



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

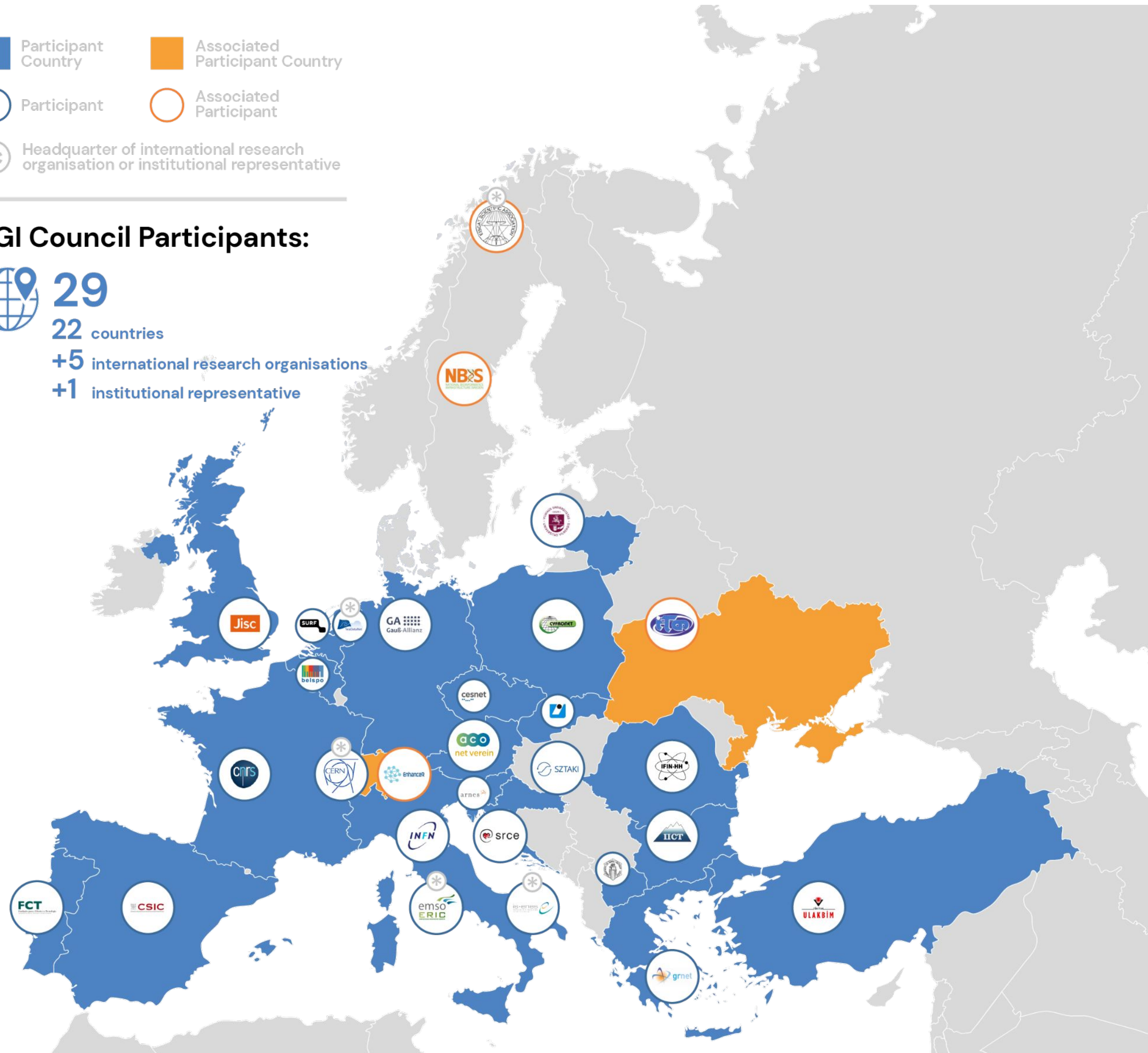
Mission of the EGI Foundation

Enable the EGI Federation to serve international research and innovation together

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

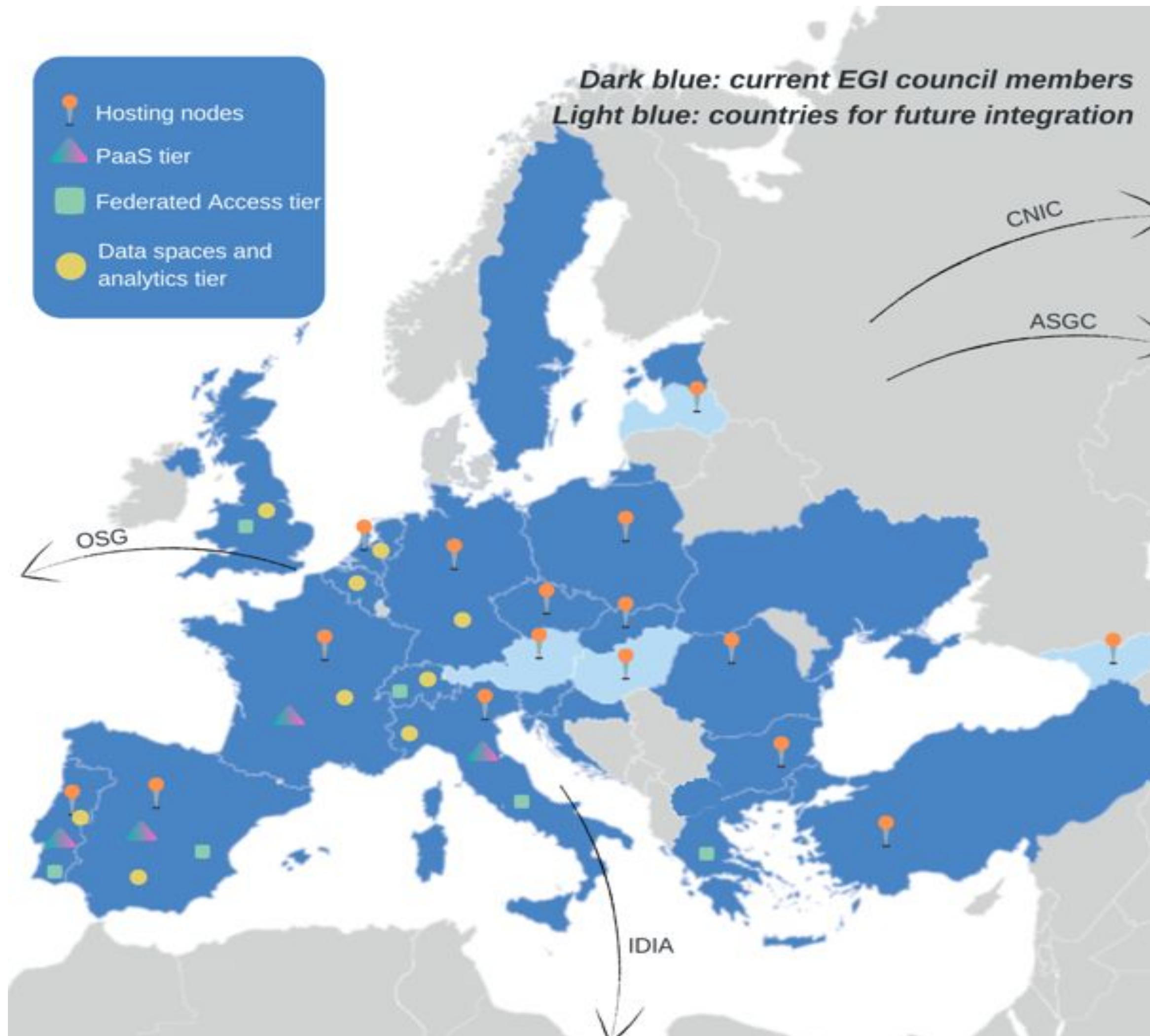
EGI Council Participants:

- 29
- 22 countries
- +5 international research organisations
- +1 institutional representative



International Partnerships with MoUs





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

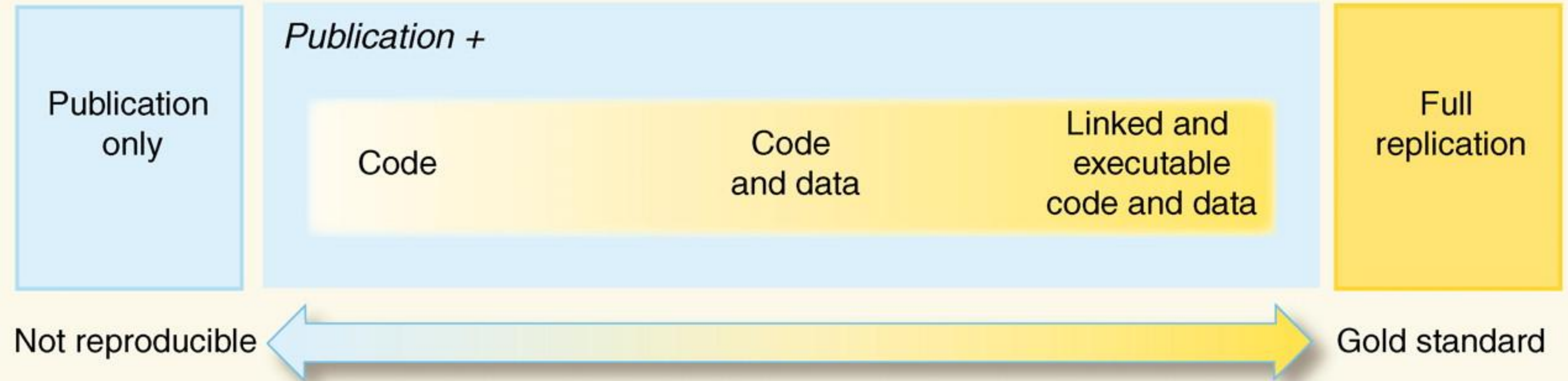
Scope:

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

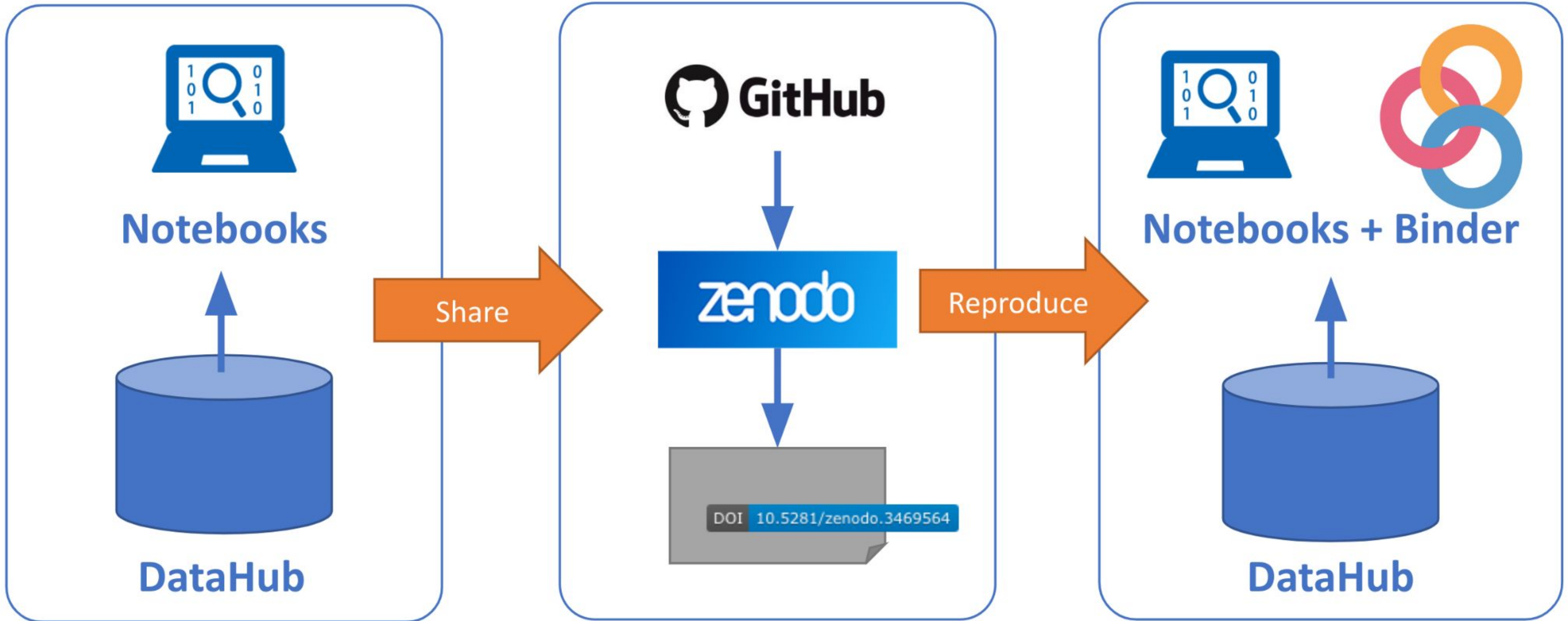
Duration:

- Jan 2021 – June 2023 (30 months)

Reproducibility Spectrum

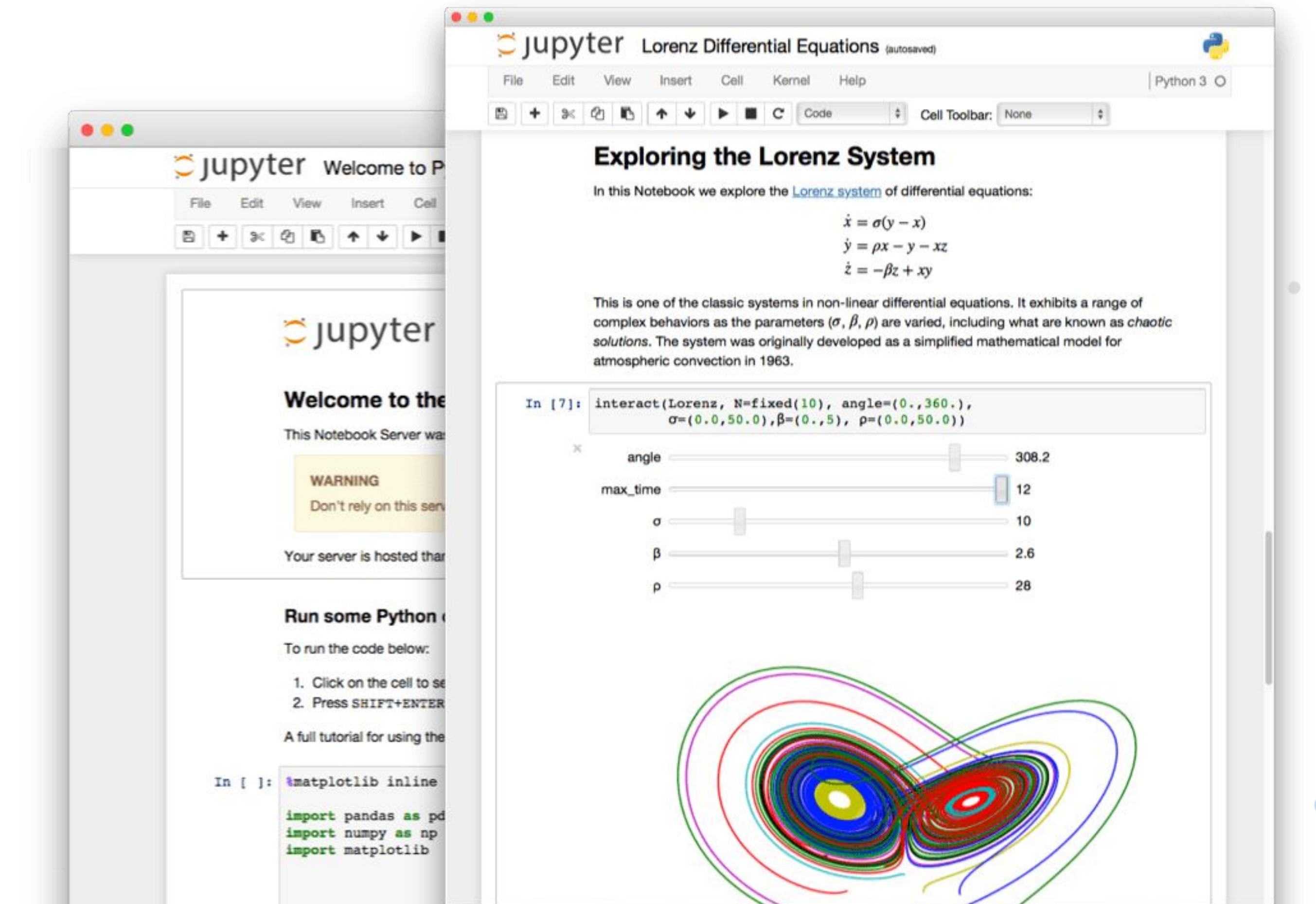


Peng, Science, 2011



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

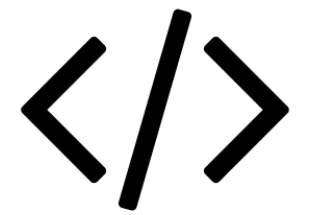
Interactive
browser based
environment



The screenshot shows a JupyterLab notebook window titled 'FirstNotebook.ipynb'. 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:

```
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)
```

The output of the code execution is a line plot showing the signal over time for different brain regions and events. The plot has a y-axis labeled 'signal' ranging from -0.1 to 0.3. The legend indicates four series: 'parietal' (blue solid line), 'frontal' (orange solid line), 'stim' (black solid line), and 'cue' (black dashed line). The 'parietal' and 'frontal' lines show a prominent peak around timepoint 10, while the 'stim' and 'cue' lines show a smaller peak at the same timepoint.



Code
Use your favourite
language



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



EGI Notebooks



Jupyter hosted in the EGI Cloud

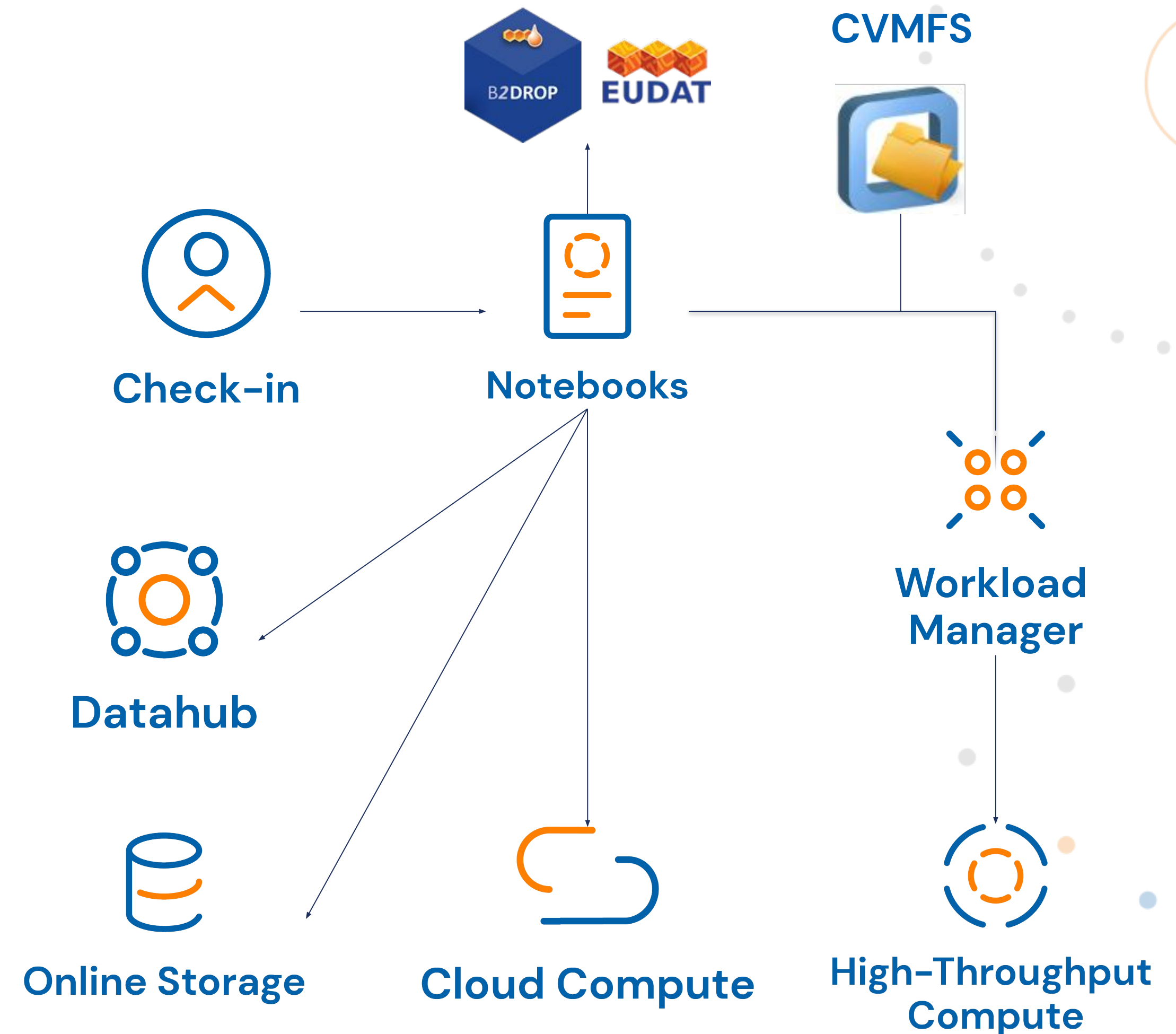
- Offers Jupyter notebooks 'as Service'
- One-click solution: login and start using

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

The screenshot shows a web browser window with the URL `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 this, a paragraph describes the service as an environment based on Jupyter and the EGI cloud service, offering a browser-based, scalable tool for interactive data analysis. A second paragraph states that access requires a valid EGI account and enrolling to the `vo.notebooks.egi.eu` VO. A prominent blue button labeled 'Continue with EGI Check-in' is centered on the page. At the bottom, there is a footer with 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



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

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

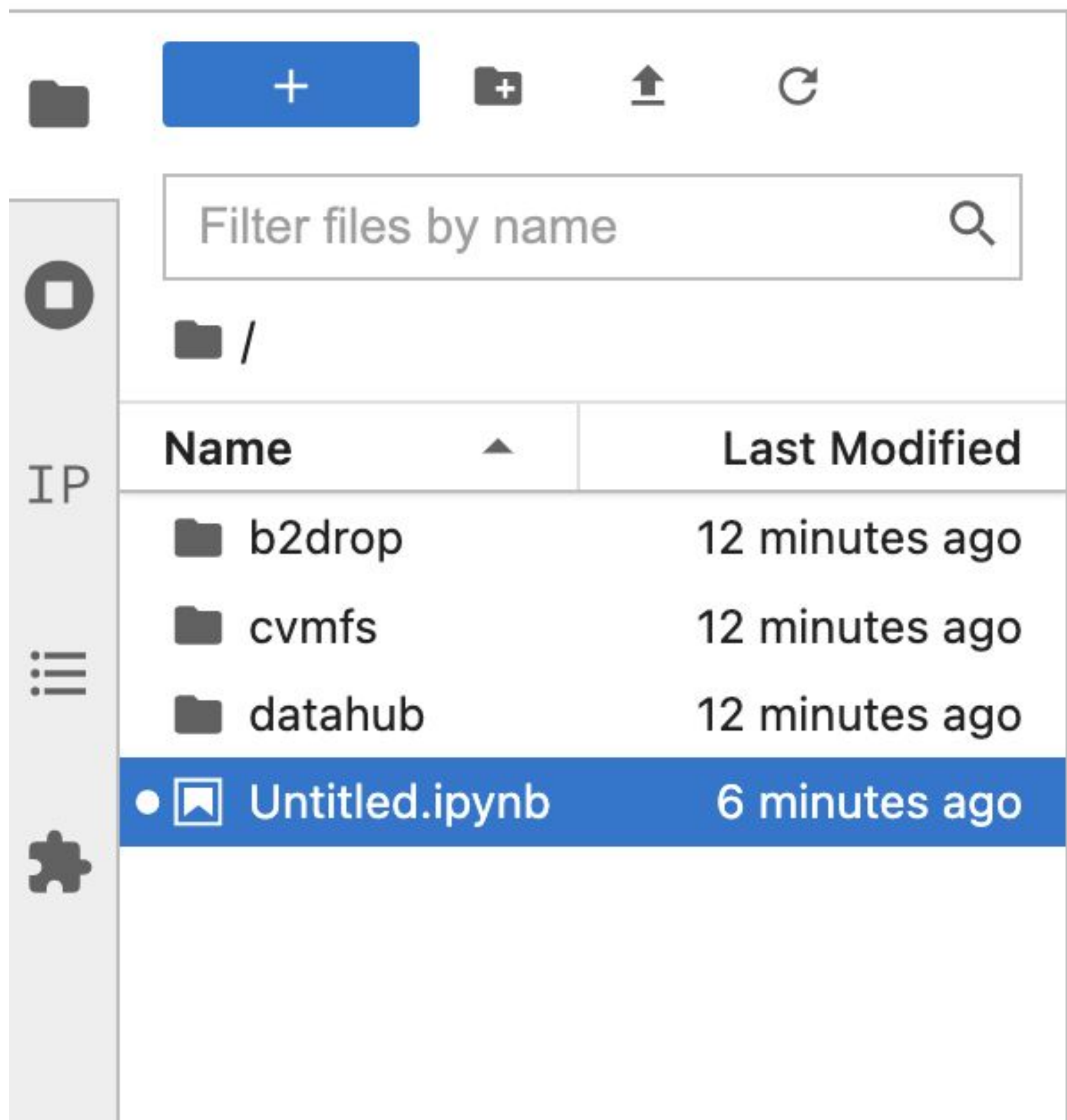
The MATLAB environment (requires a [valid license](#)), includes Python and MATLAB kernels

- 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



- **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**

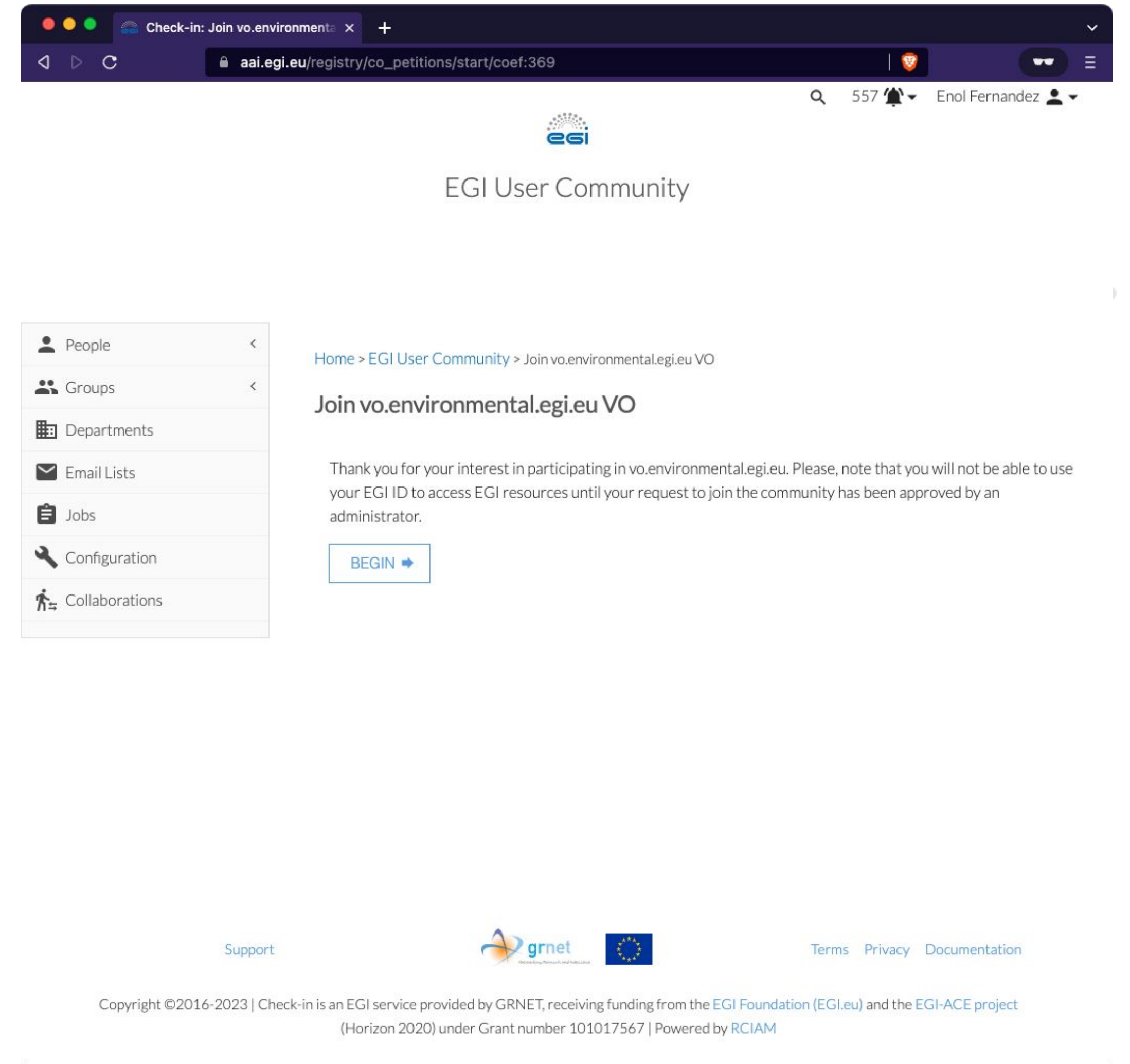
- Get code from any git repository from a single URL
- <https://hub.jupyter.org/nbgitpuller/link.html>

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](https://auger.egi.eu)
- [biomed](https://biomed.egi.eu)
- vo.reliance-project.eu
- eiscat.se
- **vo.environmental.egi.eu**
(<https://go.egi.eu/8Hspz>)

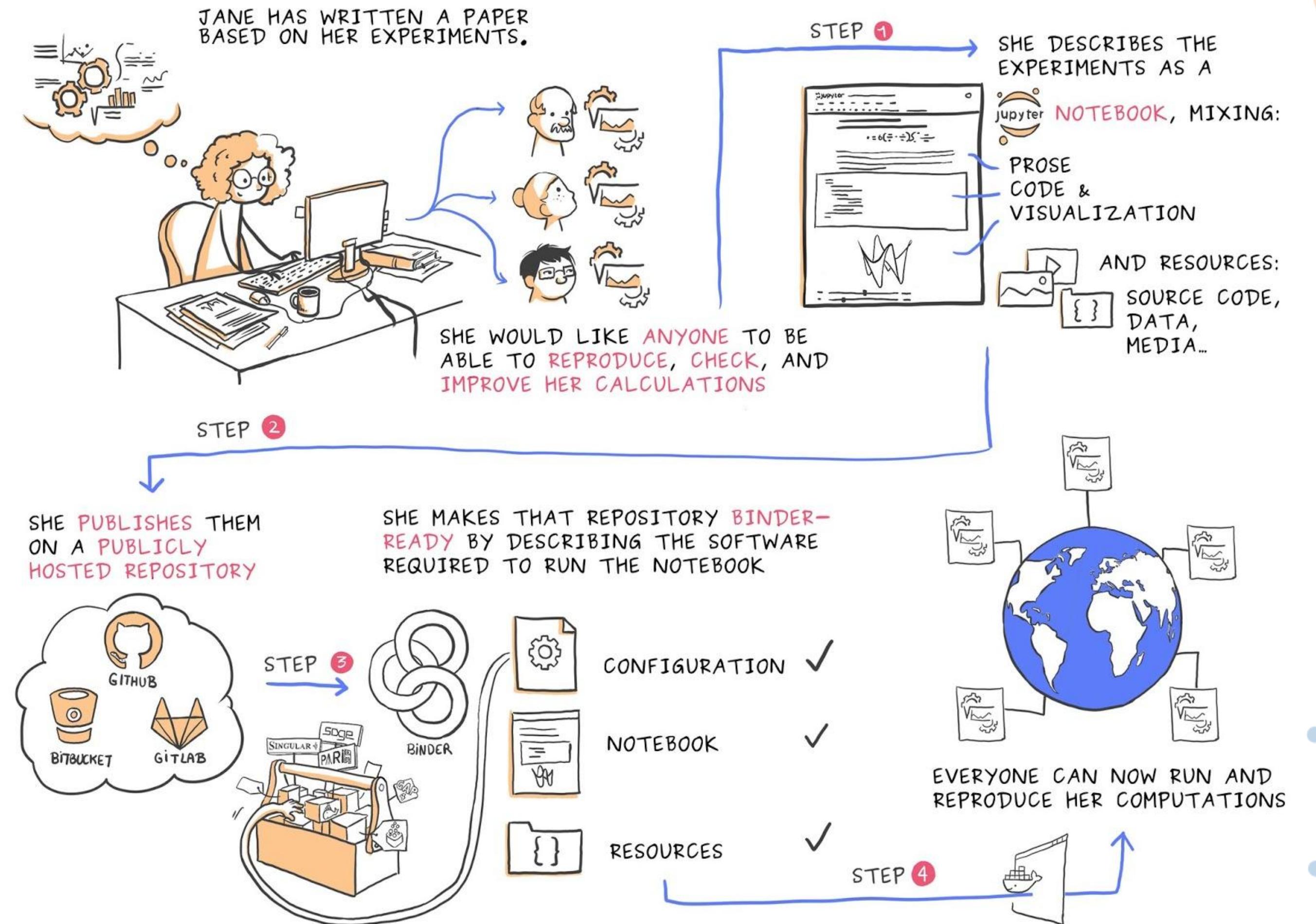


The screenshot shows a web browser window with the URL aai.egi.eu/registry/co_petitions/start/coef:369. The page title is "EGI User Community". The main content area displays "Join vo.environmental.egi.eu VO" with a "BEGIN" button. A message states: "Thank you for your interest in participating in vo.environmental.egi.eu. Please, note that you will not be able to use your EGI ID to access EGI resources until your request to join the community has been approved by an administrator." The footer includes "Support", "grnet", "Terms Privacy Documentation", and copyright information: "Copyright ©2016-2023 | Check-in is an EGI service provided by GRNET, receiving funding from the EGI Foundation (EGI.eu) and the EGI-ACE project (Horizon 2020) under Grant number 101017567 | Powered by RCIAM".

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)





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



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

GitHub repository name or URL
GitHub repository name or URL

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:
Fill in the fields to see a URL for sharing your Binder.

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

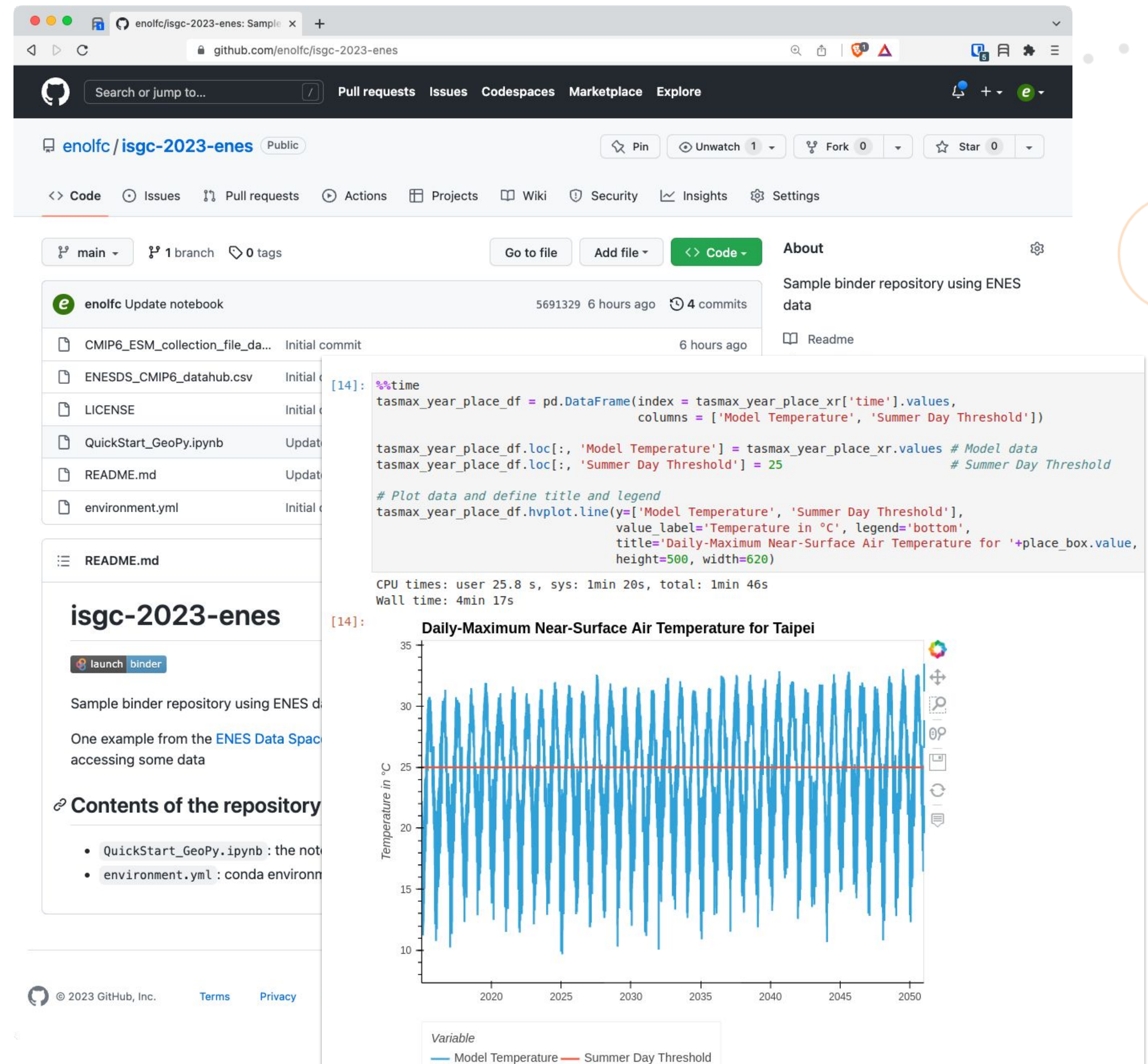
How it works

- 1** Enter your repository information
Provide in the above form a URL or a GitHub repository that contains Jupyter notebooks, as well as a branch, tag, or commit hash. Launch will build your Binder repository. If you specify a path to a notebook file, the notebook will be opened in your browser after building.
- 2** We build a Docker image of your repository
Binder will search for a dependency file, such as requirements.txt or environment.yml, in the repository's root directory ([more details on more complex dependencies in documentation](#)). The dependency files will be used to build a Docker image. If an image has already been built for the given repository, it will not be rebuilt. If a new

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>

The screenshot shows a GitHub repository page for 'enolfc/isgc-2023-enes'. The repository is public and contains a Jupyter notebook named 'Update notebook'. The notebook code is as follows:

```
[14]: %%time
tasmax_year_place_df = pd.DataFrame(index = tasmax_year_place_xr['time'].values,
                                     columns = ['Model Temperature', 'Summer Day Threshold'])

tasmax_year_place_df.loc[:, 'Model Temperature'] = tasmax_year_place_xr.values # Model data
tasmax_year_place_df.loc[:, 'Summer Day Threshold'] = 25 # Summer Day Threshold

# Plot data and define title and legend
tasmax_year_place_df.hvplot.line(y=['Model Temperature', 'Summer Day Threshold'],
                                value_label='Temperature in °C', legend='bottom',
                                title='Daily-Maximum Near-Surface Air Temperature for '+place_box.value,
                                height=500, width=620)
```

The plot shows the 'Daily-Maximum Near-Surface Air Temperature for Taipei' from 2020 to 2050. The y-axis is 'Temperature in °C' ranging from 10 to 35. The x-axis is 'Year' ranging from 2020 to 2050. The plot displays a blue line for 'Model Temperature' and a red horizontal line for 'Summer Day Threshold' at 25°C. The temperature fluctuates between approximately 10°C and 35°C, with a clear seasonal cycle. The Summer Day Threshold is consistently above the temperature, indicating that the temperature is generally below the threshold.

zenodo.org/account/settings/github

zenodo Search

Home / Account / GitHub

Settings

- Profile
- Change password
- Security
- Linked accounts
- Applications
- Shared links
- GitHub**

GitHub Rep

1 Flip the...
Select the repository to preserve, and turn on automatic software.

Enabled Repos

- EGI-Fed
DOI 10.5281/zenodo.3242074
- EGI-Fed
DOI 10.5281/zenodo.7124666
- EGI-Fed
DOI 10.5281/zenodo.3561323
- EGI-Fed
DOI 10.5281/zenodo.3475785
- c-scale
DOI 10.5281/zenodo.3475785
- egi-qc/
DOI 10.5281/zenodo.3475785
- enolfer/
DOI 10.5281/zenodo.3475785

ESI binder

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

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:

enol.fernandez@egi.eu

Edit

New version

148 views 11 downloads

See more details...

available in

GitHub

indexed in

OpenAIRE

publication date: December 4, 2019

DOI: DOI 10.5281/zenodo.3561323

subject(s): EOSC Jupyter Notebook

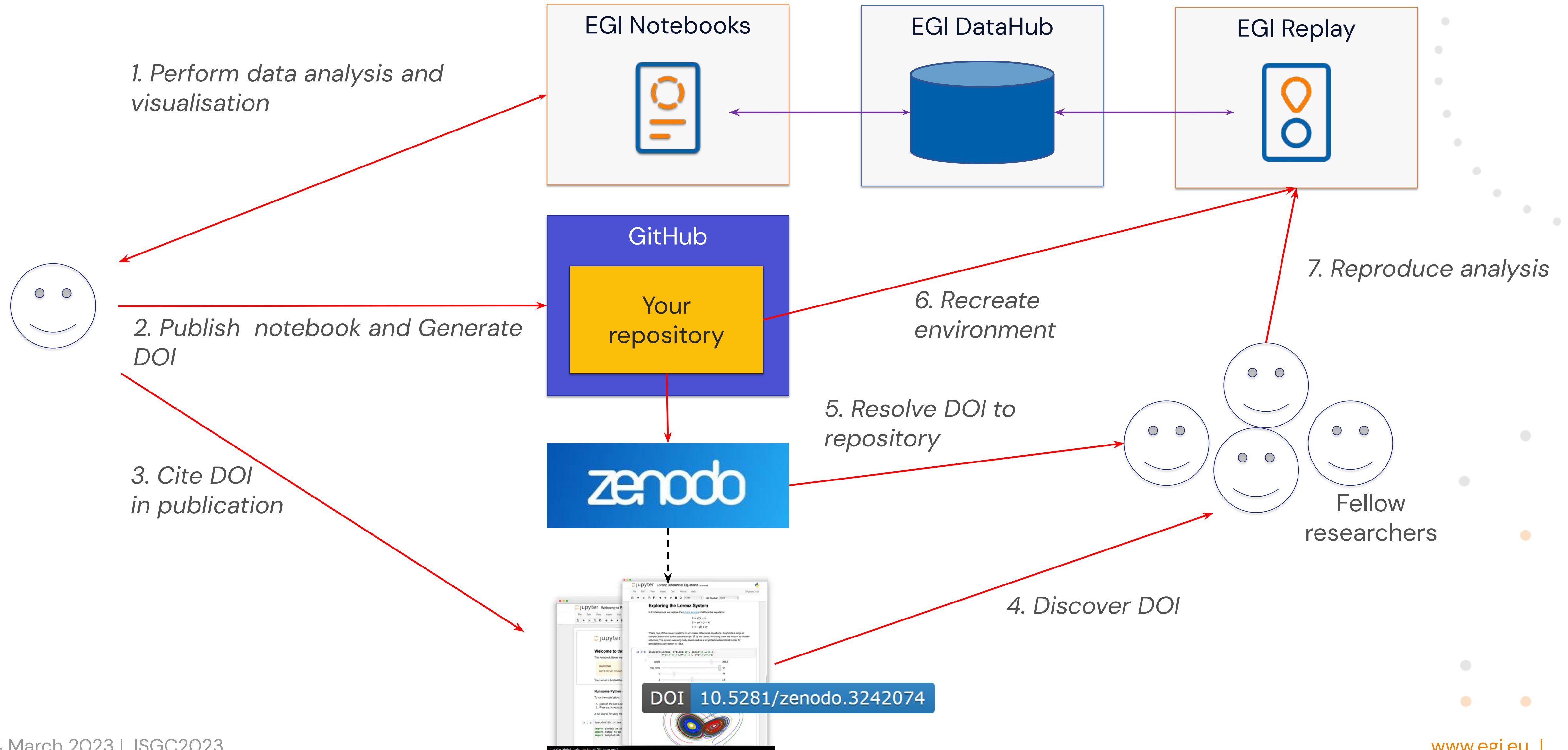
related identifiers: Supplement to <https://github.com/EGI-Foundation/training-notebooks-seadatanet/tree/0.4>

license (for files): [Other \(Open\)](#)

versions

Version 0.4	Dec 4, 2019
10.5281/zenodo.3561323	
Version 0.3	Oct 8, 2019
10.5281/zenodo.3475785	

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



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