



Reproducibility with EGI services

Enol Fernández, Andrea Manzi, Sebastián Luna-Valero

3rd October 2024
EGI Conference 2024

TLP: WHITE Public

See: <https://www.egi.eu/services/>

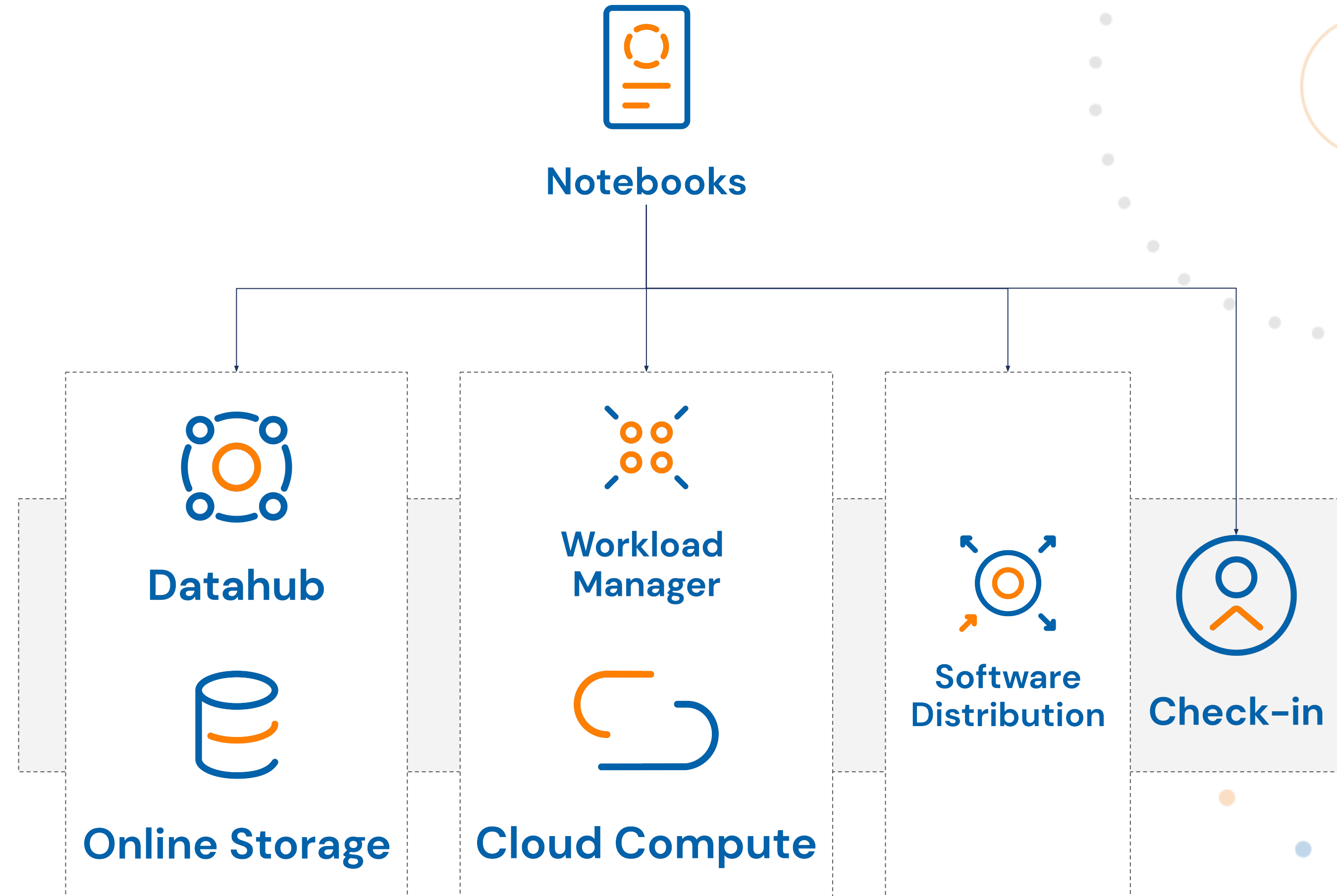
- EGI Notebooks and Replay
- EGI DataHub
- Reproducible deployments with TOSCA and Infrastructure Manager



Notebooks

Managed JupyterHub service

- Easy access: login Check-In and start using it
- Persistent storage with seamless access to DataHub, Software Distribution, and B2DROP
- Wide support for environments: Python, R (including RStudio), Julia, Octave, MATLAB, and user-installed conda environments





Managed BinderHub service

- Reproduce research on notebooks running on EGI Cloud
- No hard limits on sessions duration, customisable resource limits for users/communities
- Same access conditions and integrations as Notebooks

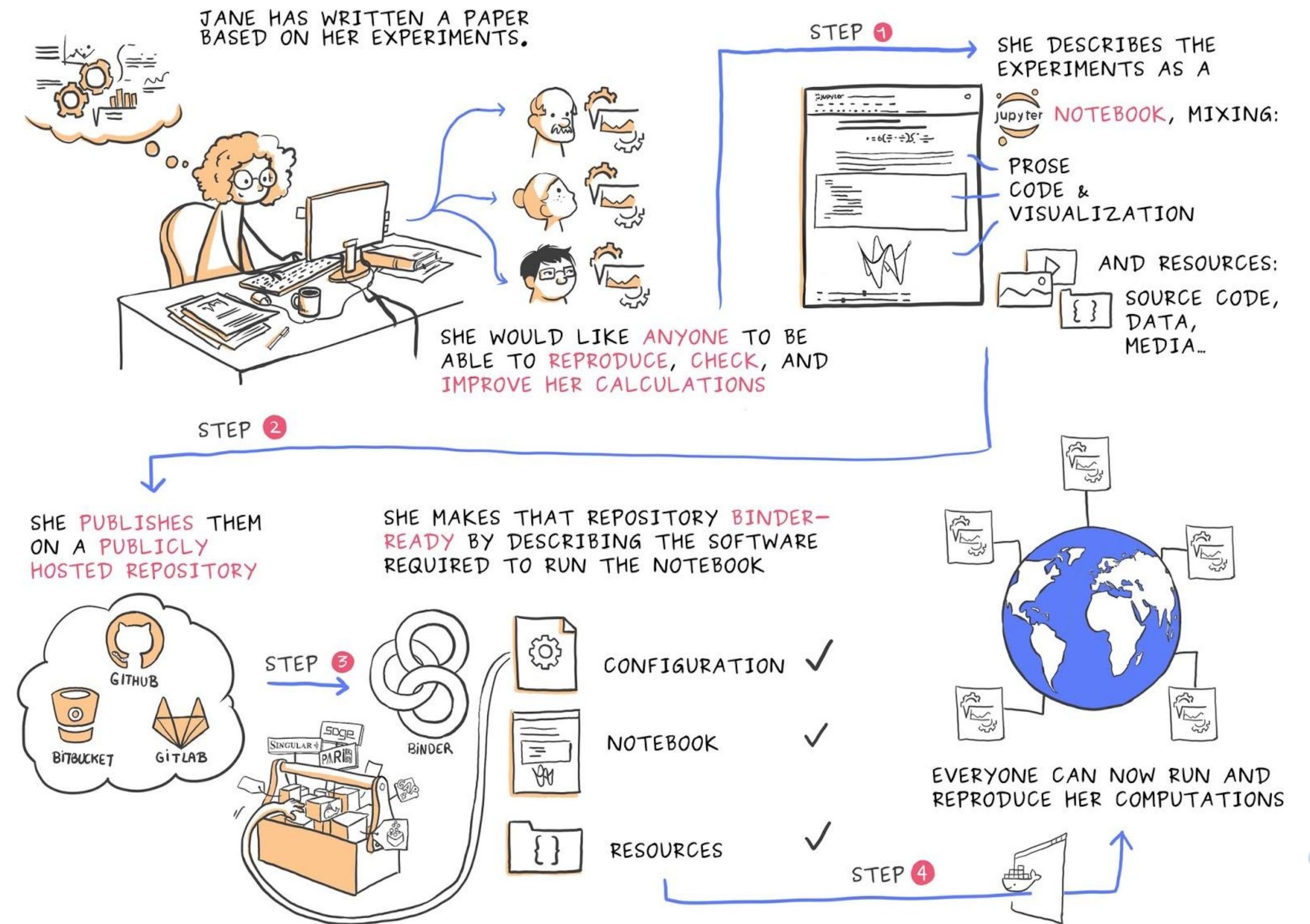
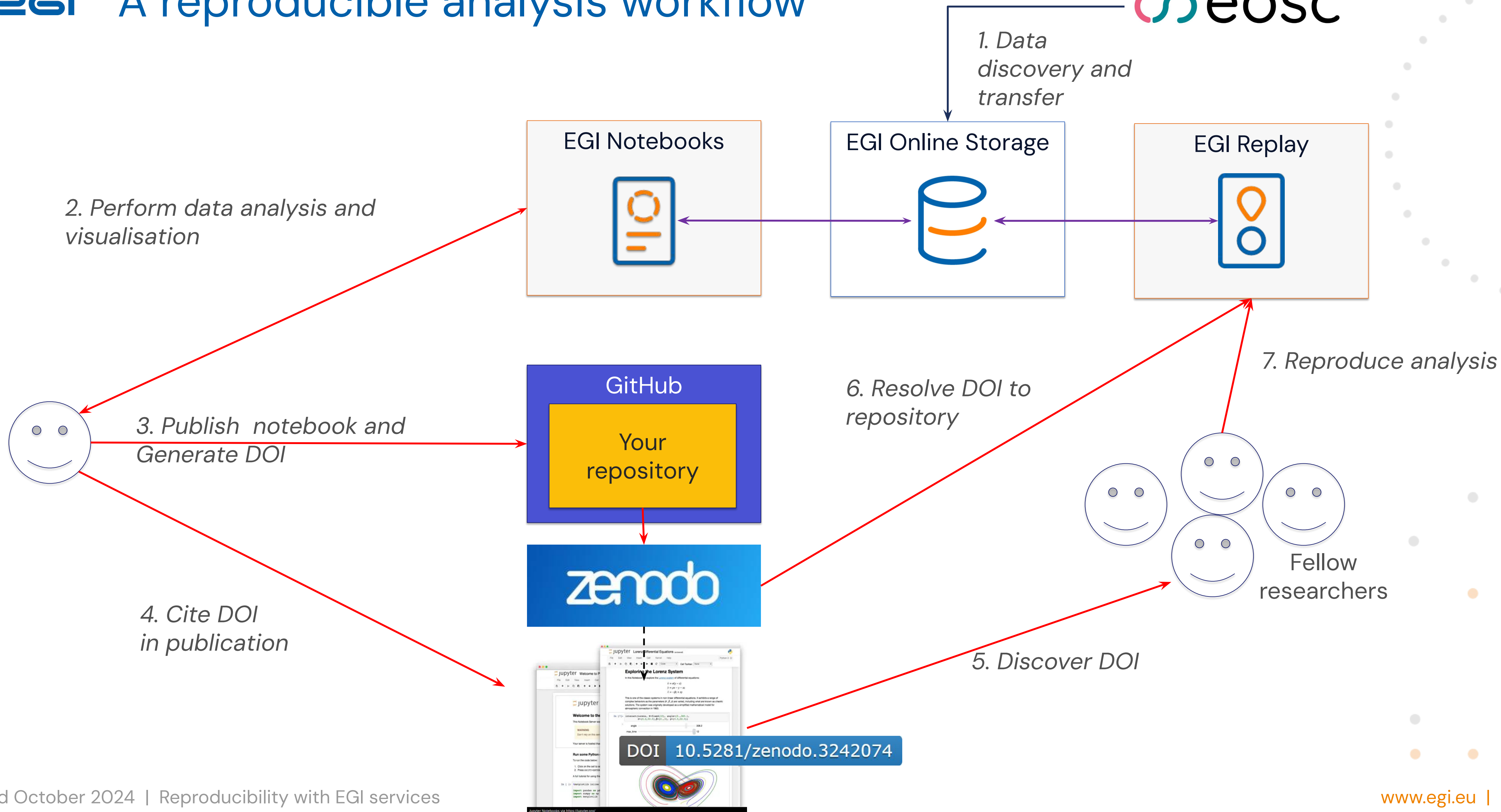


Image credit: <https://zenodo.org/record/4421040>



A&A 632, A78 (2019)
<https://doi.org/10.1051/0004-6361/201936349>
 © ESO 2019

**Astronomy
&
Astrophysics**

Evolution of compact groups from intermediate to final stages

A case study of the H I content of HCG 16*

M. G. Jones¹, L. Verdes-Montenegro¹, A. Damas-Segovia¹, S. Borthakur², M. Yun³, A. del Olmo¹, J. Perea¹,
 J. Román¹, S. Luna¹, D. Lopez Gutierrez⁴, B. Williams⁶, F. P. A. Vogt^{5, **}, J. Garrido¹, S. Sanchez¹,
 J. Cannon⁴, and P. Ramírez-Moreta¹

○ DOI:

<https://doi.org/10.1051/0004-6361/201936349>

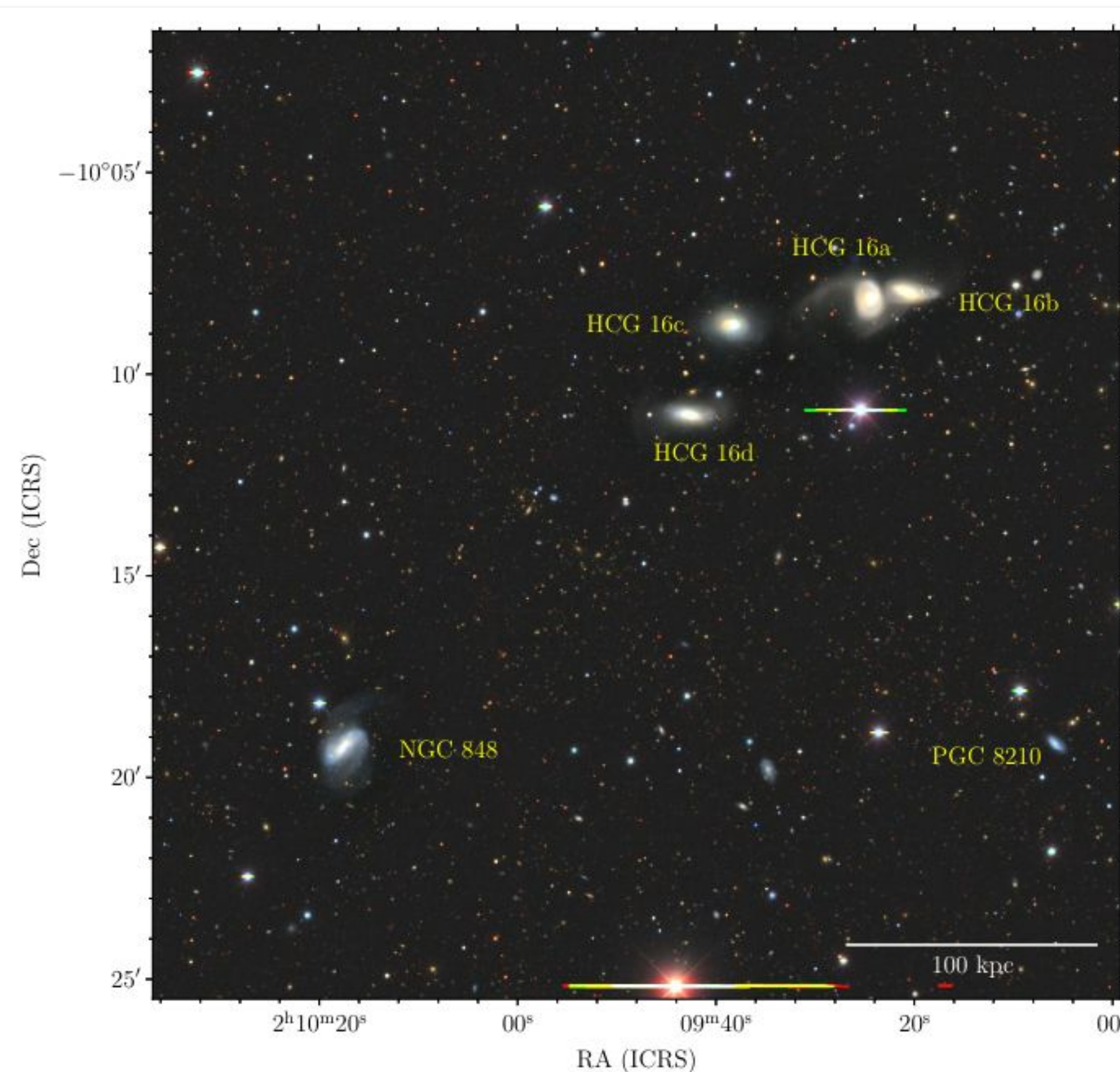
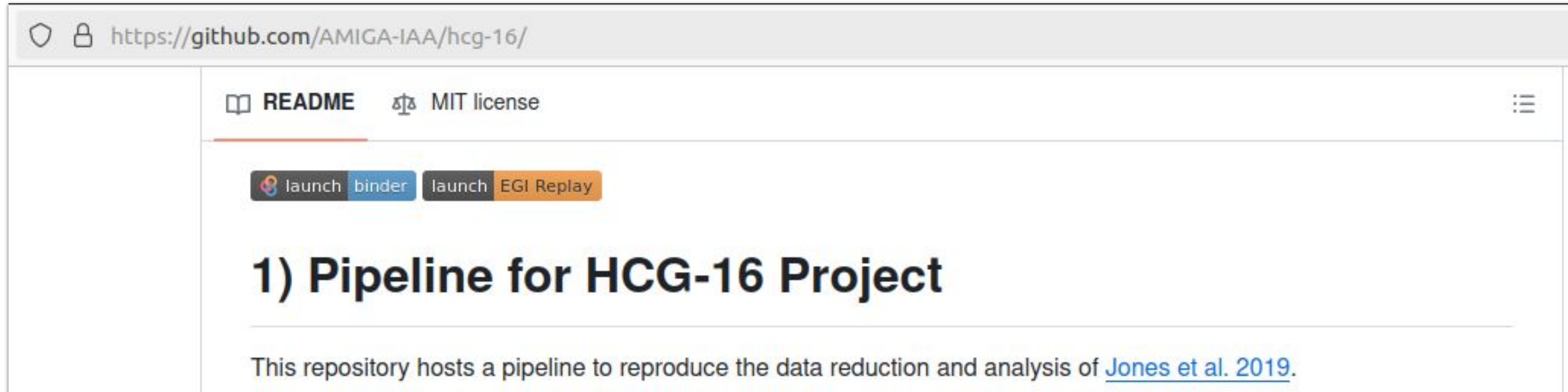


Fig. 1. DECaLS *grz* colour image of HCG 16 with the member galaxies labelled.




The screenshot shows a GitHub repository page for `AMIGA-IAA/hcg-16/`. The browser address bar displays `https://github.com/AMIGA-IAA/hcg-16/`. The repository page includes a navigation bar with `README` and `MIT license` links. Below the navigation bar, there are three buttons: `launch binder`, `launch EGI Replay`, and another `launch` button. The main heading is **1) Pipeline for HCG-16 Project**. The introductory text states: "This repository hosts a pipeline to reproduce the data reduction and analysis of [Jones et al. 2019](#)."

- Repository: <https://github.com/AMIGA-IAA/hcg-16/>



Watch the webinar

- <https://www.egi.eu/event/reproducible-open-science-with-egi-replay/>



Webinar

Reproducible Open Science with EGI Replay

17 April 2024 - 14:00 CEST

EGI
WEBINAR PROGRAMME 2024
Reproducible Open Science with EGI Replay

Sebastian Luna-Valero (EGI) | 17 April 2024 - 14:00 CET

- More webinars in <https://www.egi.eu/trainings-and-webinars/>



• Communities using EGI Replay

- The Environmental Data Science book
 - Collection of reproducible notebooks in Earth Science
 - Check out the gallery: <https://edsbook.org/notebooks/gallery>

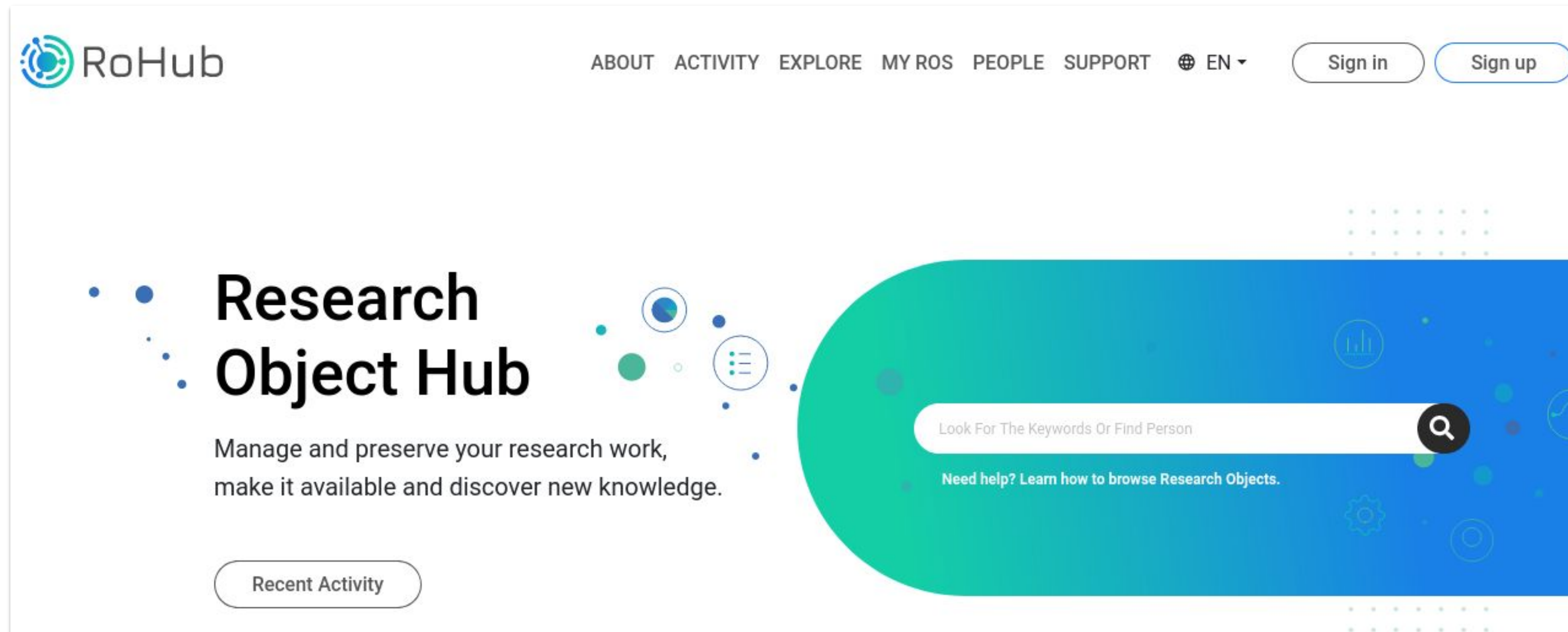
The screenshot shows the EDS book website interface. On the left is a navigation sidebar with the EDS book logo (featuring two globes and the text '1010 1010 EDS book') and links for 'Welcome', 'Preamble', 'About EDS book', 'Citation and Reuse', 'Contribute', 'Notebooks', and 'Our Notebooks'. The main content area is titled 'Gallery' and displays three notebook cards:

- Card 1:** 'Variational data assimilation with deep prior (CIRC23)'. Tags: Ocean, Modelling, Special Issue, Python. It shows two side-by-side maps labeled 'Ground truth' and 'Deep prior 4D-Var'.
- Card 2:** 'Deep learning and variational inversion for climate science (CIRC23)'. Tags: General, Modelling, Special Issue, Python. It features a line graph of GMST (°C) from 1900 to 2020, comparing 'Inversion of GHG', 'Inversion of AER', 'Inversion of NAT', and 'Obs'.
- Card 3:** 'Underlying physics of the ocean's temperature (CIRC23)'. Tags: Ocean, Modelling, Special Issue, Python. It displays a global map of ocean temperature anomalies.



Replay

- **Communities using EGI Replay**
 - The RELIANCE project
 - Research Objects integration with EGI Replay
 - See <https://www.egi.eu/case-study/reliance/>





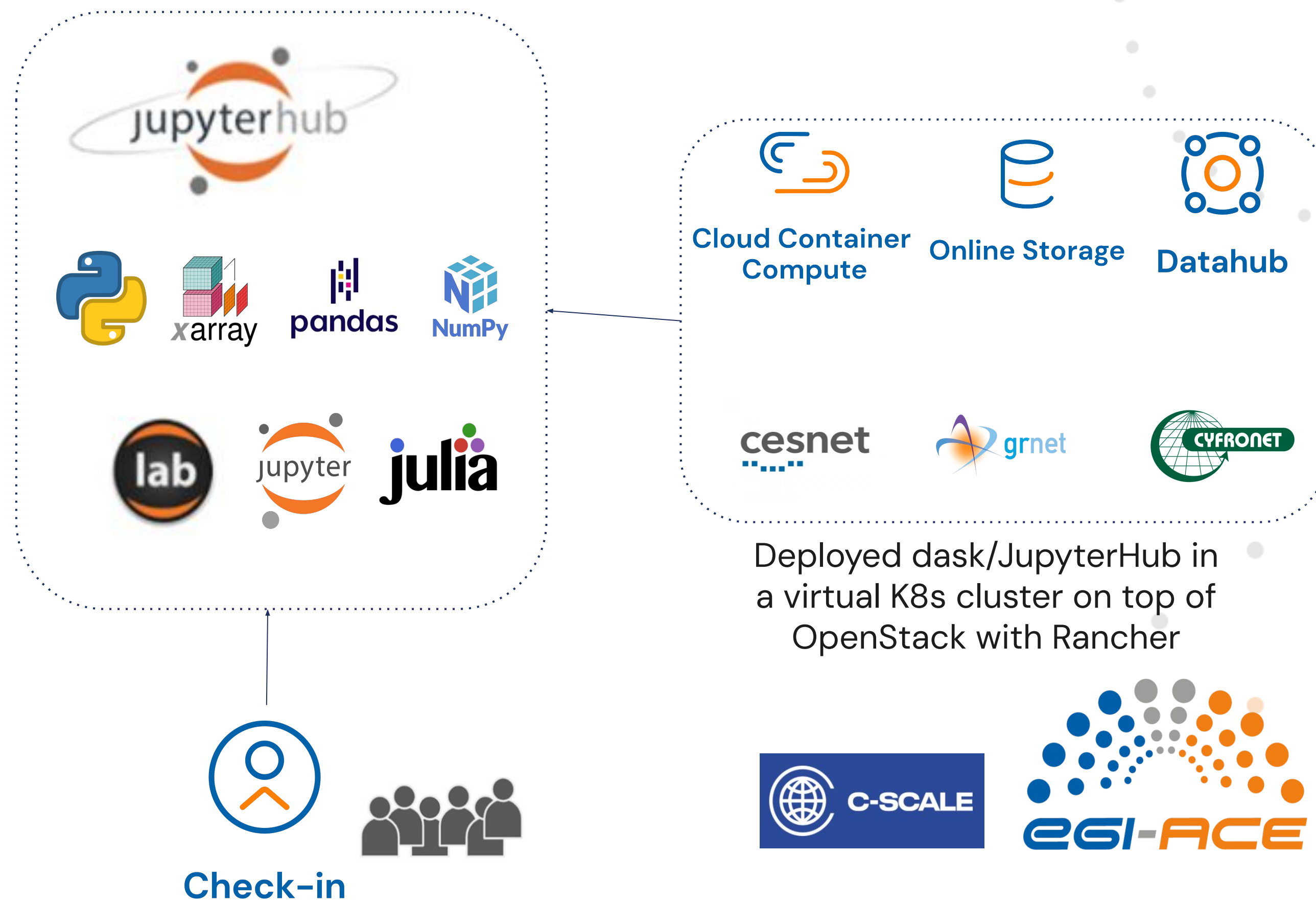
Notebooks



• Communities using EGI Notebooks

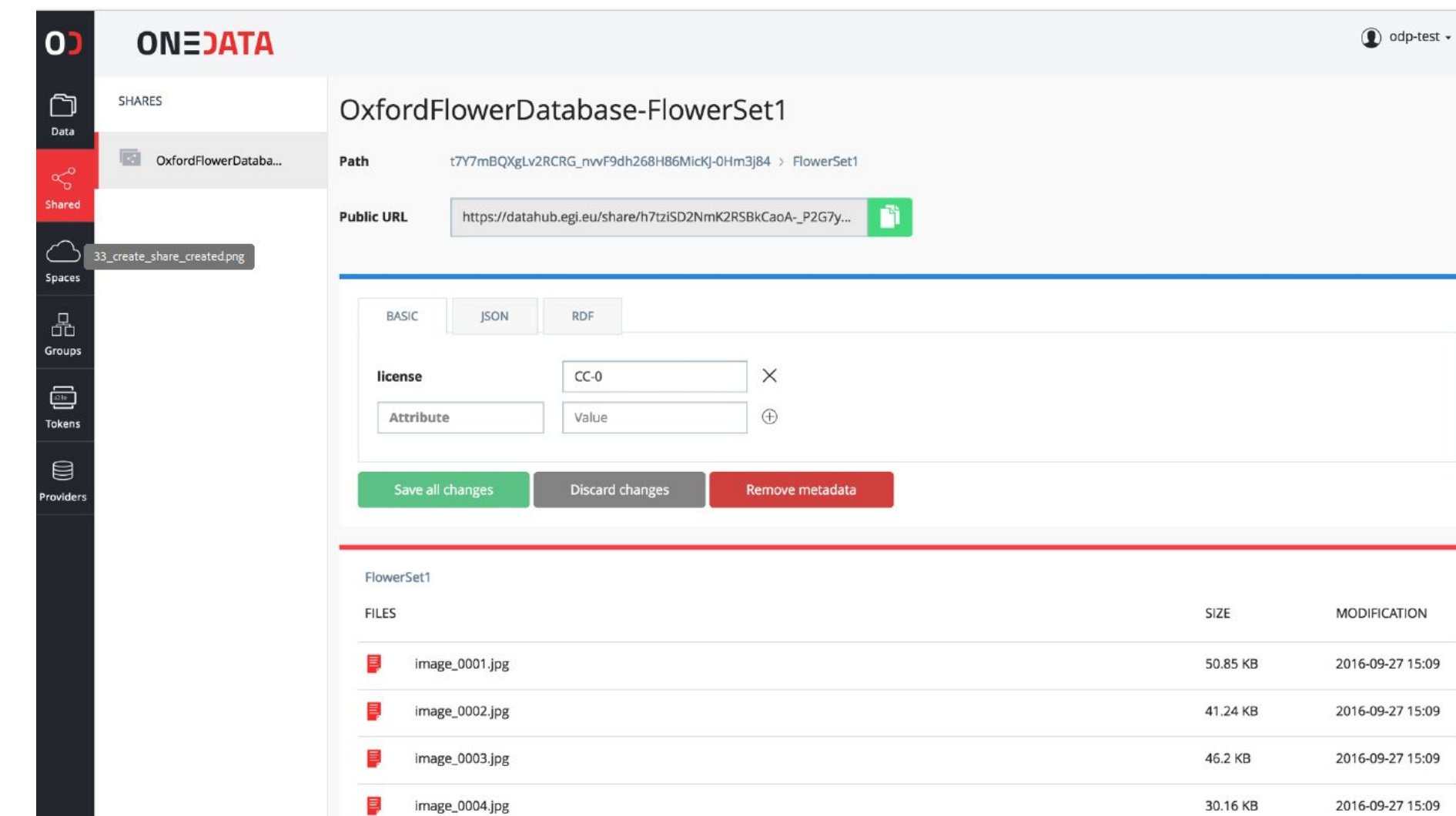
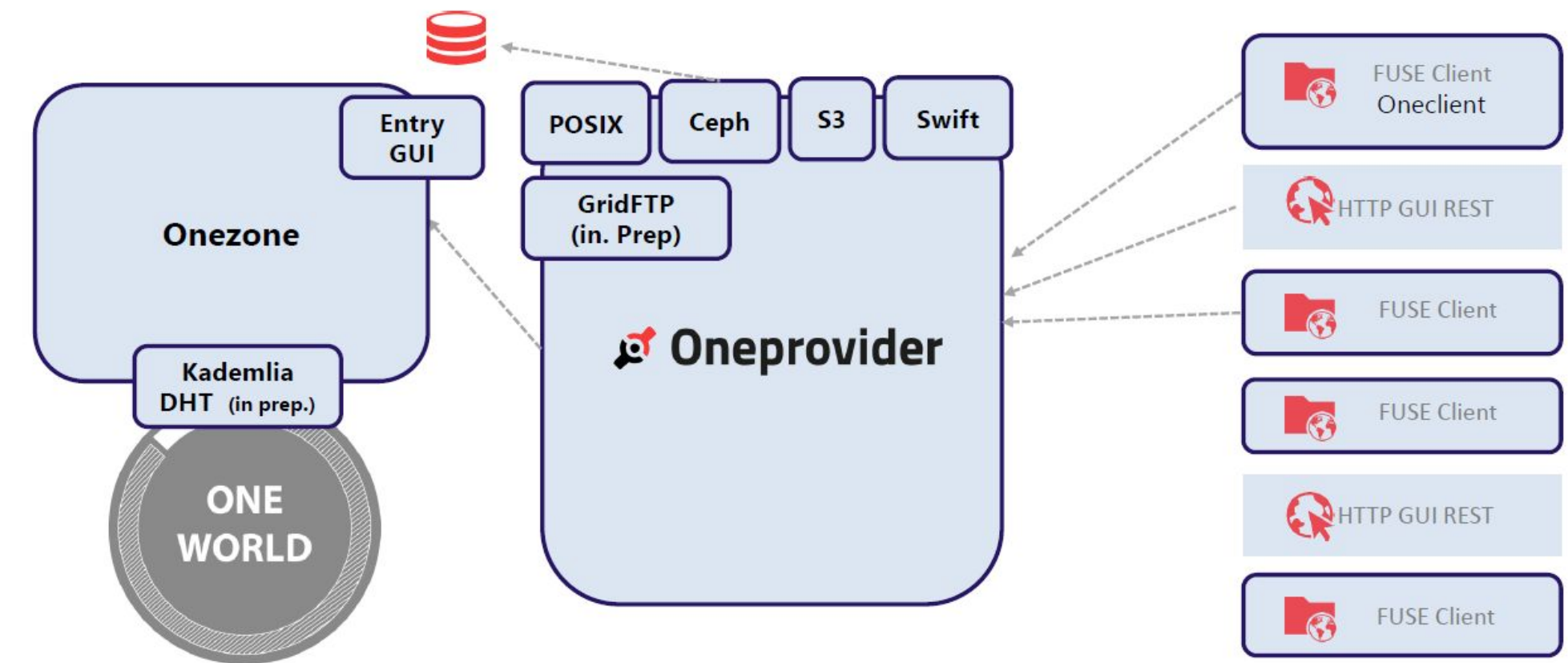
- Pangeo@EOSC

- Dedicated deployment for the Pangeo community in Europe
- Goal: Foster collaboration around best practices for open, reproducible, and open science
- 100+ researchers trained until today



Data management service

- Based on OneData.
- Datasets can be stored in **distributed nodes** transparently for the user.
- Different **types of storage** virtualised (S3, POSIX, Ceph, OpenStack Swift...).
- **Metadata management** for files and directories.
- **GUI** and **API** to query files by metadata, time, owner etc.
- Automatic retrieval of **PIDs** for open data.
- Data access **permissions** based on ACL and others.

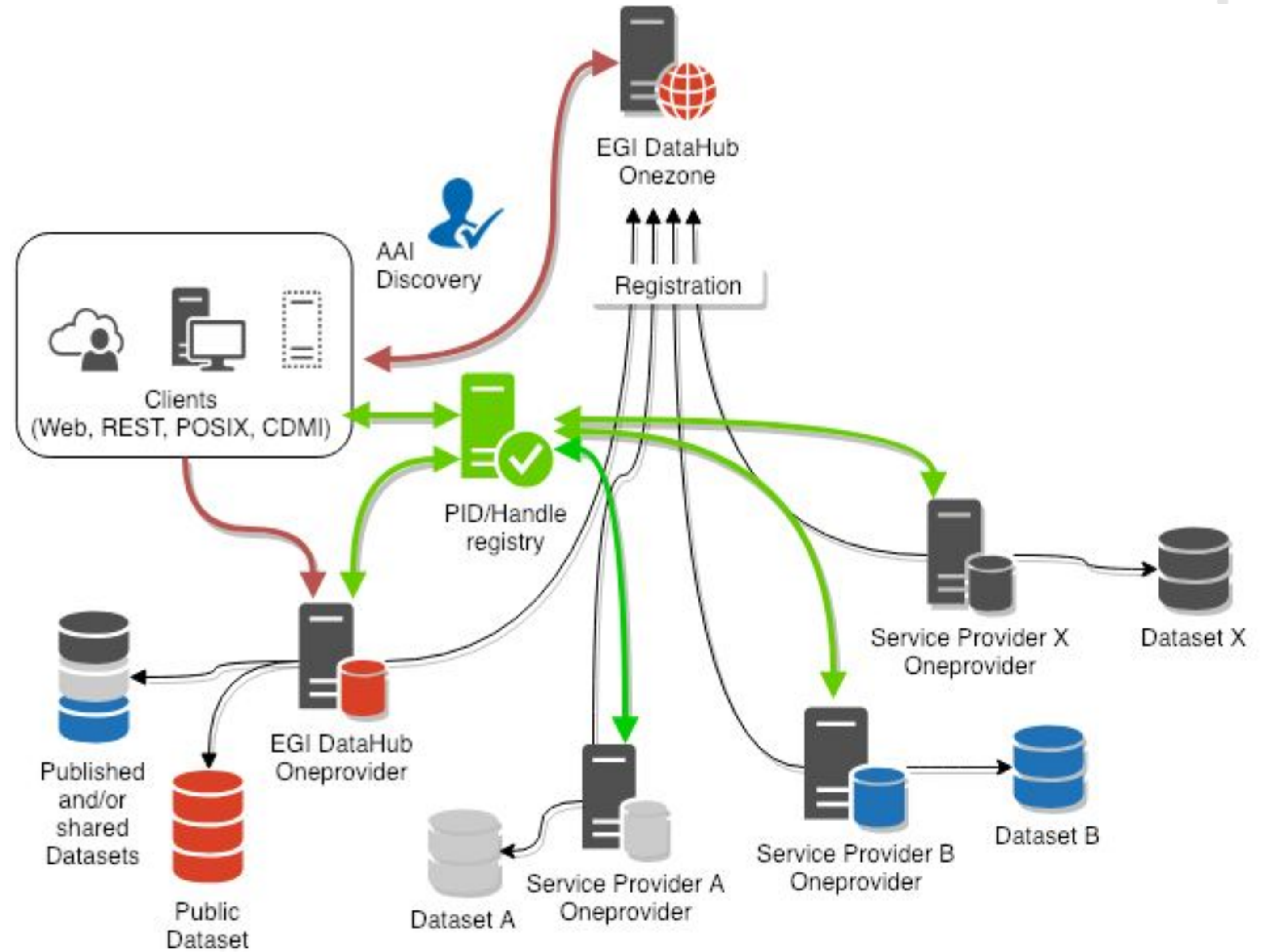


Directories can be publicly shared

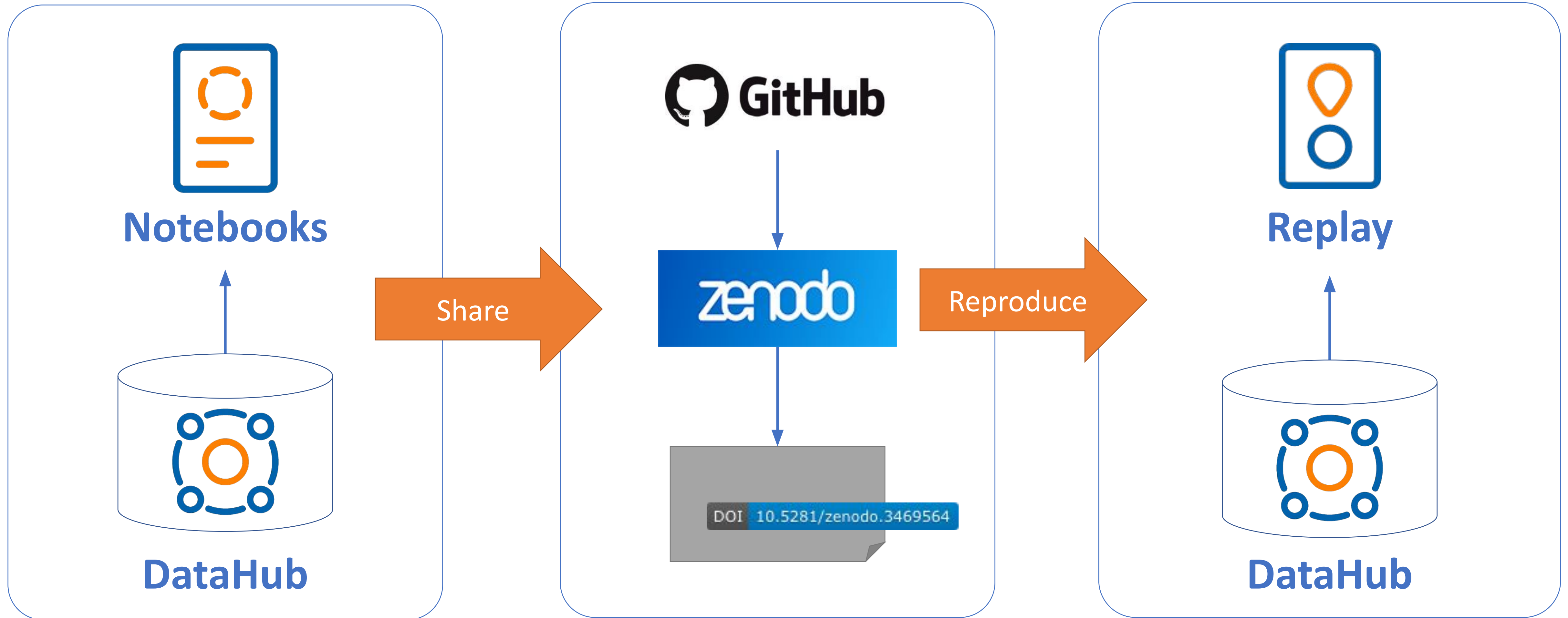
- <https://datahub.egi.eu/share/8a7d0e1de074f8ab12cfdd8f2428f4b8ch6b7b>

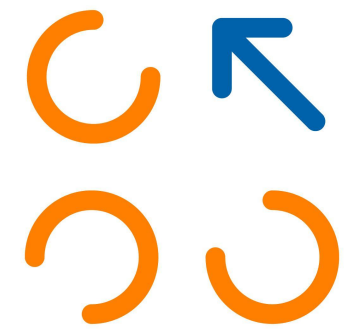
Shares can eventually be published via Handle services and have a PID assigned

- <http://hdl.handle.net/21.T15999/QBFI7Pw>



- Communities using DataHub








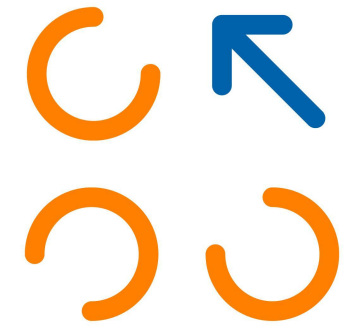


Visit: <https://im.egi.eu/>



Available templates

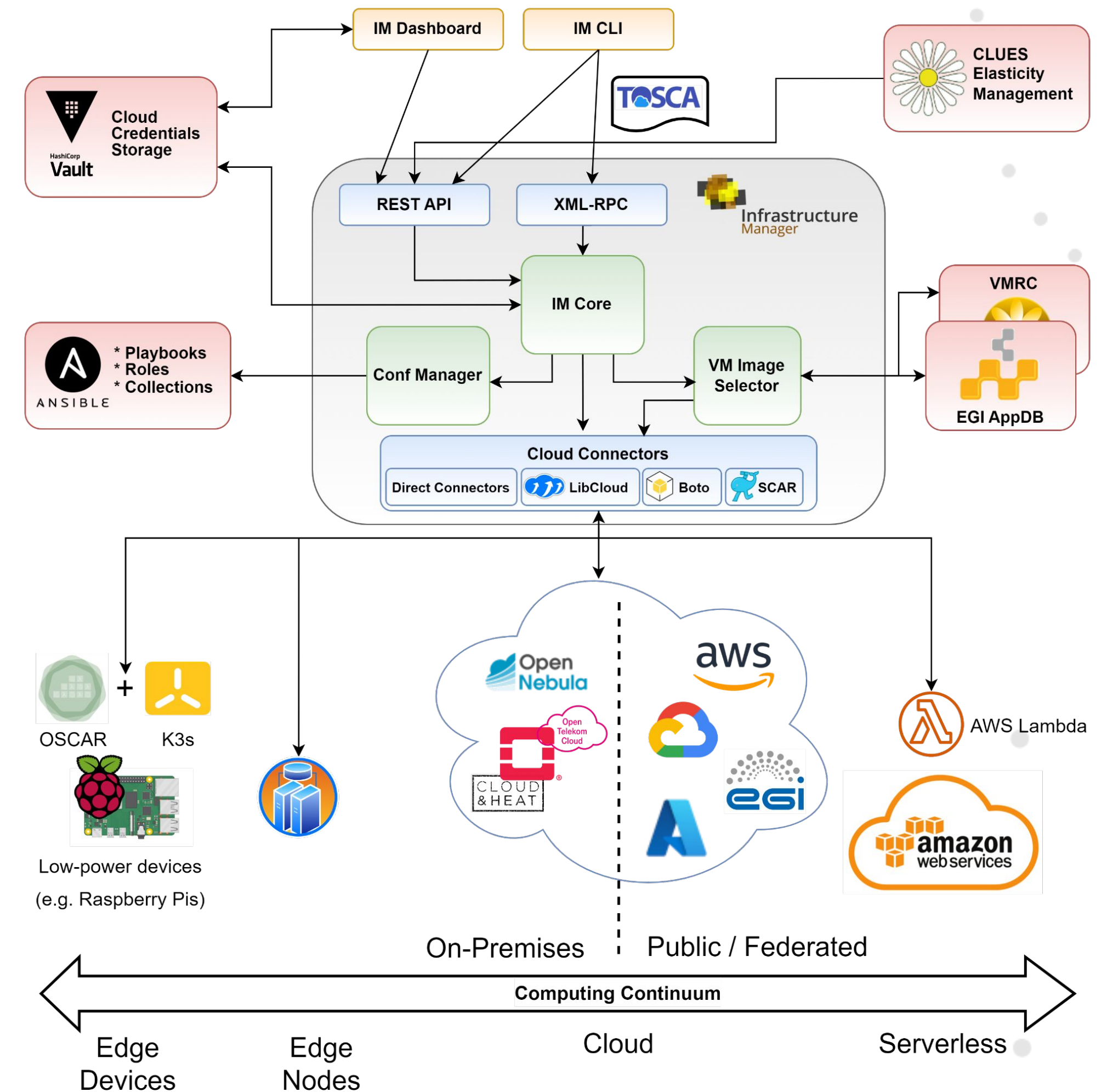
<p>Deploy a VM</p>  <p>VM</p>	<p>Install Docker + Docker Compose</p> 
<p>SLURM virtual cluster</p> 	<p>Deploy a ARC in front of a Slurm cluster</p> 
<p>Deploy a Kubernetes Virtual Cluster</p> 	<p>Deploy Kubeflow on top of a Kubernetes Virtual Cluster</p> 

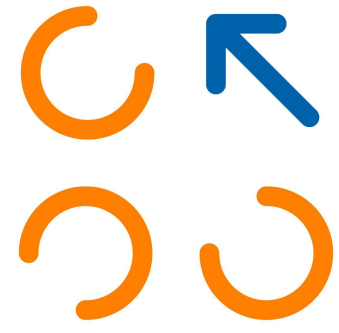


Infrastructure Manager

IM deploys virtual infrastructures on the Cloud

- **Automates** the deployment, configuration, software installation, monitoring and update of virtual infrastructures
- **Infrastructure as Code (IaC)** using **TOSCA** for infrastructure description & **Ansible** for contextualization
- Wide variety of back-ends from edge to serverless, making applications **Cloud agnostic**



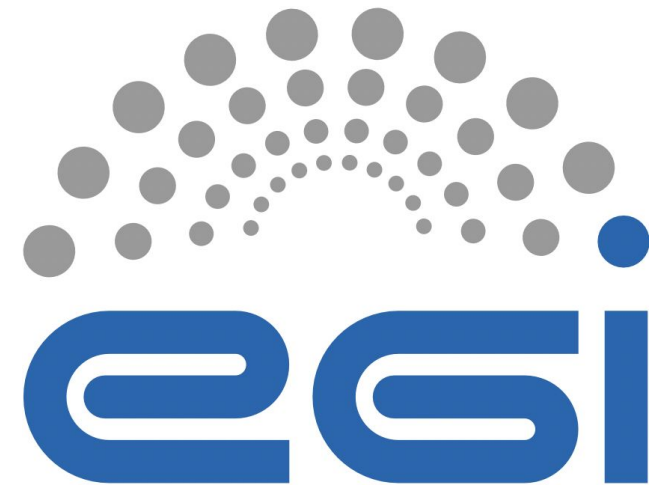


Simple TOSCA example:

```
1  tosca_definitions_version: tosca_simple_yaml_1_0
2
3  imports:
4  - grycap_custom_types: https://raw.githubusercontent.com/grycap/tosca/main/custom_types.yaml
5
6  topology_template:
7    node_templates:
8      simple_node:
9        type: tosca.nodes.indigo.Compute
10       capabilities:
11         endpoint:
12           properties:
13             network_name: PUBLIC
14       host:
15         properties:
16           num_cpus: 2
17           mem_size: 4 GB
18       os:
19         properties:
20           image: appdb://TR-FC1-ULAKBIM/egi.ubuntu.24.04?vo.access.egi.eu
21     outputs:
22       node_ip:
23         value: { get_attribute: [ simple_node, public_address, 0 ] }
24       node_creds:
25         value: { get_attribute: [ simple_node, endpoint, credential, 0 ] }
```




What would you like to see?



Contact us

contact@egi.eu

Let's talk. Or
meet in person

Get in touch with us

www.egi.eu



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