EGI Tutorial at ISGC 2023

Giuseppe La Rocca

Community Support Team Lead.

TLP: GREEN Limited disclosure

About me

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/

Agenda

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

Introduction about the EGI and the EGI Infrastructure

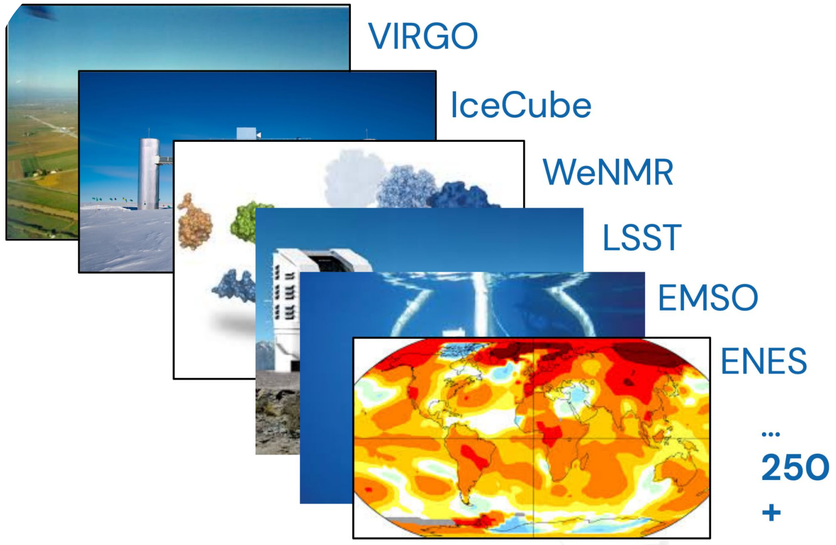
TLP: GREEN Limited disclosure

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

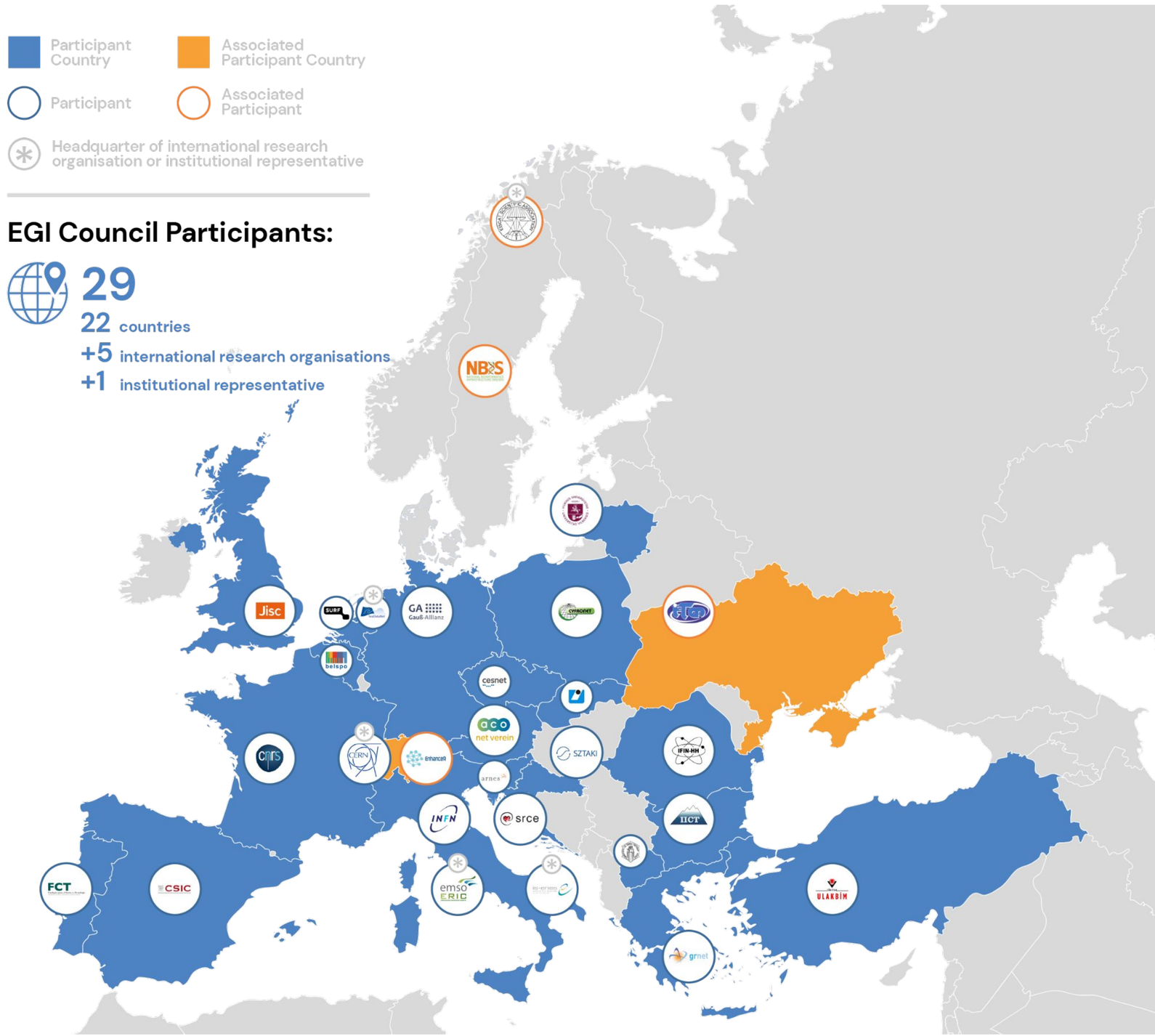
All researchers have seamless access to services, resources and expertise to collaborate and conduct world-class research and innovation

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



International Partnerships

The EGI Infrastructure

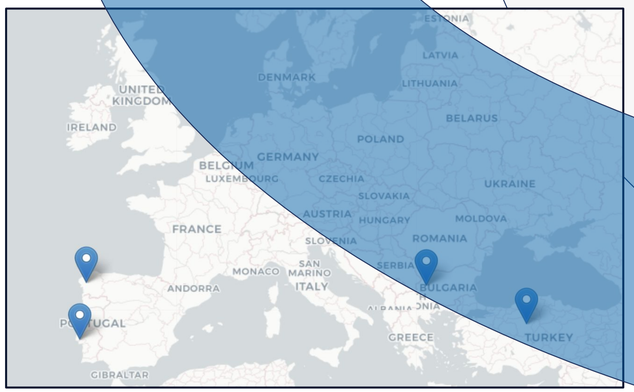
Platform and community specific services built on top:
Scientific gateways, Virtual Research Environments, Application portals, etc.

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

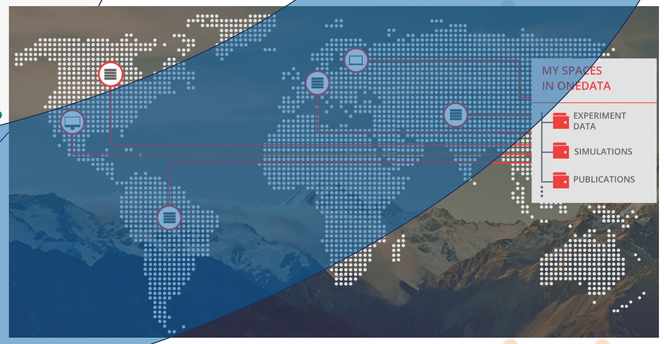
Cloud Compute
(based on OpenStack. Since 2013)

EGI Internal services

- Service registry
- Monitoring
- Accounting
- Helpdesk
- Single sign-in
- Security oversight
- Capacity management
- etc.



High Performance Compute
(since 2022)

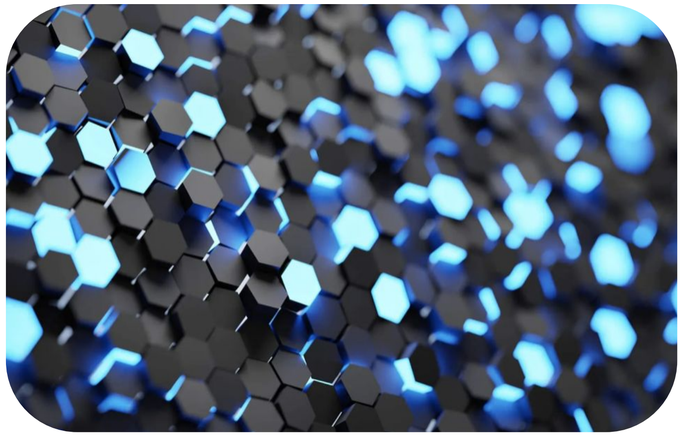


Data
(based on OneData)

EGI Value Propositions

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

Type, number, size, cost, availability, etc.

Users' representative

Instance requirements

Negotiator

Foundation

Conditions

Applic. provider

Storage provider

Cloud provider

HTC provider

Training

Support

Regular satisfaction interviews

Ensure satisfaction, capture achievements, record feedback

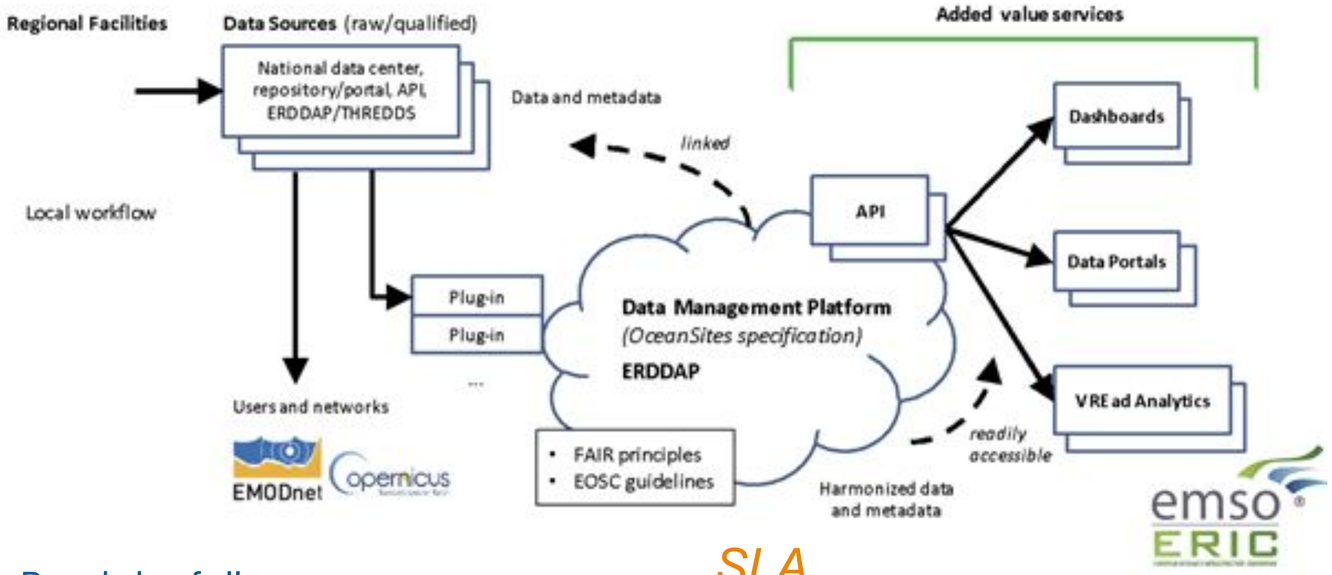
Service Level Agreement

Operation Level Agreement

Ensuring that agree targets are met

Regular service delivery reports

An SLA example



Read the full story: https://www.egi.eu/case-study/emso-eric/

Check the SLA-OLA documents and reports: https://documents.egi.eu/document/3539

SLA

OLA-1 OLA-2

CESGA Centro de Supercomputación de Galicia

INFN Istituto Nazionale di Fisica Nucleare Sezione di Bari

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

EGI VO

VICE 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)

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]:

5

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/

6

- Planned maintenance windows or service interruptions ("scheduled downtimes"[8]) will be notified via email in a timely manner i.e. 24 hours before the start of the outage[9].
- Downtime periods exceeding 24 hours need justification.

3 Support

Support is provided via EGI Service Desk[10]. Access requires a valid X.509 or the login via a EGI SSO account[11]. 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

EGI allocations for the Asian Pacific (AP)

TLP: GREEN Limited disclosure

ISGC 2023

EGI resource pool for Asia Pacific

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

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

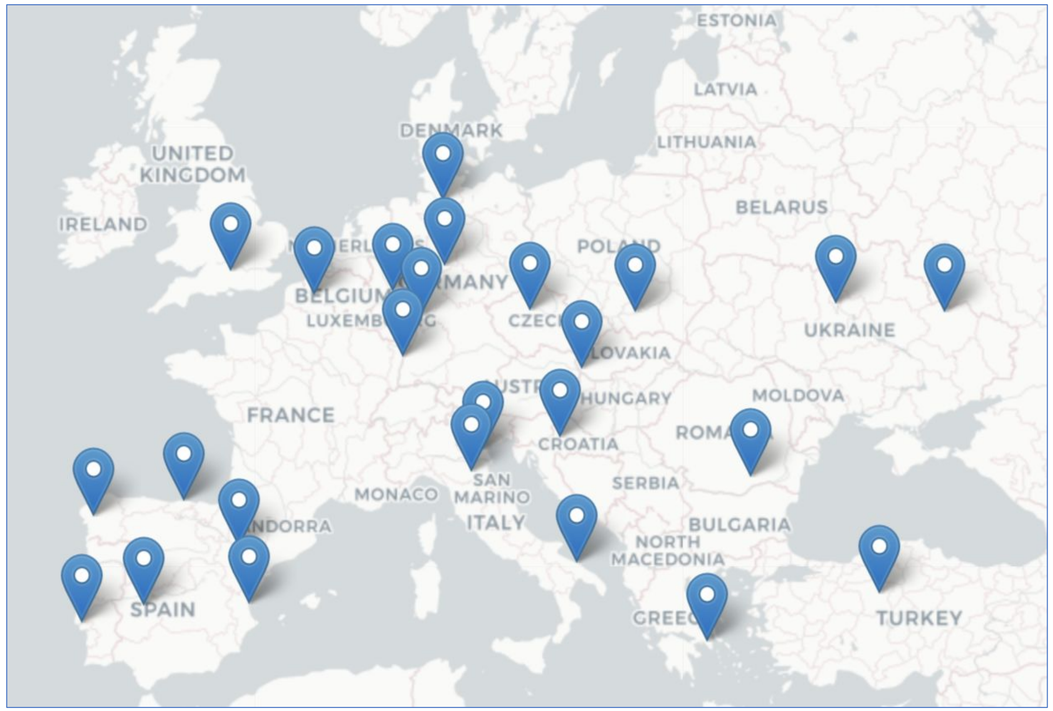
Documentation

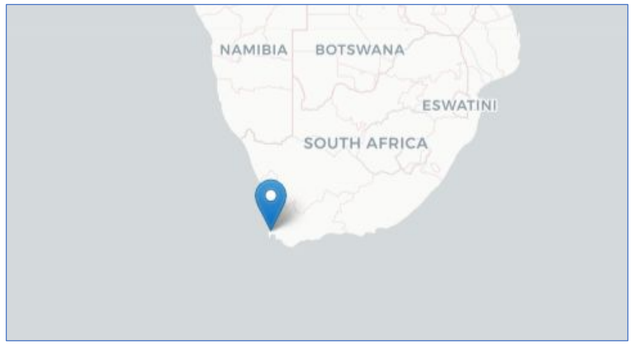
https://docs.egi.eu/users/getting-started/communities/dmcc/

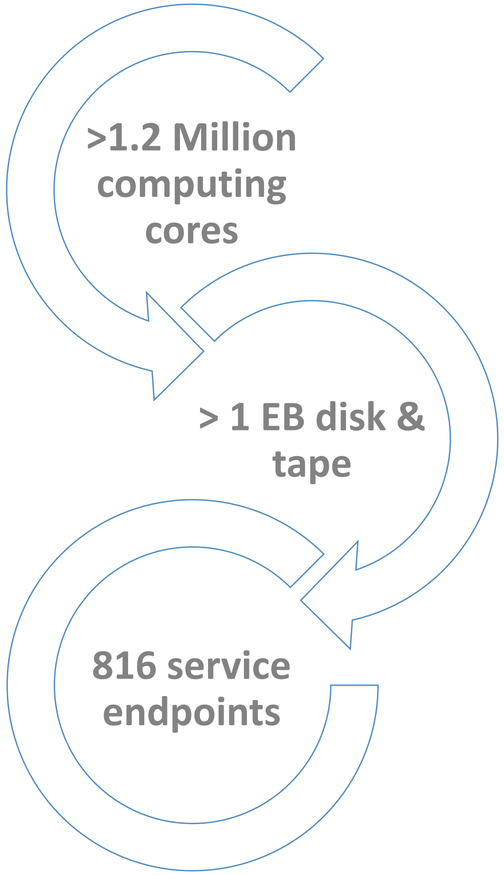
Demo: How to use the EGI Cloud resources pool for Asia Pacific

TLP: GREEN Limited disclosure

The EGI Cloud Federation







- 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

Kubernetes (Containers) in the EGI Cloud

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

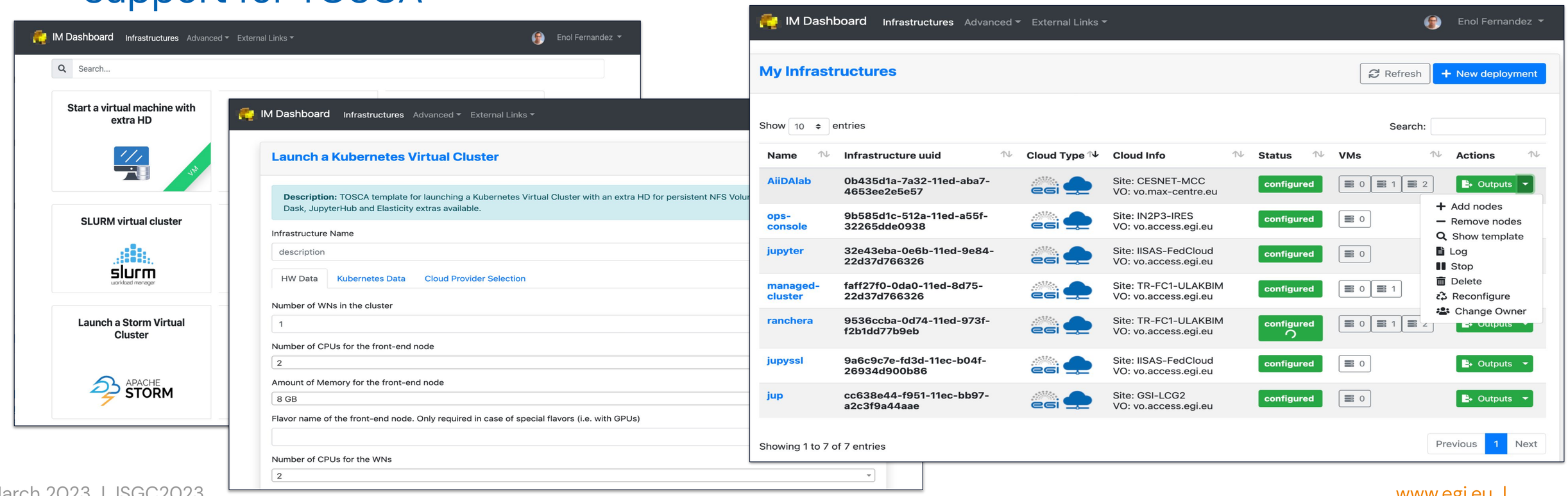
Rancher-based access to managed clusters (under development)

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

Federation services for Cloud

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



AppDB: an application store for EGI Cloud

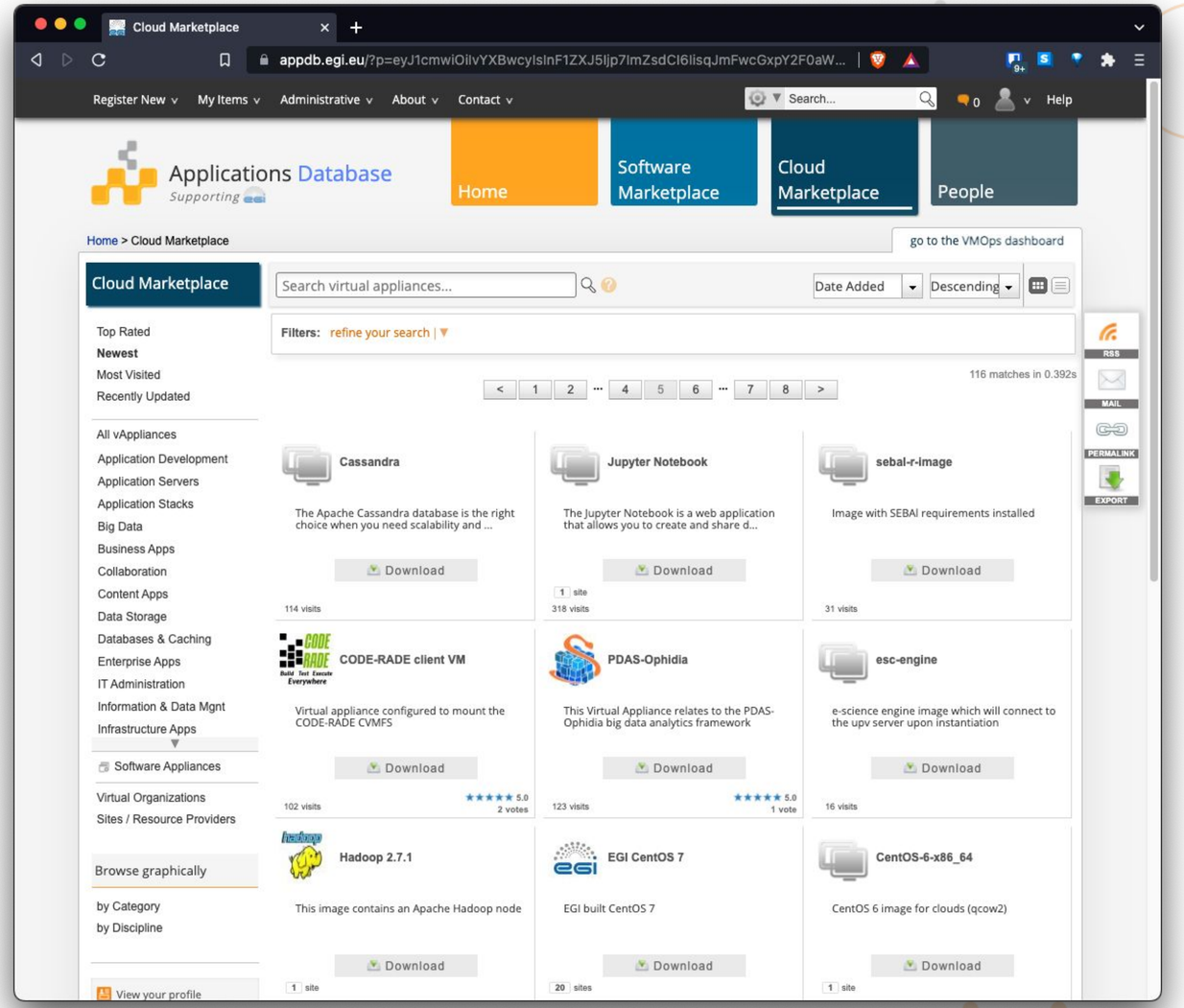
Common registry for Virtual Appliances (VA)

- VM image + metadata

Available for running at the EGI cloud or on any hypervisor

Community-level management of VAs

- Automatic distribution to providers

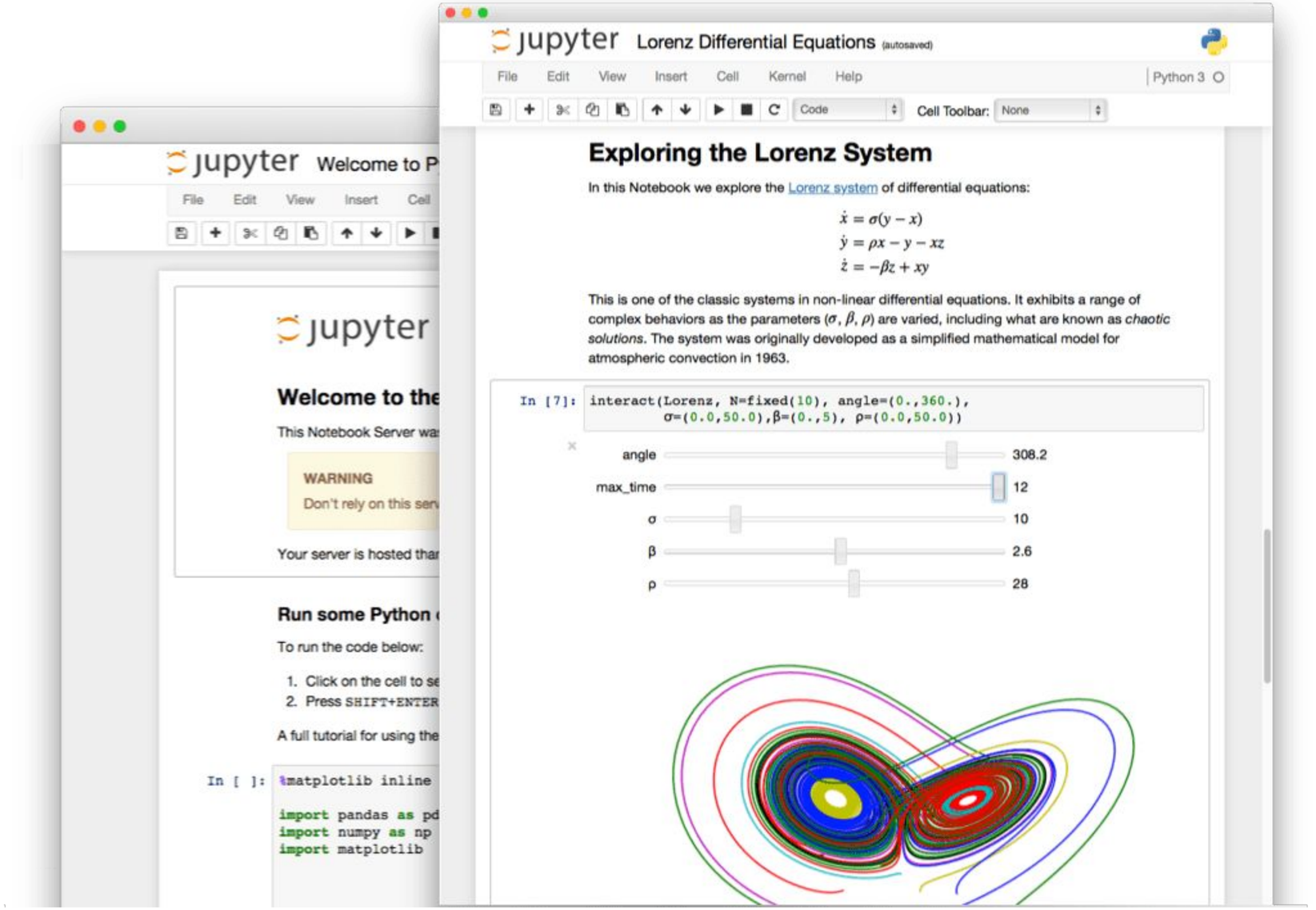


EGI Notebooks in the Asia Pacific VO

TLP: GREEN Limited disclosure

ISGC 2023

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.



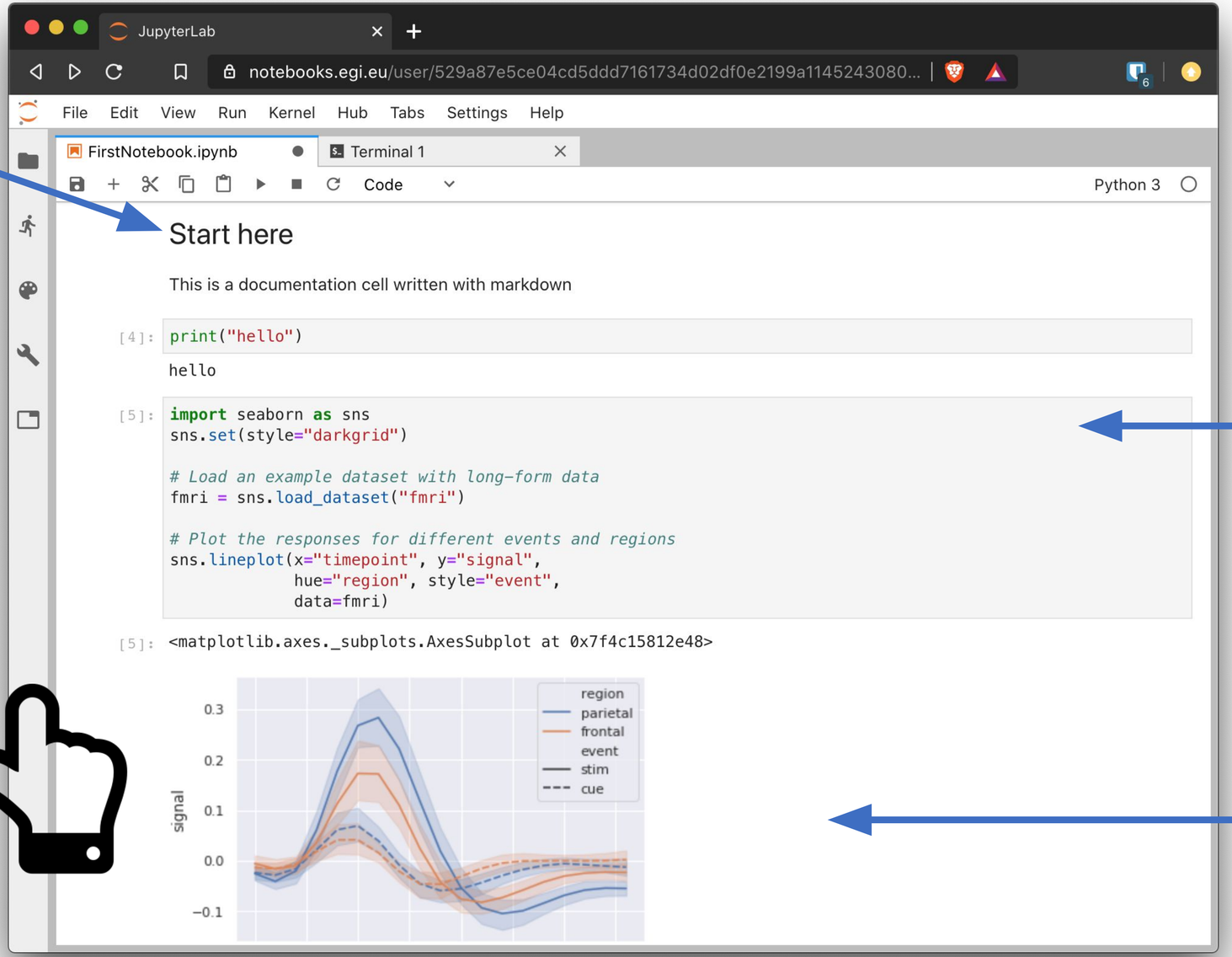
Jupyter

Documentation
Text formatted using Markdown/LaTeX

Interactive
browser based environment

Code
Use your favourite language

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

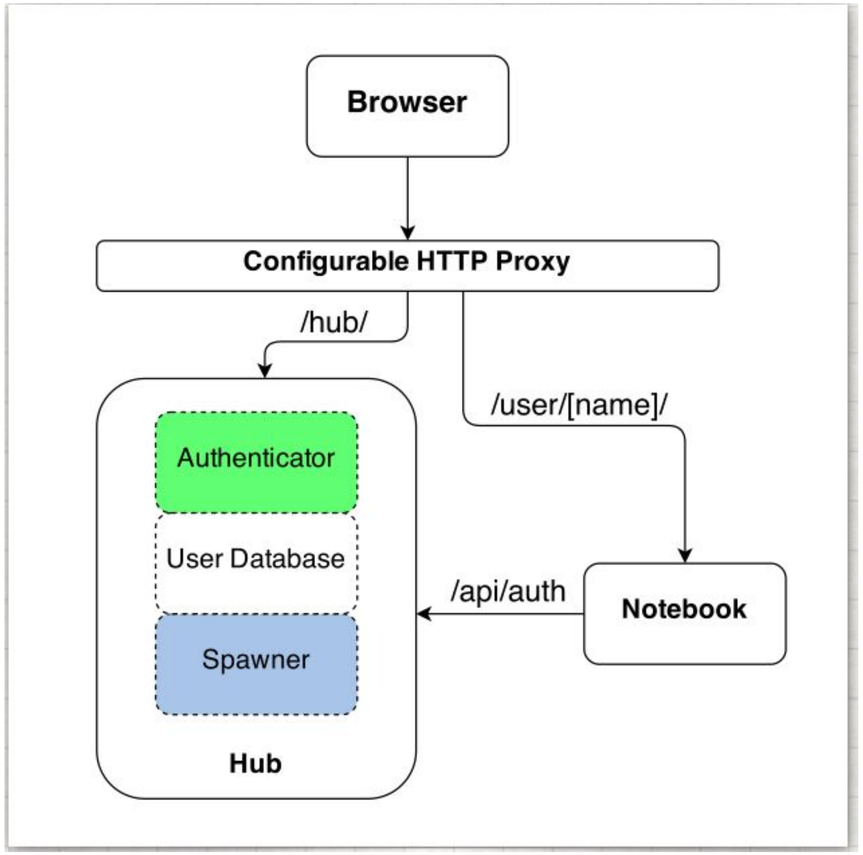


JupyterHub

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



EGI Notebooks

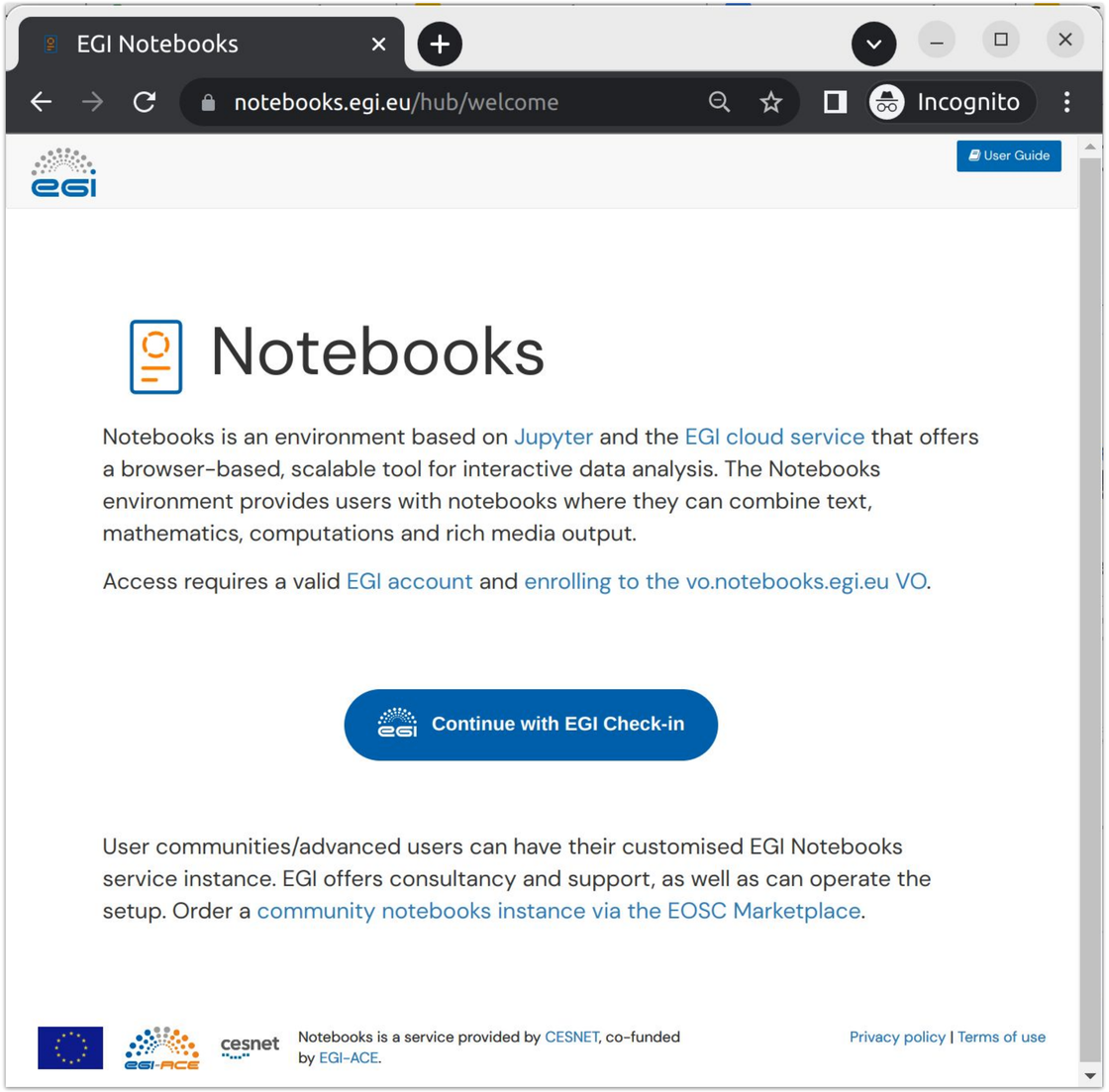
https//notebooks.egi.eu

JupyterHub hosted in the EGI Cloud

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

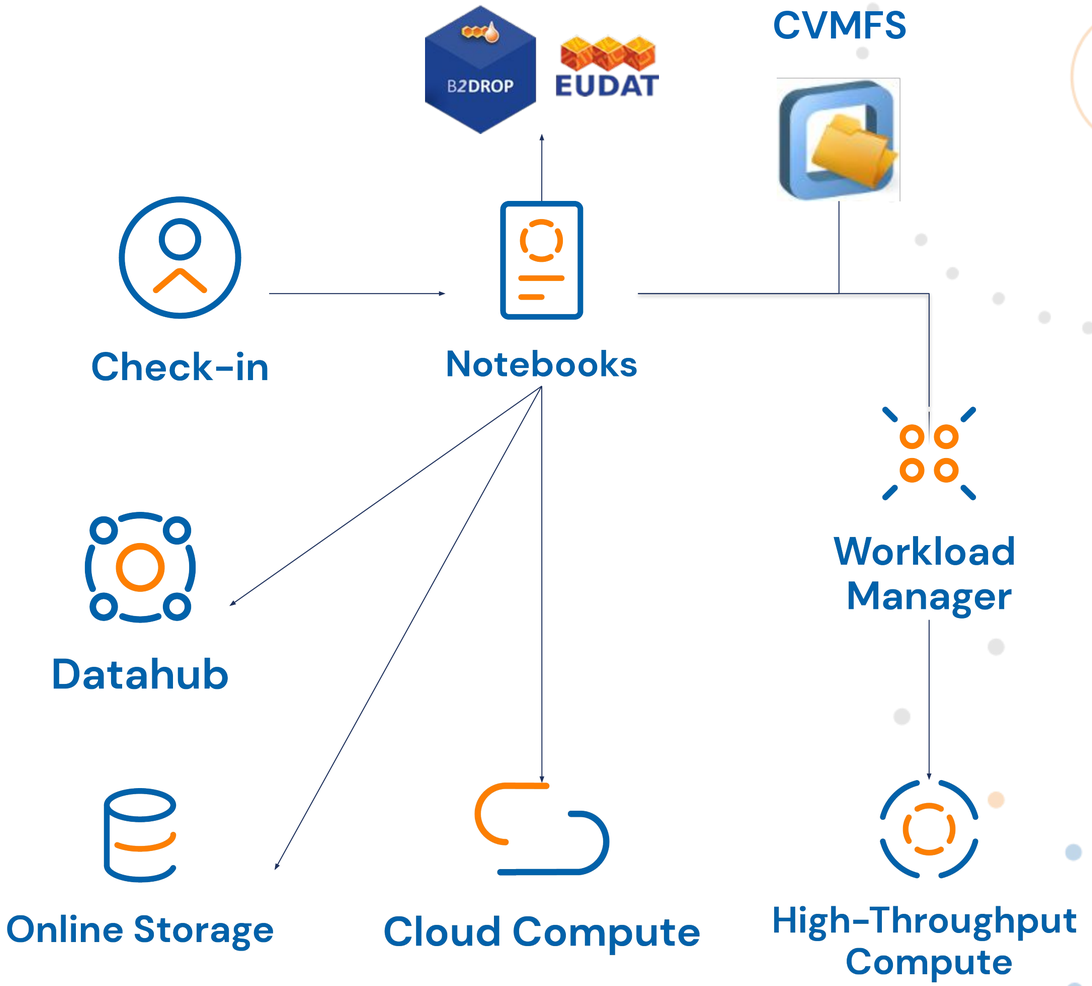
Main Features:

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



EGI Notebooks integrations

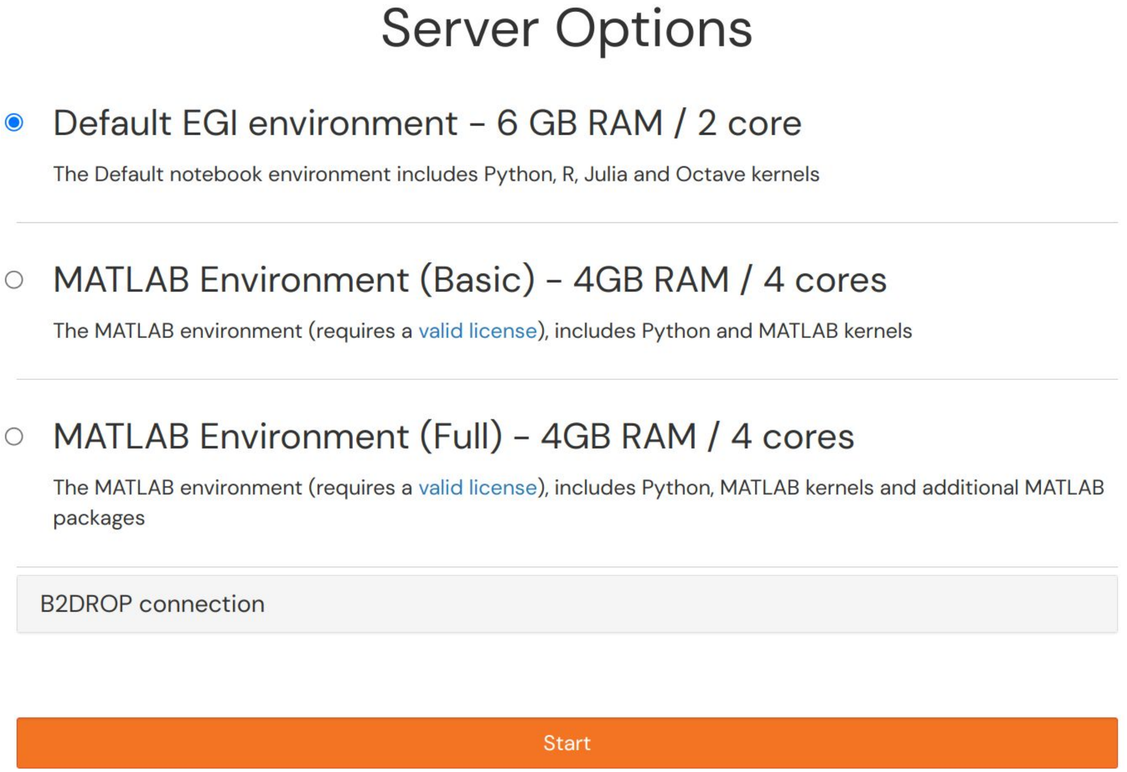
1. Runs on EGI Cloud providers (Cloud Compute / Online Storage)
2. Uses Check-in for authentication + access to other services
3. Makes user-level software available from CVMFS
   - DIRAC client for submission of jobs to Workload Manager / High Throughput Compute
   - Fedcloudclient for interaction with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



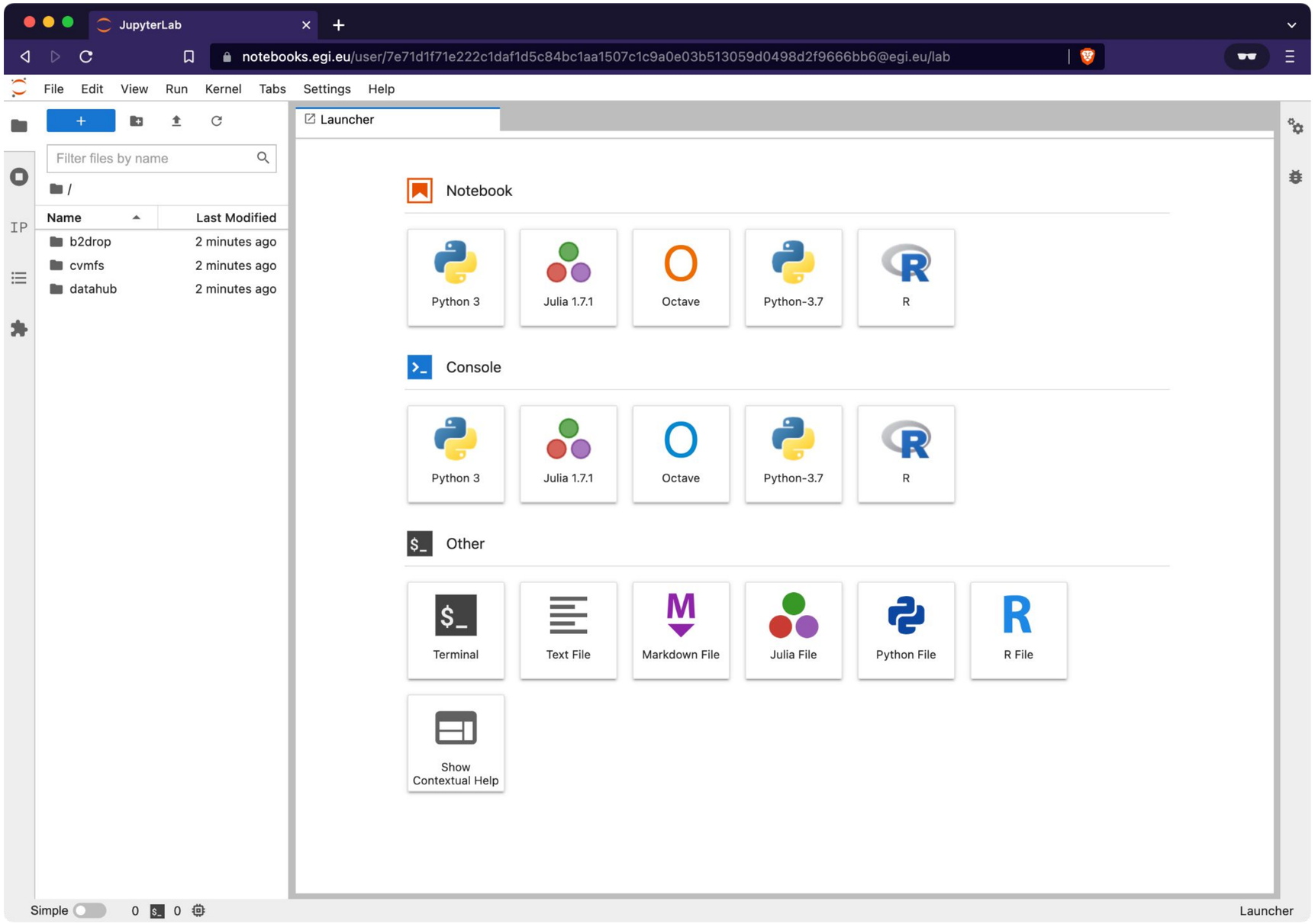
Starting your server

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!)

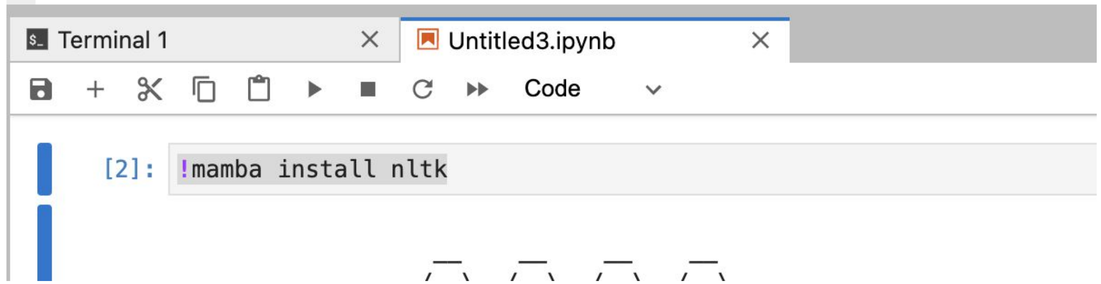


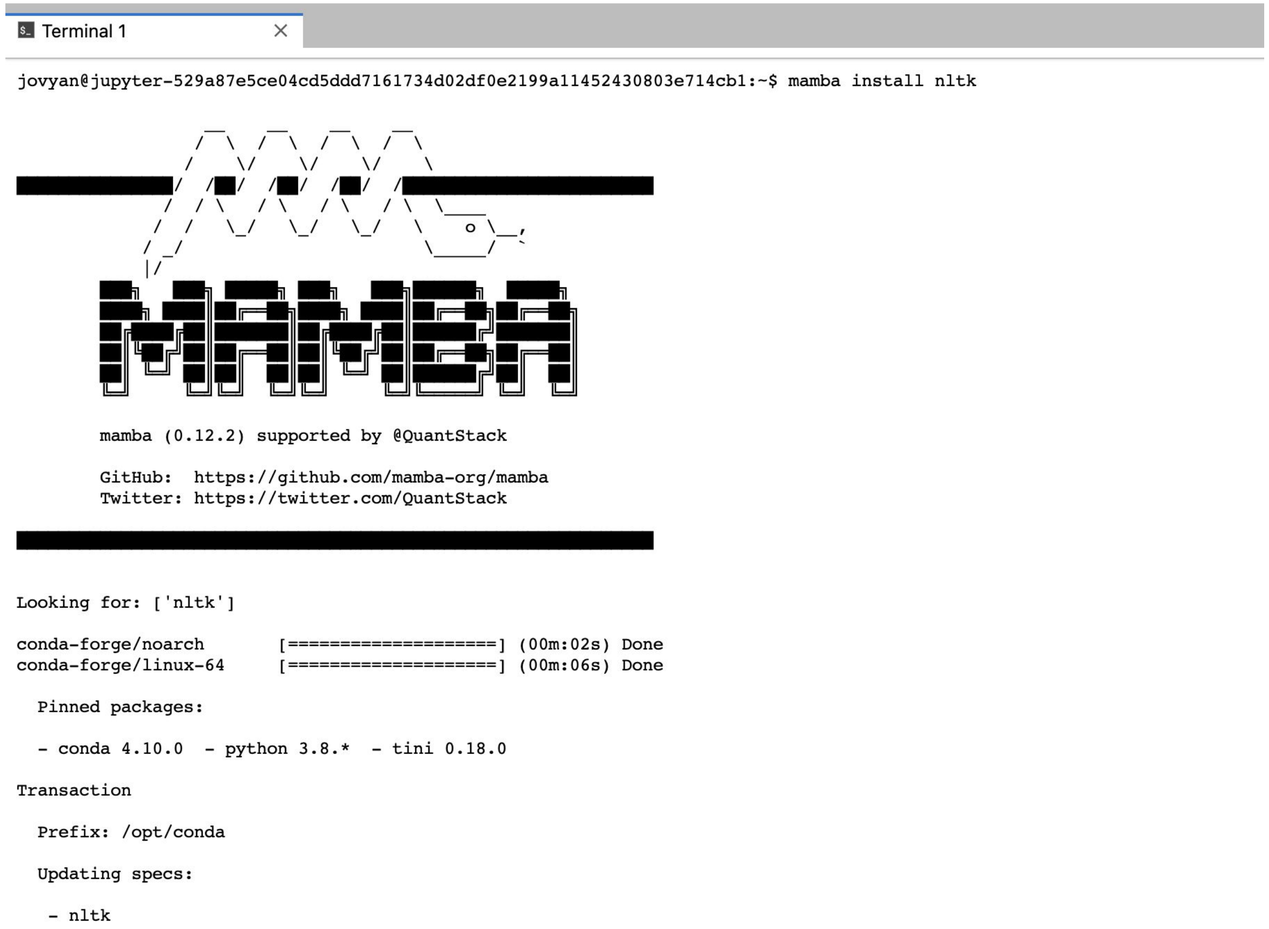
JupyterLab interface



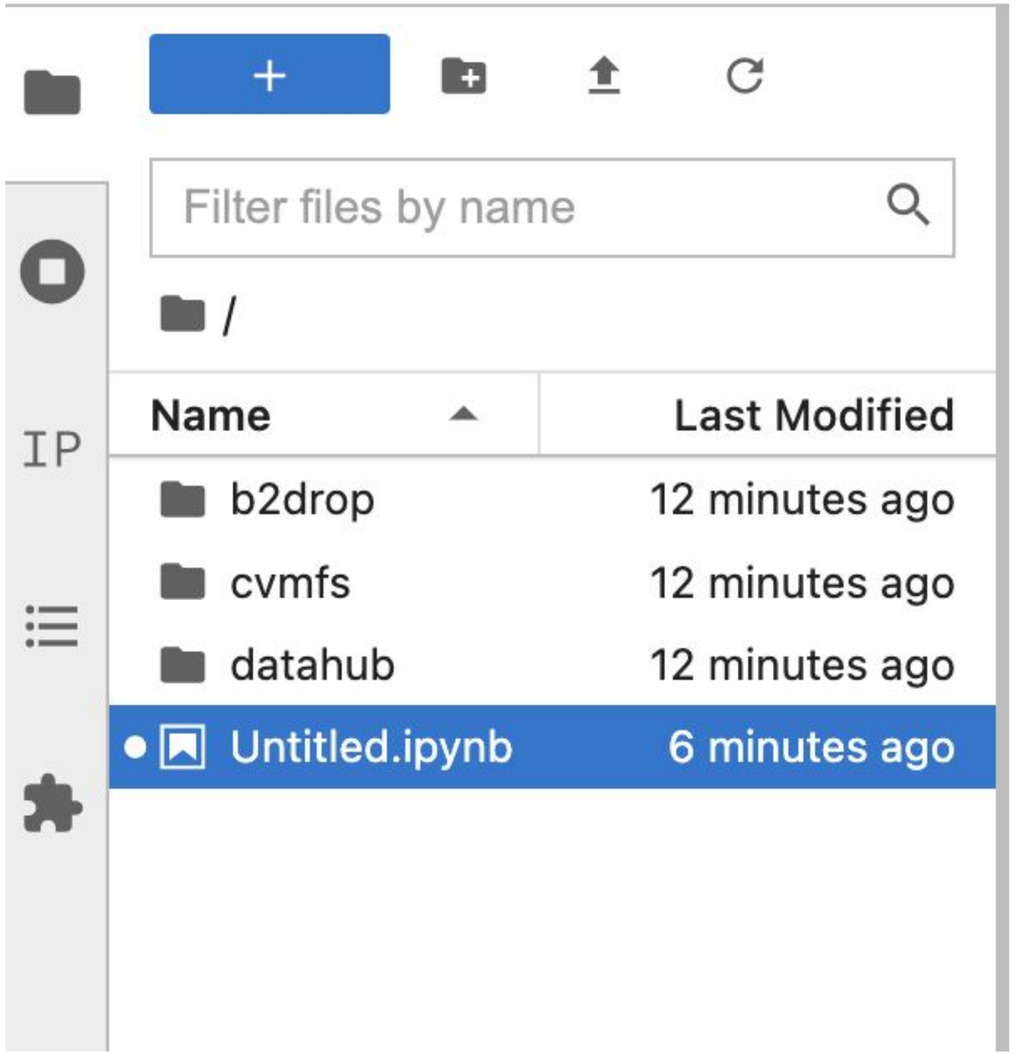
Installing Software

- (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!





Accessing your data and code



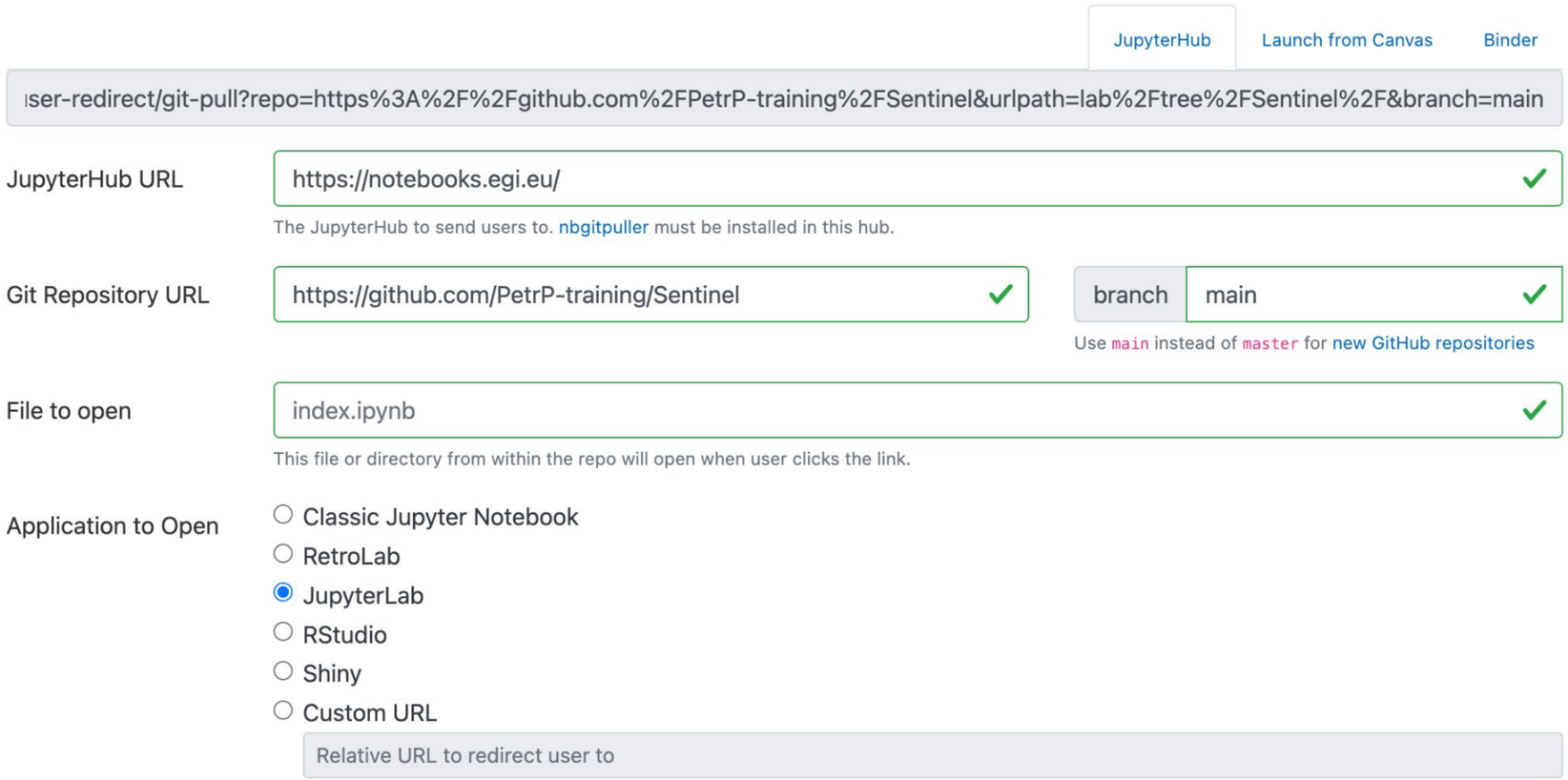
- Persistent home
  - Can be used to store data (10GB limit)
  - Files will be kept even if the notebook server dies
- nbgitpuller
  - Get code from any git repository from a single URL
  - https://hub.jupyter.org/nbgitpuller/link.html
- CMVFS
  - Selected CVMFS repositories available
  - Easy to add 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
  - ....

nbgitpuller

https://hub.jupyter.org/nbgitpuller/link.html

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

Try it: https://go.egi.eu/x8oui



CernVM-FS

- 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**.

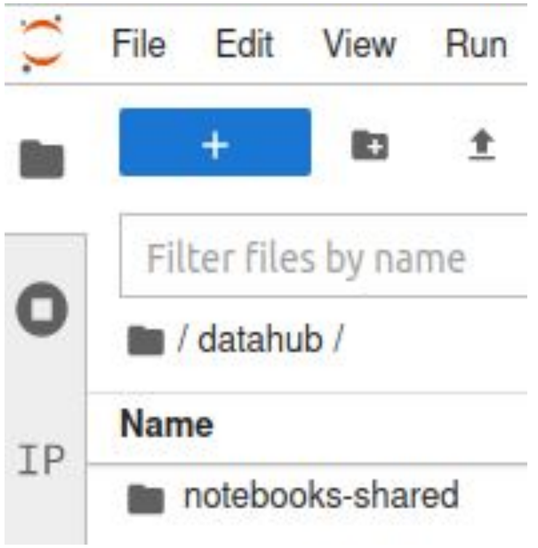
Main features

- 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/

EGI DataHub



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)

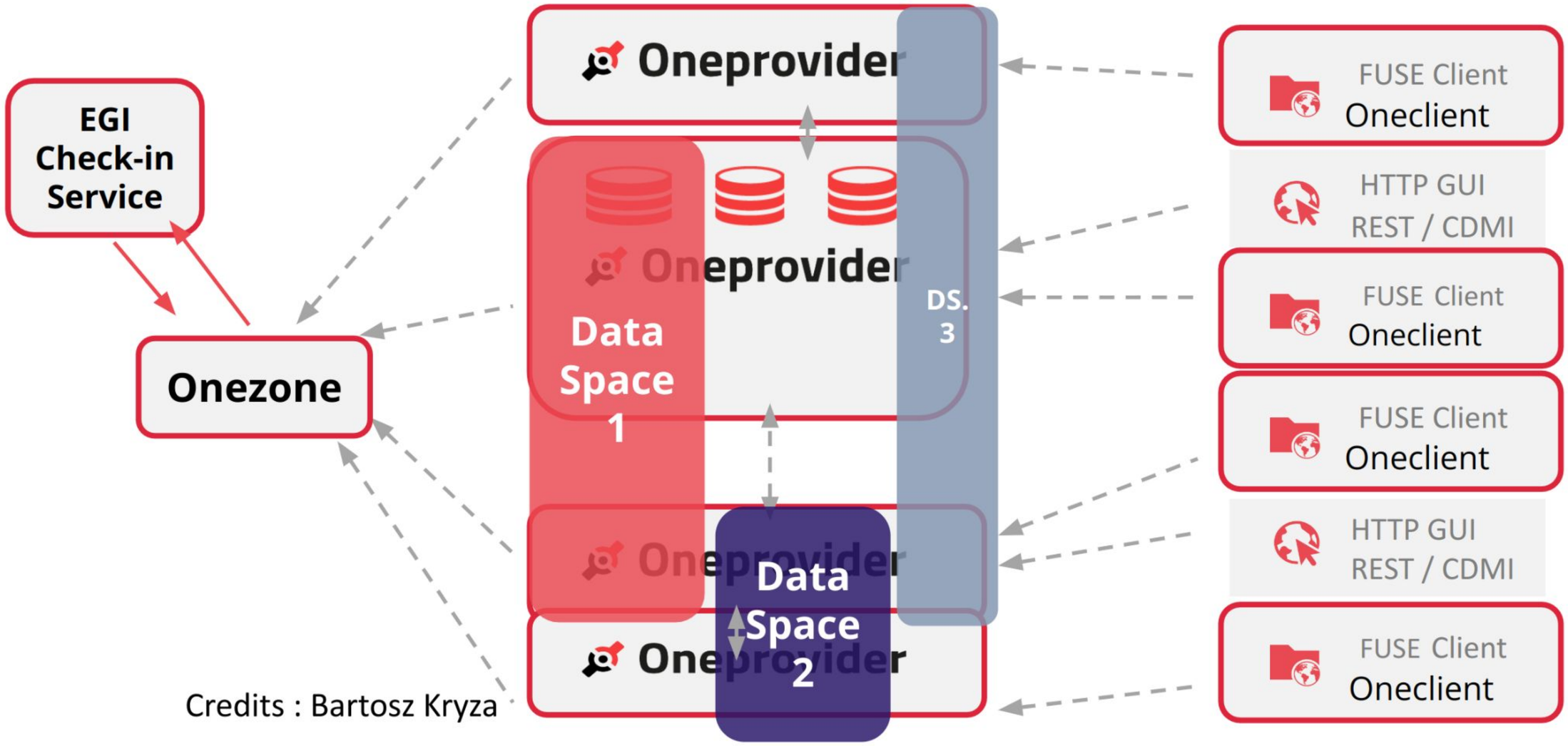
Data can be accessed via a **GUI** or **APIs**

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



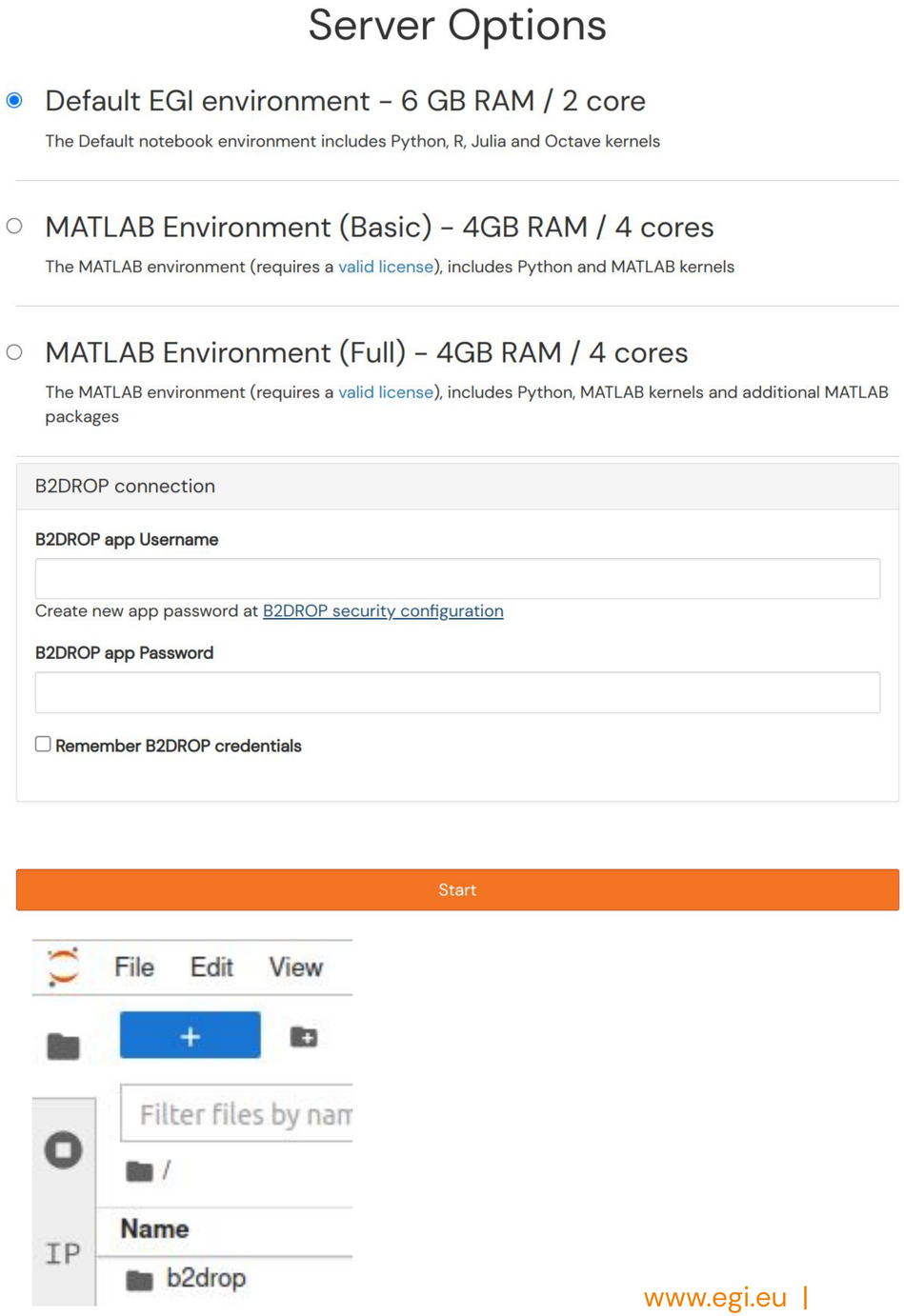
B2DROP

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



Getting access with the EGI Notebooks in the Asia Pacific VO

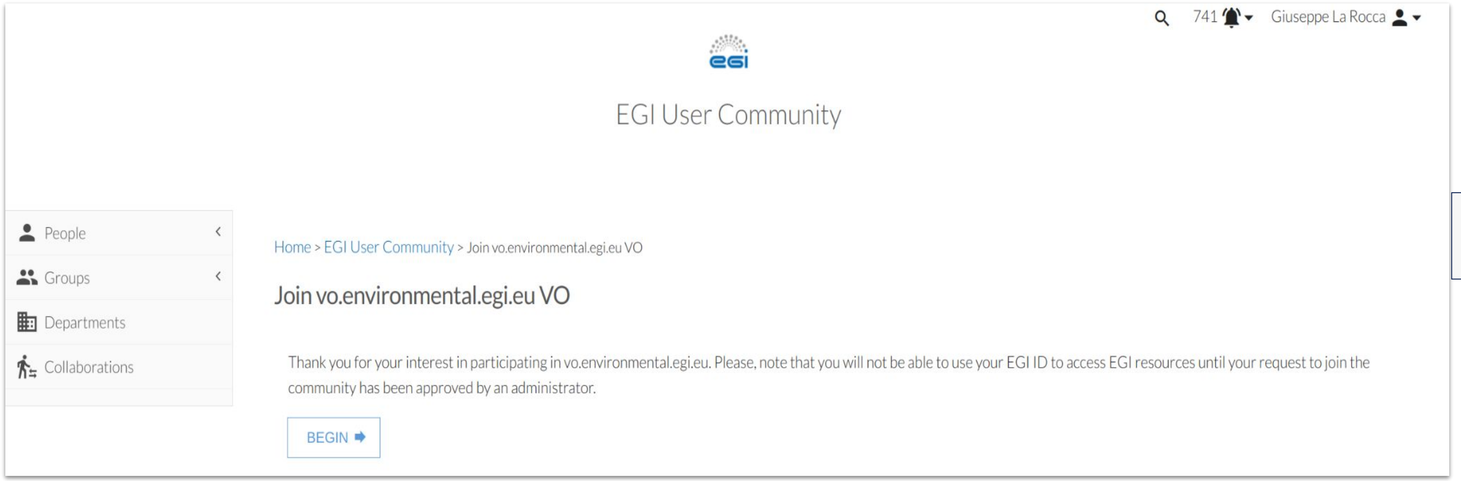
TLP: GREEN Limited disclosure

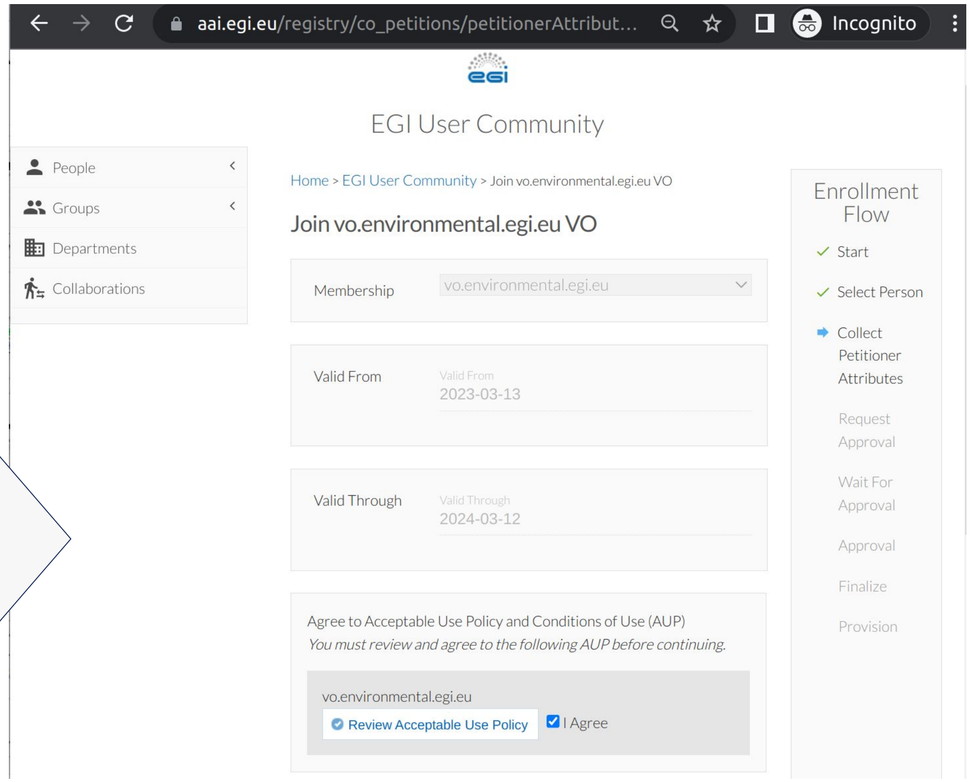
ISGC 2023

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.

HandsOn





Hands-on: Running your first notebooks

TLP: GREEN Limited disclosure

A very simple notebook

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()
```

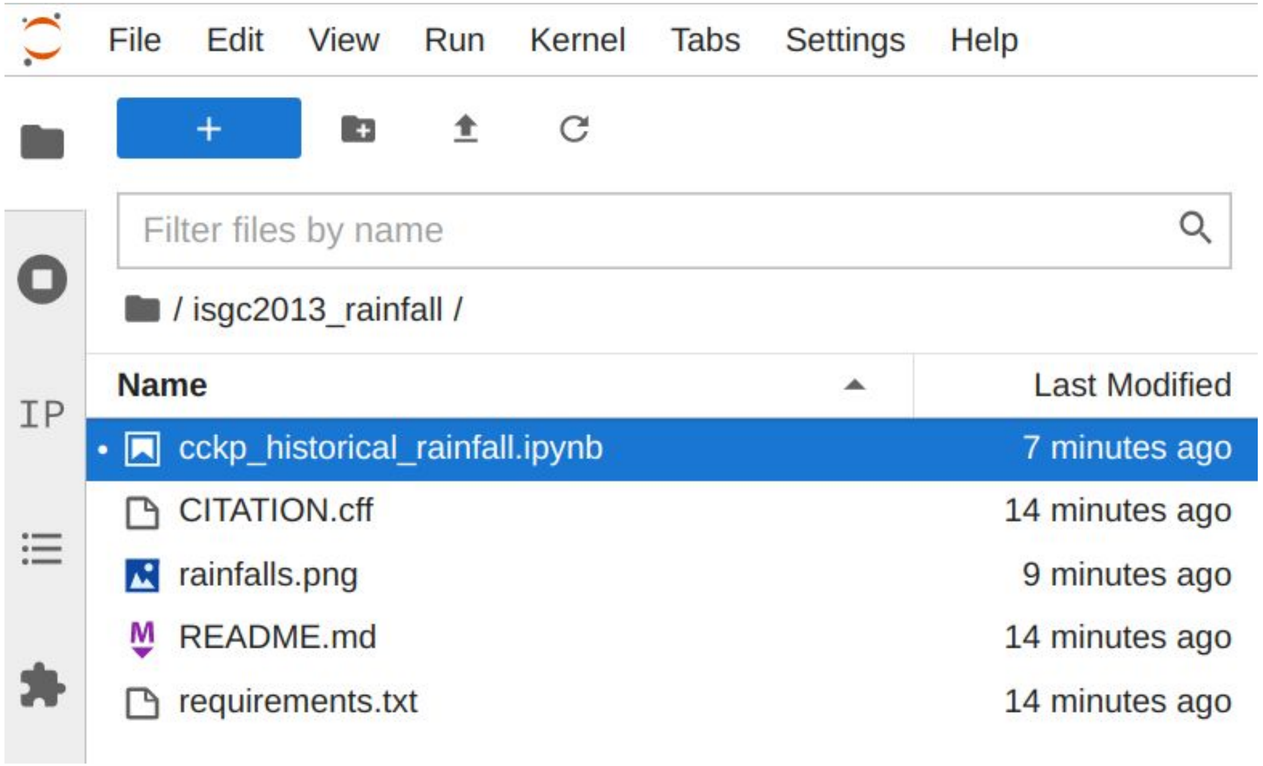
To launch the notebook ⇒ go.egi.eu/xABiu

00-first-notebook.ipynb

- GitHub repo for this notebook
- **Notebook fetched with the nbgitpuller**

HandsOn

Climate Change Knowledge Portal (CCKP)



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"

HandsOn

Coffee break

EGI Check-in

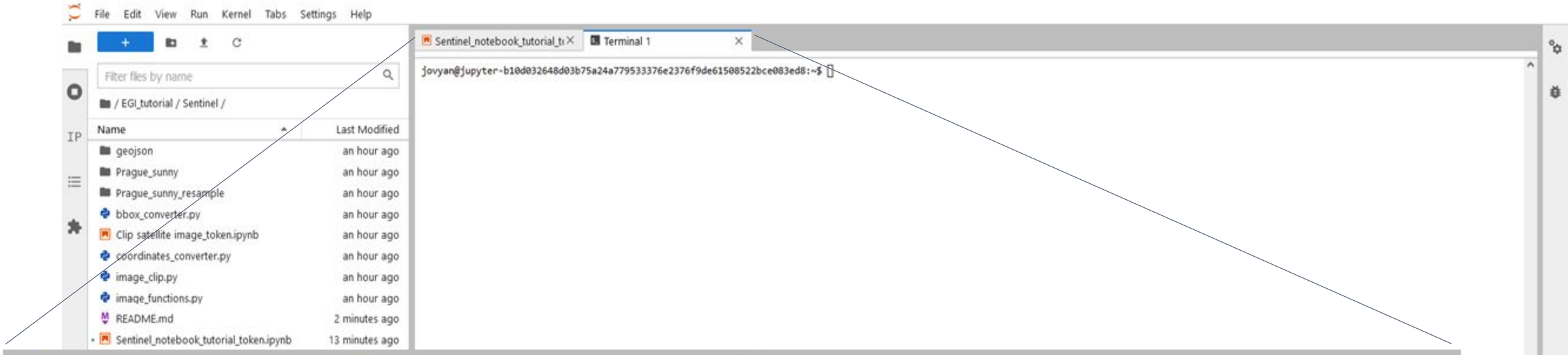
Check-in provides authentication, authorization and user management for the EOSC Compute Platform

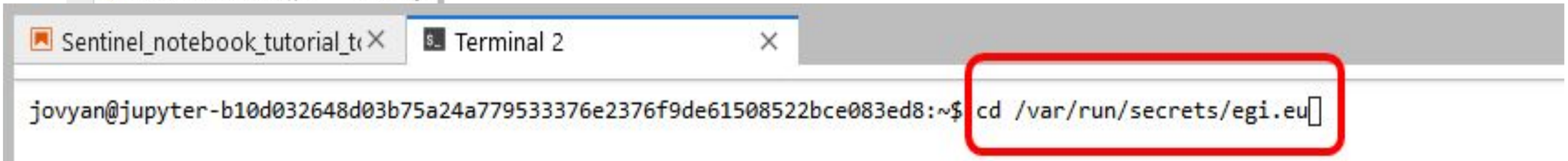
Standards based:
- SAML 2.0 / OpenID Connect 1.0 / OAuth 2.0 / LDAP

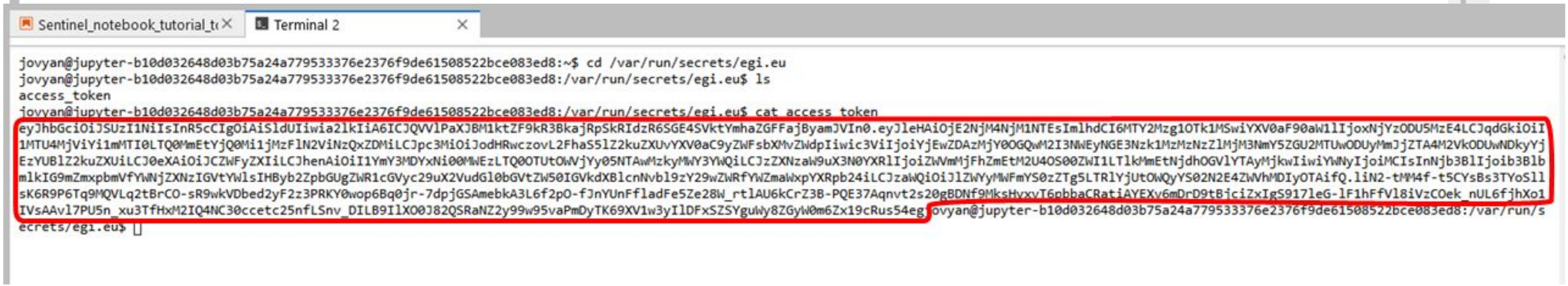
Interoperable:
- AARC and EOSC AAI compliant
- Support for legacy X.509 services via MasterPortal

Community management:
- Comanage and Perun supported
- Other Community AAIs pluggable

Where to found the token in the EGI Notebooks







Get info about yourself

04-check-in.ipynb

Code

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])
```

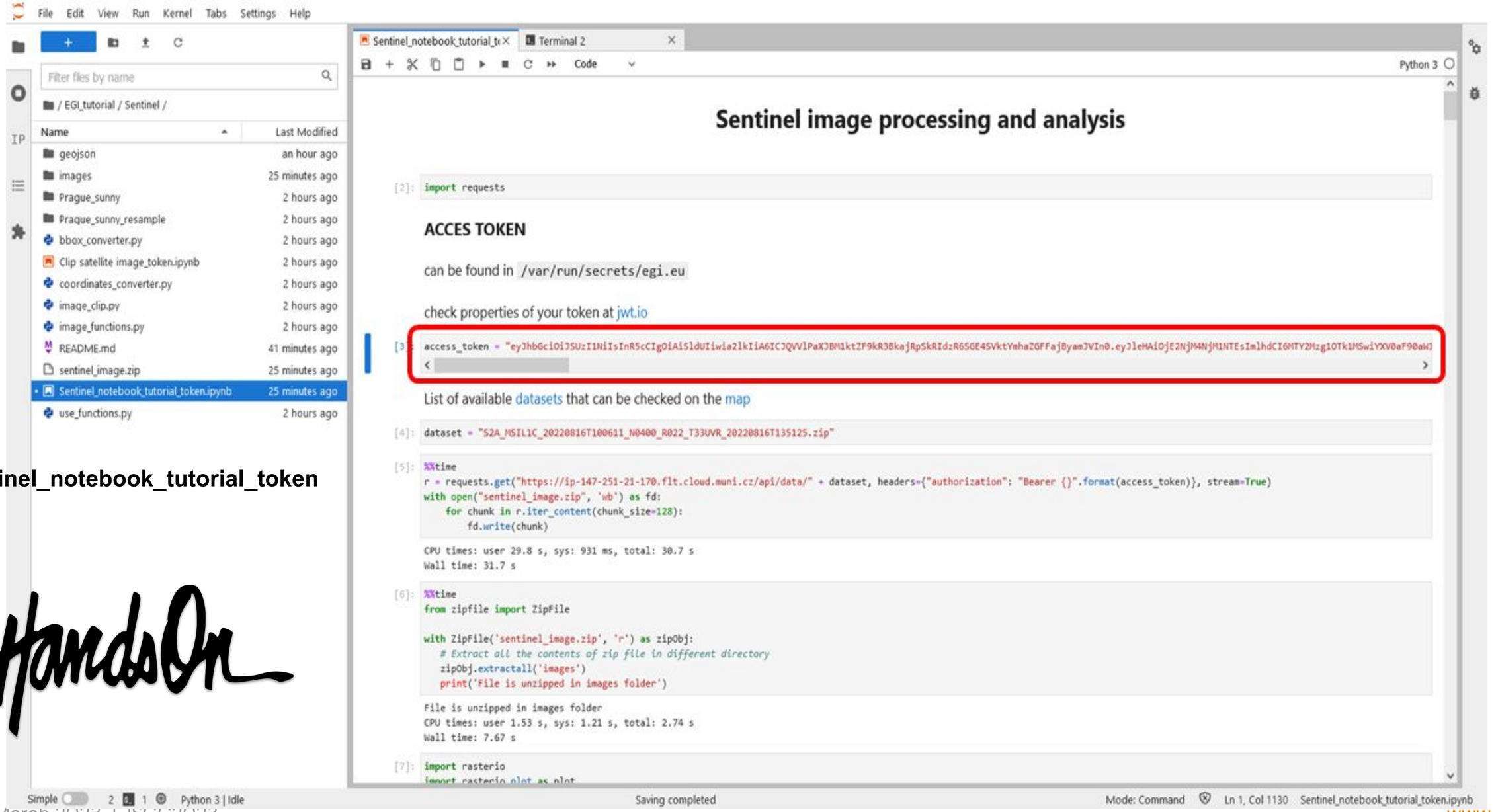
```
Got an check-in access token: eyJhbGciOi...
```

[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
```

HandsOn

Sentinel image processing and analysis



Sentinel_notebook_tutorial_token

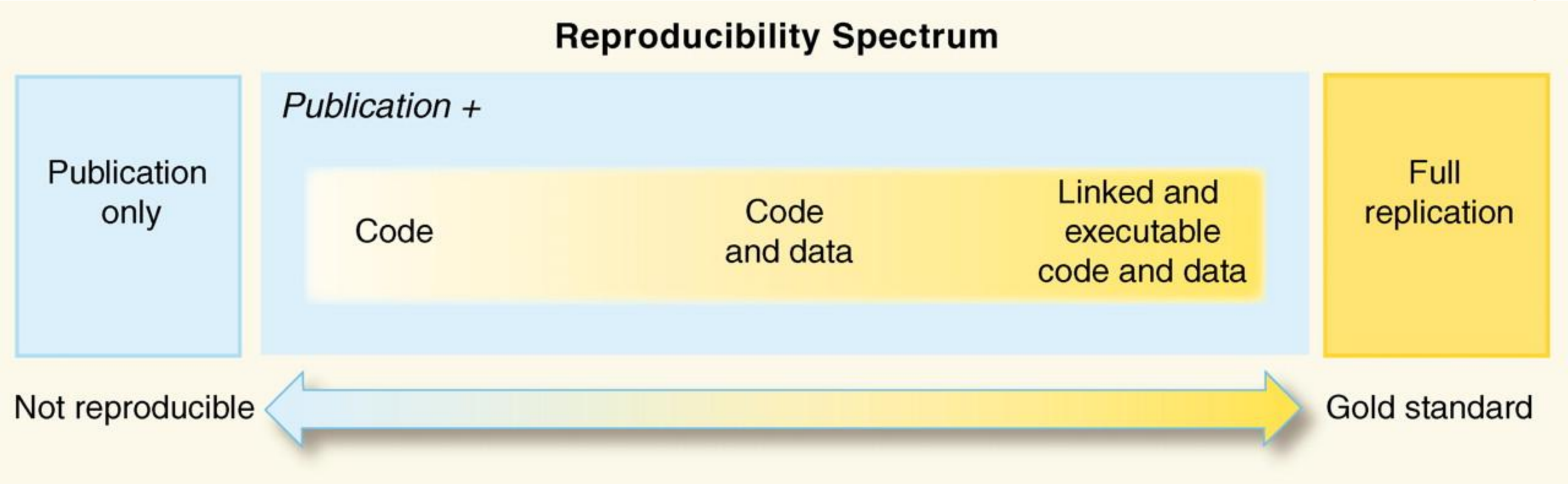
HandsOn

Approach to reproducible Open Science with EGI and EOSC

TLP: GREEN Limited disclosure

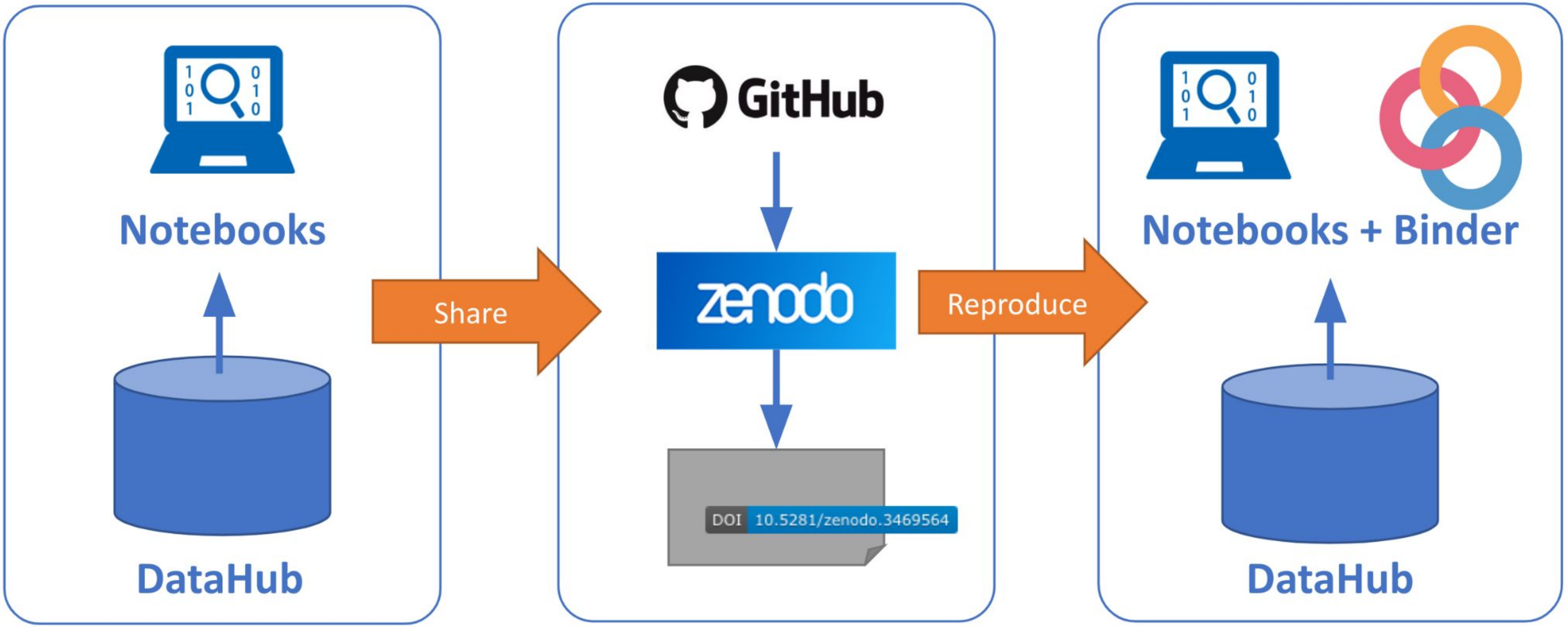
ISGC 2023

Reproducibility: beyond sharing code and data



Peng, Science, 2011

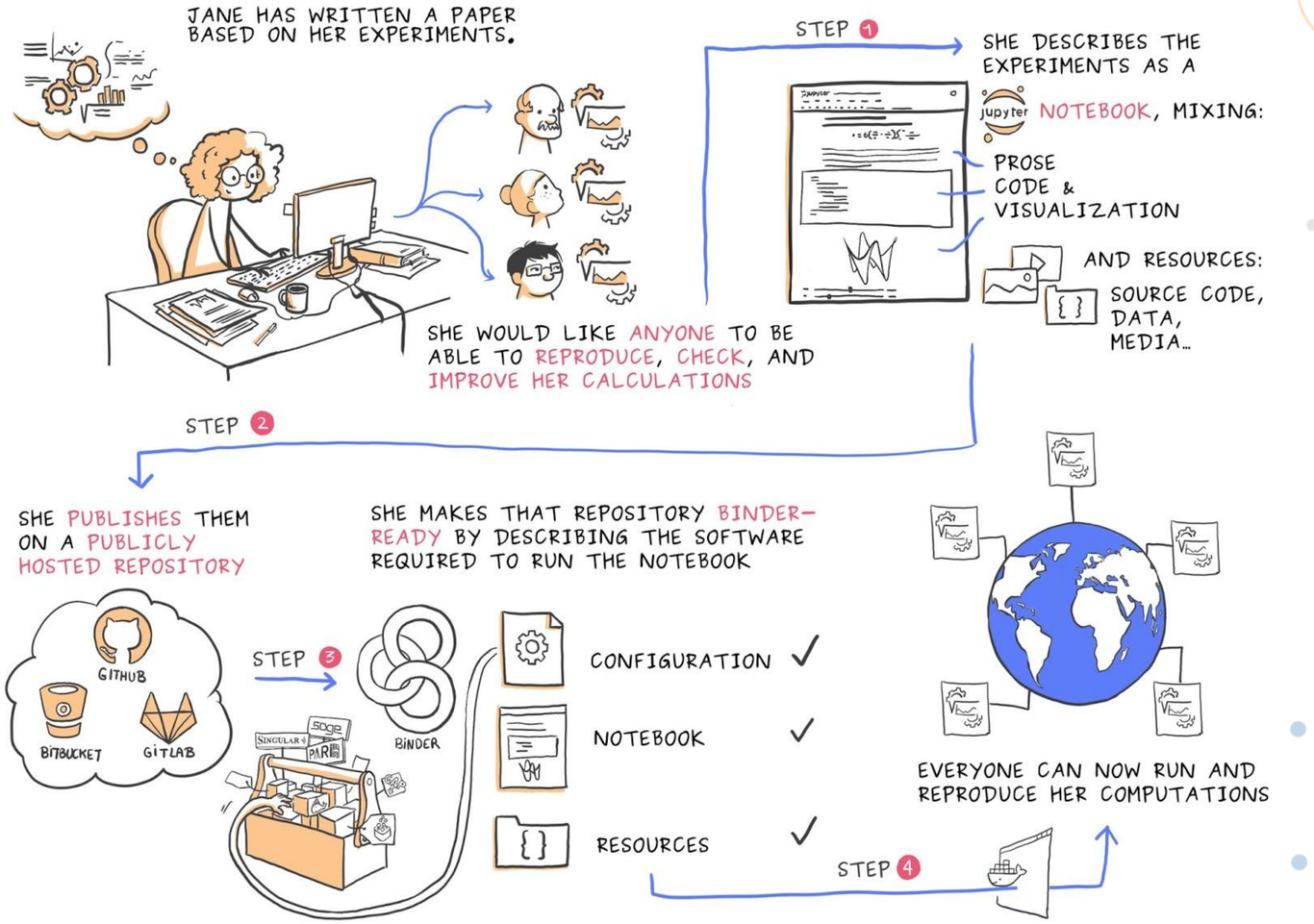
Reproducible Open Science in EGI/EOSC



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)



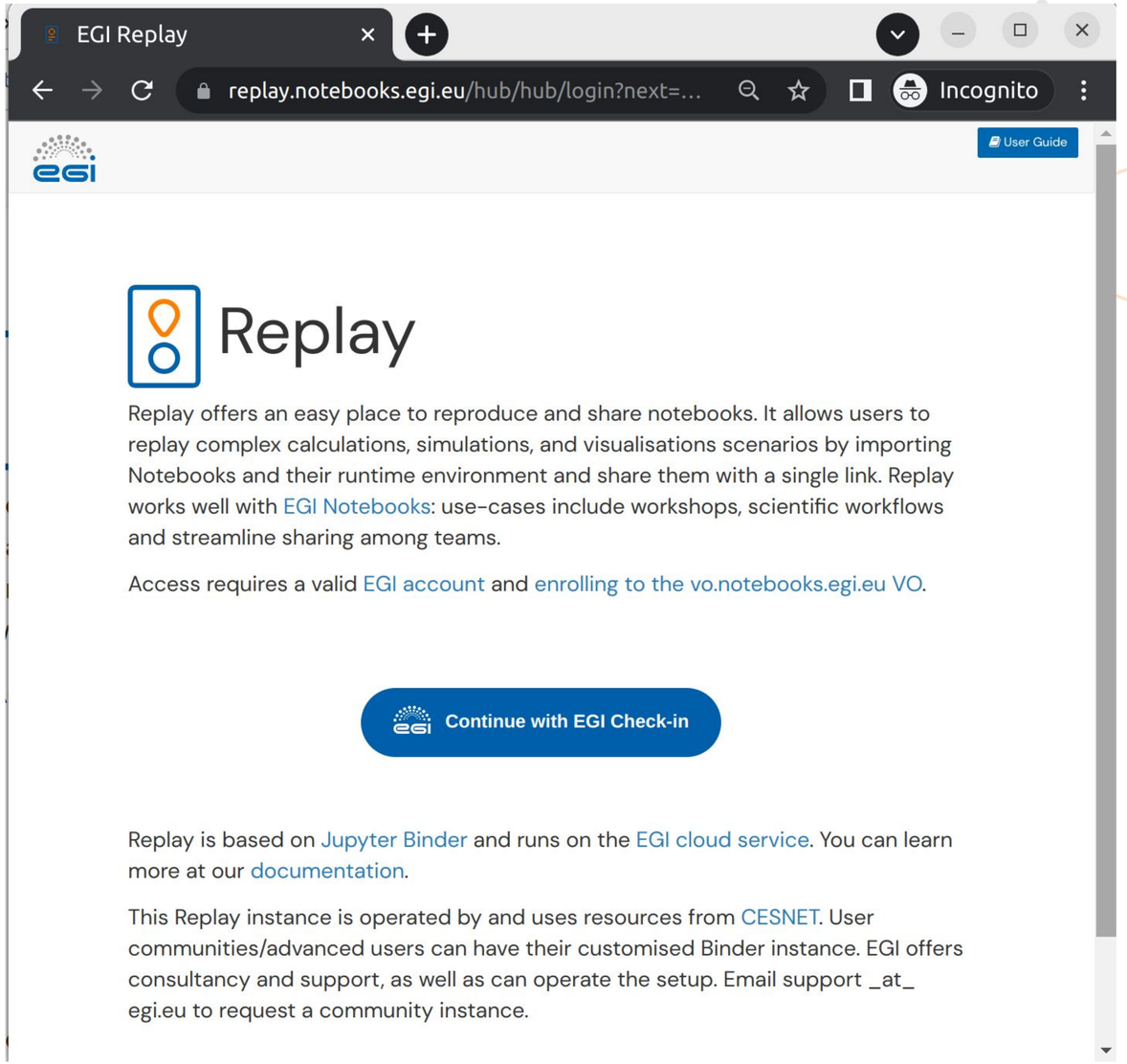
EGI Replay

BinderHub hosted by EGI

- Offered ‘as Service’
- Same access conditions as EGI Notebooks

Main Features:

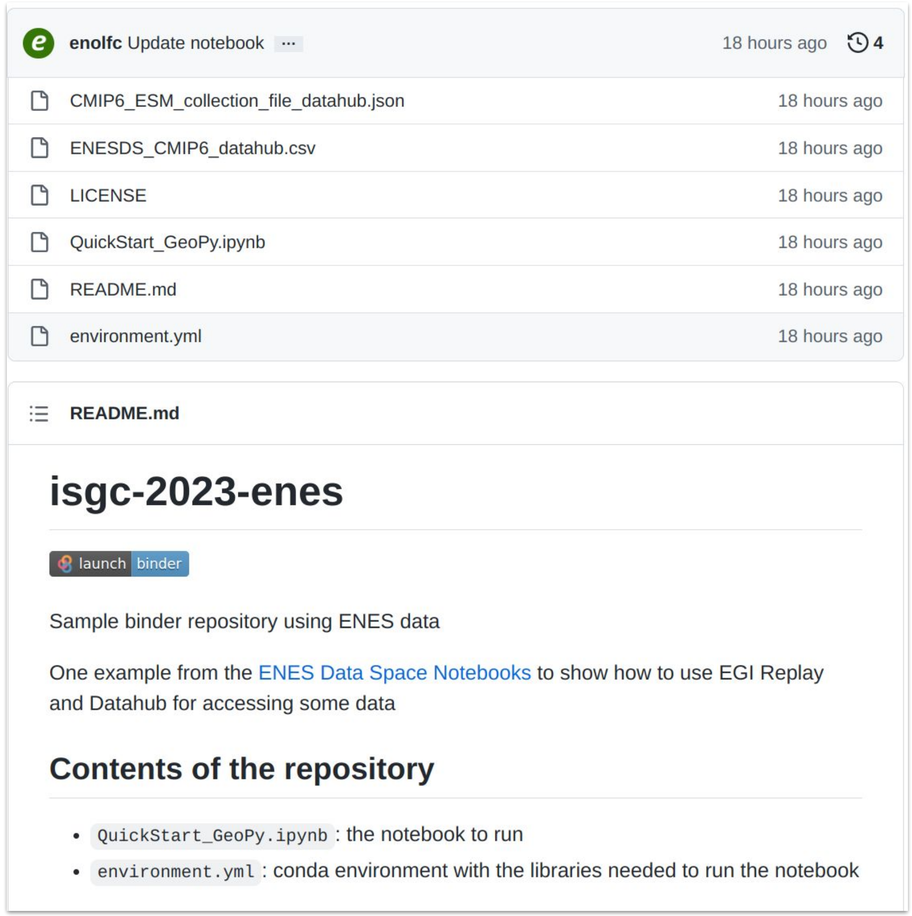
- Use any binder-compatible repository
- Reproduce your notebooks with access to EGI resources (e.g. DataHub)
- No hard limits on sessions duration, customisable resource limits for users/communities



Example: Number of Summer Days in Taipei with EGI Replay

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

launch binder



QuickStart_GeoPy

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

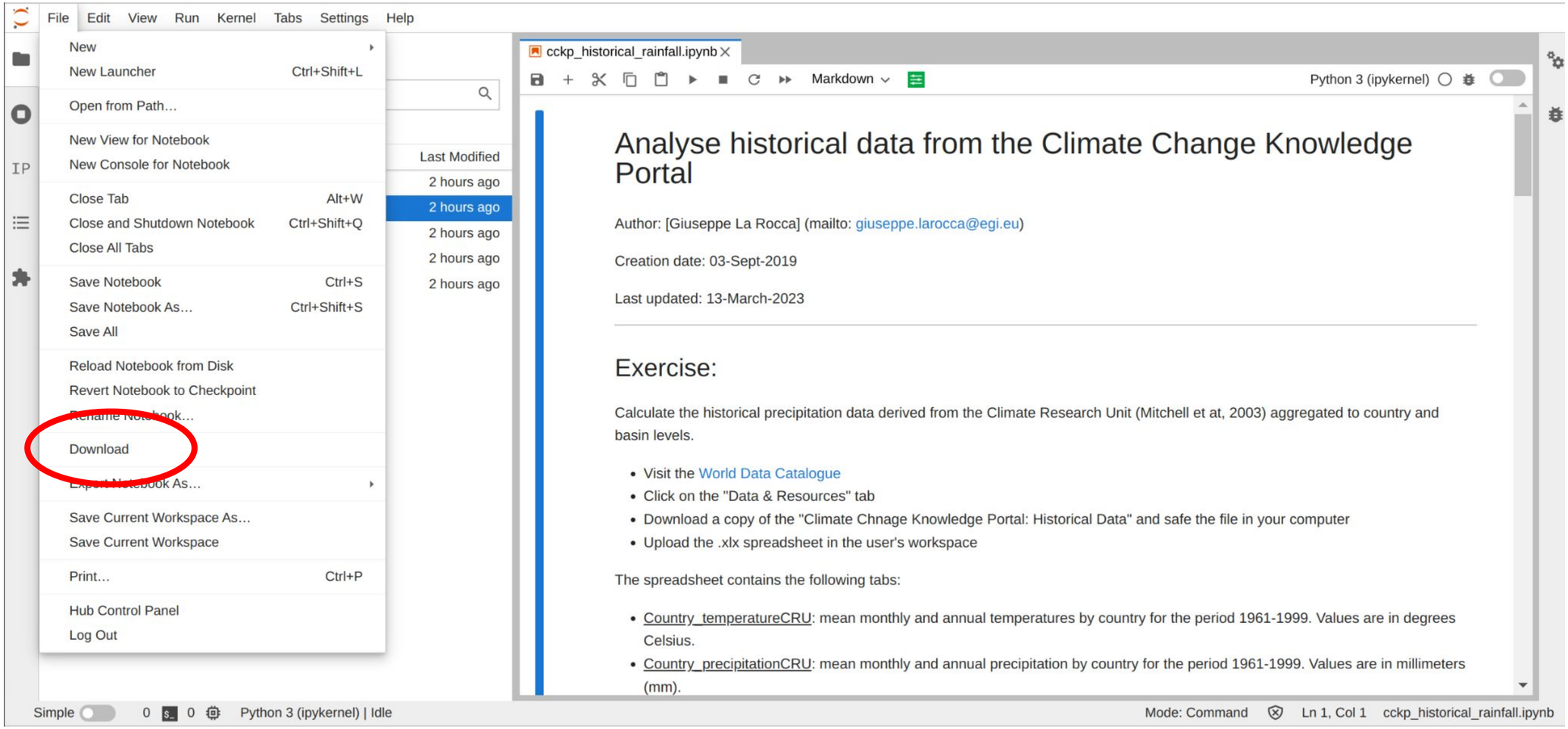
See: "How to make your datasets available in DataHub"

HandsOn

How to make notebooks Reproducible and Shareable

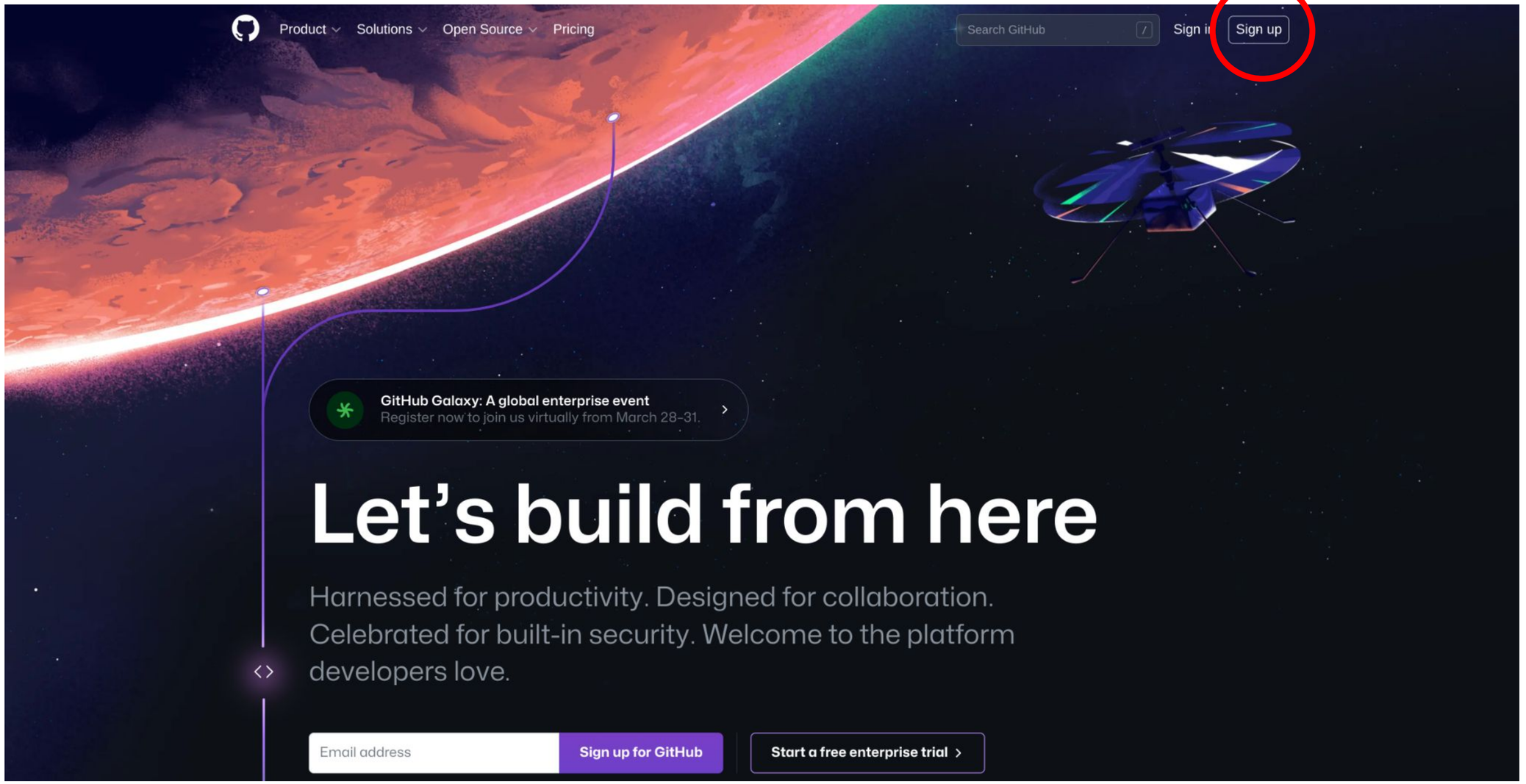
TLP: GREEN Limited disclosure

1. Create your first sharable notebook



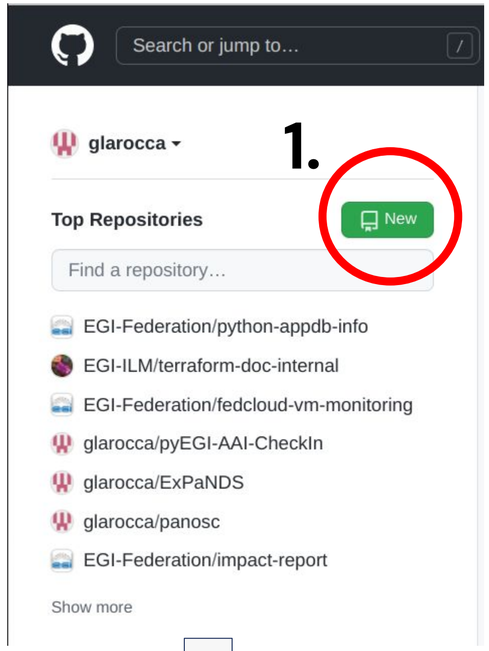
Download your first notebook in your laptop (*.ipynb)

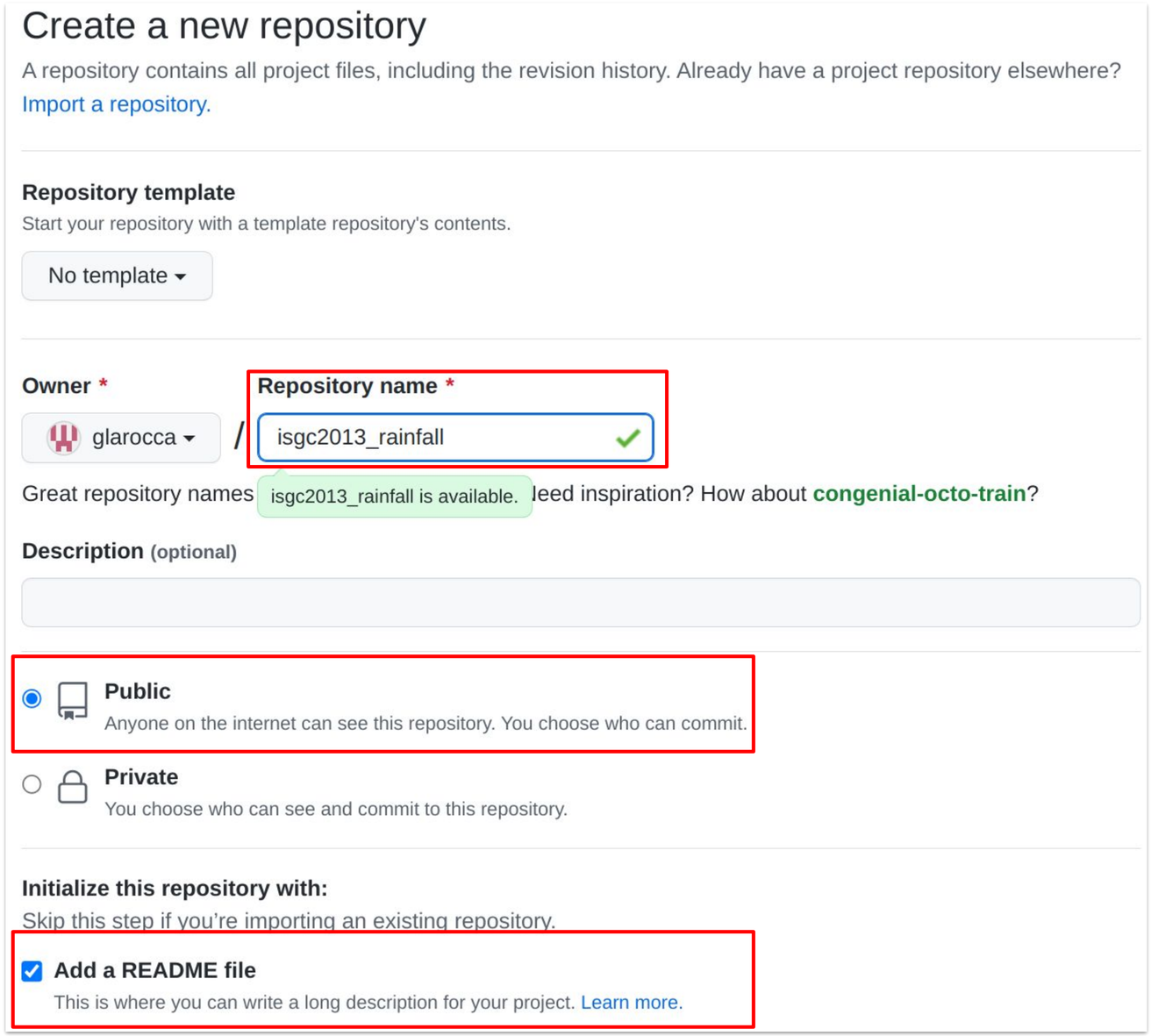
2. Getting a GitHub account

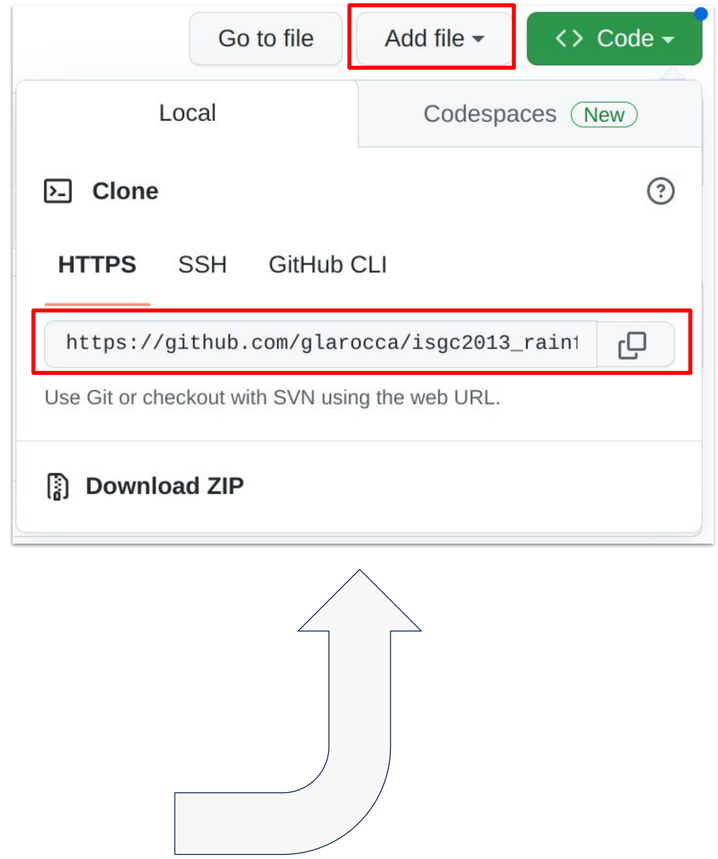


Sign-up if you don't have an account already

3. Creating a new GitHub repository







4. Adding dependencies

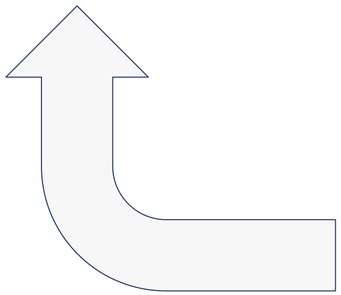
Create the ‘requirements.txt’ file in your repo

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 ▾ | https://github.com/glarocca/isgc2013_rainfall.git

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/gh/glarocca/isgc2013_rainfall.git/HEAD
```

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

m

```
[![Binder](https://replay.notebooks.egi.eu/badge_logo.svg)](https://replay.notebooks.egi.eu/v2/gh/g
```

.rst

```
.. image:: https://replay.notebooks.egi.eu/badge_logo.svg
 :target: https://replay.notebooks.egi.eu/v2/gh/glarocca/isgc2013_rainfall.git/HEAD
```

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-glaroc
ca5 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.fedclo
ud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8
a0018cd410f"
2023-03-13T16:10:56Z [Normal] Successfully pulled image "docker-notebooks.fedclo
ud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb07248485
0d231d2542c8a0018cd410f" 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:sh
a-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
Successfully pushed docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5fr
ainfall-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/binde
r-glarocca-2disgc2013-5frainfall-674ed6:dc2b600eb072484850d231d2542c8a0018cd410f"
```

Unable to launch the server

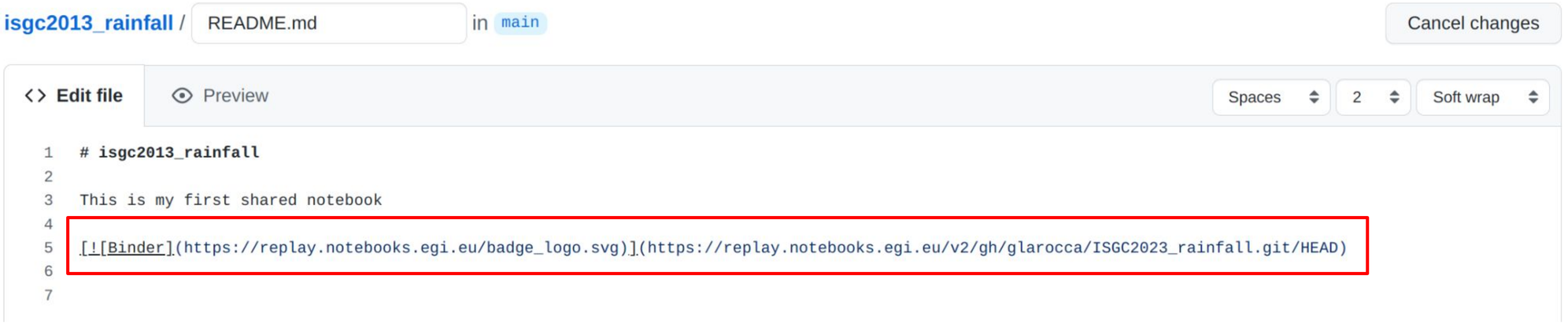
https://replay.notebooks.egi.eu/

Build logs view raw hide

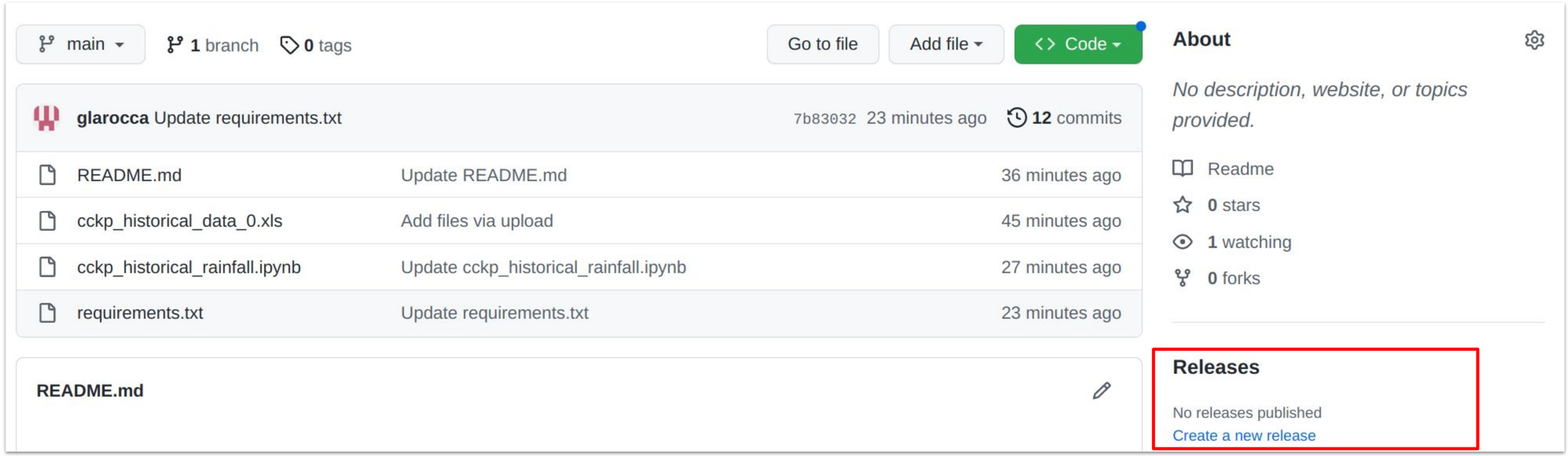
```
Step 49/51 : COPY /repo2docker-entrypoint /usr/local/bin/repo2docker-entrypoint
 ---> 3dda7421129e
Step 50/51 : ENTRYPOINT ["/usr/local/bin/repo2docker-entrypoint"]
 ---> Running in 96ff2f556184
Removing intermediate container 96ff2f556184
 ---> b9b8b51524cc
Step 51/51 : CMD ["jupyter", "notebook", "--ip", "0.0.0.0"]
 ---> Running in 2be6bdc29474
Removing intermediate container 2be6bdc29474
 ---> 46d5422a2fab
{"aux": {"ID": "sha256:46d5422a2fabb7693b3701aca6ef150b8298e0b31fb150c1eb3905585f3590a9"}}Su
ccessfully built 46d5422a2fab
Successfully tagged docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5fra
infall-674ed6:e842592725541fa91123f33e9ddbfdef1dcd1d73
Pushing image
Successfully pushed docker-notebooks.fedcloud-tf.fedcloud.eu/binder-glarocca-2disgc2013-5fr
infall-674ed6:e842592725541fa91123f33e9ddbfdef1dcd1d73Built 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.
```

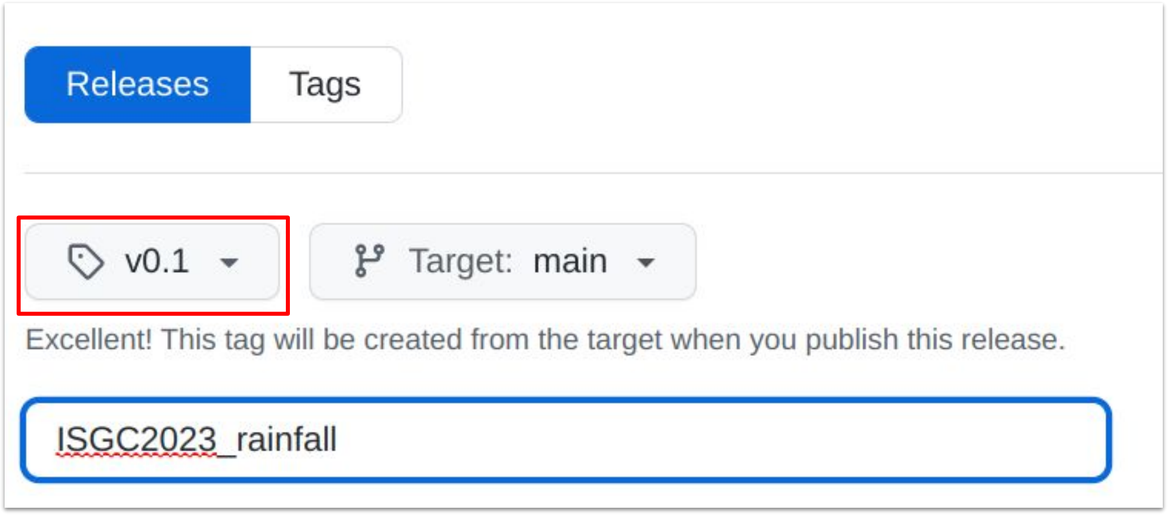
Only 1 server per user is allowed!
Stop server at: https://replay.notebooks.egi.eu/hub/hub/home

6. Adding Binder badge in the GitHub repository



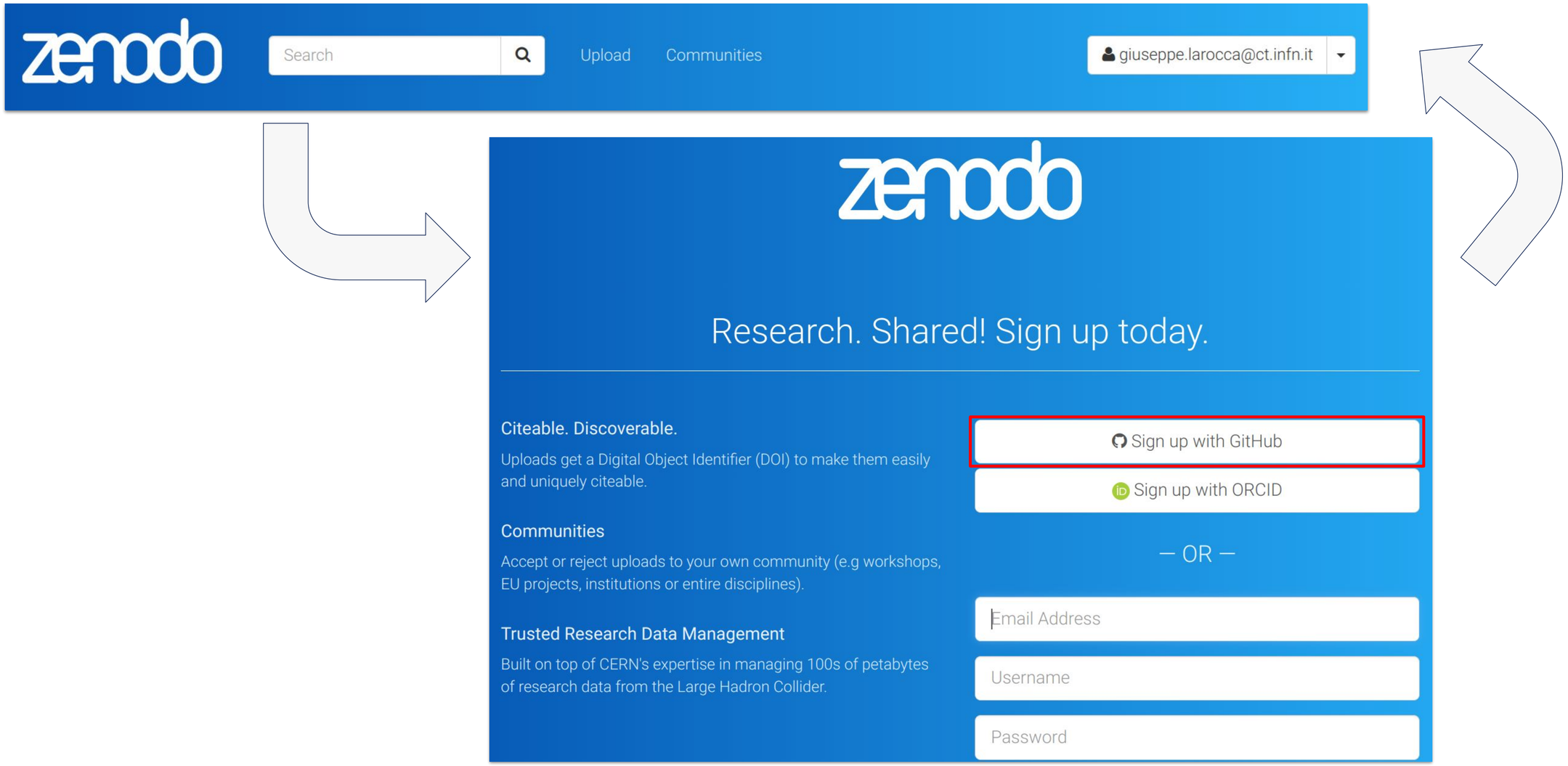
7. Create a new Release for the repository



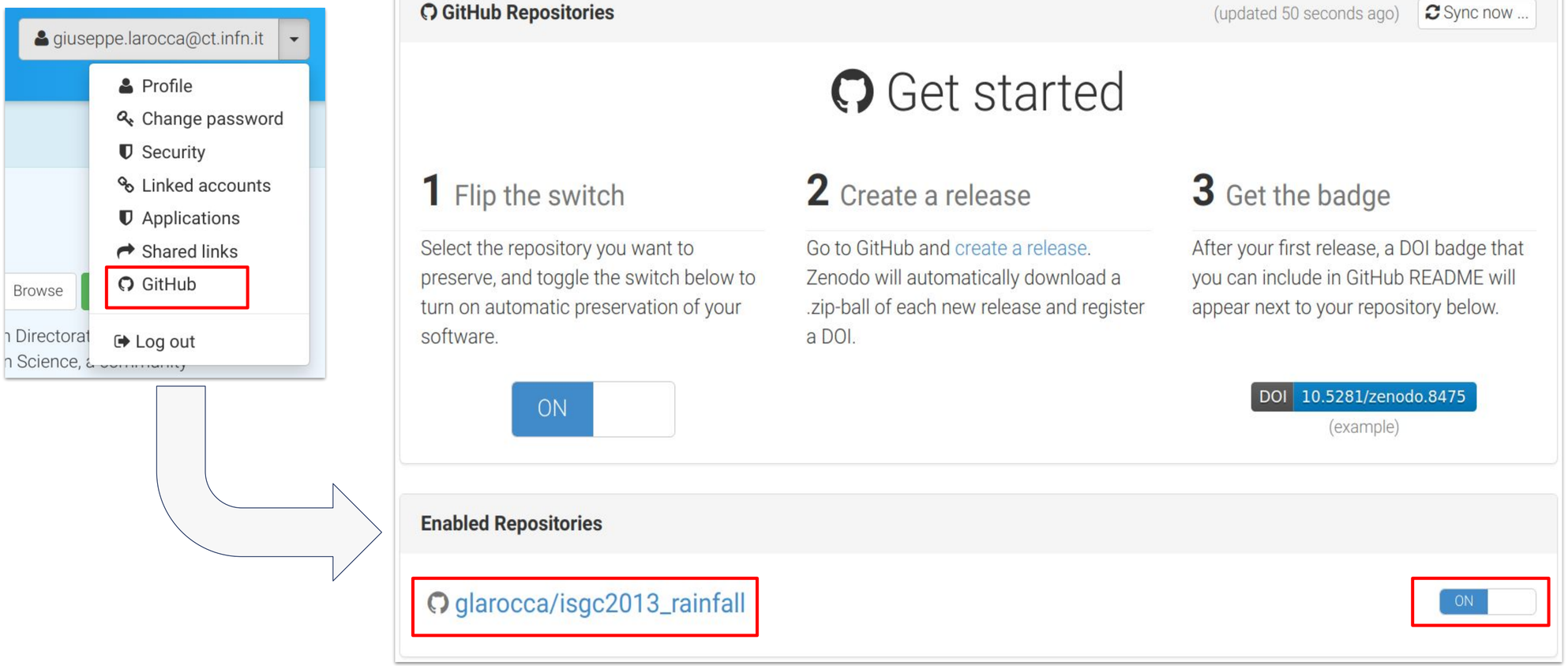


8. Making repository citable

https://zenodo.org/

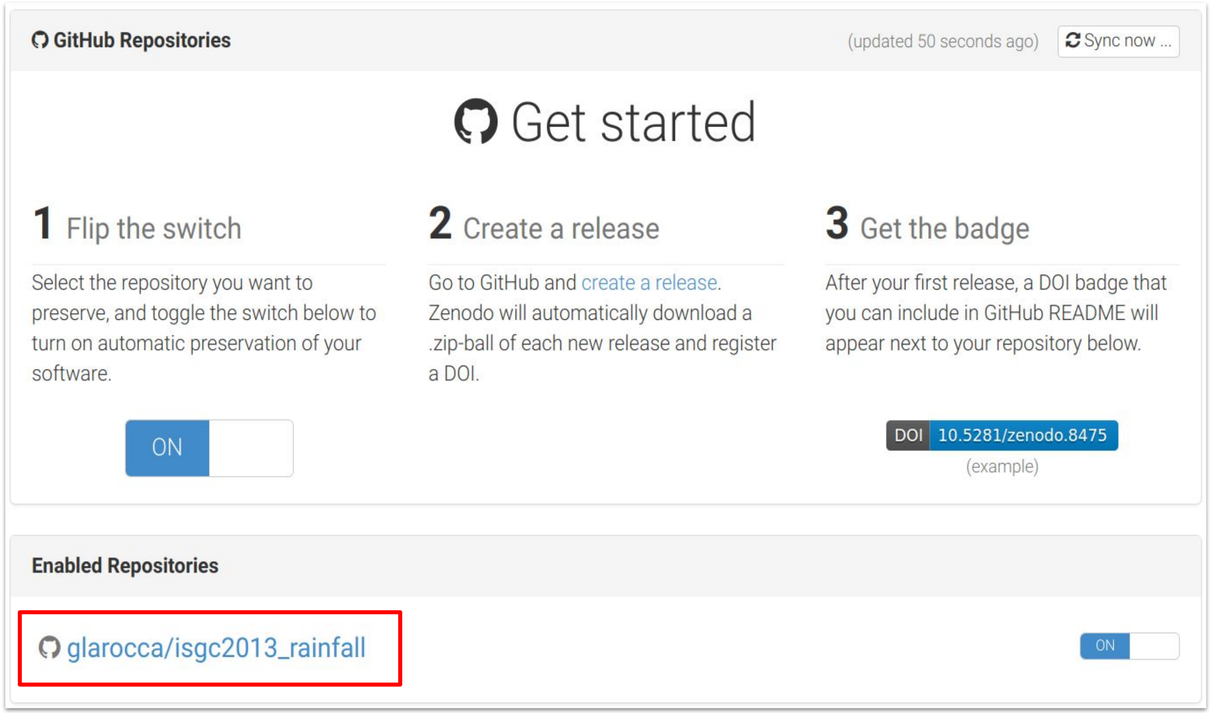


9. Synch Zenodo with repository



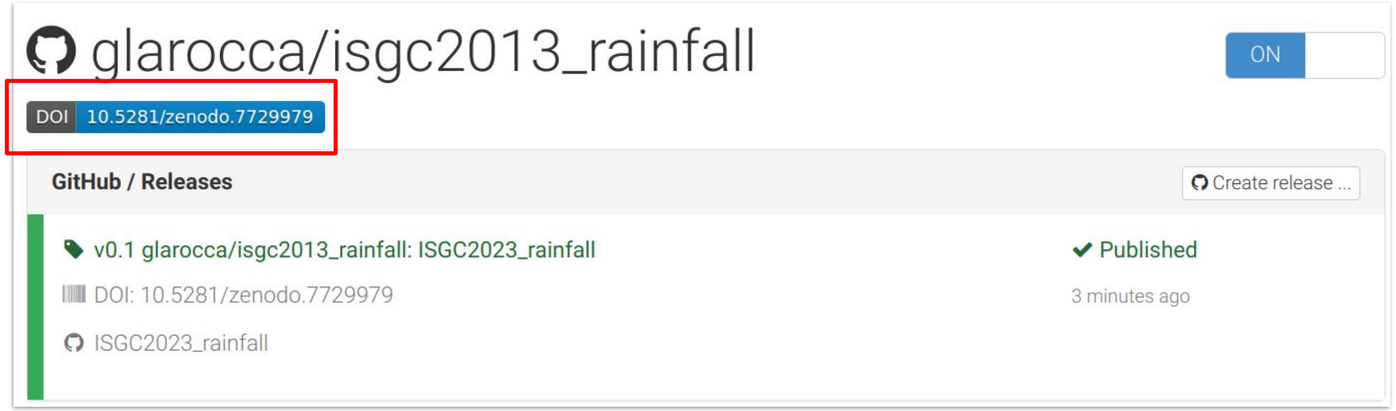
10. Get Zenodo DOI

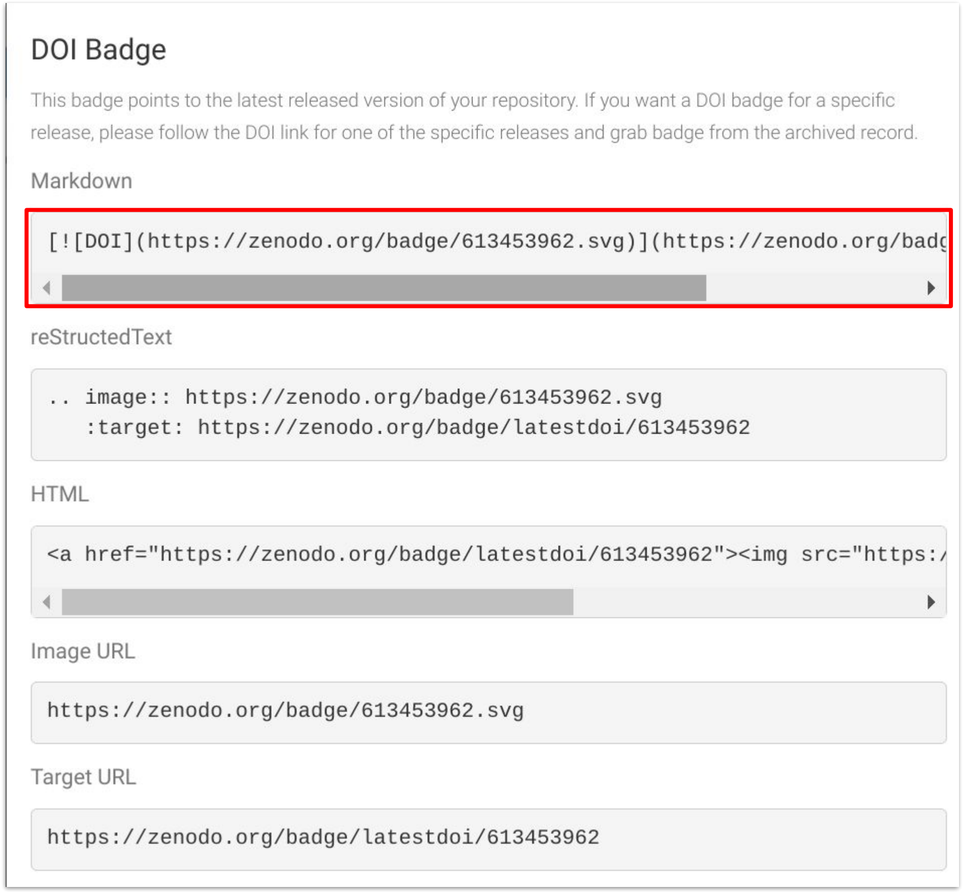
Click on the Enabled Repository



11. Add Zenodo badge in repository

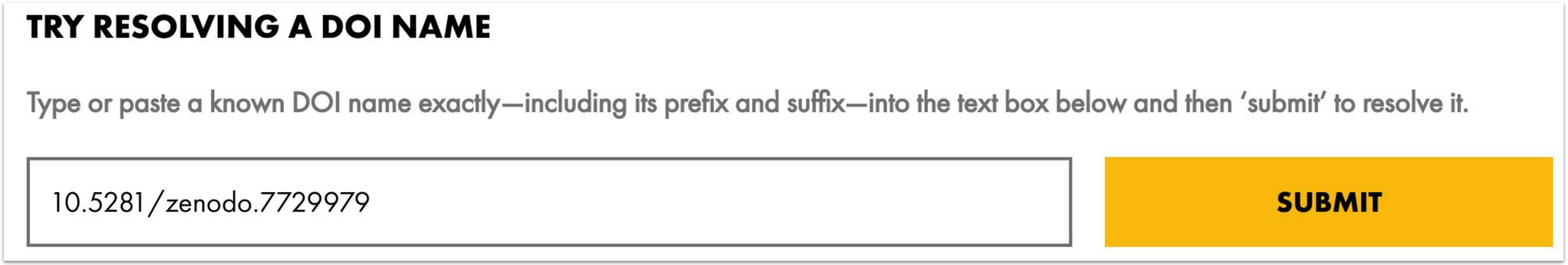
Click on the DOI icon to view details





12. Check if the DOI is valid

https://www.doi.org/



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) | **Path to a notebook file (optional)**

HEAD | 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/`

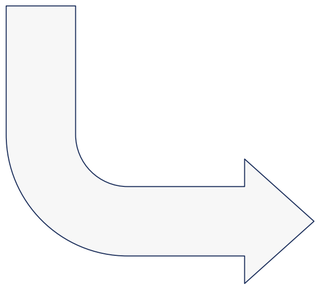
Expand to see the text below, paste it into your README to show a binder badge: launch binder

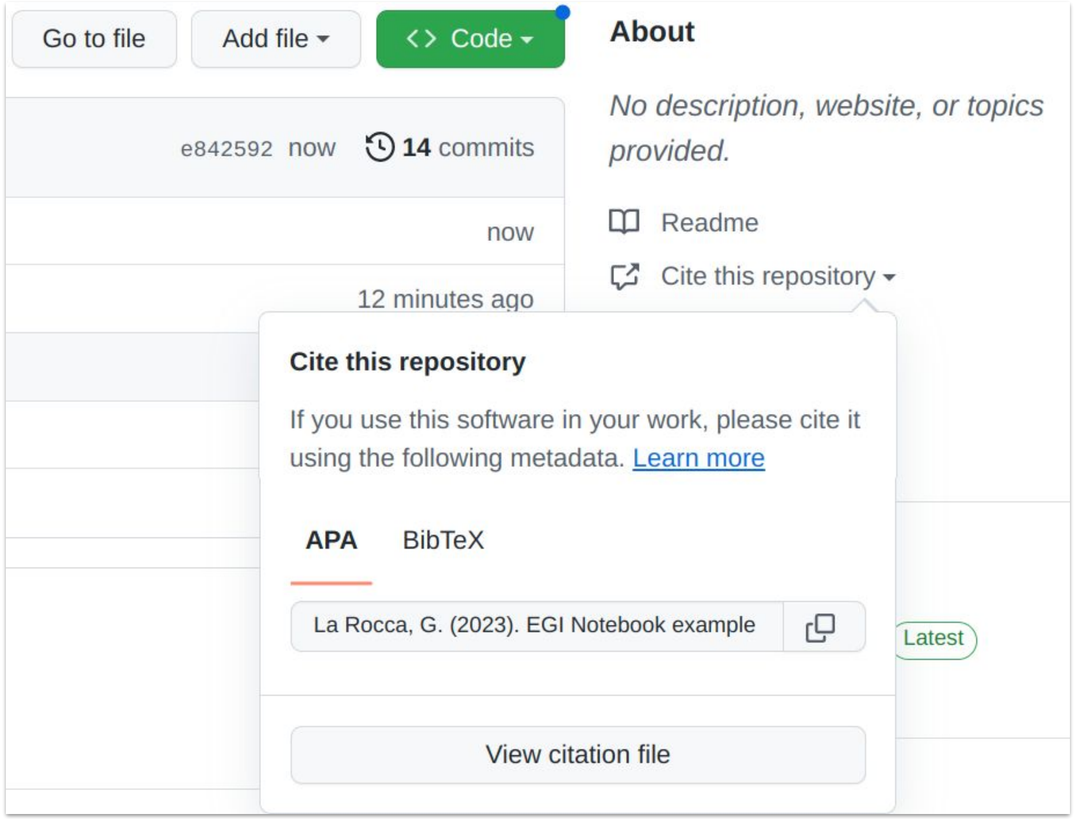
14. About CITATION files to your GitHub repository

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"
```





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/

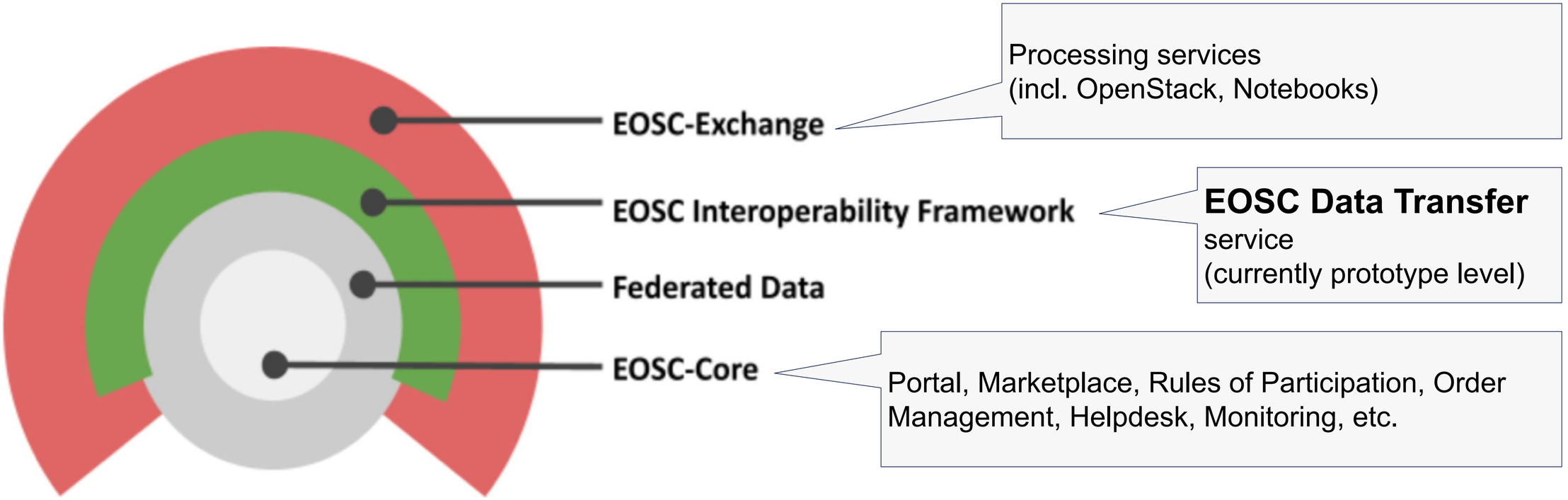
The **European Open Science Cloud (EOSC)** is an environment for hosting and processing research data to support EU science.

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 Architecture



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

EOSC Data Transfer

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

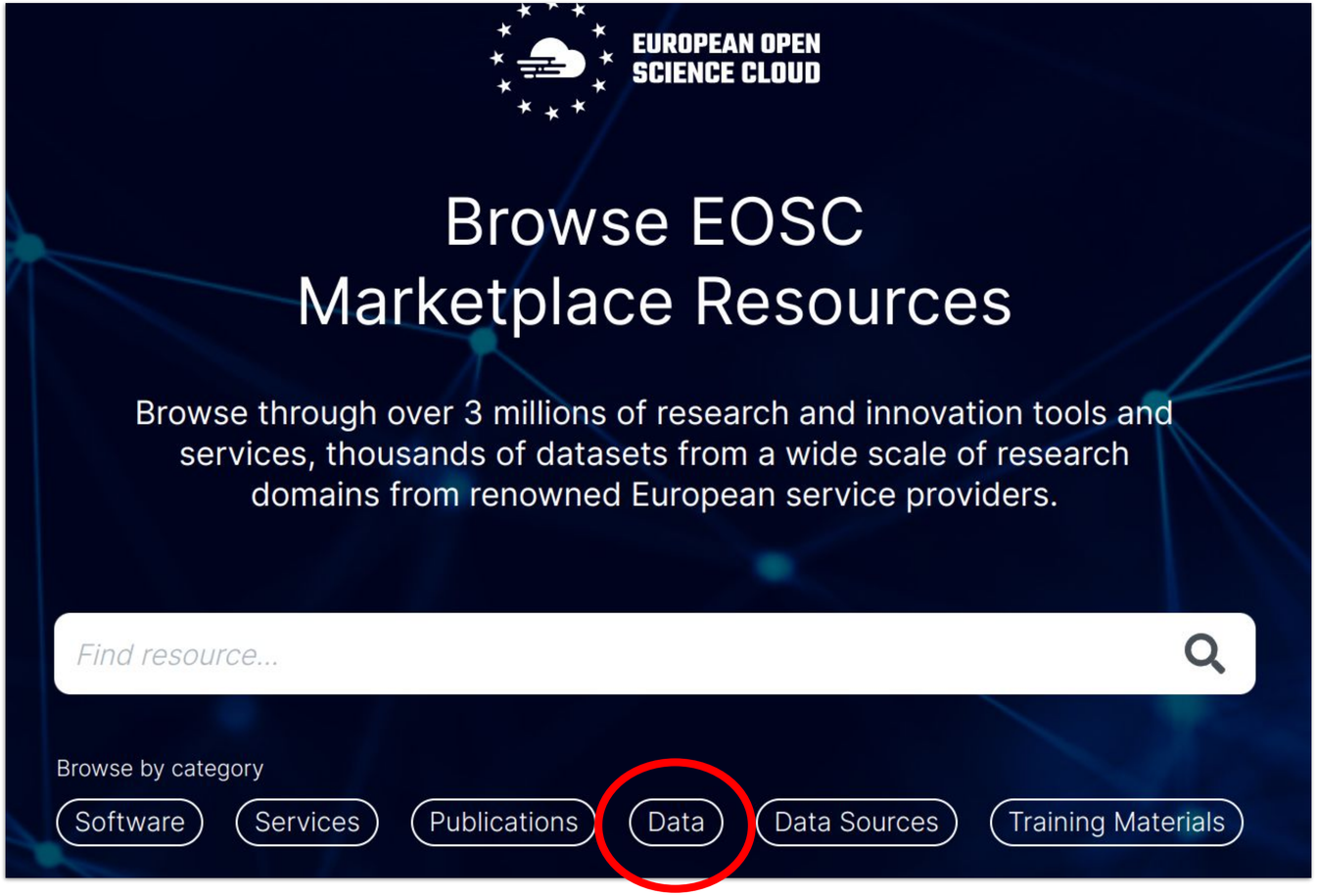
6.) Use the AWS CLI to check the files transferred to the destination

EOSC Data Transfer

DEMO

1.) Find the dataset in the EOSC Marketplace

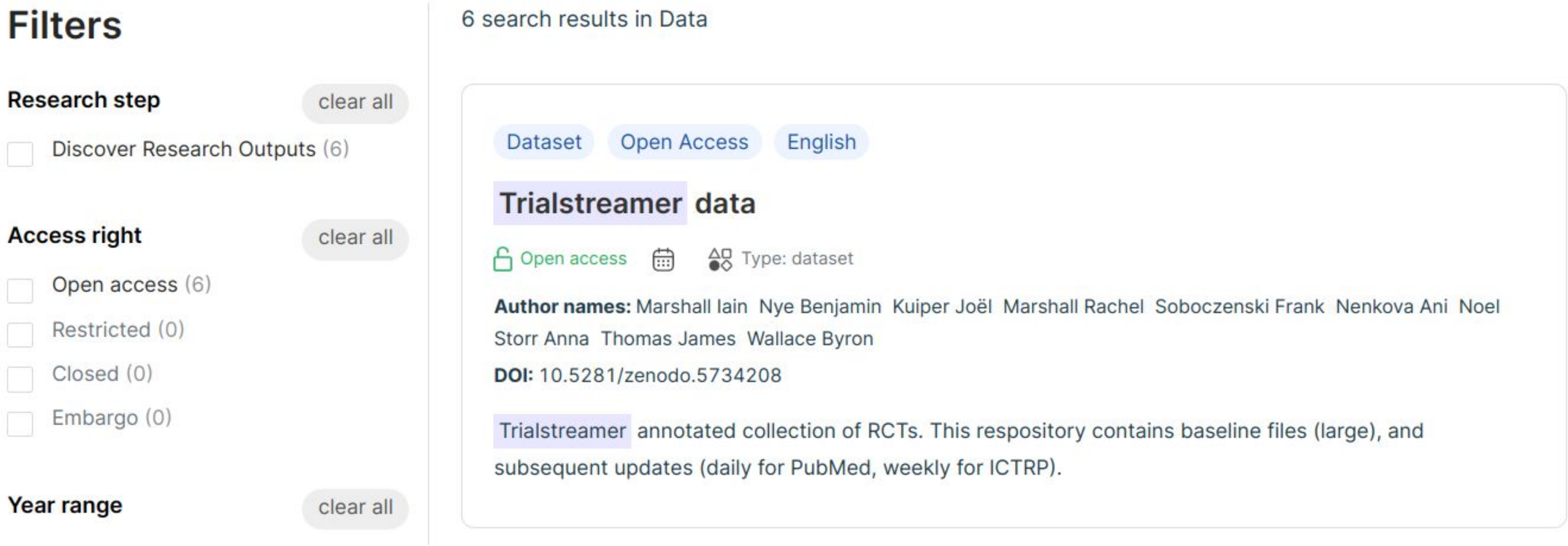
- Access the EOSC Marketplace: https://marketplace.eosc-portal.eu/
- Select Data as “Category”



EOSC Data Transfer

DEMO

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

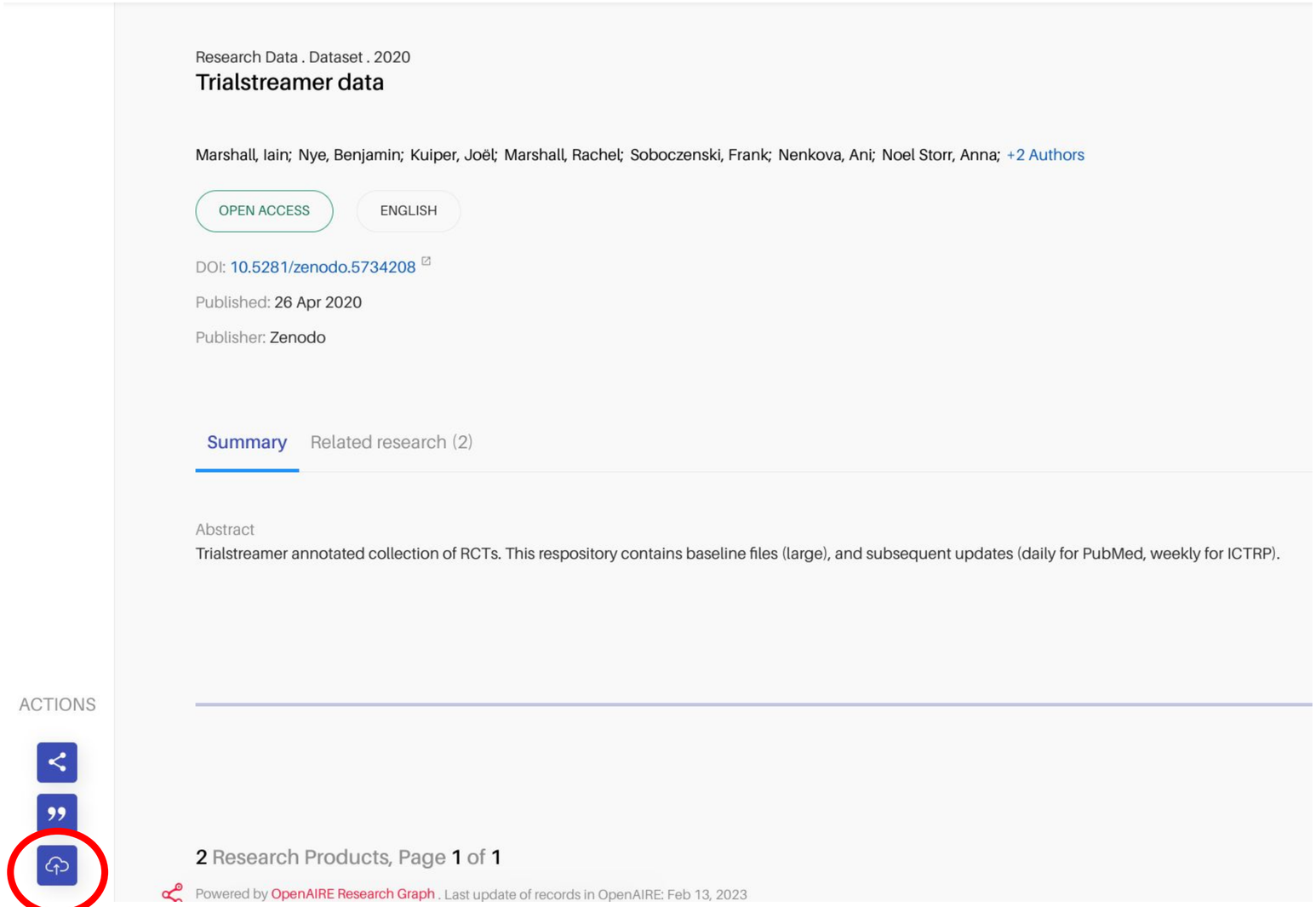


DOI: 10.5281/zenodo.5734208

EOSC Data Transfer

DEMO

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



WARNING

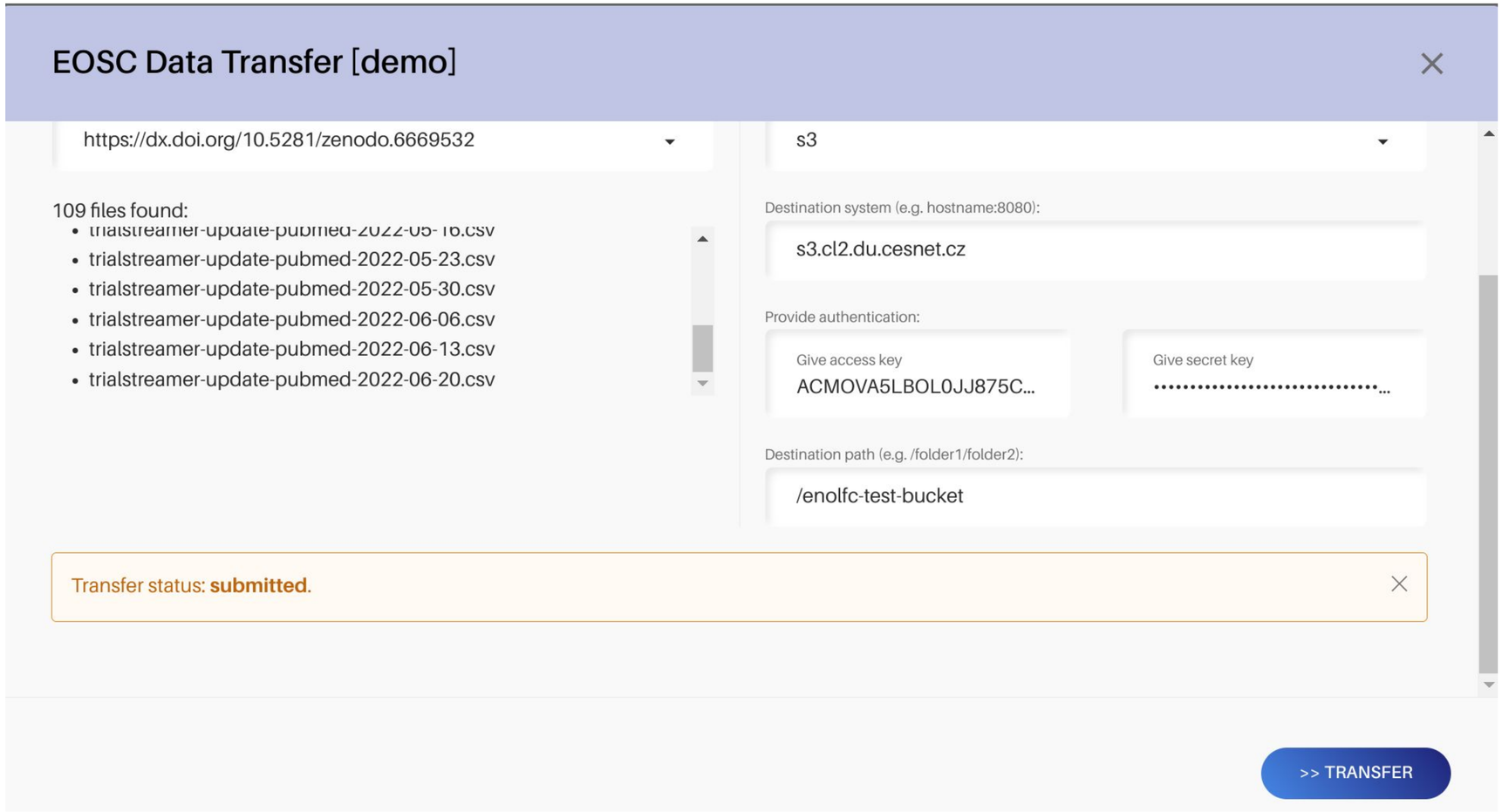
– Datasets must be stored in Zenodo!

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

EOSC Data Transfer

DEMO

4.) Start the data transfer using the dev instance



5.) Click on “Transfer” to initiate the data transfer

EOSC Data Transfer

DEMO

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 tasmax_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20150101-20391231.nc
2023-03-13 18:07:42 1150771873 tasmax_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20400101-20641231.nc
2023-03-14 07:52:41 1148226564 tasmax_day_CMCC-ESM2_ssp585_r1i1p1f1_gn_20650101-20891231.nc
2023-03-14 07:55:09  504796434 tasmax_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