

Reproducible Open Science within EOSC: expanding EGI Notebooks with B2DROP access and reproducibility with Binder

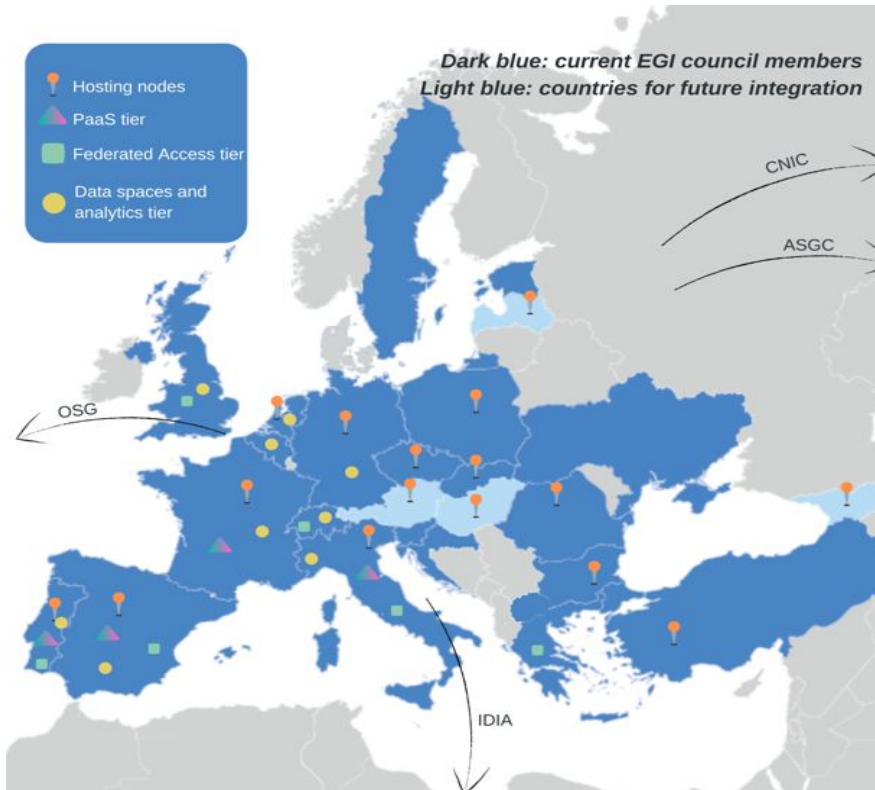
Giuseppe La Rocca,
Community Support Lead. at EGI Foundation

giuseppe.larocca@egi.eu

EOSC Symposium, 14-17 November, 2022, Prague



Project Overview



EGI Advanced Computing for EOSC Grant agreement ID: 101017567

Budget

- Total budget: € 12,009,988
- EC budget: € 8,000,000

Consortium

- Coordinator - Stichting EGI
- 33 Partners, 23 third parties

Effort

- 1472 PMs, 48 FTEs
- **49% Virtual Access** (35 services, 38 providers)

Duration

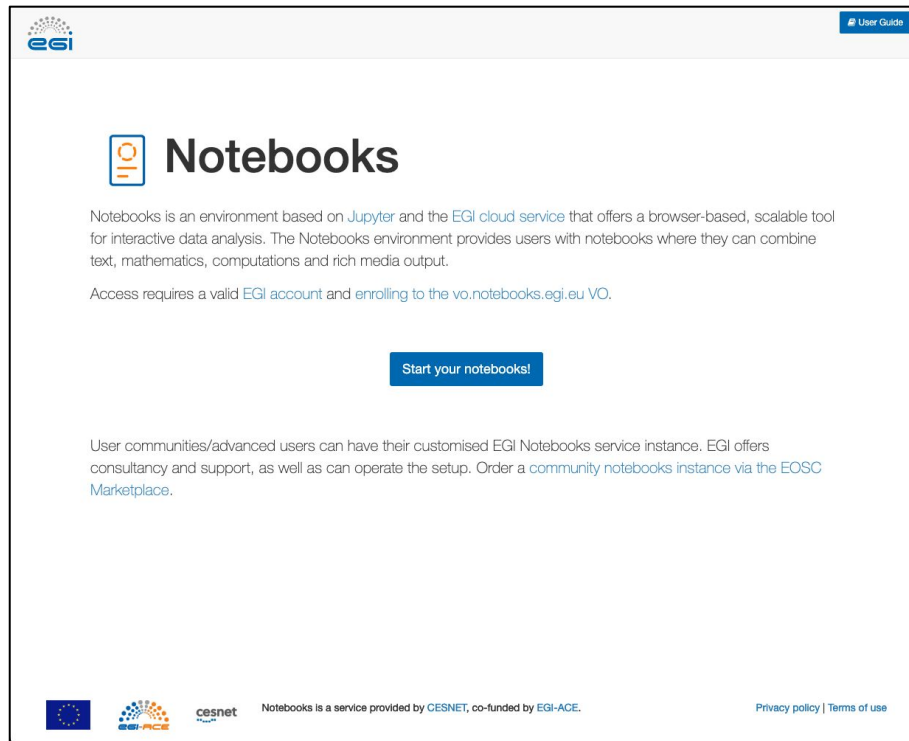
- Jan 2021 - June 2023 (30 months)

Multi-user Jupyter @ 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 the EGI Notebooks user guide page. At the top left is the EGI logo, and at the top right is a "# User Guide" link. The main heading is "Notebooks" with a notebook icon. Below the heading is a paragraph describing the service: "Notebooks is an environment based on Jupyter and the EGI cloud service that offers a browser-based, scalable tool for interactive data analysis. The Notebooks environment provides users with notebooks where they can combine text, mathematics, computations and rich media output." This is followed by a paragraph stating: "Access requires a valid EGI account and enrolling to the [vo.notebooks.egi.eu VO](#)." A prominent blue button labeled "Start your notebooks!" is centered on the page. Below the button, another paragraph states: "User communities/advanced users can have their customised EGI Notebooks service instance. EGI offers consultancy and support, as well as can operate the setup. Order a [community notebooks instance via the EOSC Marketplace](#)." At the bottom of the page, there are logos for the European Union, EGI-ACE, and CESNET. To the right of these logos, it says "Notebooks is a service provided by CESNET, co-funded by EGI-ACE." Further right, there are links for "Privacy policy" and "Terms of use".

EGI Notebooks (based on JupyterHub)



Documentation
Text formatted using
Markdown/LaTeX



Data
Import from
local/remote disk

Interactive
browser based
environment with
federated login



The screenshot shows a JupyterLab notebook with the following content:

```
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>
```

The plot shows signal values for different regions (parietal, frontal) and events (stim, cue) over time. The y-axis is labeled 'signal' and ranges from -0.1 to 0.3. The x-axis is labeled 'timepoint'. The legend indicates: region (parietal, frontal), event (stim, cue).



Code
Use your favourite
language



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

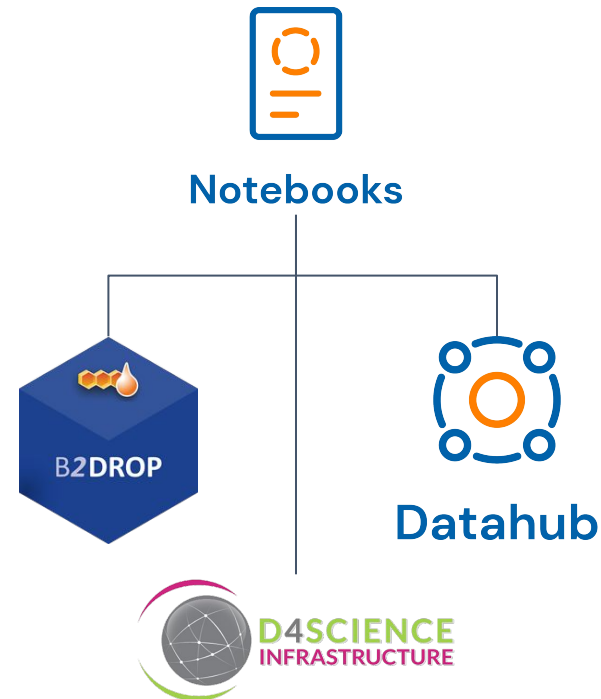
Accessing data from Notebooks

(Small) data in your user space

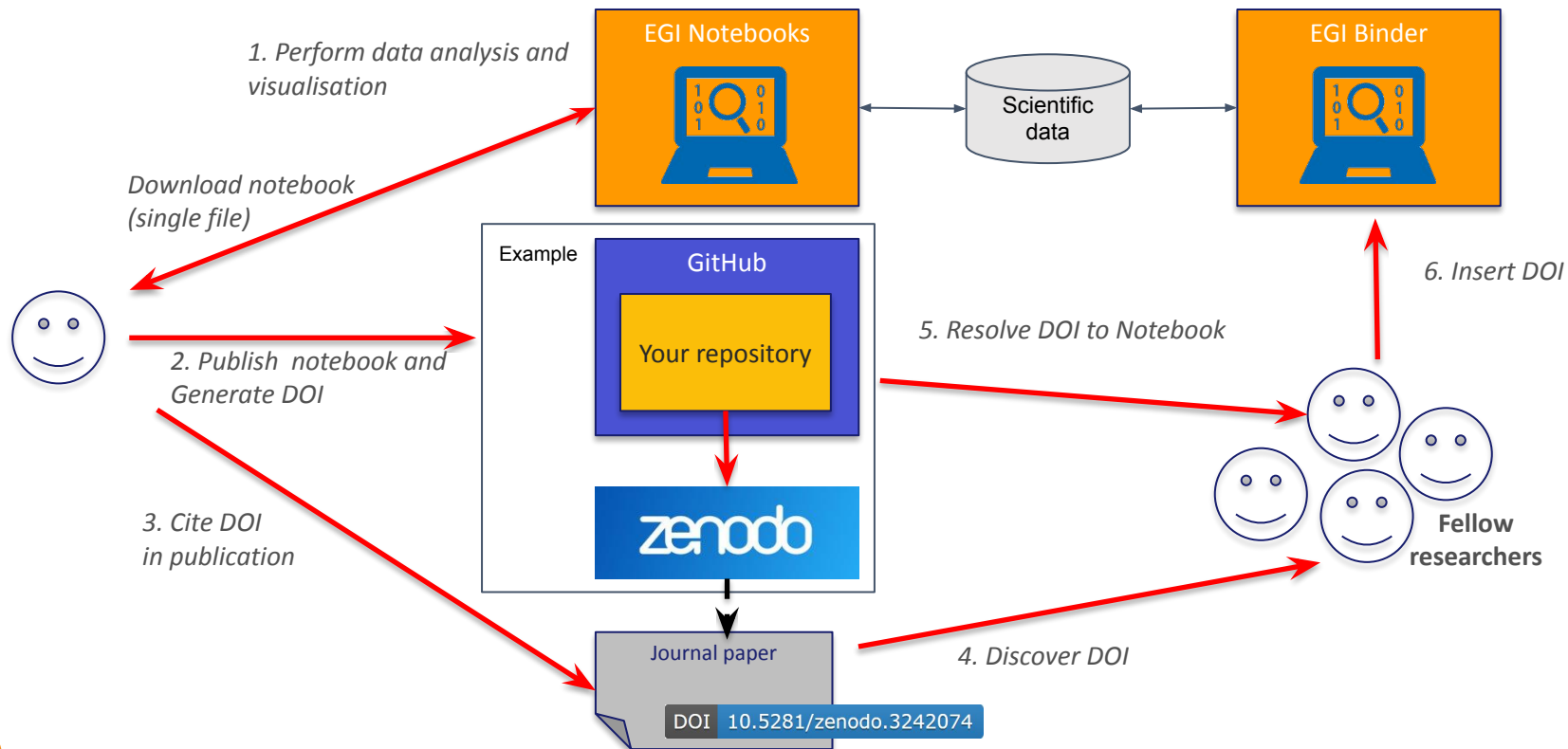
- Fetch data from publicly accessible repositories using code in notebooks or well-known tools as Rclone
- User credentials for EGI Check-in protected repositories configured in the environment

(Big) data integrations:

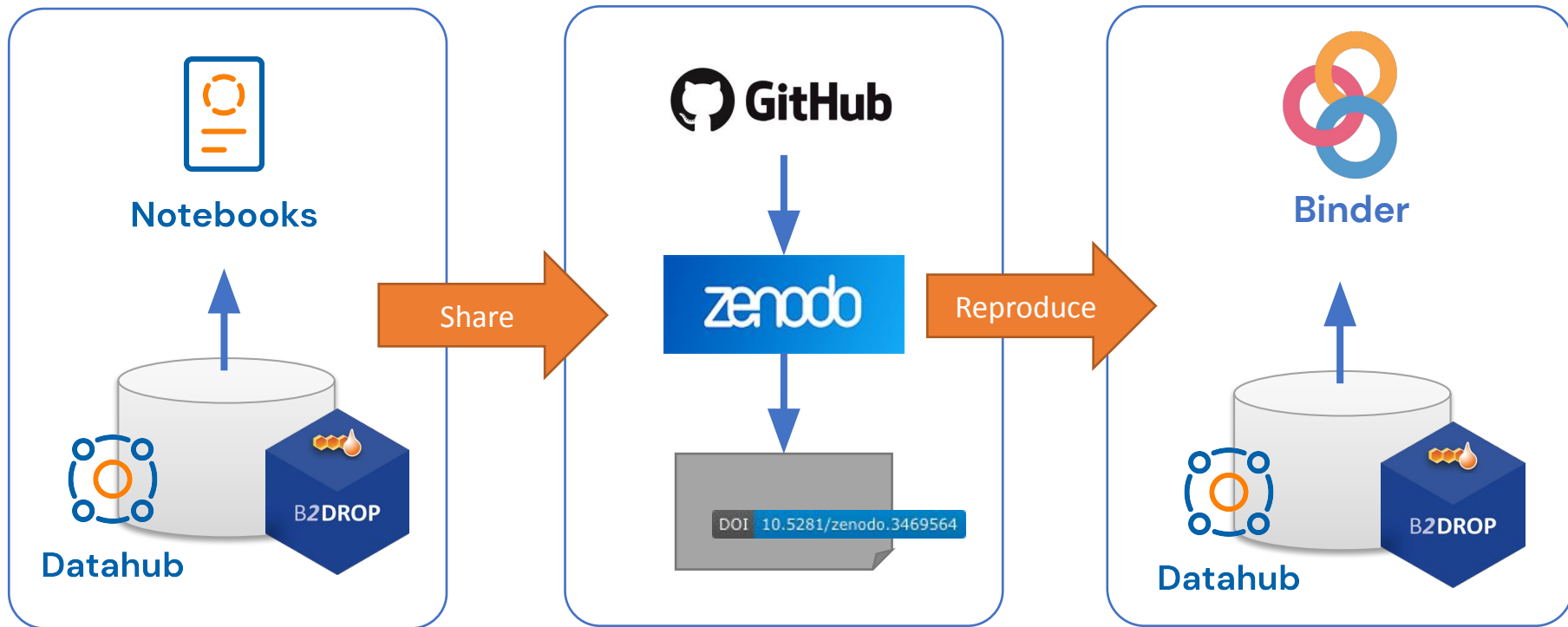
- **EGI DataHub** spaces available transparently as a local folder (available for all users)
- **EUDAT B2DROP** also available as local folder. Can be activated by users
- **D4Science Workspaces** for notebooks integrated into D4Science VREs
- We're open for more integration options. Let us know your needs!



Elements for sharing/reproducible research



Binder: Reproducible Open Science in EOSC



An example

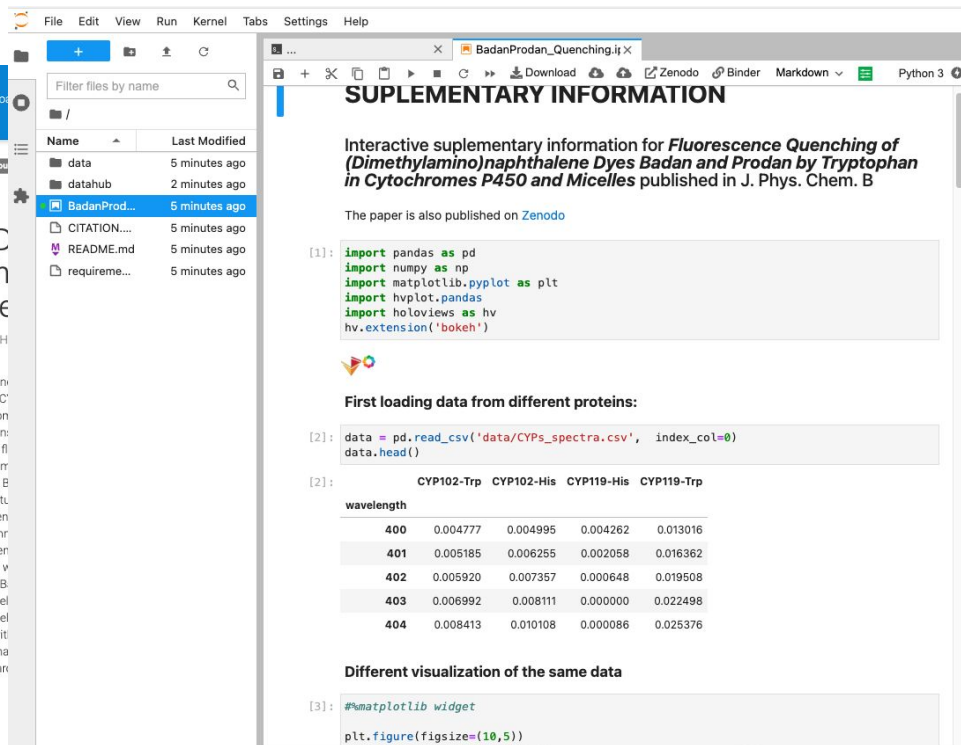


July 31, 2014

Fluorescence Quenching of (Dimethylamino)naphthalene Dyes Badan and Prodan by Tryptophan in Cytochromes P450 and Micelles

Pospišil, Petr; Luxem, Katja E; Ener, Maria; Šykora, Jan; Kocábová, Jana; Gray, H Hof, Martin

Fluorescence of 2-(*N,N*-dimethylamino)-6-propionyl-naphthalene dyes Badan and Prodan by tryptophan in Brij 58 micelles as well as in two cytochrome P450 proteins (CYP102 and CYP119) was studied. The dyes are covalently attached to a cysteine residue. Formation of nonemissive complex between dye molecule and tryptophan accounts for about 76% of the fluorescence intensity. The rest is due to diffusive encounters. In the absence of tryptophan, fluorescently labeled cytochromes decay with triexponential kinetics characterized by lifetimes of 700–800 ps, and 3 ns. Site mutation of a histidine residue in the vicinity of the Brij 58 tryptophan results in shortening of all three decay lifetimes. The relative amplitude of the fastest component increases at the expense of the two slower ones. The average quenching rate constants are $4.5 \times 10^8 \text{ s}^{-1}$ (CYP102) and $3.7 \times 10^8 \text{ s}^{-1}$ (CYP119), at 288 K. Cyclic voltammogram in MeCN shows a reversible reduction peak at -1.85 V vs NHE that becomes shifted positively upon addition of water. A quasireversible reduction at -0.88 V vs NHE in aqueous buffer (pH 7.3). The excited-state reduction potential of Prodan (and Badan) varies from about $+0.6 \text{ V}$ (vs NHE) in polar aprotic media (MeCN) to approximately $+0.4 \text{ V}$ (vs NHE) in aqueous media. Tryptophan quenching of Badan/Prodan fluorescence in CYPs and Brij 58 micelles involves tryptophan oxidation by excited Badan/Prodan, coupled with tryptophan reduction by the reduced dye and water. Photoreduction is a new quenching mechanism for fluorescently labeled dyes that are often used as solvatochromic FRET donors and acceptors, as well as reporters of solvation dynamics.



The screenshot shows a Jupyter Notebook titled "BadanProdan_Quenching.ipynb". The notebook content includes:

- A title: "SUPPLEMENTARY INFORMATION"
- Text: "Interactive supplementary information for *Fluorescence Quenching of (Dimethylamino)naphthalene Dyes Badan and Prodan by Tryptophan in Cytochromes P450 and Micelles* published in J. Phys. Chem. B"
- Text: "The paper is also published on Zenodo"
- Code cell [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import hvplot.pandas
import holoviews as hv
hv.extension('bokeh')
```
- Text: "First loading data from different proteins:"
- Code cell [2]:

```
data = pd.read_csv('data/CYPs_spectra.csv', index_col=0)
data.head()
```
- Table output for [2]:

	CYP102-Trp	CYP102-His	CYP119-His	CYP119-Trp
wavelength				
400	0.004777	0.004995	0.004262	0.013016
401	0.005185	0.006255	0.002058	0.016362
402	0.005920	0.007357	0.000648	0.019508
403	0.006992	0.008111	0.000000	0.022498
404	0.008413	0.010108	0.000086	0.025376
- Text: "Different visualization of the same data"
- Code cell [3]:

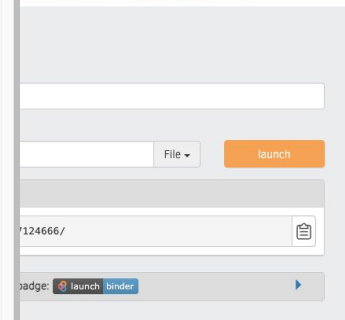
```
!matplotlib widget
plt.figure(figsize=(10,5))
```

Under

collection of interactive books

Under, open those notebooks in an executable and reproducible by anyone, anywhere.

Under binder tutorial in Julia, Python or R.



The screenshot shows a Binder interface with a file upload area containing a file named "124666/". Below the upload area is a "File" dropdown menu and a "launch" button. At the bottom, there is a "badge" section with "launch" and "binder" options.

Reproducible Open Science in EOSC with EGI-ACE



Summary

EGI Notebooks: an EOSC data analytics platform

- Integrated with EGI and other EOSC services
- Transparent Access to data in DataHub and B2DROP
- Easy to share existing code from git repositories

EGI Binder (to be renamed EGI Replay)

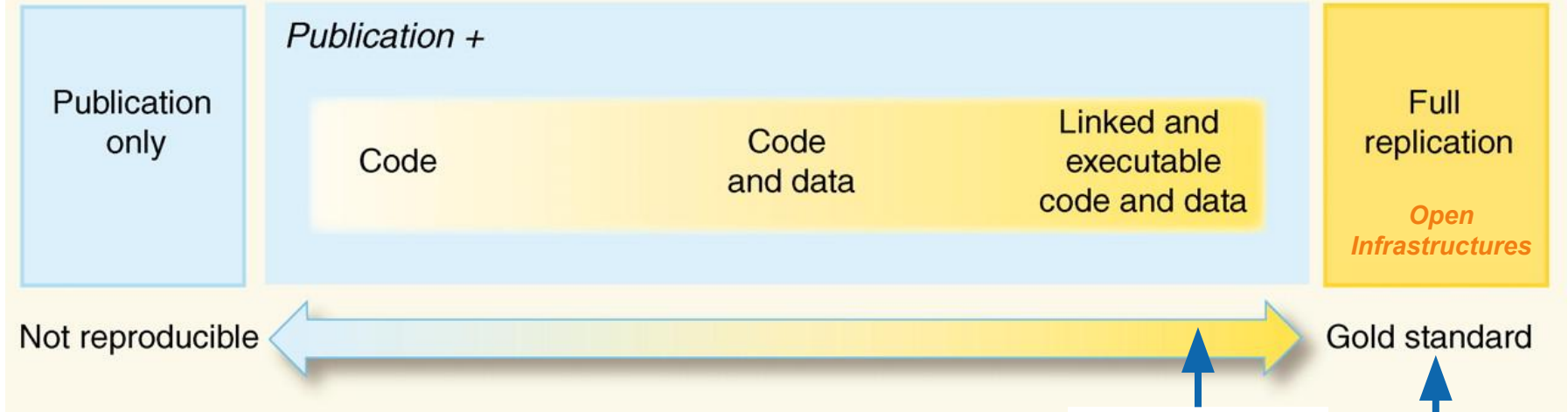
- Easy to share research ensuring reproducible software environment
- Integrated with Zenodo DOIs
- Access to same data repositories as in EGI Notebooks: DataHub and B2DROP

We offer different service modes depending on your needs:

- **Notebooks for researchers:**
 - Individual users can use the centrally operated service from EGI
- **Notebooks for community:**
 - Community specific deployment to provide notebooks for all the users of a community. Allows further customisation to meet the community needs (e.g. shared storage).

Reproducible Science - YES YOU CAN!

Reproducibility Spectrum



How to get access:

- <https://notebooks.egi.eu/hub/login> (Notebooks for researchers)
- <https://www.egi.eu/service-contact/> (Notebooks for community)



Thank you!

Contact: egi-ace-po@mailman.egi.eu

Website: www.egi.eu/projects/egi-ace



[EGI Foundation](#)



[@EGI_eInfra](#)



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.