Reproducible Open Science with EGI Notebooks and Replay services

Giuseppe La Rocca

Community Support Team Lead. / EGI Foundation

24 March 2023

ISGC 2023

TLP: GREEN Limited disclosure

Vision

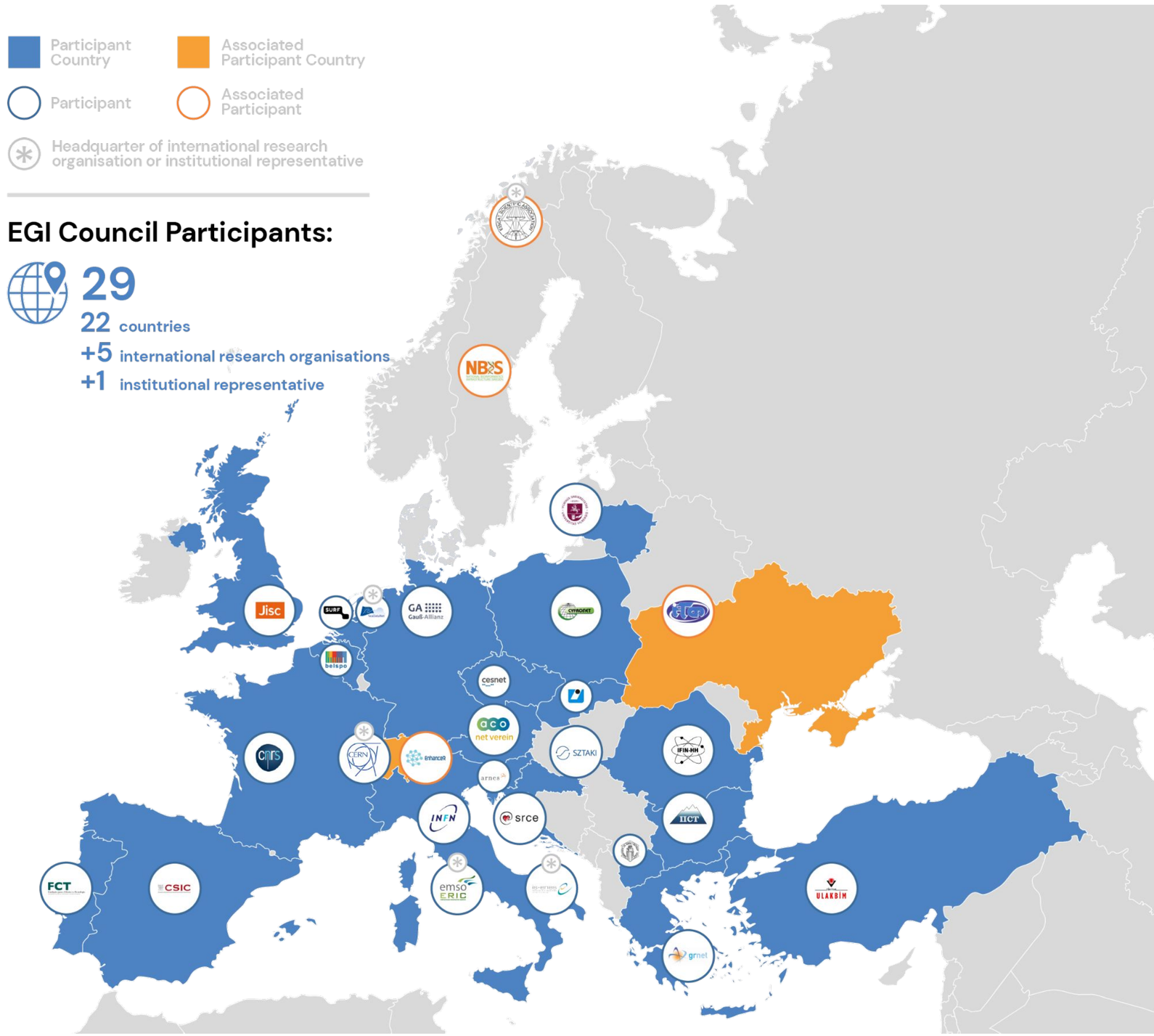
All researchers have seamless access to services, resources and expertise to collaborate and conduct world-class research and innovation

Mission of the EGI Federation

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

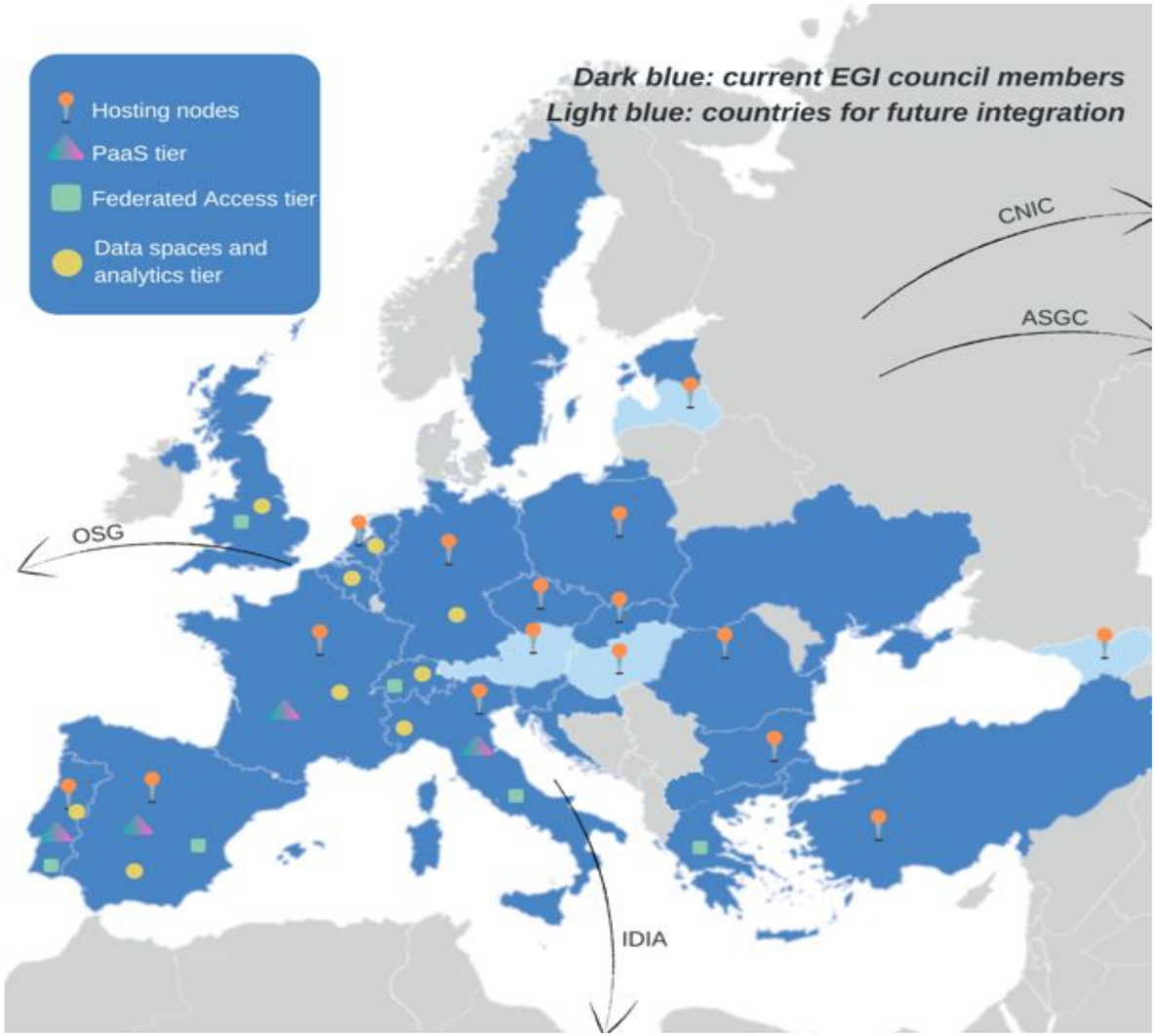
Mission of the EGI Foundation

Enable the EGI Federation to serve international research and innovation together



International Partnerships with MoUs

EGI-ACE - EGI flagship project



Consortium:

- Coordinator – Stichting EGI
- 33 Partners, 23 third parties
- Most of EGI members + Several RIs

Services:

- EGI Services for Research
- EGI Services for Federation
- EGI Services for Business

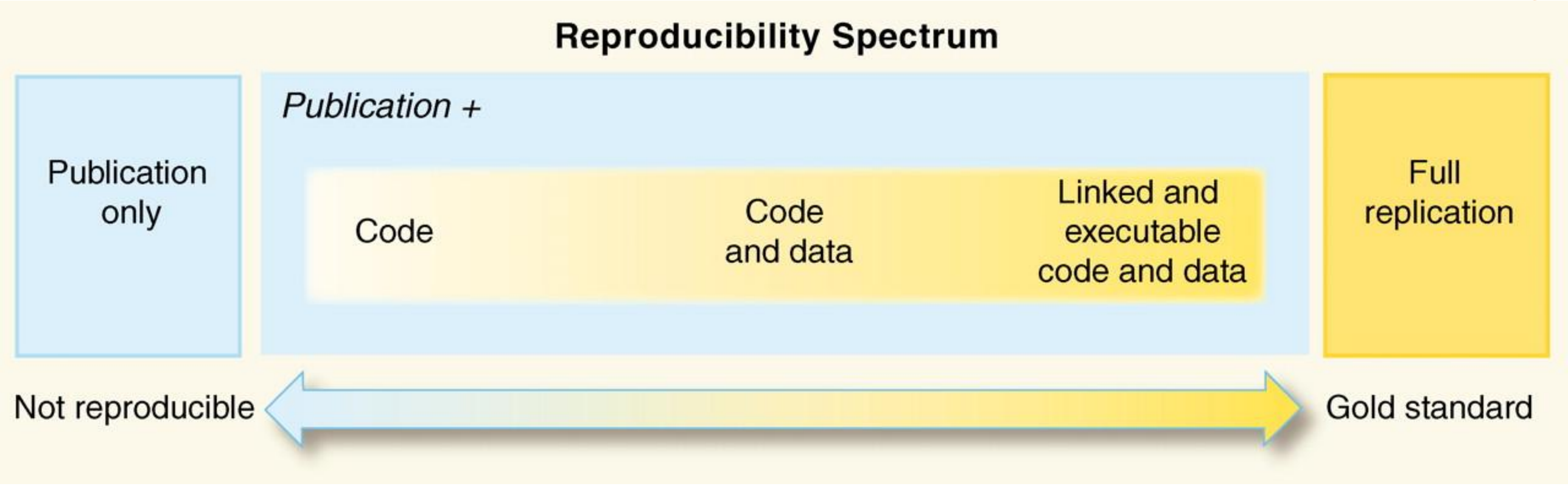
Scope:

- **49% service delivery (Virtual Access)**
- Co-development of services with RIs

Duration:

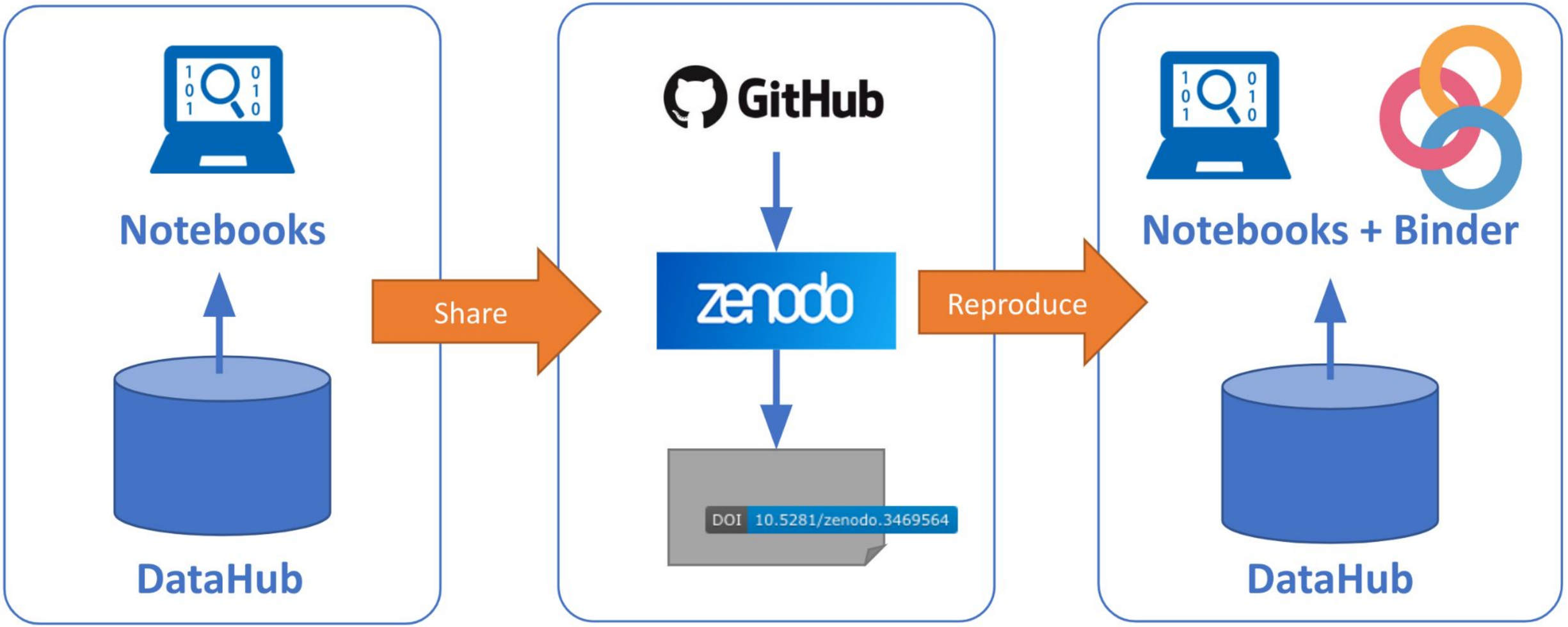
- Jan 2021 - June 2023 (30 months)

Reproducibility: beyond sharing code and data



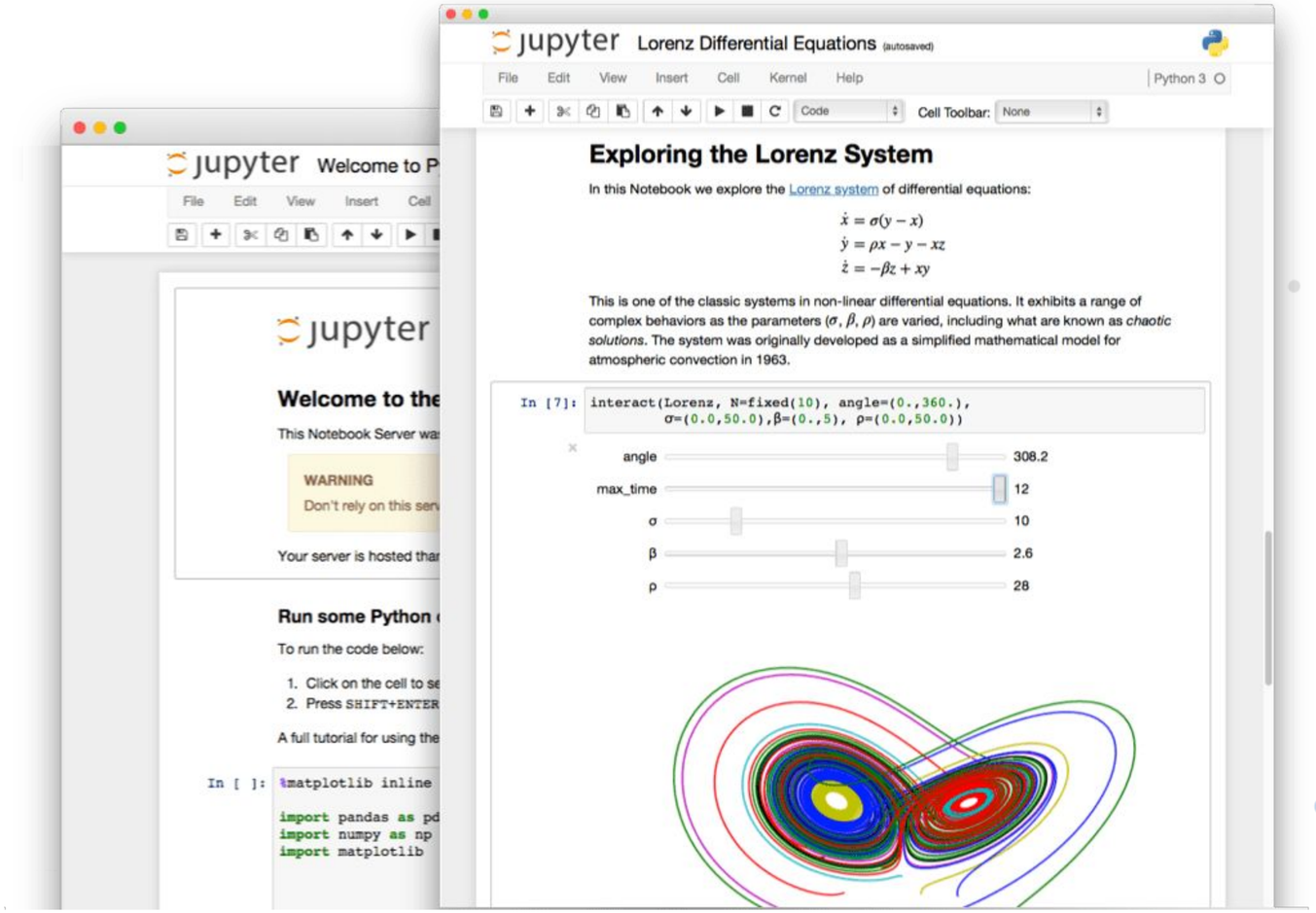
Peng, Science, 2011

Reproducible Open Science in EGI/EOSC



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.

Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



Documentation
Text formatted using Markdown/LaTeX

Interactive
browser based environment

JupyterLab

notebooks.egi.eu/user/529a87e5ce04cd5ddd7161734d02df0e2199a1145243080...

File Edit View Run Kernel Hub Tabs Settings Help

FirstNotebook.ipynb | Terminal 1

Code | Python 3

Start here

This is a documentation cell written with markdown

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

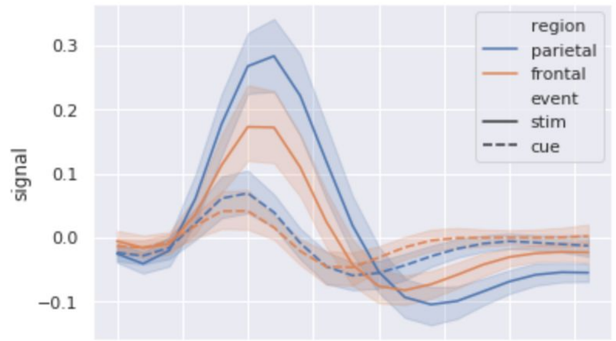
```
hello
```

```
[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)
```

```
[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7f4c15812e48>
```



Code
Use your favourite language

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

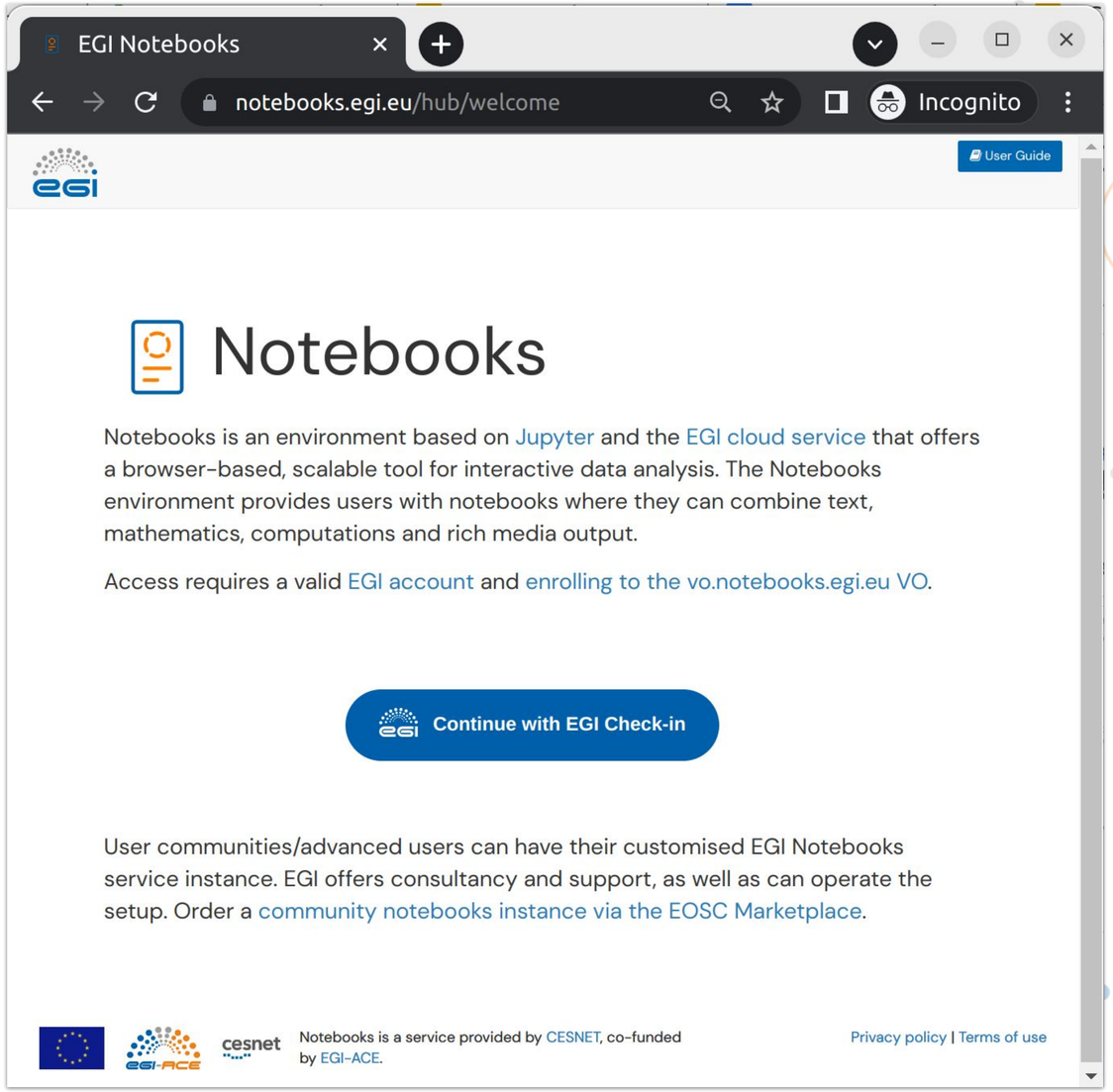
EGI Notebooks

Jupyter hosted in the EGI Cloud

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

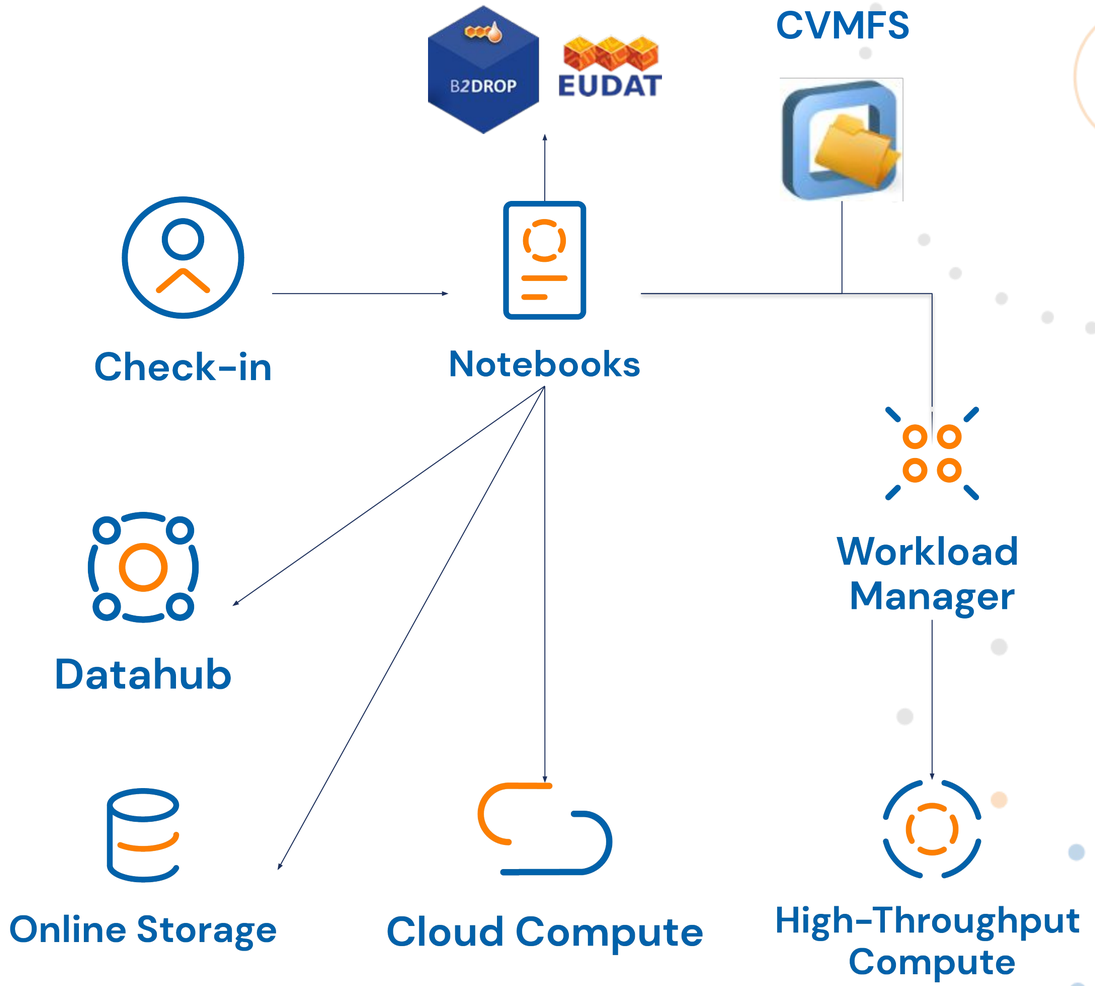
Main Features:

- 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

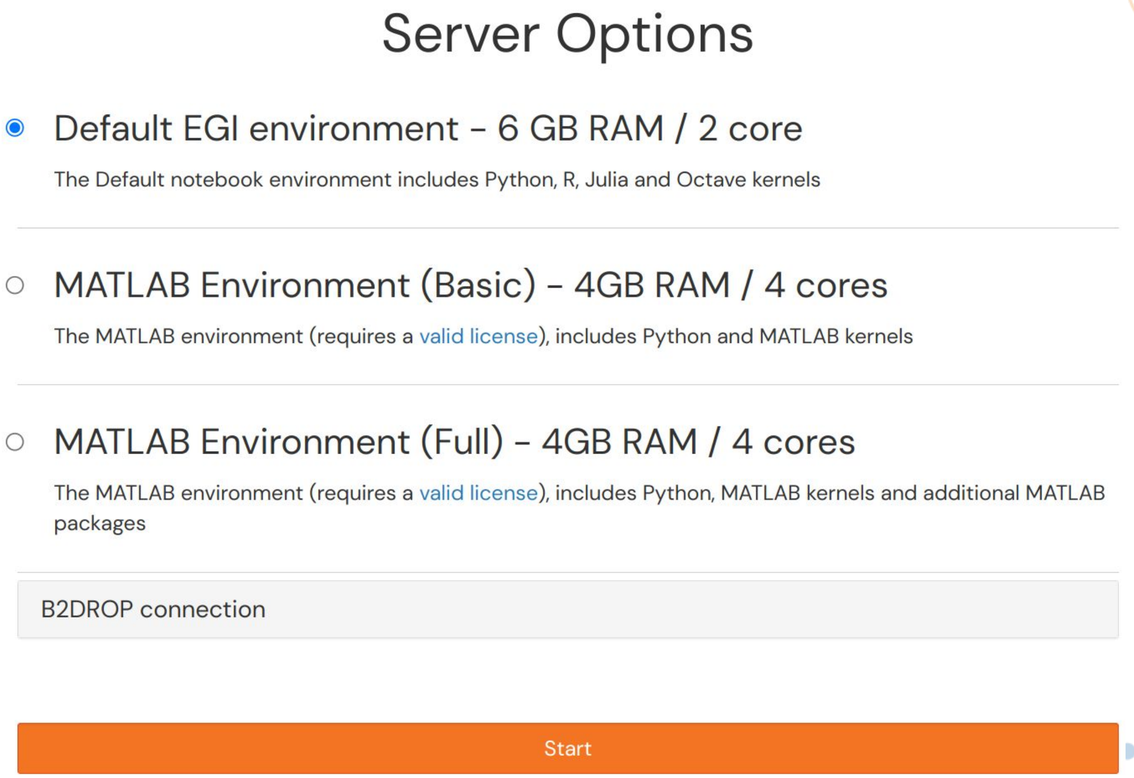
1. Runs on EGI Cloud providers (Cloud Compute / Online Storage)
2. Uses Check-in for authentication + access to other services
3. Makes user-level software available from CVMFS
   - DIRAC client for submission of jobs to Workload Manager / High Throughput Compute
   - Fedcloudclient for interaction with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



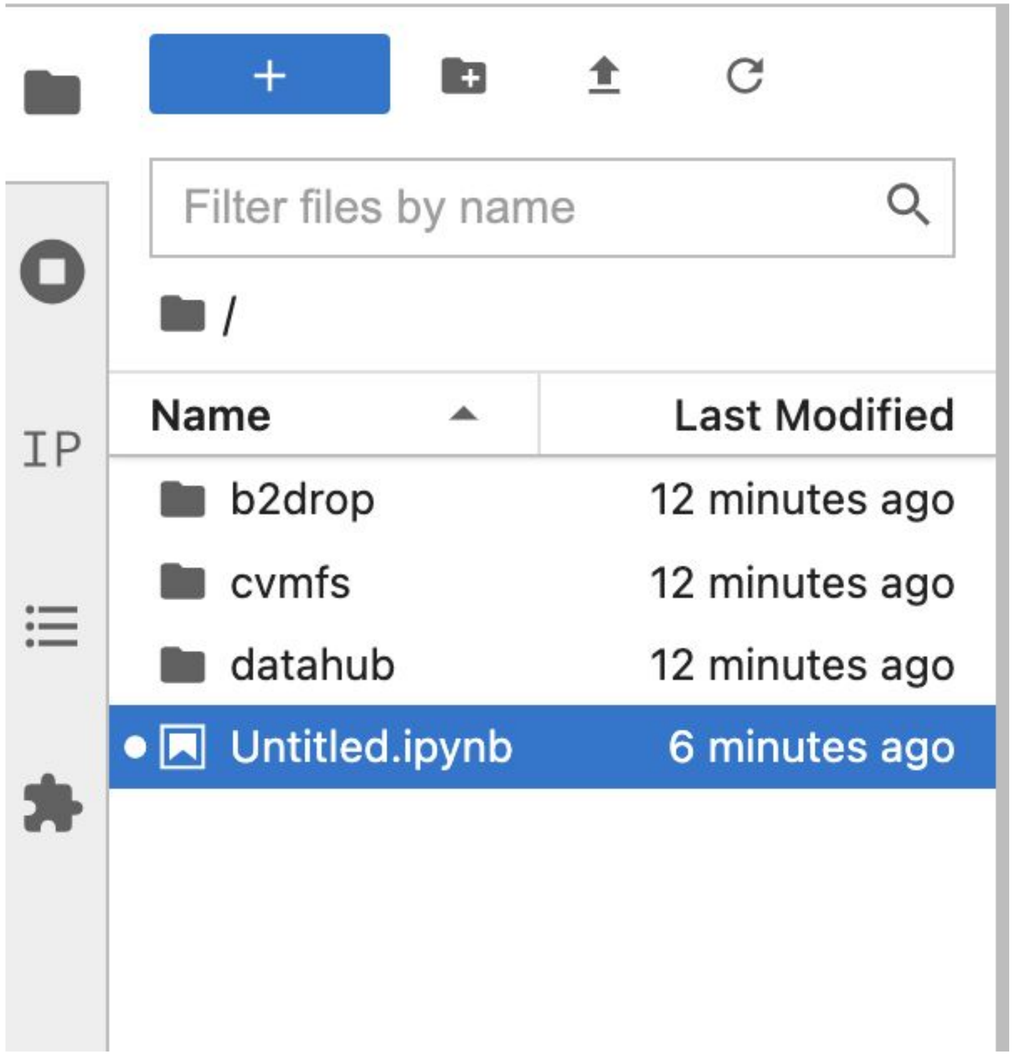
Customisable options

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!)
- **Community environments**: tuned to meet the needs of specific user communities



Accessing data and code

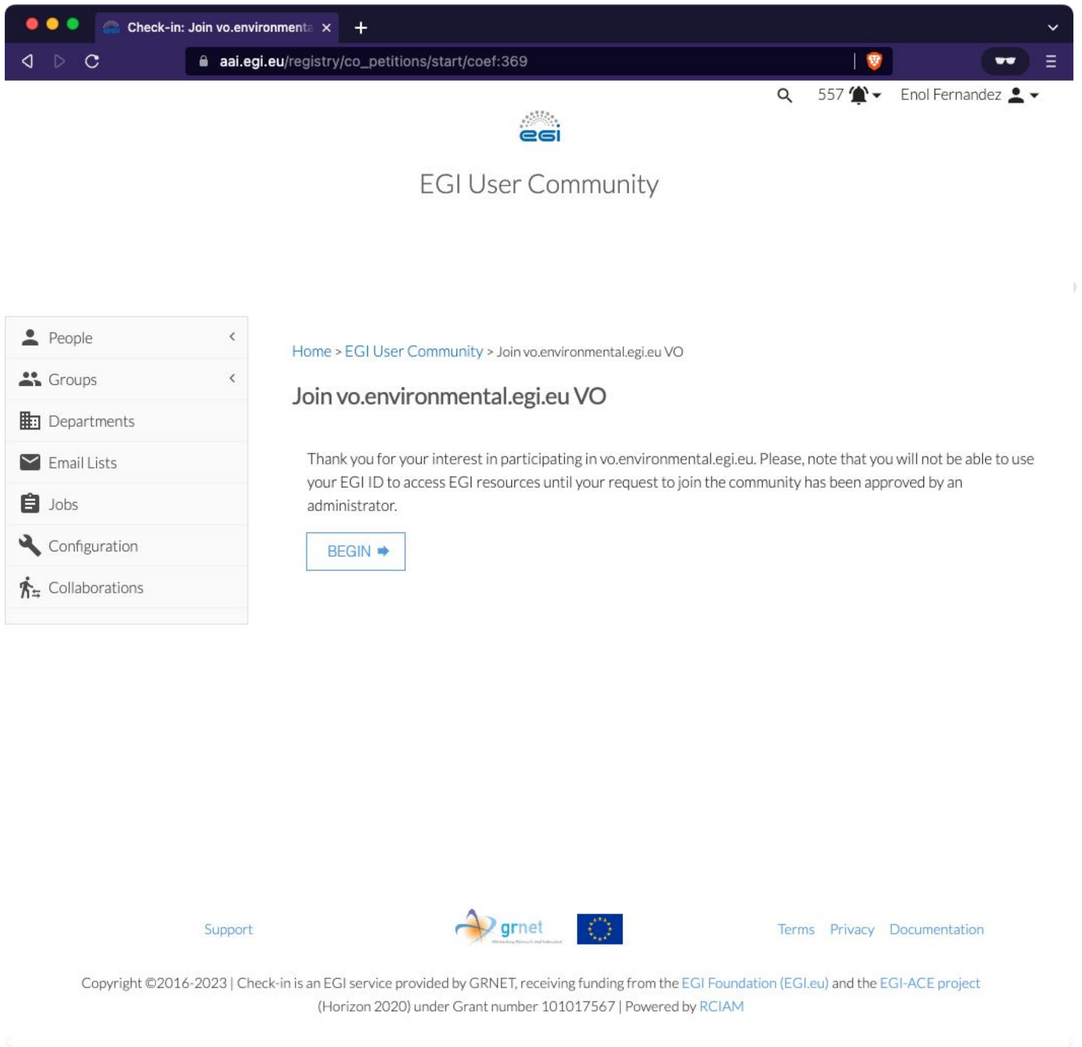


- Persistent home
  - Can be used to store data (10GB limit)
  - Files will be kept even if the notebook server dies
- DataHub
  - Access to your accessible spaces in datahub
  - Share data and assign PIDs to shared spaces
- B2DROP
  - Access data stored in EUDAT's B2DROP service
- CVMFS
  - Selected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

Getting access

1. **Get an EGI account** (https://aai.egi.eu/signup)

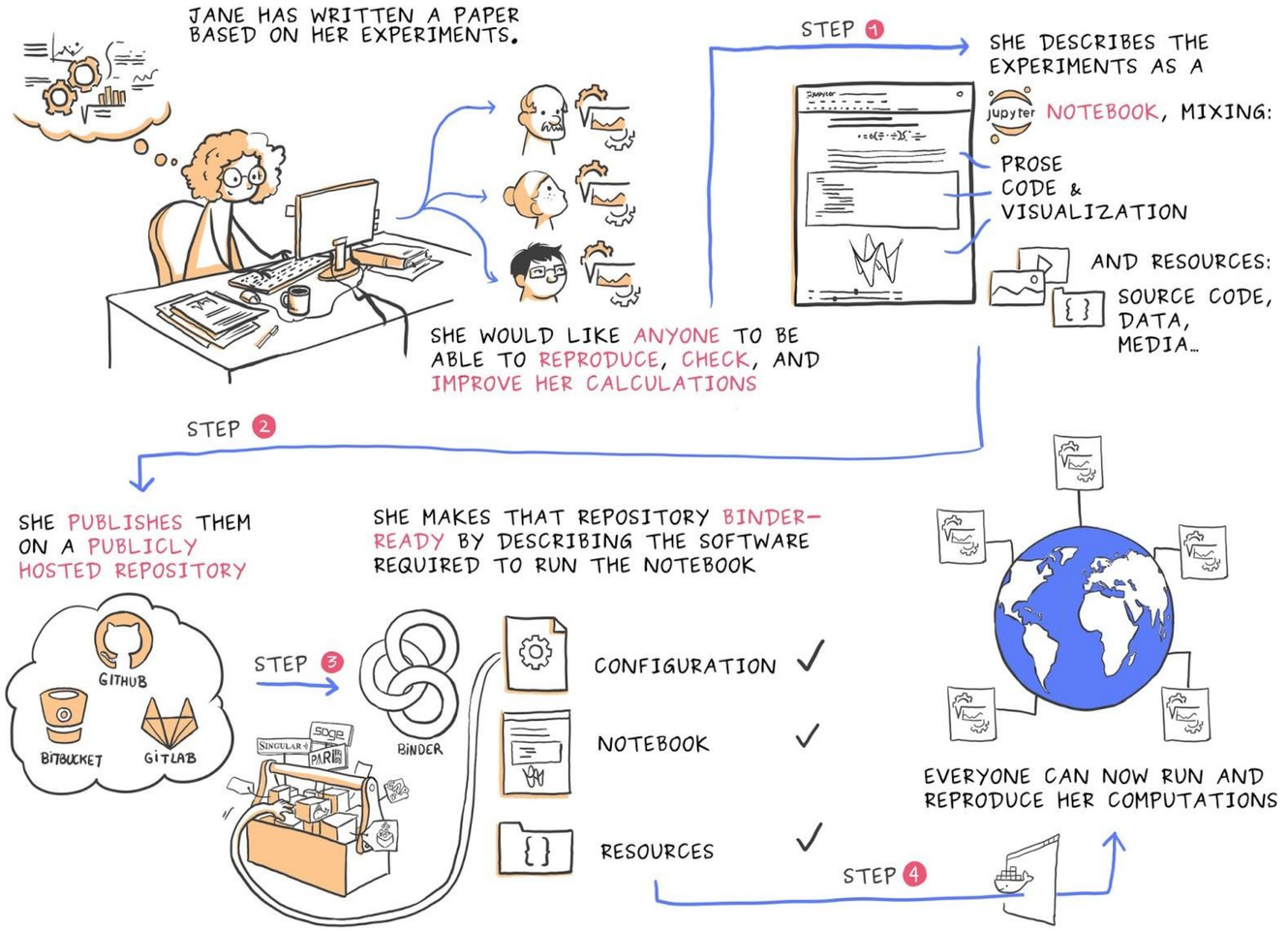
2. **Enroll to one of the supported VOs (Virtual Organisations):**
   - vo.notebooks.egi.eu
   - vo.access.egi.eu
   - auger
   - biomed
   - vo.reliance-project.eu
   - eiscat.se
   - **vo.environmental.egi.eu** (https://go.egi.eu/8Hspz)



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)



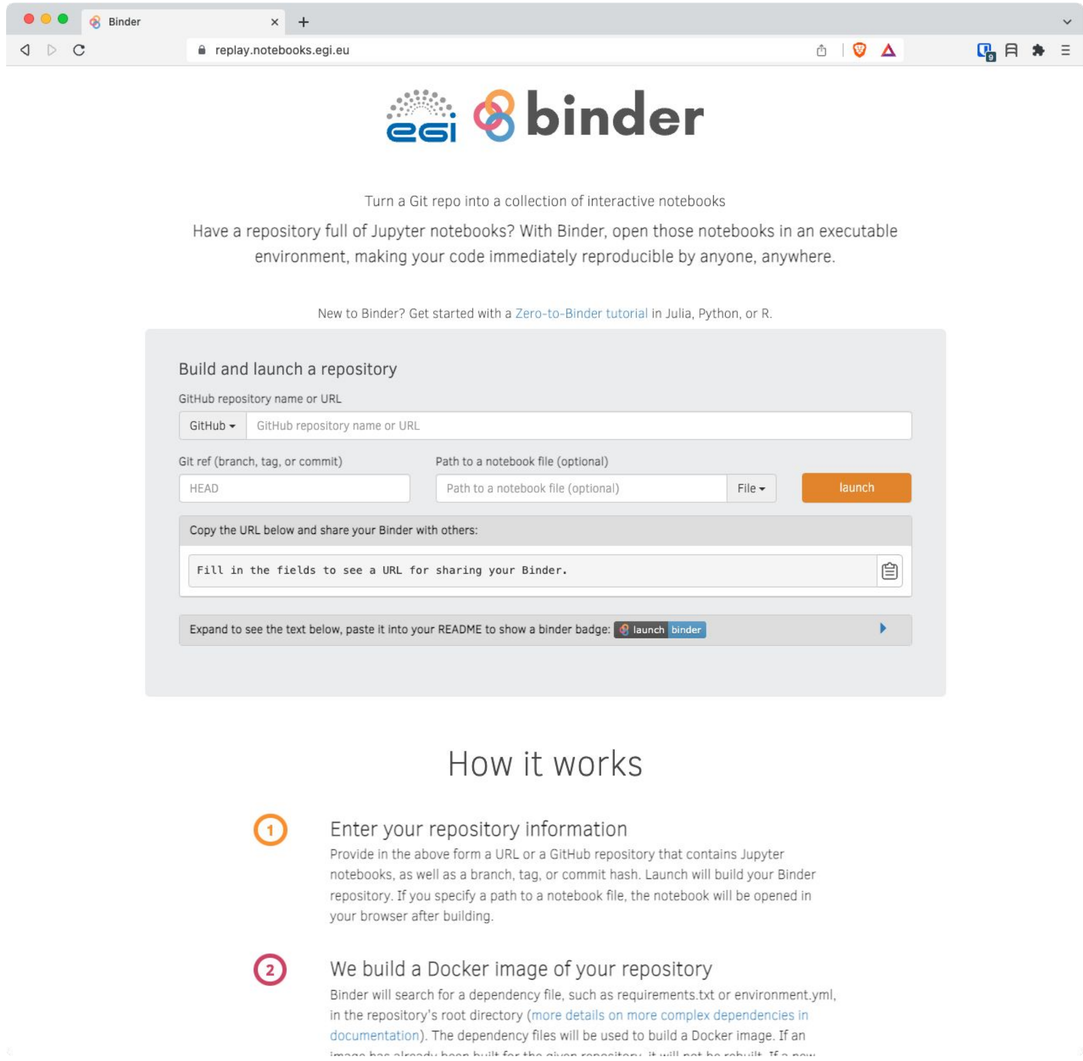
EGI Replay

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



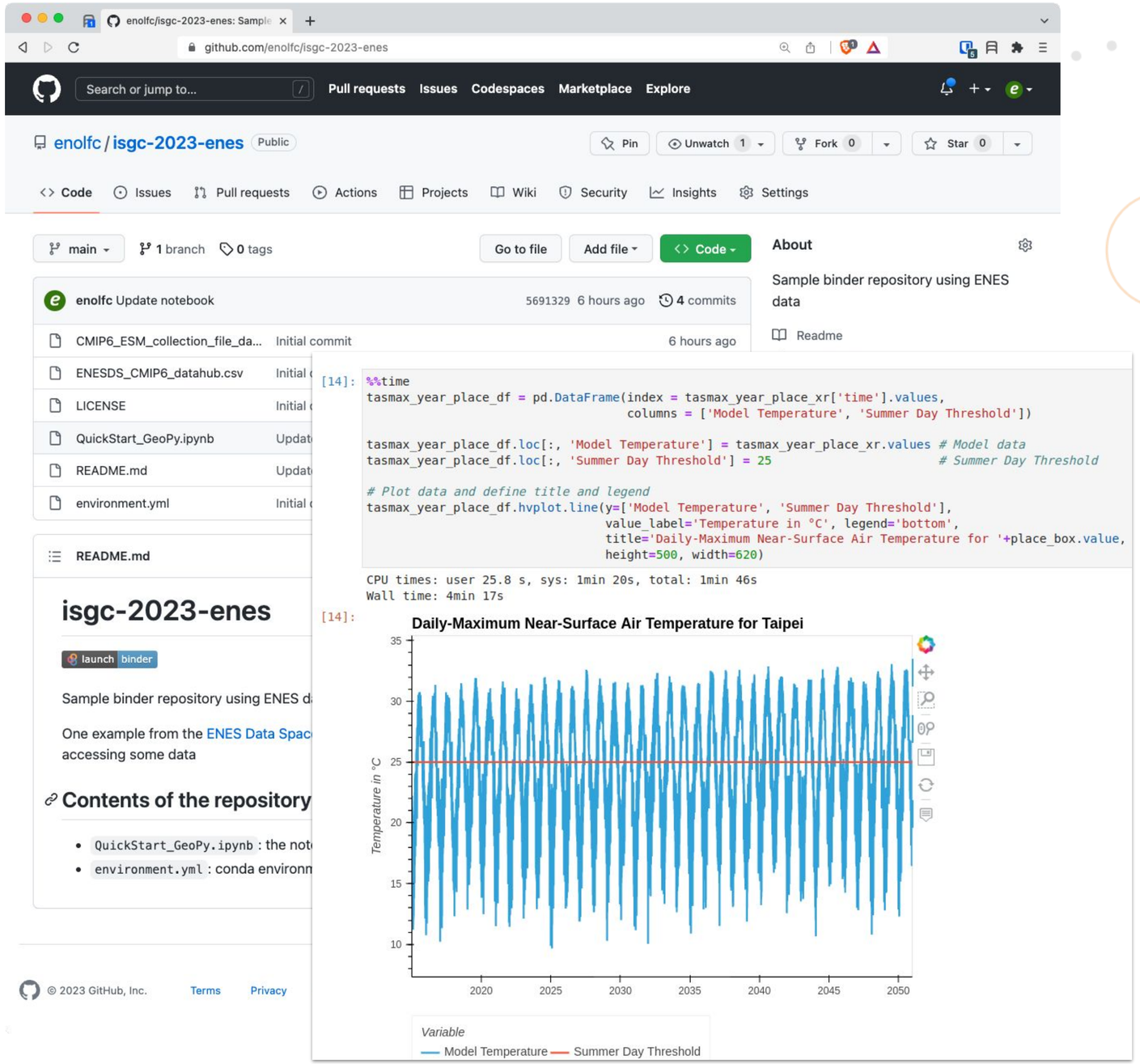
A Binder repository

A code repository that contains:

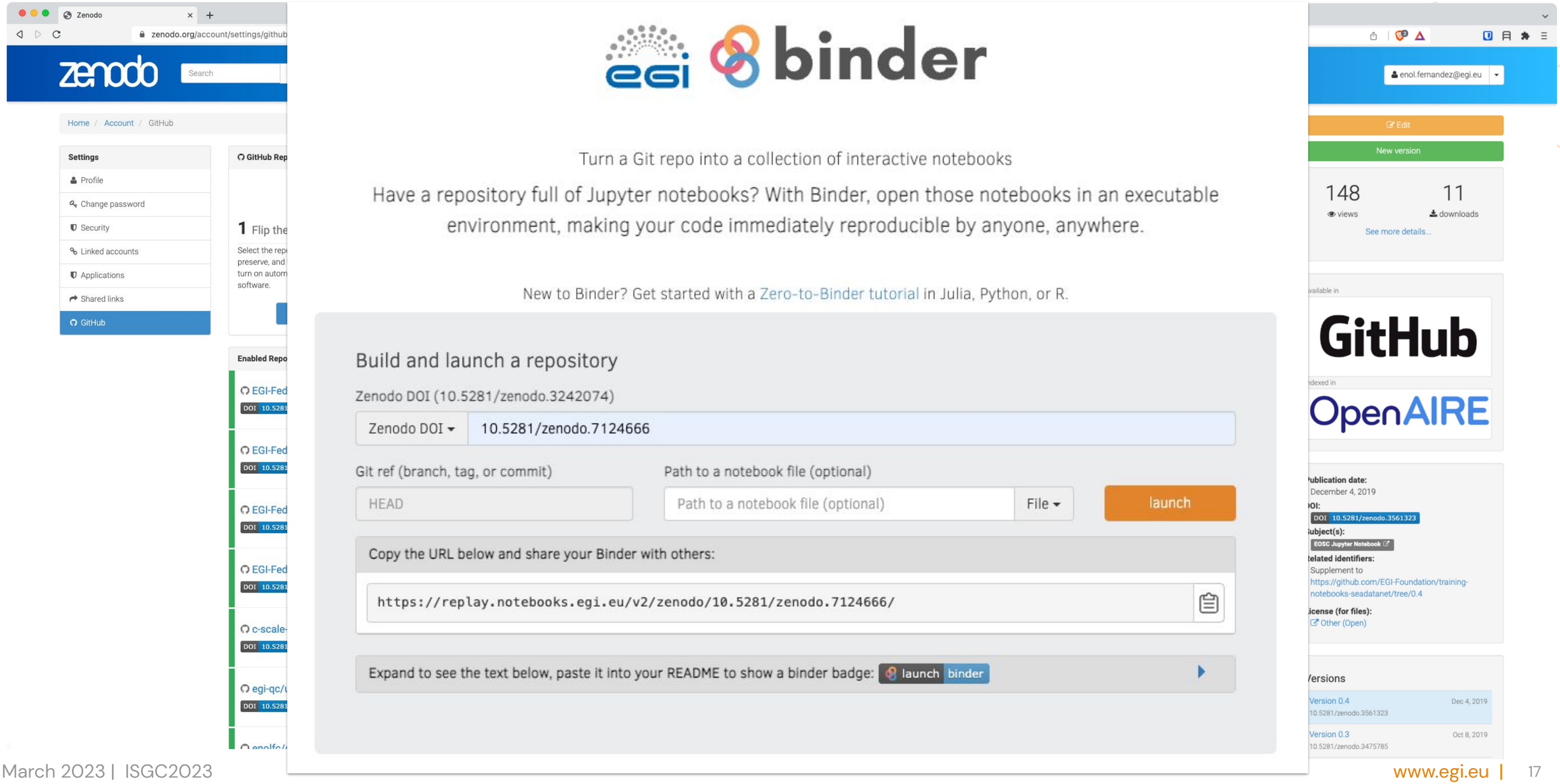
- Code to reproduce (i.e. set of notebooks)
- + description of the software runtime (e.g. a conda environment)
- + any auxiliary files needed to run the code

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

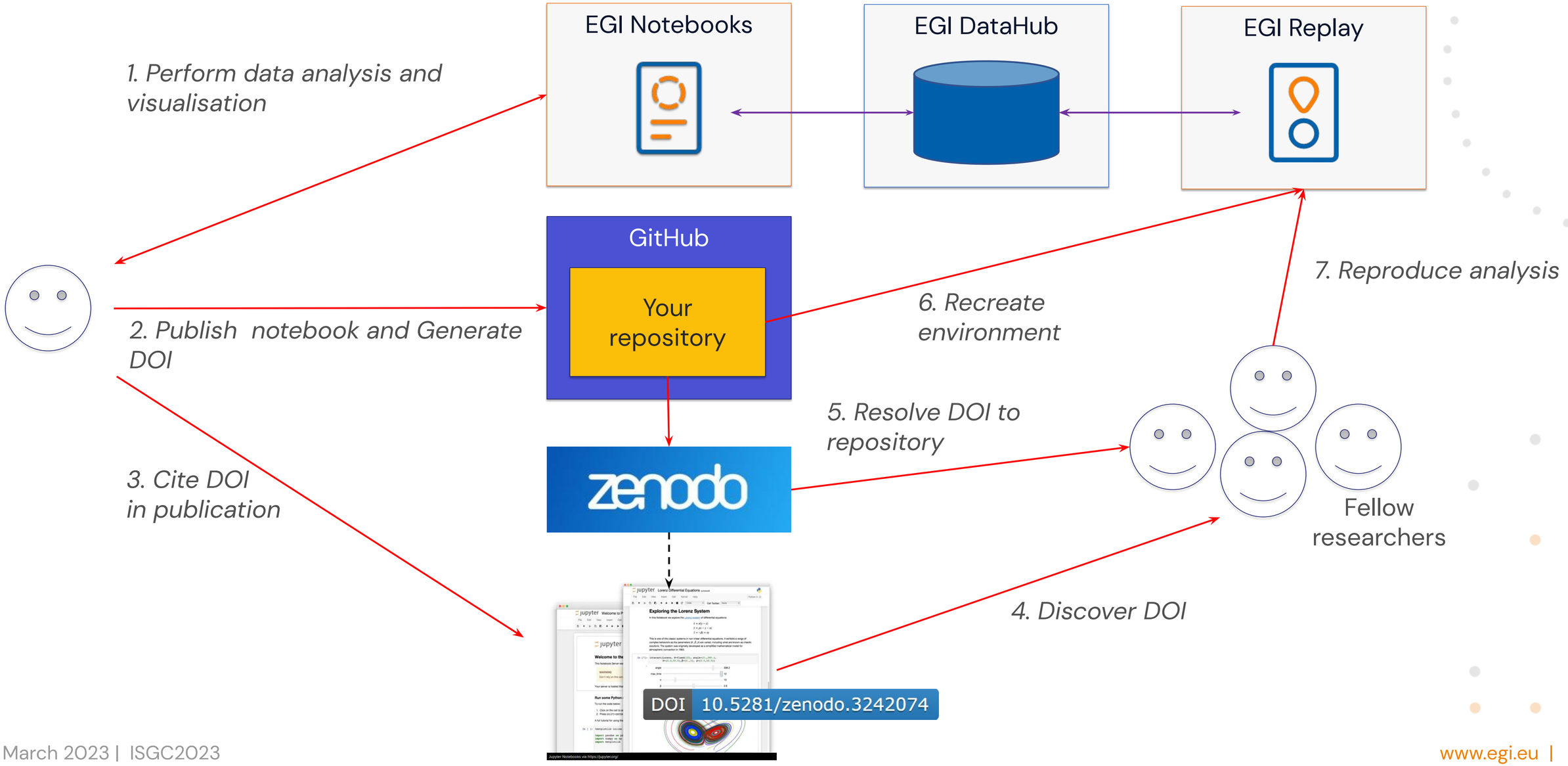
launch binder



Making repositories citable



EGI Notebooks and EGI Replay services



Start using the services!

https://notebooks.egi.eu/

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

Documentation: https://docs.egi.eu/users
Support: support@egi.eu

Join us live to learn more

ISGC 2023 - 21st. March at 14:00

[EGI Tutorial: Regional infrastructure for reproducible open science in Asia Pacific]

June 19th - June 23rd 2023, Poznań, Poland

Call for Contributions is open!
https://indico.egi.eu/event/6071/abstracts/

Deadline extension to March 30th.

Thank you

Giuseppe La Rocca

e giuseppe.larocca@egi.eu

www.egi.eu

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