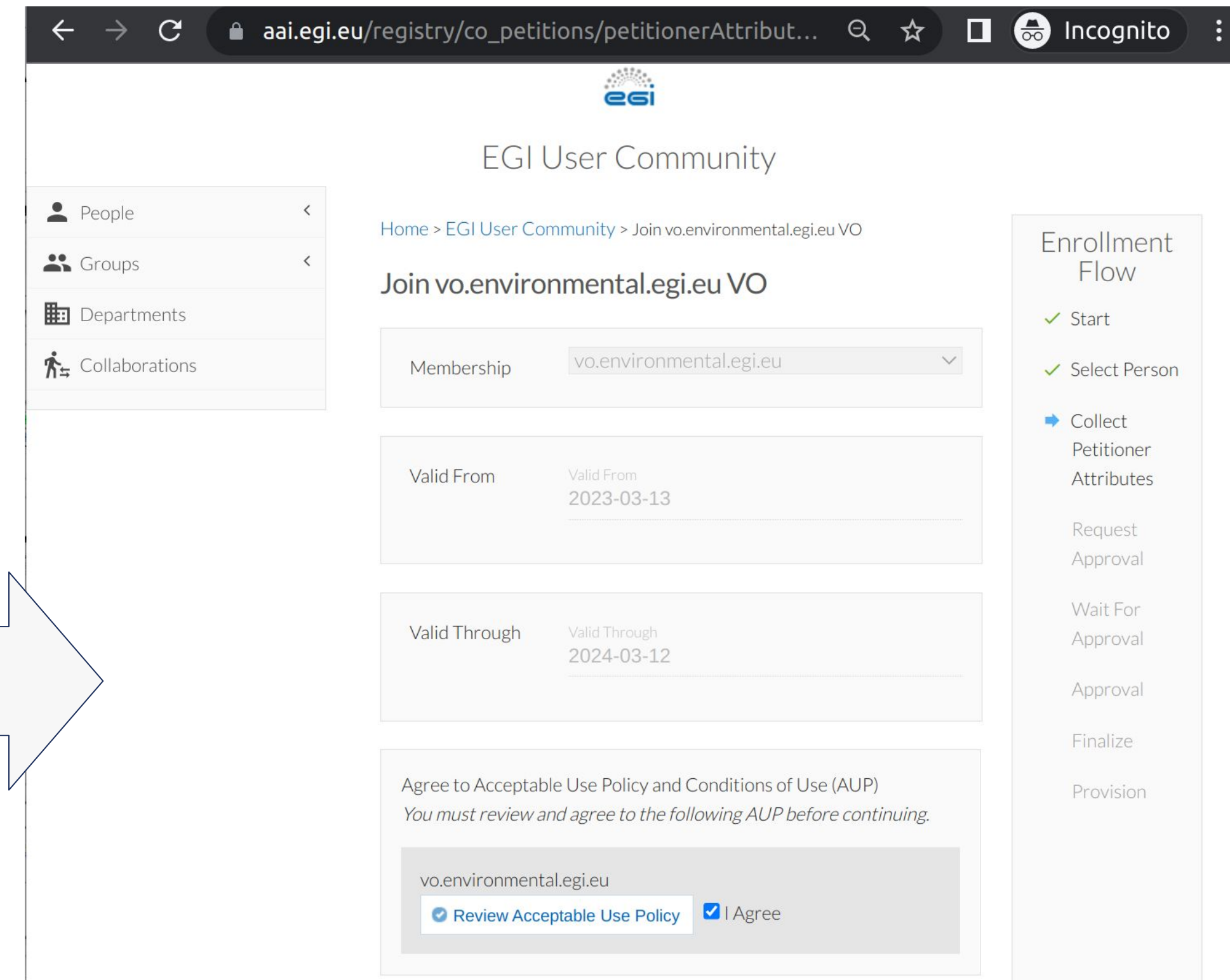
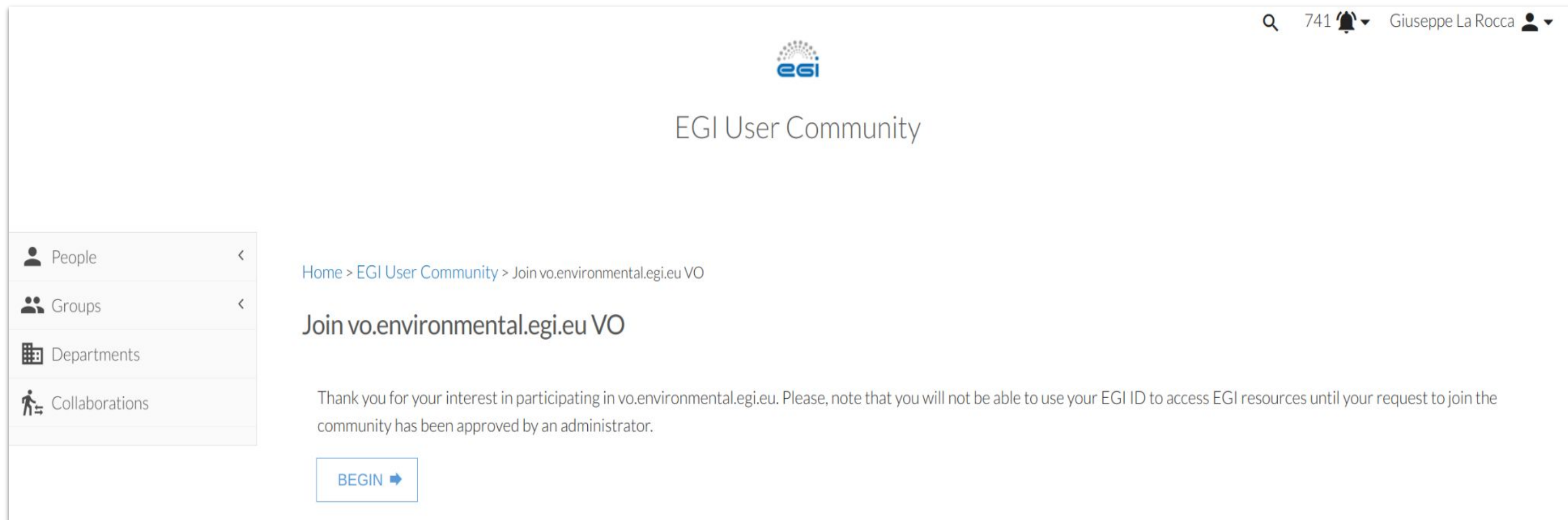
TLP: GREEN Limited disclosure



Sign-up and join the VO for the AP region

1. **Sign-up** (<https://aai.egi.eu/signup>) for an EGI Check-In account
2. **Subscribe** the **vo.environmental.egi.eu** Virtual Organisation (VO) by visiting the enrollment URL (<https://go.egi.eu/8Hspz>) with your EGI Check-In account.
 - The subscription requires approval from the VO Manager.

Hands On





Hands-on: Running your first notebooks

TLP: GREEN Limited disclosure

Your first notebook

This is a markdown cell, you can **format** text using [Markdown](#).

```
In [1]: a = 5
        b = 6
        a + b
```

Out[1]: 11

Variables

The variables defined in previous cells are available in following cells

```
In [2]: a = a + 1
        print(b)
        a
```

6
Out[2]: 6

Plotting

Output of cells can be more than just text

Code taken from [Matplotlib tutorial subplot example](#)

```
In [3]: %matplotlib inline
import matplotlib.pyplot as plt
import numpy as np

np.random.seed(19680801)
data = np.random.randn(2, 100)

fig, axs = plt.subplots(2, 2, figsize=(5, 5))
axs[0, 0].hist(data[0])
axs[1, 0].scatter(data[0], data[1])
axs[0, 1].plot(data[0], data[1])
axs[1, 1].hist2d(data[0], data[1])

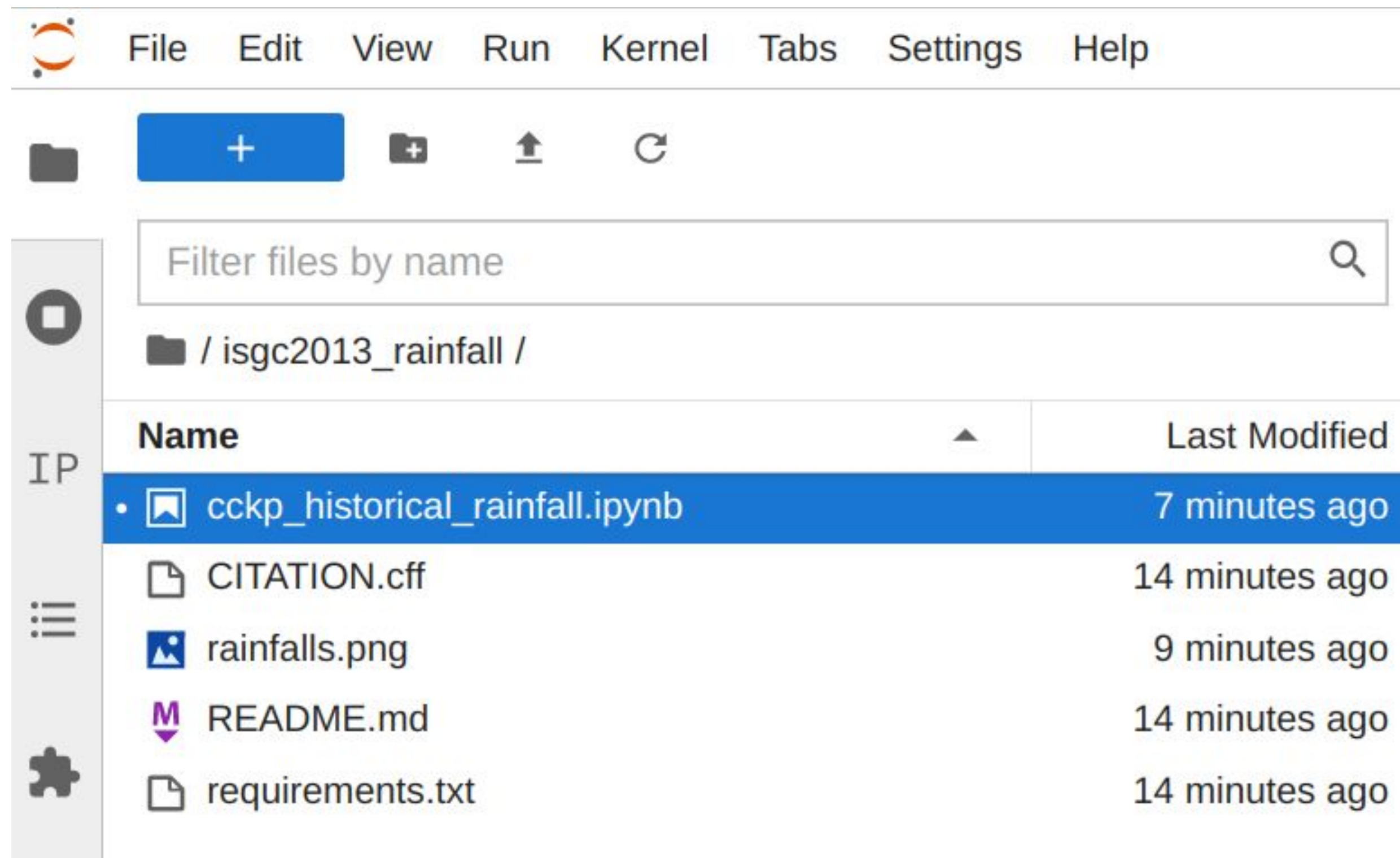
plt.show()
```

To launch the notebook \Rightarrow go.egi.eu/xABiu

00-first-notebook.ipynb

- GitHub repo for this notebook
- Notebook fetched with the [nbgitpuller](#)





cckp_historical_rainfall.ipynb

- Monthly and annual precipitation by country for the period 1961–1999.
- Values are in millimeters (mm).
- **Datasets are stored in CVMFS**

To launch the notebook ⇒ <https://go.egi.eu/KWIsA>

See: [“How to make datasets and application in CVMFS”](#)





Coffee break



Check-in provides authentication, authorization and user management for the EOSC Compute Platform

Standards based:

- SAML 2.0 / OpenID Connect 1.0 / OAuth 2.0 / LDAP

Interoperable:

- AARC and EOSC AAI compliant
- Support for legacy X.509 services via MasterPortal

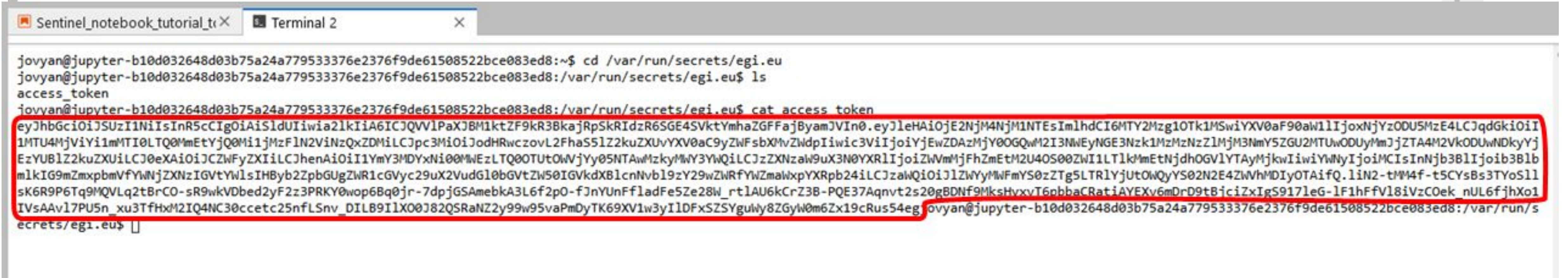
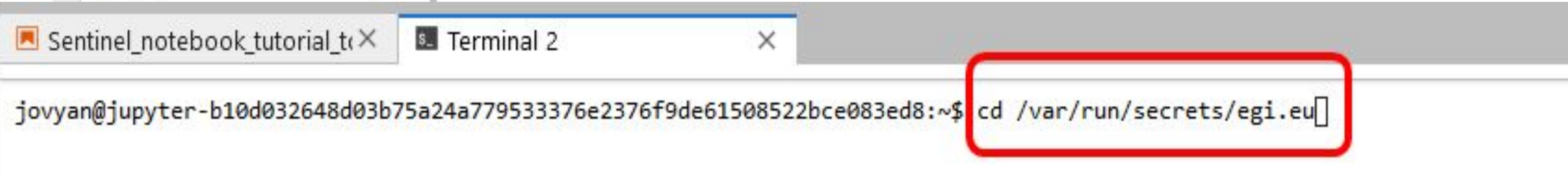
Community management:

- Comanage and Perun supported
- Other Community AAls pluggable





Where to found the token in the EGI Notebooks



```
04-check-in.ipynb Python 3
```

Interacting with the infrastructure with your own credentials

In this example we use the token to get our own information from Check-in

```
[9]: import requests
```

```
[10]: with open("/var/run/secrets/egi.eu/access_token") as f:
      access_token = f.read()

      print("Got an check-in access token: %s..." % access_token[:10])
```

Got an check-in access token: eyJhbGciOi...

```
[11]: userinfo_url = 'https://aai.egi.eu/auth/realms/egi/protocol/openid-connect/userinfo'

      r = requests.get(userinfo_url,
                       headers={'Authorization': 'Bearer %s' % access_token})

      userinfo = r.json()

      print("My user info:")
      print("Name: %s" % userinfo['name'])
      print("email: %s" % userinfo['email'])
```

My user info:
Name: Enol Fernandez
email: enol.fernandez@egi.eu

Hands On

The screenshot shows a Jupyter Notebook titled "Sentinel notebook tutorial_token.ipynb" in a Python 3 environment. The left sidebar displays a file explorer for the directory "/ EGI_tutorial / Sentinel /". The notebook content includes the following code cells:

```
[2]: import requests
```

ACCES TOKEN

can be found in `/var/run/secrets/egi.eu`

check properties of your token at jwt.io

```
[3]: access_token = "eyJhbGciOiJIUzU1IiwiaXN5IjoiInR5cCI6ImF1dGUiIiwia2kiOiIiA6ICJQVWVlPaXJBm1ktZF9kr3BkaJRpSkRIIdzR6SGE4SVktYmhaZGFFajByamJVIj0.eyJleHAiOiJlE2NjM4NjM1NTESIm1hdCI6MTY2MzgzOTk1MSwiYXV0aF90aW1" < >
```

List of available [datasets](#) that can be checked on the [map](#)

```
[4]: dataset = "S2A_MSIL1C_20220816T100611_N0400_R022_T33UVR_20220816T135125.zip"
```

```
[5]: %%time
r = requests.get("https://ip-147-251-21-170.flt.cloud.muni.cz/api/data/" + dataset, headers={"authorization": "Bearer {}".format(access_token)}, stream=True)
with open("sentinel_image.zip", 'wb') as fd:
    for chunk in r.iter_content(chunk_size=128):
        fd.write(chunk)
```

CPU times: user 29.8 s, sys: 931 ms, total: 30.7 s
Wall time: 31.7 s

```
[6]: %%time
from zipfile import ZipFile

with ZipFile('sentinel_image.zip', 'r') as zipObj:
    # Extract all the contents of zip file in different directory
    zipObj.extractall('images')
    print('File is unzipped in images folder')
```

File is unzipped in images folder
CPU times: user 1.53 s, sys: 1.21 s, total: 2.74 s
Wall time: 7.67 s

```
[7]: import rasterio
import rasterio.plot as plot
```

Sentinel_notebook_tutorial_token

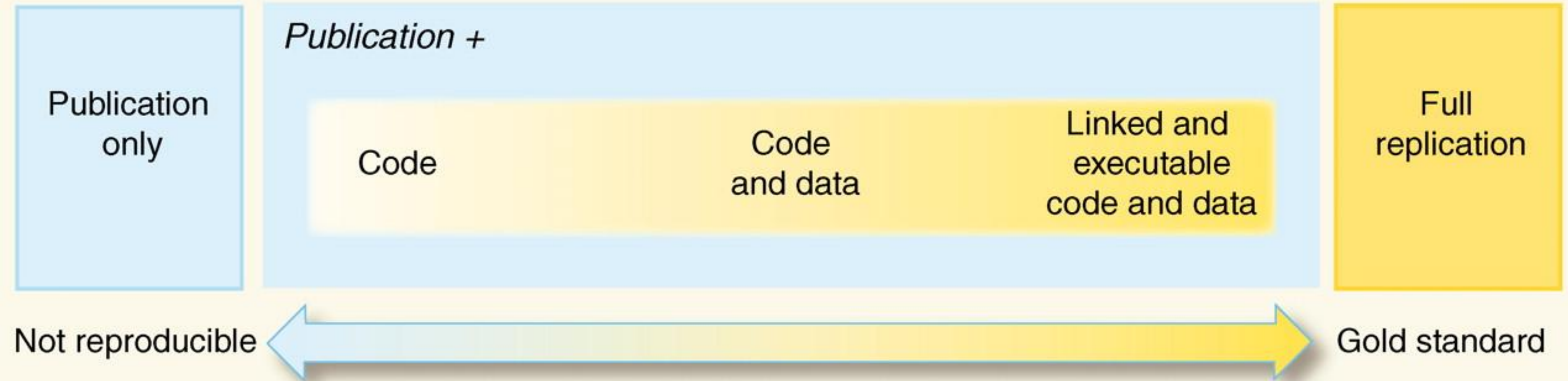
Hands On



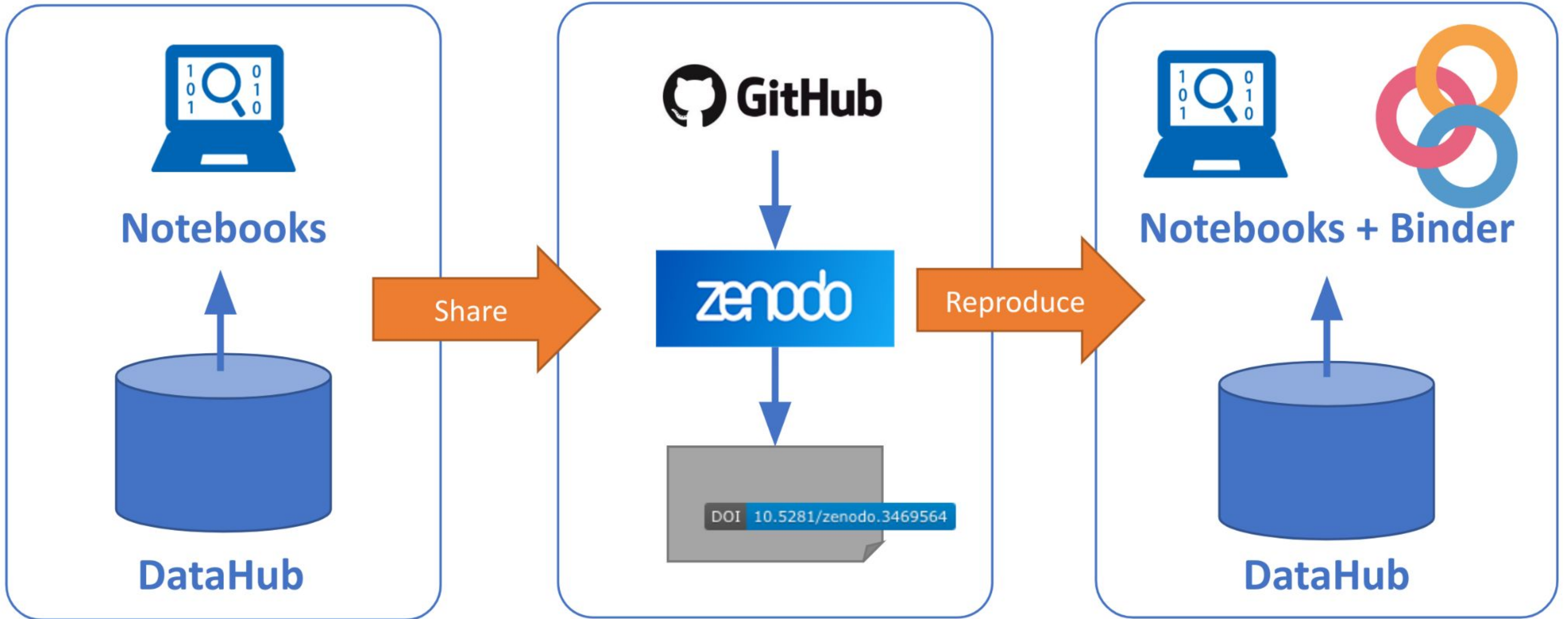
Approach to reproducible Open Science with EGI and EOSC

TLP: GREEN Limited disclosure

Reproducibility Spectrum



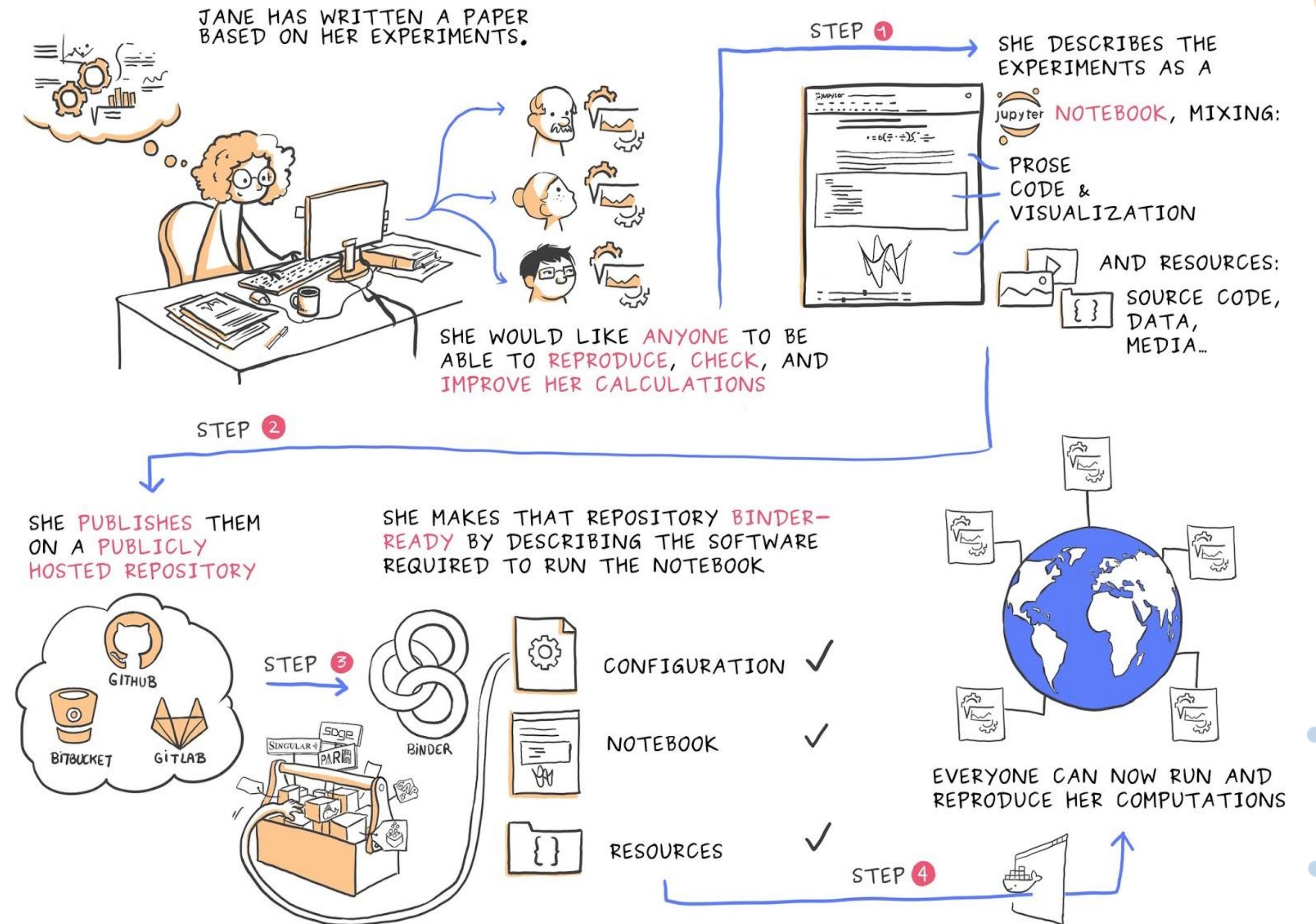
Peng, Science, 2011



Binder: reproducing execution environments

An open-source web application to turn repositories in interactive notebooks

It uses modern technology in cloud orchestration (Kubernetes), interactive computing (Jupyter), scientific computing (the open-science ecosystem)





EGI Replay



BinderHub hosted by EGI

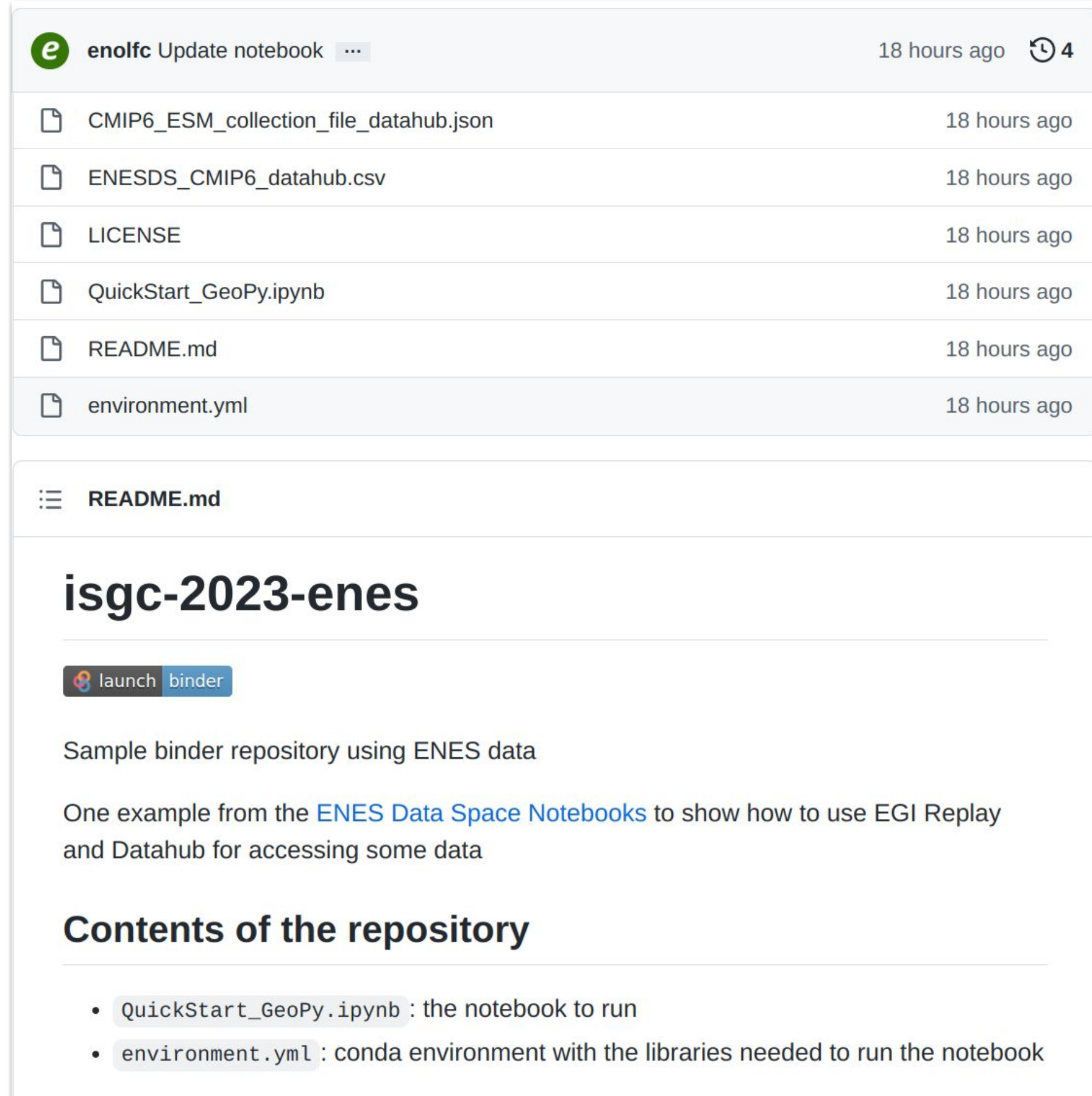
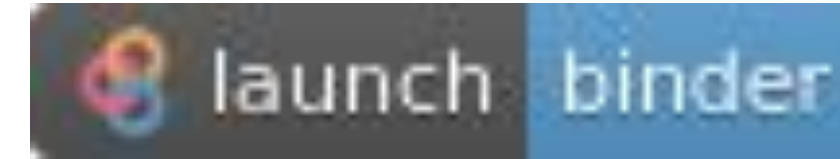
- Offered 'as Service'
- Same access conditions as EGI Notebooks

Main Features:

- Use any binder-compatible repository
- Reproduce your notebooks with access to EGI resources (e.g. DataHub)
- No hard limits on sessions duration, customisable resource limits for users/communities

The screenshot shows a web browser window with the URL `replay.notebooks.egi.eu/hub/hub/login?next=...`. The page features the EGI logo and the title "Replay". The main text describes the service: "Replay offers an easy place to reproduce and share notebooks. It allows users to replay complex calculations, simulations, and visualisations scenarios by importing Notebooks and their runtime environment and share them with a single link. Replay works well with EGI Notebooks: use-cases include workshops, scientific workflows and streamline sharing among teams." A blue button labeled "Continue with EGI Check-in" is prominently displayed. Below this, there is a note: "Access requires a valid EGI account and enrolling to the vo.notebooks.egi.eu VO." Further down, it states: "Replay is based on Jupyter Binder and runs on the EGI cloud service. You can learn more at our documentation." The final paragraph mentions: "This Replay instance is operated by and uses resources from CESNET. User communities/advanced users can have their customised Binder instance. EGI offers consultancy and support, as well as can operate the setup. Email support _at_ egi.eu to request a community instance."

From <https://github.com/enolfc/isgc-2023-enes>



The screenshot shows a GitHub repository page for 'enolfc Update notebook'. The repository contains several files, all updated 18 hours ago: CMIP6_ESM_collection_file_datahub.json, ENESDS_CMIP6_datahub.csv, LICENSE, QuickStart_GeoPy.ipynb, README.md, and environment.yml. Below the file list is the README.md content, which includes a 'launch binder' button, a description of the repository as a sample binder repository using ENES data, and a list of repository contents: QuickStart_GeoPy.ipynb (the notebook to run) and environment.yml (conda environment with the libraries needed to run the notebook).

QuickStart_GeoPy

- Calculate the number of summer days in the specific location
- From the ENES Data Spaces
- **Datasets are stored in DataHub**

See: [“How to make your datasets available in DataHub”](#)

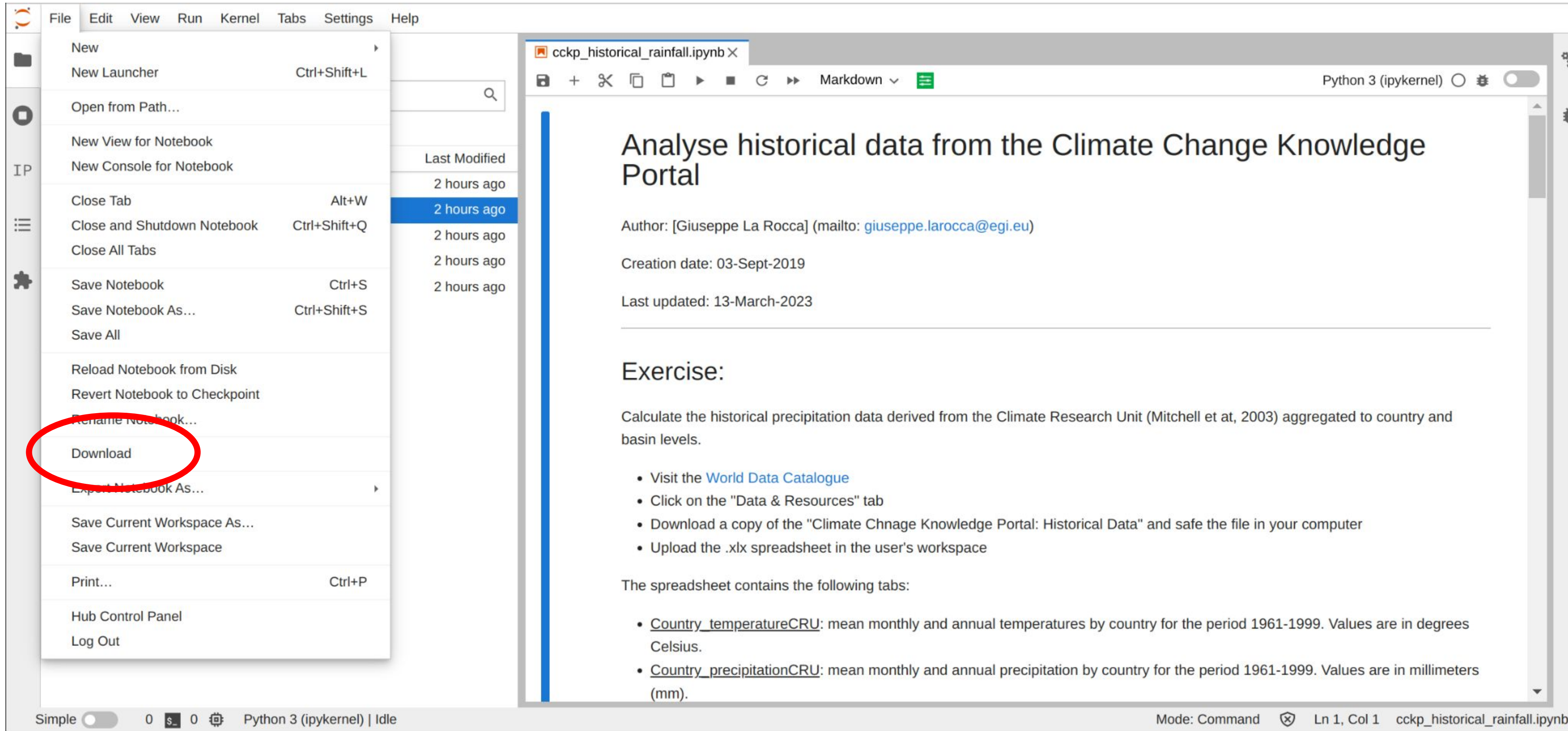




How to make notebooks Reproducible and Shareable

TLP: GREEN Limited disclosure

1. Create your first sharable notebook



The screenshot shows a Jupyter Notebook interface. The 'File' menu is open, and the 'Download' option is circled in red. The notebook content is as follows:

Analyse historical data from the Climate Change Knowledge Portal

Author: [Giuseppe La Rocca] (mailto: giuseppe.larocca@egi.eu)

Creation date: 03-Sept-2019

Last updated: 13-March-2023

Exercise:

Calculate the historical precipitation data derived from the Climate Research Unit (Mitchell et al, 2003) aggregated to country and basin levels.

- Visit the [World Data Catalogue](#)
- Click on the "Data & Resources" tab
- Download a copy of the "Climate Chnage Knowledge Portal: Historical Data" and safe the file in your computer
- Upload the .xlsx spreadsheet in the user's workspace

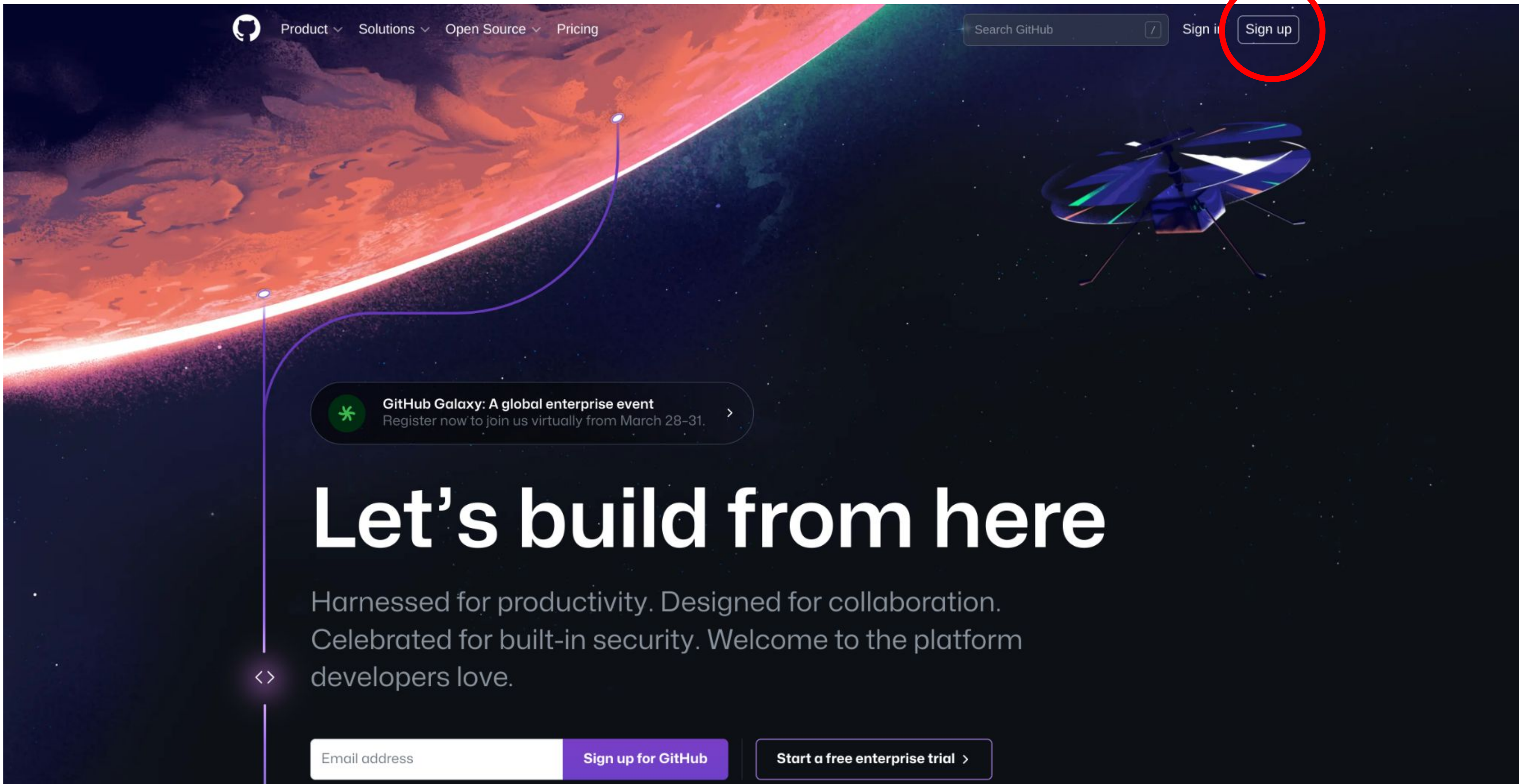
The spreadsheet contains the following tabs:

- Country_temperatureCRU: mean monthly and annual temperatures by country for the period 1961-1999. Values are in degrees Celsius.
- Country_precipitationCRU: mean monthly and annual precipitation by country for the period 1961-1999. Values are in millimeters (mm).

Download your first notebook in your laptop (*.ipynb)

2. Getting a GitHub account

Sign-up if you don't have an account already

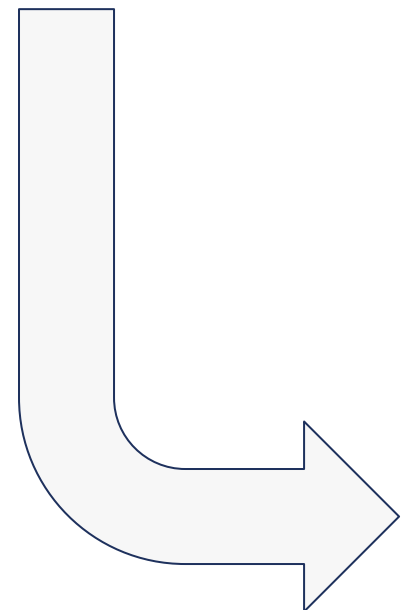


The screenshot shows the GitHub homepage with a dark background and a red planet in the upper left. The navigation bar includes the GitHub logo, 'Product', 'Solutions', 'Open Source', and 'Pricing'. A search bar and 'Sign in' link are on the right, with the 'Sign up' button circled in red. A banner for 'GitHub Galaxy: A global enterprise event' is visible. The main heading reads 'Let's build from here' with a subtext: 'Harnessed for productivity. Designed for collaboration. Celebrated for built-in security. Welcome to the platform developers love.' At the bottom, there is an email input field, a 'Sign up for GitHub' button, and a 'Start a free enterprise trial' button.

3. Creating a new GitHub repository

1.

The screenshot shows the GitHub user profile for 'glarocca'. A red circle highlights the 'New' button in the top right corner of the profile header. Below the header, there is a search bar for repositories and a list of top repositories.



Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository.](#)

Repository template
Start your repository with a template repository's contents.

Owner * / **Repository name *** ✓

Great repository names is available. Need inspiration? How about [congenial-octo-train?](#)

Description (optional)

Public
Anyone on the internet can see this repository. You choose who can commit.

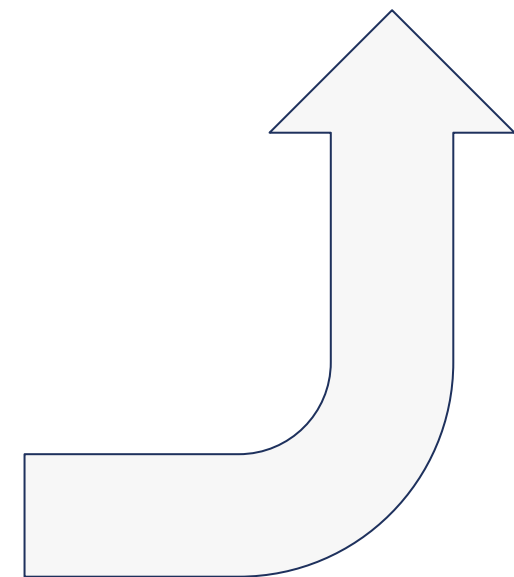
Private
You choose who can see and commit to this repository.

Initialize this repository with:
Skip this step if you're importing an existing repository.

Add a README file
This is where you can write a long description for your project. [Learn more.](#)



2.


The screenshot shows the 'Add file' dropdown menu on a GitHub repository page. The 'Add file' button is highlighted with a red box. Below it, the 'Clone' section is visible, with the 'HTTPS' option selected. The URL 'https://github.com/glarocca/isgc2013_rainfall' is highlighted with a red box, and a copy icon is visible to its right. Below the URL, there is a note: 'Use Git or checkout with SVN using the web URL.' The 'Download ZIP' option is also visible.







4. Adding dependencies

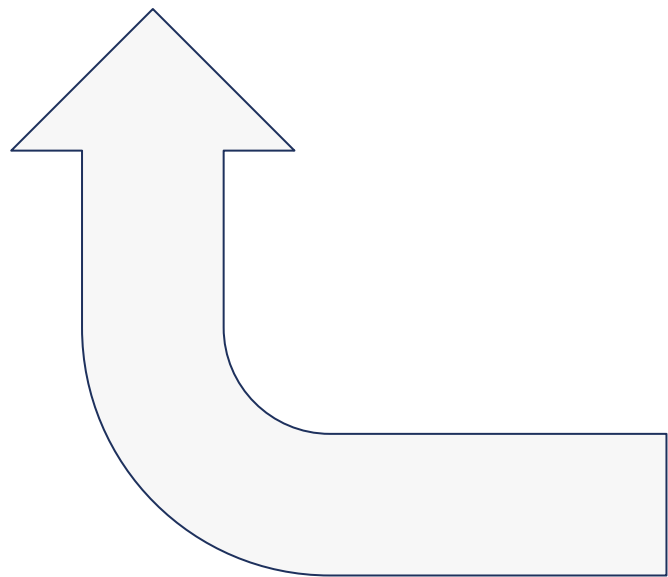
Create the 'requirements.txt' file in your repo

 **glarocca** Update requirements.txt Latest commit 7b83032 yesterday  History

 1 contributor

3 lines (3 sloc) | 23 Bytes [Raw](#) [Blame](#)    

```
1 pandas
2 matplotlib
3 xlrd
```



Import necessary libraries

```
[1]: import pandas as pd
      from pandas import DataFrame
      import matplotlib.pyplot as plt
```

Provide the ISO_3DIGIT of the country you are interested to analyse

```
[2]: ISO_3DIGIT="ITA"
```

Load historical datasets from local and create a DataFrame object

```
[3]: raw_data = pd.read_excel('/cvmfs/notebooks.egi.eu/isgc-2023/cckp/cckp_historical_data_0.xls', sheet_name='Country_precipitationCRU')
```


5. Executing Notebooks in Replay

<https://replay.notebooks.egi.eu/>

Build and launch a repository

GitHub repository name or URL

GitHub

Git ref (branch, tag, or commit)

Path to a notebook file (optional)

Copy the URL below and share your Binder with others:

Expand to see the text below, paste it into your README to show a binder badge:

```
Already built! Launching
Build logs view raw hide
Found built image, launching...
Launching server...
Server requested
2023-03-13T16:09:49.236416Z [Normal] Successfully assigned binder/jupyter-glarocca5 to k8s-w-ingress
2023-03-13T16:09:52Z [Normal] Container image "jupyterhub/k8s-network-tools:2.0.0" already present on machine
2023-03-13T16:09:53Z [Normal] Created container block-cloud-metadata
2023-03-13T16:09:53Z [Normal] Started container block-cloud-metadata
2023-03-13T16:09:55Z [Normal] Pulling image "docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8a0018cd410f"
2023-03-13T16:10:56Z [Normal] Successfully pulled image "docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8a0018cd410f" in 1m1.414306408s
2023-03-13T16:10:57Z [Normal] Created container notebook
2023-03-13T16:10:58Z [Normal] Started container notebook
2023-03-13T16:10:59Z [Normal] Container image "eginotebooks/oneclient-sidecar:sha-8172f71" already present on machine
2023-03-13T16:10:59Z [Normal] Created container oneclient
2023-03-13T16:11:00Z [Normal] Started container oneclient
2023-03-13T16:13:01Z [Normal] Stopping container oneclient
```

```
Build logs view raw hide
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Pushing image
Successfully pushed docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8a0018cd410fBuilt image, launching...
Launching server...
Server requested
2023-03-13T16:09:49.236416Z [Normal] Successfully assigned binder/jupyter-glarocca5 to k8s-w-ingress
2023-03-13T16:09:52Z [Normal] Container image "jupyterhub/k8s-network-tools:2.0.0" already present on machine
2023-03-13T16:09:53Z [Normal] Created container block-cloud-metadata
2023-03-13T16:09:53Z [Normal] Started container block-cloud-metadata
2023-03-13T16:09:55Z [Normal] Pulling image "docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8a0018cd410f"
```

Unable to launch the server

<https://replay.notebooks.egi.eu/>

```
Build logs view raw 
Step 49/51 : COPY /repo2docker-entrypoint /usr/local/bin/repo2docker-entrypoint
---> 3dda7421129e
Step 50/51 : ENTRYPOINT ["/usr/local/bin/repo2docker-entrypoint"]
---> Running in 96ff2f556184
Removing intermediate container 96ff2f556184
---> b9b8b51524cc
Step 51/51 : CMD ["jupyter", "notebook", "--ip", "0.0.0.0"]
---> Running in 2be6bdc29474
Removing intermediate container 2be6bdc29474
---> 46d5422a2fab
{"aux": {"ID": "sha256:46d5422a2fab7693b3701aca6ef150b8298e0b31fb150c1eb3905585f3590a9"}}Successfully built 46d5422a2fab
Successfully tagged docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:e842592725541fa91123f33e9ddbdfdef1dcd1d73
Pushing image
Successfully pushed docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:e842592725541fa91123f33e9ddbdfdef1dcd1d73Built image, launching...
Launching server...
Launch attempt 1 failed, retrying...
Launch attempt 2 failed, retrying...
Launch attempt 3 failed, retrying...
User glarocca5 already has a running server.
```

Only 1 server per user is allowed!

Stop server at: <https://replay.notebooks.egi.eu/hub/hub/home>

6. Adding Binder badge in the GitHub repository

isgc2013_rainfall / README.md in main

Cancel changes

<> Edit file Preview

Spaces 2 Soft wrap

```
1 # isgc2013_rainfall
2
3 This is my first shared notebook
4
5 [!][Binder](https://replay.notebooks.egi.eu/badge_logo.svg)](https://replay.notebooks.egi.eu/v2/gh/glarocca/ISGC2023_rainfall.git/HEAD)
6
7
```

7. Create a new Release for the repository

main 1 branch 0 tags

Go to file Add file <> Code

About

No description, website, or topics provided.

Readme 0 stars 1 watching 0 forks

glarocca Update requirements.txt	7b83032 23 minutes ago	12 commits
README.md	Update README.md	36 minutes ago
cckp_historical_data_0.xls	Add files via upload	45 minutes ago
cckp_historical_rainfall.ipynb	Update cckp_historical_rainfall.ipynb	27 minutes ago
requirements.txt	Update requirements.txt	23 minutes ago

README.md

Releases

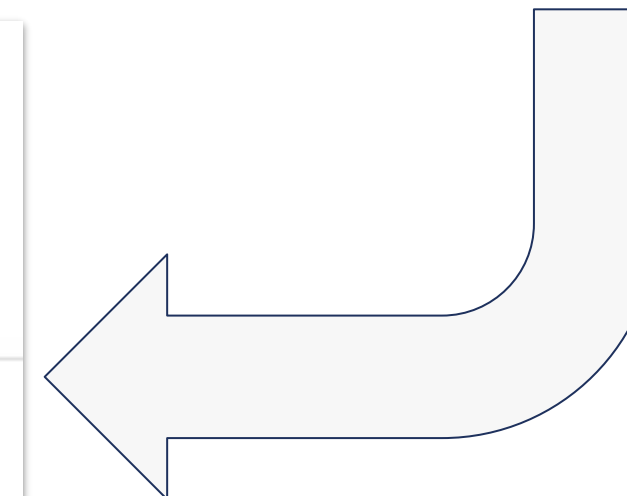
No releases published
[Create a new release](#)

Releases Tags

v0.1 Target: main

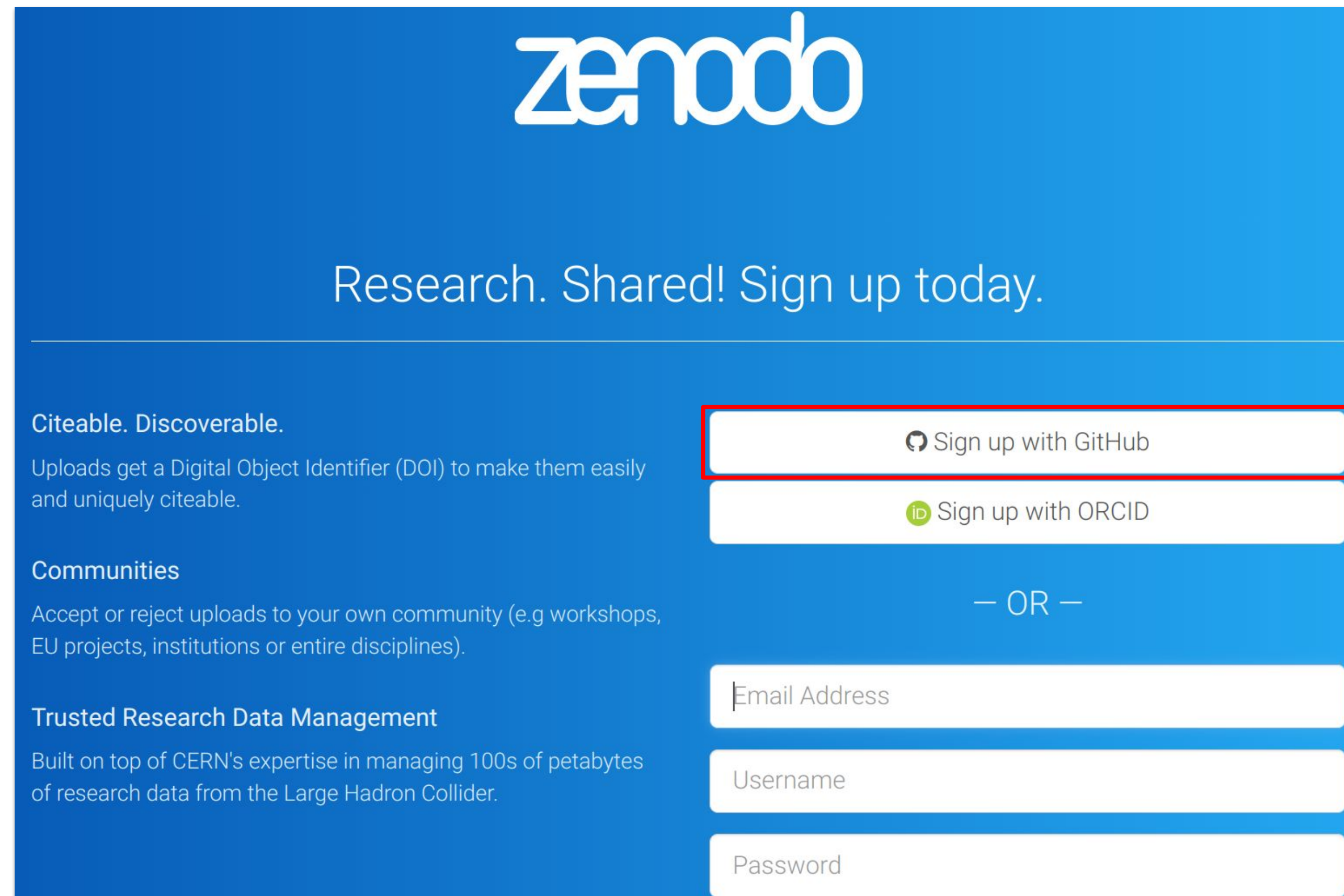
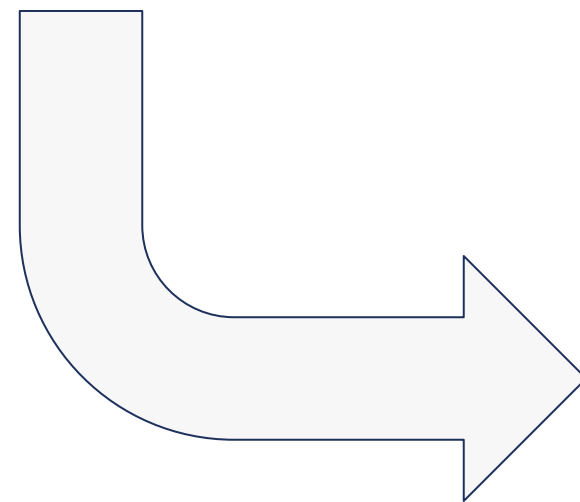
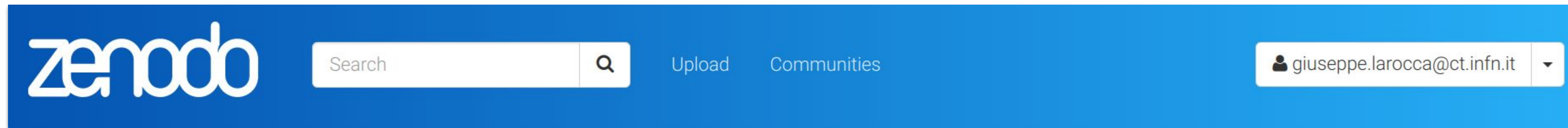
Excellent! This tag will be created from the target when you publish this release.

ISGC2023_rainfall



8. Making repository citable

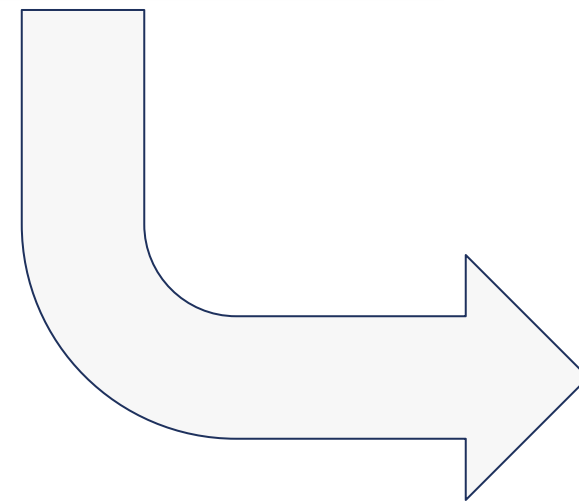
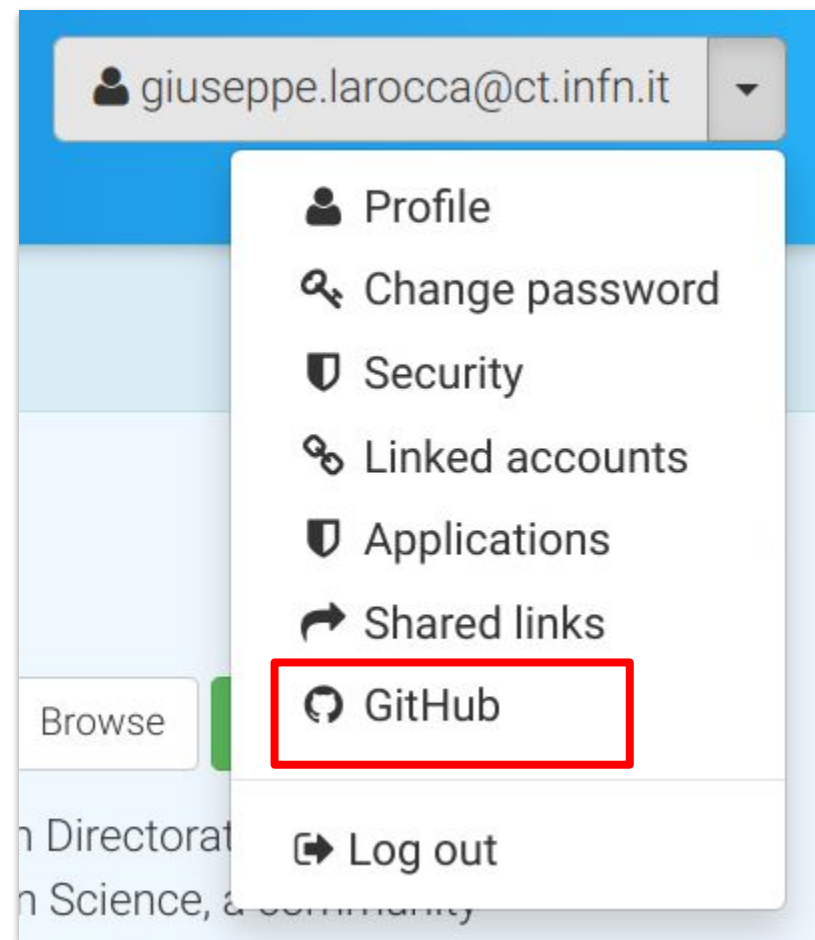
<https://zenodo.org/>



The main content area of the Zenodo website, featuring the Zenodo logo at the top. Below the logo is the text 'Research. Shared! Sign up today.' followed by a horizontal line. The page is divided into two columns. The left column contains three sections: 'Citeable. Discoverable.' with a subtext 'Uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable.', 'Communities' with a subtext 'Accept or reject uploads to your own community (e.g workshops, EU projects, institutions or entire disciplines).', and 'Trusted Research Data Management' with a subtext 'Built on top of CERN's expertise in managing 100s of petabytes of research data from the Large Hadron Collider.' The right column contains a sign-up form. At the top of the form are two buttons: 'Sign up with GitHub' (highlighted with a red border) and 'Sign up with ORCID'. Below these buttons is the text '— OR —'. The form includes three input fields: 'Email Address', 'Username', and 'Password'.



9. Synch Zenodo with repository



GitHub Repositories

(updated 50 seconds ago) [Sync now ...](#)

Get started

- 1 Flip the switch**

Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.

ON
- 2 Create a release**

Go to GitHub and [create a release](#). Zenodo will automatically download a .zip-ball of each new release and register a DOI.
- 3 Get the badge**

After your first release, a DOI badge that you can include in GitHub README will appear next to your repository below.

DOI [10.5281/zenodo.8475](#)
(example)

Enabled Repositories

glarocca/isgc2013_rainfall	<input checked="" type="checkbox"/>
--	-------------------------------------

10. Get Zenodo DOI

Click on the Enabled Repository

GitHub Repositories (updated 50 seconds ago) Sync now ...

Get started

- 1 Flip the switch**

Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.

ON
- 2 Create a release**

Go to GitHub and [create a release](#). Zenodo will automatically download a .zip-ball of each new release and register a DOI.
- 3 Get the badge**

After your first release, a DOI badge that you can include in GitHub README will appear next to your repository below.

DOI [10.5281/zenodo.8475](#)
(example)

Enabled Repositories

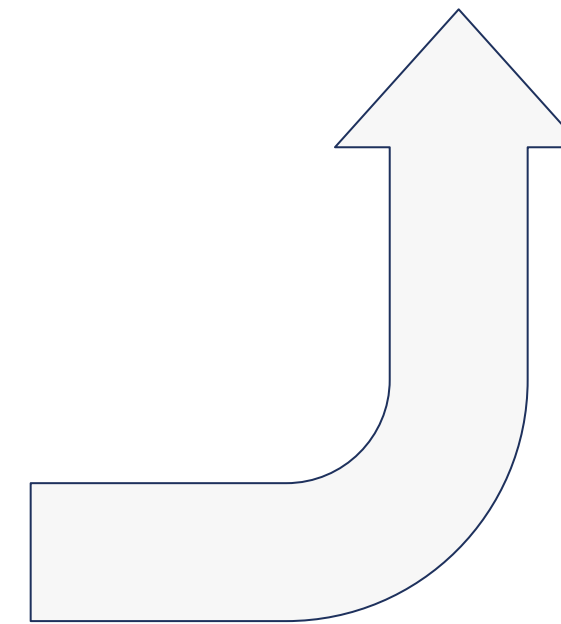
[glarocca/isgc2013_rainfall](#) ON

glarocca/isgc2013_rainfall ON

DOI [10.5281/zenodo.7729979](#)

GitHub / Releases

- v0.1 glarocca/isgc2013_rainfall: ISGC2023_rainfall Published
DOI: [10.5281/zenodo.7729979](#)
3 minutes ago
- ISGC2023_rainfall



11. Add Zenodo badge in repository

Click on the DOI icon to view details

glarocca/isgc2013_rainfall ON

DOI 10.5281/zenodo.7729979

GitHub / Releases Create release ...

v0.1 glarocca/isgc2013_rainfall: ISGC2023_rainfall ✓ Published
 DOI: 10.5281/zenodo.7729979 3 minutes ago
 ISGC2023_rainfall

DOI Badge

This badge points to the latest released version of your repository. If you want a DOI badge for a specific release, please follow the DOI link for one of the specific releases and grab badge from the archived record.

Markdown

```
[![DOI](https://zenodo.org/badge/613453962.svg)](https://zenodo.org/badge/613453962)
```

reStructuredText

```
.. image:: https://zenodo.org/badge/613453962.svg
   :target: https://zenodo.org/badge/latestdoi/613453962
```

HTML

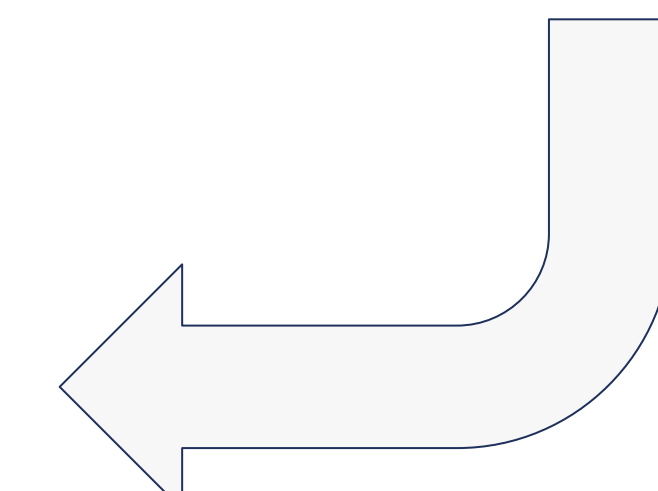
```
<a href="https://zenodo.org/badge/latestdoi/613453962">
```

Image URL

```
https://zenodo.org/badge/613453962.svg
```

Target URL

```
https://zenodo.org/badge/latestdoi/613453962
```



isgc2013_rainfall / README.md in main

Edit file Preview Spaces 2

```
1 # isgc2013_rainfall
2
3 This is my first shared notebook
4
5 [![Binder](https://replay.notebooks.egi.eu/badge_logo.svg)](https://replay.notebooks.egi.eu/v2/gh/glarocca/ISGC2023_rainfall.git/HEAD)
6
7 [![DOI](https://zenodo.org/badge/613453962.svg)](https://zenodo.org/badge/latestdoi/613453962)
8 |
```




12. Check if the DOI is valid

<https://www.doi.org/>

TRY RESOLVING A DOI NAME

Type or paste a known DOI name exactly—including its prefix and suffix—into the text box below and then 'submit' to resolve it.

SUBMIT



13. Use the DOI to reproduce the repository

<https://replay.notebooks.egi.eu/>



Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

New to Binder? Get started with a [Zero-to-Binder tutorial](#) in Julia, Python, or R.

Build and launch a repository

Zenodo DOI (10.5281/zenodo.3242074)

Zenodo DOI ▾ 10.5281/zenodo.7729979

Git ref (branch, tag, or commit)

HEAD

Path to a notebook file (optional)

Path to a notebook file (optional)

File ▾

launch

Copy the URL below and share your Binder with others:

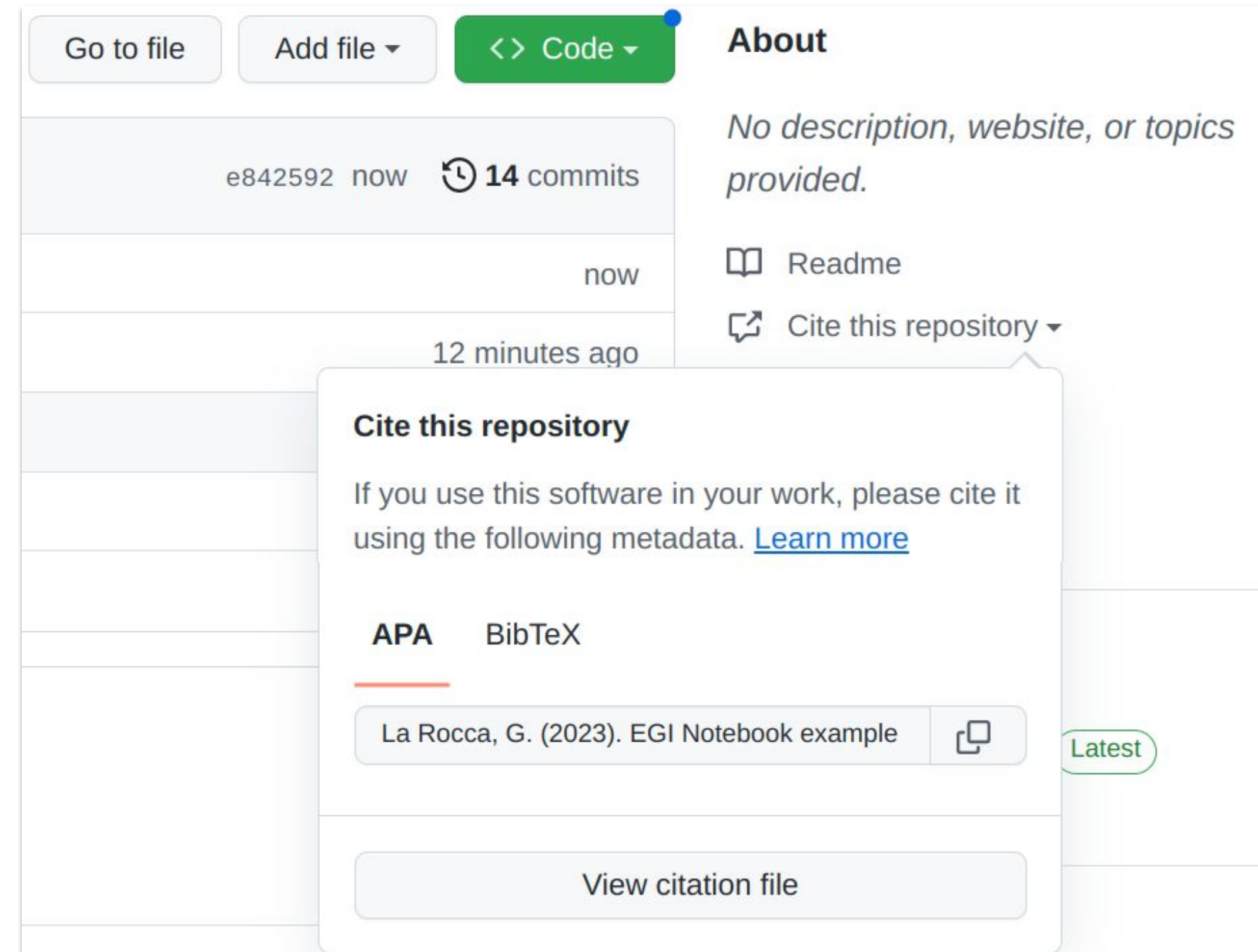
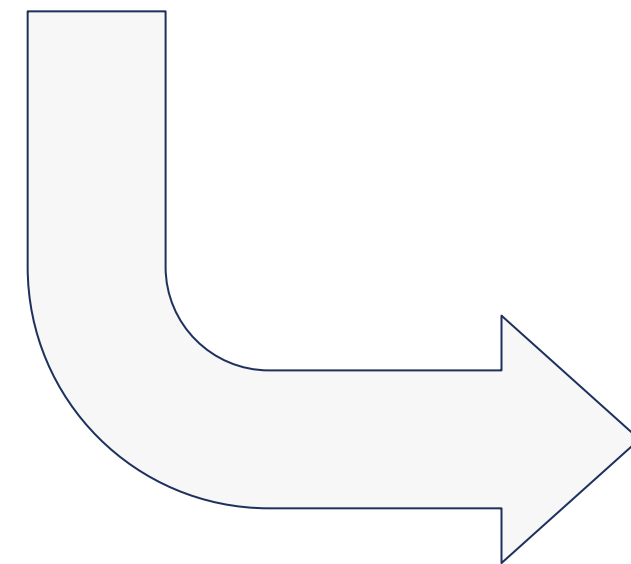
<https://replay.notebooks.egi.eu/v2/zenodo/10.5281/zenodo.7729979/>

Expand to see the text below, paste it into your README to show a binder badge: ▶

14. About CITATION files to your GitHub repository

Add a **CITATION.cff** file to the root of a repository

```
<> Edit new file Preview  
1 cff-version: 1.2.0  
2 message: "If you use this software, please cite it as below."  
3 authors:  
4 - family-names: "La Rocca"  
5   given-names: "Giuseppe"  
6   orcid: "https://orcid.org/0000-0002-8011-1450"  
7 title: "EGI Notebook example"  
8 version: 2.0.4  
9 doi: 0.5281/zenodo.7729979  
10 date-released: 2023-03-13  
11 url: "https://github.com/github/linguist"
```



The screenshot shows a GitHub repository page with a modal open for citing the repository. The modal displays the following information:

- Cite this repository**
- If you use this software in your work, please cite it using the following metadata. [Learn more](#)
- Two citation styles are shown: **APA** (selected) and **BibTeX**.
- The citation text is: `La Rocca, G. (2023). EGI Notebook example` with a copy icon.
- A **Latest** badge is visible on the right side of the modal.
- A **View citation file** button is at the bottom of the modal.



Future outlook: How to access and process data from the European Open Science Cloud

TLP: GREEN Limited disclosure



European Open Science Cloud

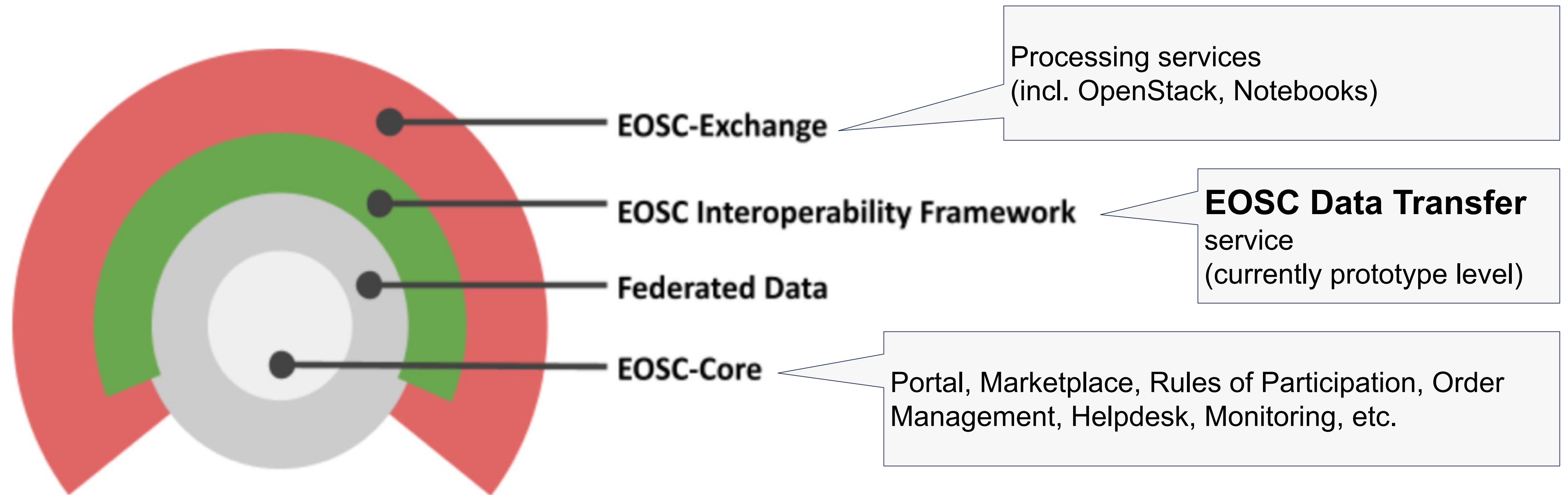
<https://eosc.eu/>

The **European Open Science Cloud (EOSC)** is an environment for hosting and processing research data to support EU science.

The **ambition** of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes.

This environment **will operate** under well-defined conditions to ensure trust and safeguard the public interest.

EOSC ultimately **aims to develop** a Web of FAIR Data and services for science in Europe upon which a wide range of value-added services can be built. These range from visualisation and analytics to long-term information preservation or the monitoring of the uptake of open science practices.



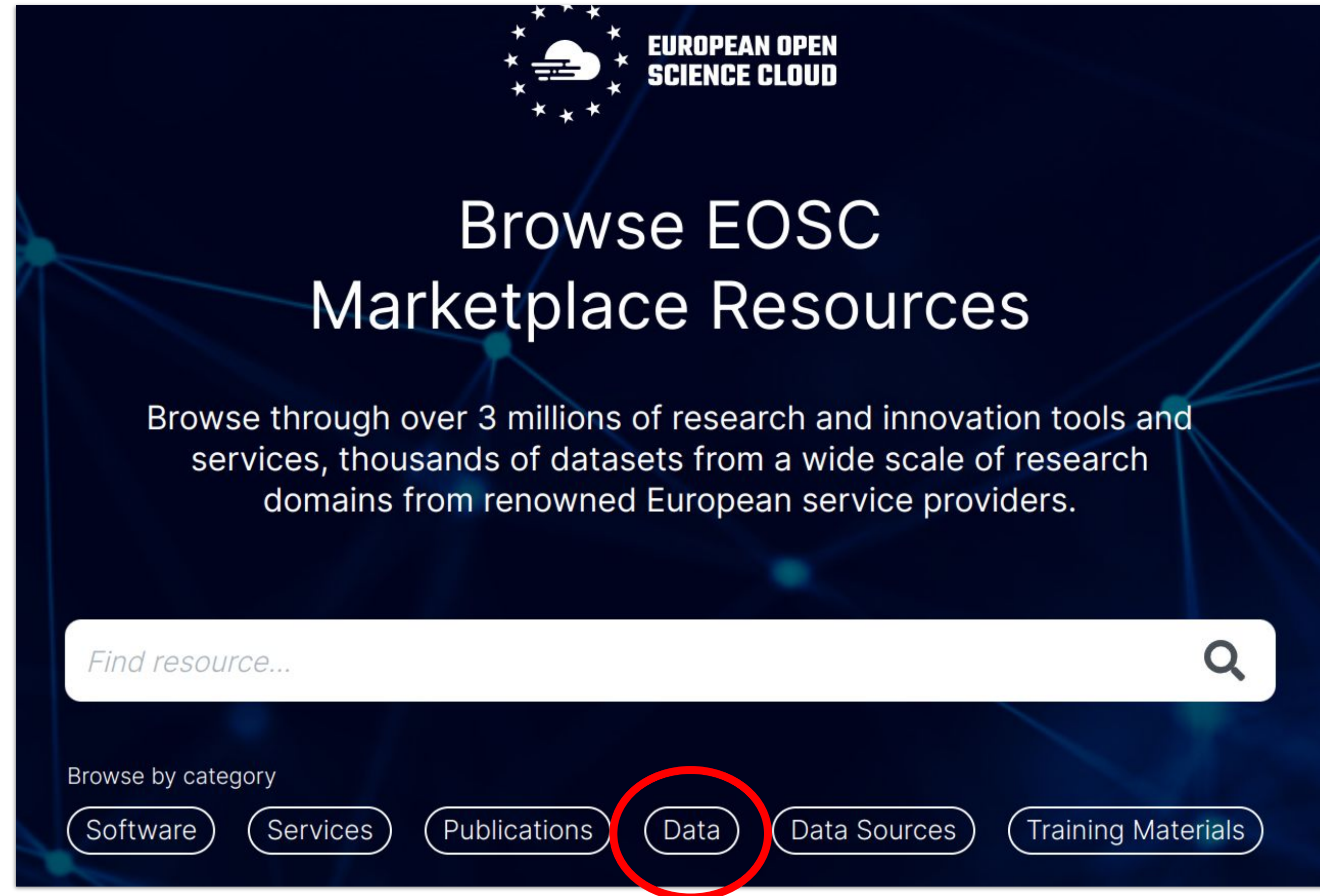
EOSC Multi-Annual Roadmap 2023-2024 (May 2022)

What we will see in this demo:

- 1.) Use the EOSC Marketplace to find a dataset
- 2.) Get the DOI of the dataset to be transferred
- 3.) Click on the dataset and open the EOSC Explore interface
- 4.) Select the settings for the data transfer (using the dev instance!)
- 5.) Initiate the data transfer
- 6.) Use the AWS CLI to check the files transferred to the destination

1.) Find the dataset in the EOSC Marketplace

- Access the EOSC Marketplace:
<https://marketplace.eosc-portal.eu/>
- Select Data as “Category”



2.) Get the dataset to be transferred (e.g.: Trialstreamer)

Filters

Research step

clear all

Discover Research Outputs (6)

Access right

clear all

Open access (6)

Restricted (0)

Closed (0)

Embargo (0)

Year range

clear all

6 search results in Data

Dataset

Open Access

English

Trialstreamer data

 Open access   Type: dataset

Author names: Marshall Iain Nye Benjamin Kuiper Joël Marshall Rachel Soboczinski Frank Nenkova Ani Noel Storr Anna Thomas James Wallace Byron

DOI: 10.5281/zenodo.5734208

Trialstreamer annotated collection of RCTs. This repository contains baseline files (large), and subsequent updates (daily for PubMed, weekly for ICTRP).

DOI: 10.5281/zenodo.5734208

3.) Click on the dataset to access the file in the EOSC Explorer

Research Data . Dataset . 2020

Trialstreamer data

Marshall, Iain; Nye, Benjamin; Kuiper, Joël; Marshall, Rachel; Soboczenski, Frank; Nenkova, Ani; Noel Storr, Anna; [+2 Authors](#)

OPEN ACCESS

ENGLISH

DOI: [10.5281/zenodo.5734208](https://doi.org/10.5281/zenodo.5734208)

Published: 26 Apr 2020

Publisher: Zenodo

Summary Related research (2)

Abstract

Trialstreamer annotated collection of RCTs. This repository contains baseline files (large), and subsequent updates (daily for PubMed, weekly for ICTRP).

ACTIONS



2 Research Products, Page 1 of 1

Powered by [OpenAIRE Research Graph](#). Last update of records in OpenAIRE: Feb 13, 2023

WARNING

- Datasets must be stored in Zenodo!

- S3 support in EOSC Data Transfer portal is not fully working. It will be available in production in April 2023.



4.) Start the data transfer using the dev instance

EOSC Data Transfer [demo] ✕

<https://dx.doi.org/10.5281/zenodo.6669532>

109 files found:

- trialstreamer-update-pubmed-2022-05-16.csv
- trialstreamer-update-pubmed-2022-05-23.csv
- trialstreamer-update-pubmed-2022-05-30.csv
- trialstreamer-update-pubmed-2022-06-06.csv
- trialstreamer-update-pubmed-2022-06-13.csv
- trialstreamer-update-pubmed-2022-06-20.csv

s3

Destination system (e.g. hostname:8080):
s3.cl2.du.cesnet.cz

Provide authentication:

Give access key
ACMOVA5LBOL0JJ875C...

Give secret key
.....

Destination path (e.g. /folder1/folder2):
/enolfc-test-bucket

Transfer status: **submitted.** ✕

>> TRANSFER

5.) Click on "Transfer" to initiate the data transfer

6.) Use the AWS client to check the files transferred in the S3 bucket

```
larocca@aktarus:~$ aws s3 --endpoint-url https://s3.cl2.du.cesnet.cz ls s3://enolfc-test-bucket
2023-03-10 11:36:02      1146 ENESDS_CMIP6.csv
2023-03-14 14:39:42   10280772 emojis.zip
2023-03-13 18:02:58 1152716971 tasmx_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20150101-20391231.nc
2023-03-13 18:07:42 1150771873 tasmx_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20400101-20641231.nc
2023-03-14 07:52:41 1148226564 tasmx_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20650101-20891231.nc
2023-03-14 07:55:09   504796434 tasmx_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20900101-21001231.nc
2023-03-16 11:03:27      179 trialstreamer-update-pubmed-2021-01-04.csv
2023-03-16 11:03:29   25688160 trialstreamer-update-pubmed-2021-08-23.csv
2023-03-16 11:00:24   23538134 trialstreamer-update-pubmed-2021-11-22.csv
2023-03-16 11:02:27      179 trialstreamer-update-pubmed-2022-01-24.csv
2023-03-16 11:01:17      179 trialstreamer-update-pubmed-2022-01-31.csv
2023-03-16 11:03:35 154835212 trialstreamer-update-pubmed-2022-02-07.csv
2023-03-16 11:00:22   37251724 trialstreamer-update-pubmed-2022-03-21.csv
2023-03-16 11:01:19      179 trialstreamer-update-pubmed-2022-04-11.csv
2023-03-16 11:00:30   92992913 trialstreamer-update-pubmed-2022-04-18.csv
2023-03-16 11:00:21   46631156 trialstreamer-update-pubmed-2022-05-02.csv
2023-03-16 10:58:11   16030071 trialstreamer-update-pubmed-2022-06-13.csv
2023-03-14 14:39:42   20608411 trialstreamer.csv
```



Take-away message and next steps

TLP: GREEN Limited disclosure



Key messages

- EGI established a cloud + Notebooks + Replay services to facilitate big data science in the Asia Pacific region
 - Current provider: CESNET
 - Documentation: <https://docs.egi.eu/users/getting-started/communities/dmcc/>
- Use the EGI Notebooks and the EGI Replay services to facilitate applications sharing and reproducible
- Use the EGI DataHub, B2DROP, CVMFS services to store and share datasets

- Open to scientists to run simulations
- Open for application developers to share big data applications
- Open for data providers to share data
- Open for OpenStack providers to share compute/storage

Upcoming EGI Conference

<https://indico.egi.eu/event/6071/abstracts/>



Call for Contributions is open!



Thank you!

Q & A

Feedback form:

<https://survey.egi.eu/576676>

www.egi.eu



This work is partially funded by the EU research and innovation programme