Open Science and reproducibility at SRCs

Julián Garrido Sánchez (@tetrarquis)
on behalf of AMIGA group
Instituto de Astrofísica de Andalucía (CSIC)

AENEAS meeting, Utrecht
13 November 2019

OPEN

Knowledge is open if anyone is free to access, use, modify, and share it

OPEN KNOWLEDGE FOUNDATION

A WORLD WHERE KNOWLEDGE CREATES POWER FOR THE MANY, NOT THE FEW.
THIS IS THE WORLD WE CHOOSE.

Open Definition, Open Knowledge Foundation, https://okfn.org
Open Science, European Commission, 2017, doi: 10.2777/75255

OPEN

Knowledge is open if anyone is free to access, use, modify, and share it

OPEN KNOWLEDGE FOUNDATION

A WORLD WHERE KNOWLEDGE CREATES POWER FOR THE MANY, NOT THE FEW.
THIS IS THE WORLD WE CHOOSE.

Open Science represents an approach to research that is collaborative, transparent and accessible

Open Definition, Open Knowledge Foundation, https://okfn.org

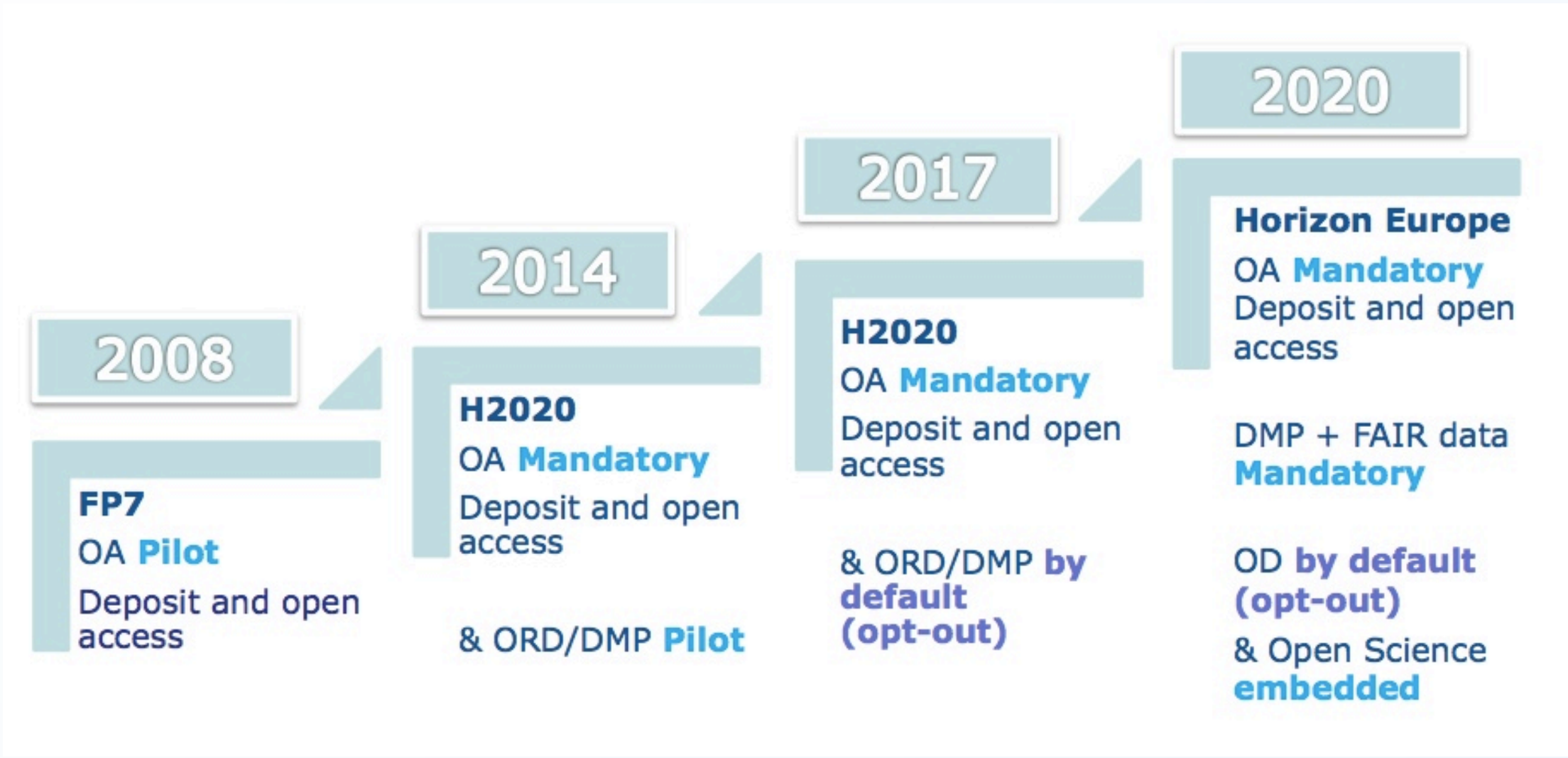
Open Science

Other activities that come under the umbrella of Open Science:

- Open access publishing
- Open data
- Open Source
- Open peer review
- Open research
- Citizen science + stakeholder engagement

Open Science, European Commission, 2017, doi: 10.2777/75255

OPEN ACCESS IN THE FPS

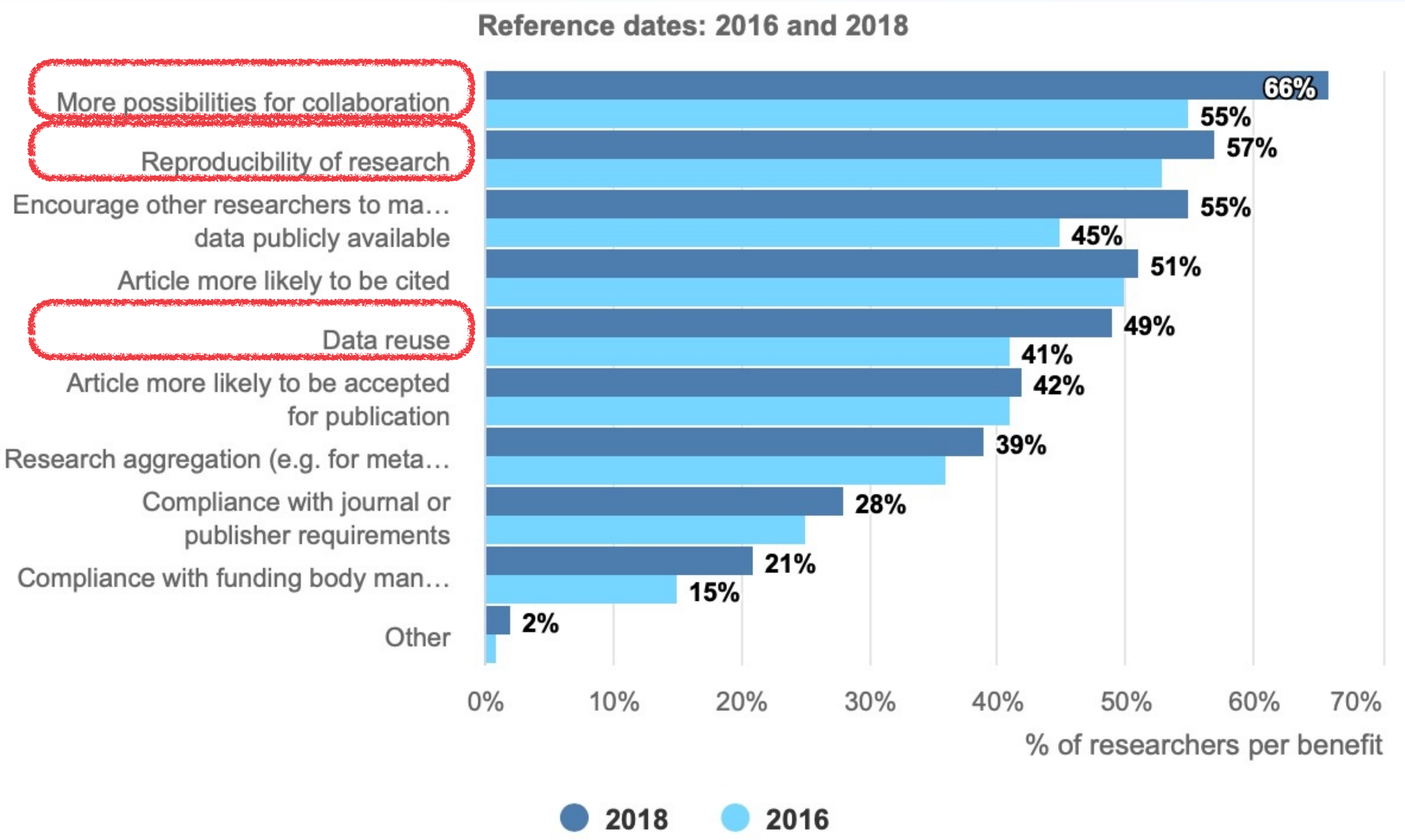


SRCs: EU + SKA policies

Source: https://ec.europa.eu/research/mariecurieactions/sites/mariecurie2/files/05_open_access_itn_2018.pdf

ATTITUDE TOWARDS DATA SHARING

Benefits of sharing research data

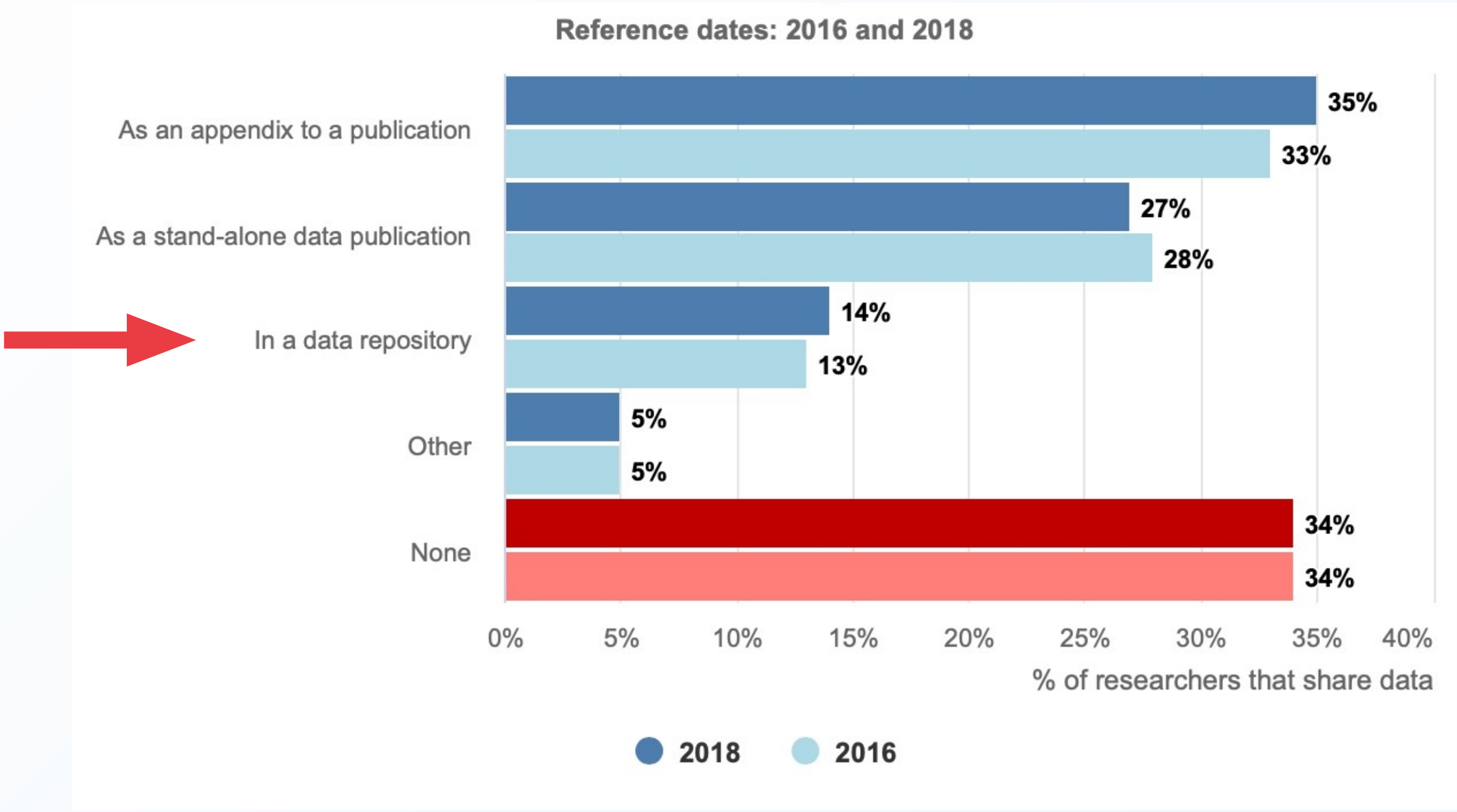


% of researchers per benefit

Source: Open science monitor

ATTITUDE TOWARDS DATA SHARING

% of researchers that share data



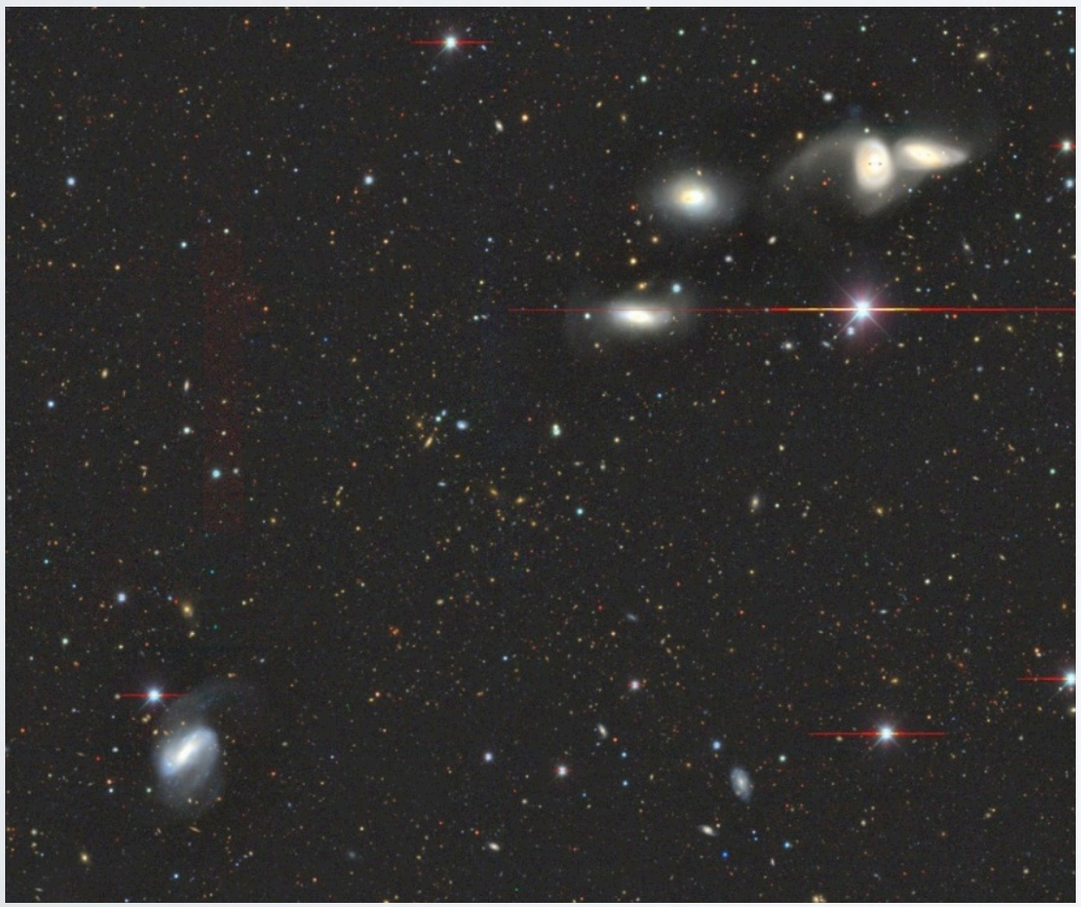
Source: Open science monitor

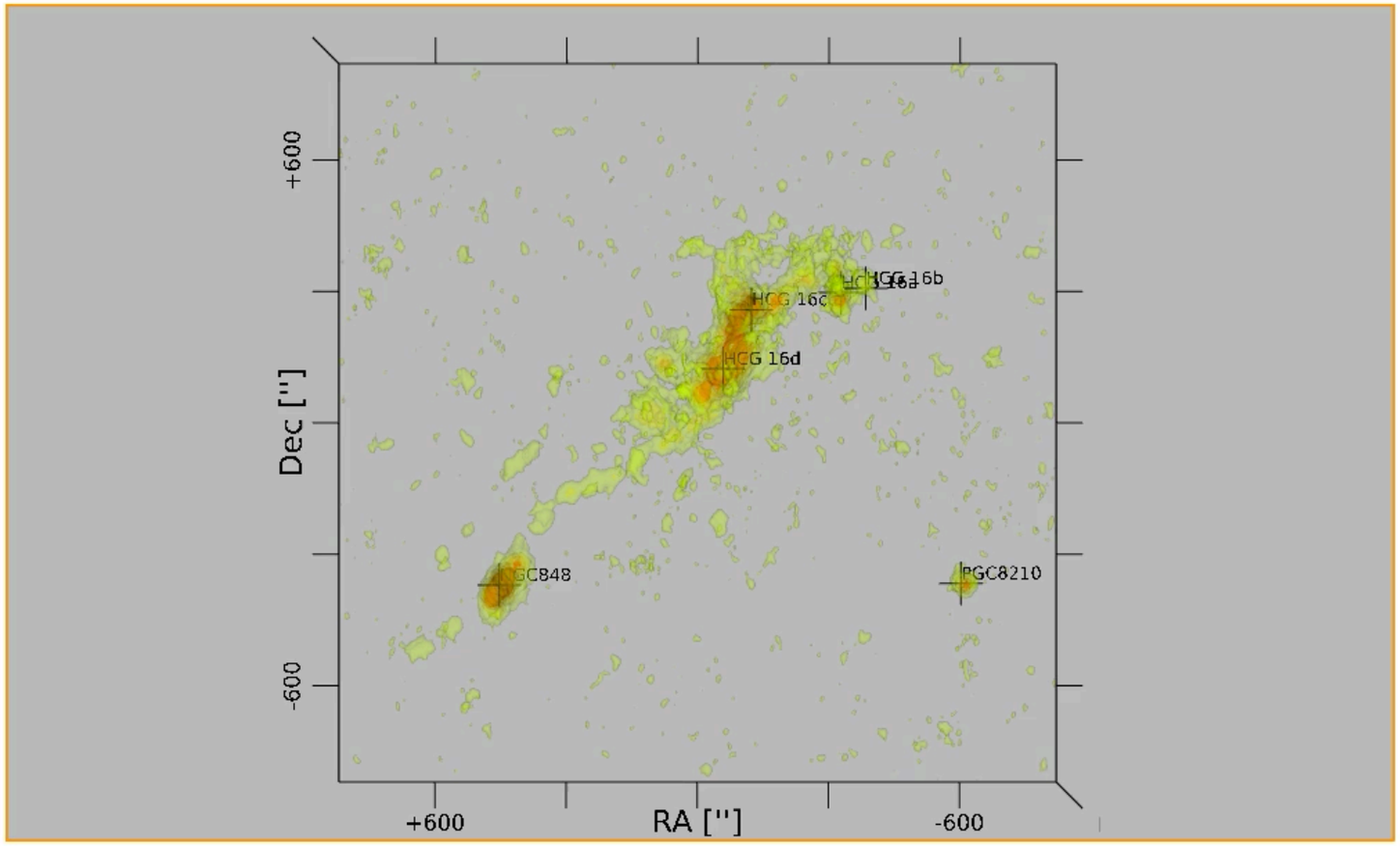
Open Science in a real scientific experiment

http://littlepandabears.blogspot.com/2018/09/galaxies.html
https://searchengineland.com/figz/wp-content/seloads/2014/08/lab-test-experiment-ss-1920.jpg

SPECIFIC EXAMPLE: HI IN HCG 16

- HCG 16 is complex compact group with starburst galaxies, AGN, tidal tails, etc. The main goal of this project is to is to study the HI content of the group and to determine which on-going processes are causing it to change.
- **Collaborators:** L. Verdes-Montenegro, A. Damas, S. Borthakur, M. Yun, A. del Olmo, J. Perea, B. Williams, D. Lopez Gutierrez, F. Vogt, S. Luna, J. Román, J. Garrido, S. Sanchez, J. Cannon & P. Ramírez





Viewpoints: Reset View | R.A. - Dec. | V - Dec. | V - R.A. | Perspective View | Next View

HI layers: 12-sigma | 9-sigma | 6-sigma | 3-sigma

BEING FAIR

FAIR (www.go-fair.org) is a multi-disciplinary bottom-up initiative to make scientific data reusable. The FAIR principles state that scientific data should be:

- **Findable:** Data have sufficient metadata and unique, persistent identifiers in a searchable database.
- **Accessible:** Data is stored in trusted/standard repository. Metadata and data can be understood by machines/people.
- **Interoperable:** Metadata use a standard language, external connections to other data/resources are qualified.
- **Reusable:** Data have sufficient provenance information and clear licenses.

BEING UN-FAIR

Common astronomy examples of un-FAIR practices:

- The final data are "available", but you **need to request** them by email.
- The **raw** data are in an **archive** but the final, **reduced data** and images are only publicly available in the paper **PDF**.
- There are some **scripts** for processing the data on a server somewhere, but no one remembers how to run them.
- The code is on **github**, but good luck trying to install/execute it.

I'm not pointing fingers here, we are all guilty of these things, myself included. We need to improve as a community.

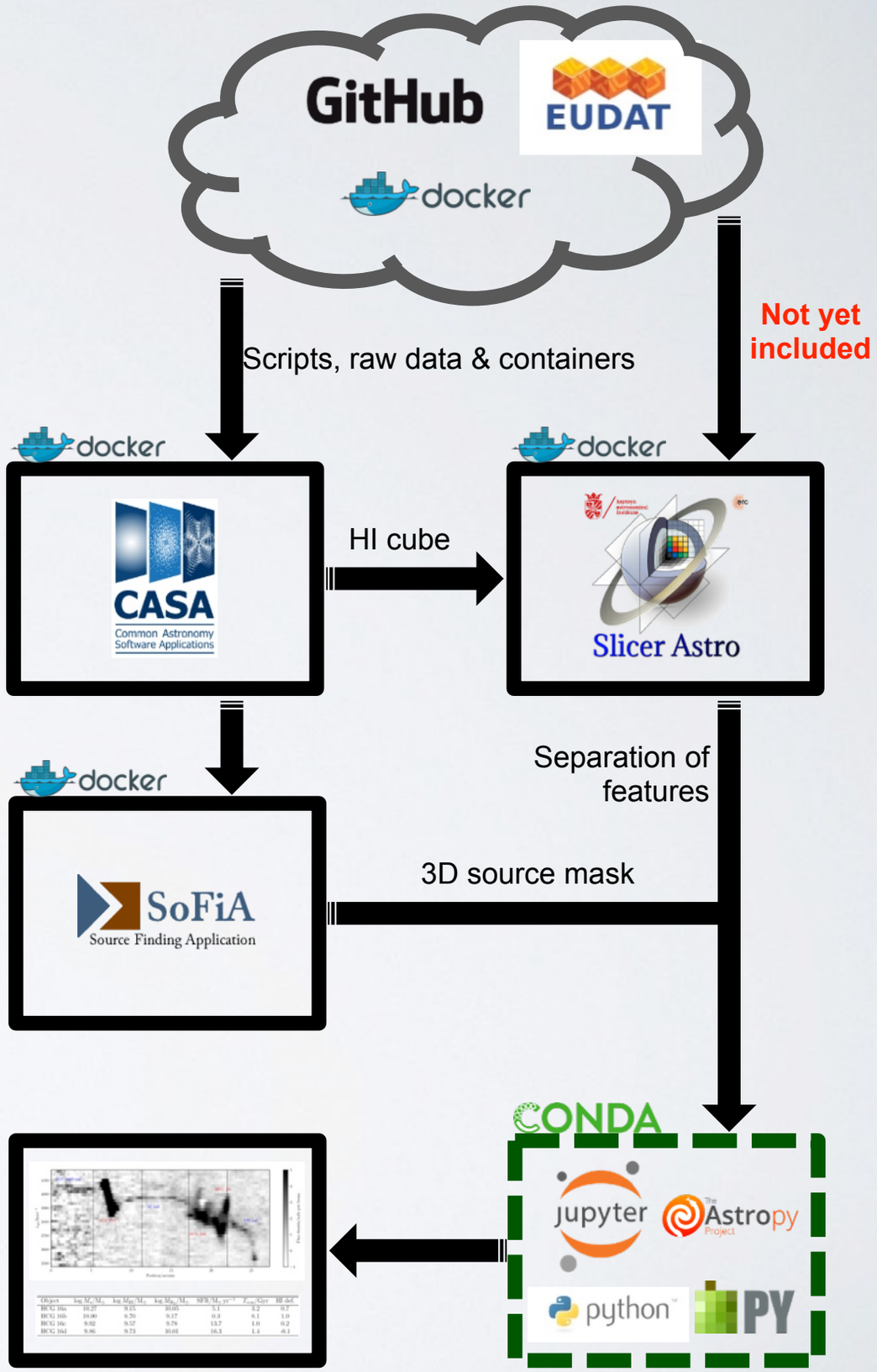
Our Workflow

- FAIR focuses on the **data,** we want to go beyond this and include also the **methods**.

- It is executed entirely within **Docker** containers and **Conda** environments. So it can be run on any platform with Docker and Conda, using a single bash script.

- The code and data are publicly available in **github** and **EUDAT**.

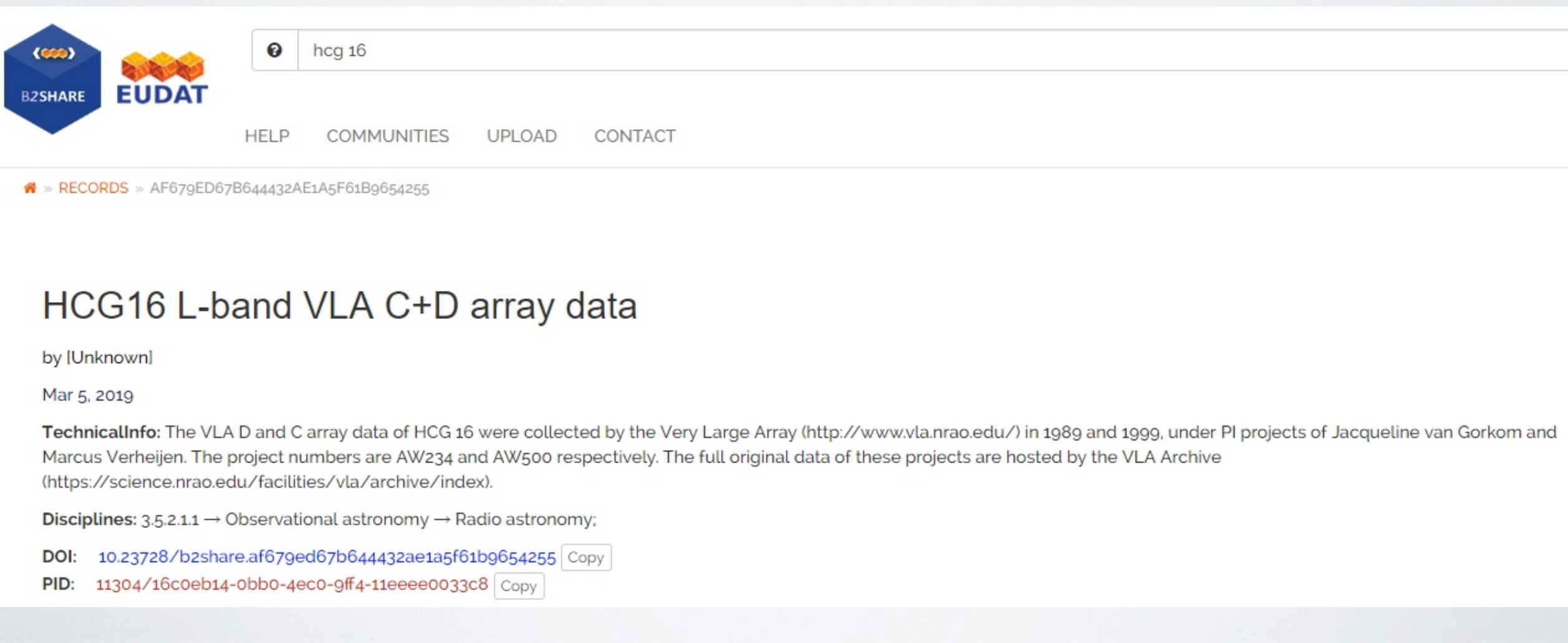
- The workflow can also be executed in EOSC



Our Workflow

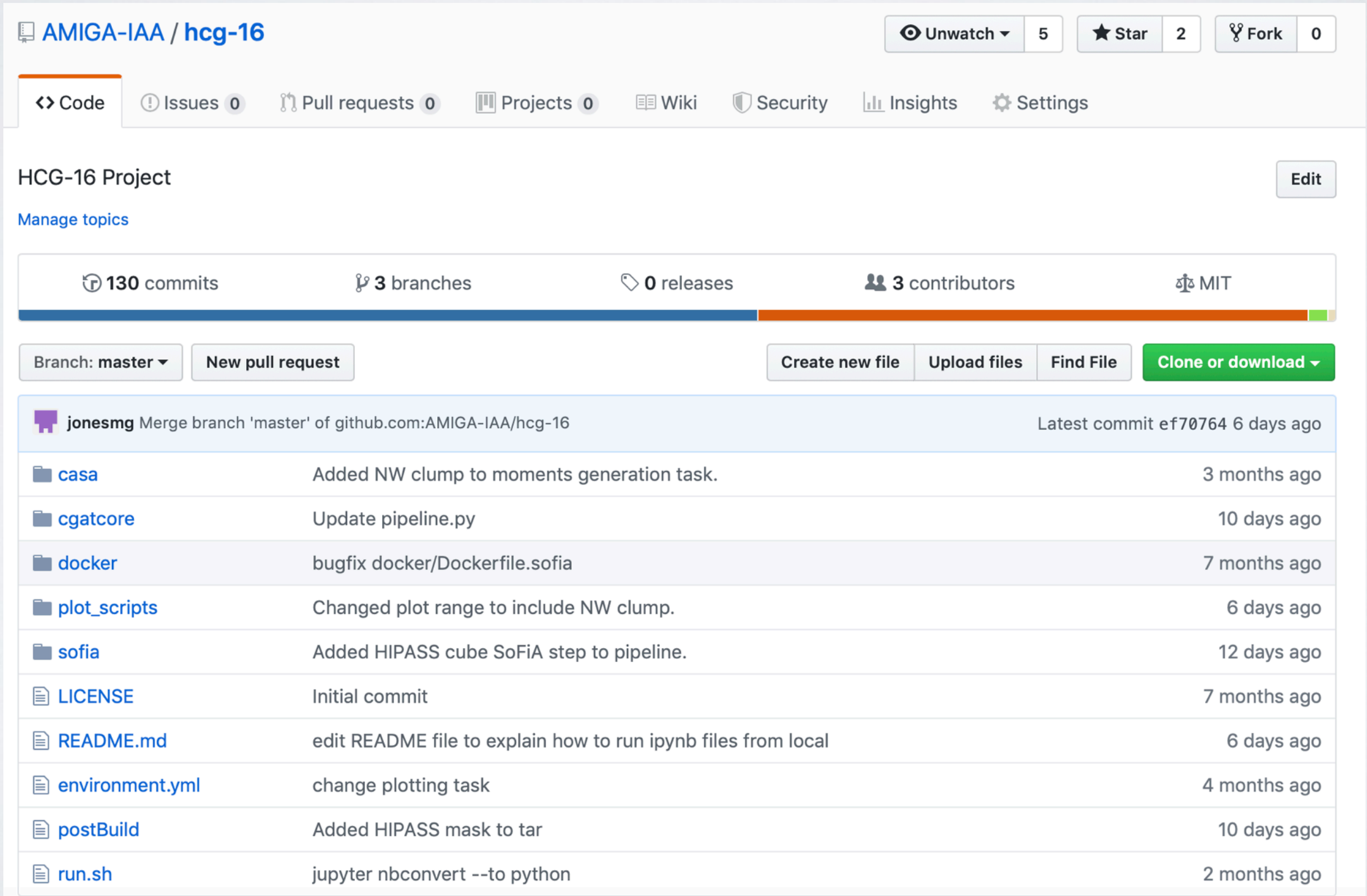
The raw data are hosted on a the EUDAT service, which provides:

- Cloud storage
- Persistent identifiers (DOI)
- Access (can download with wget)
- Basic metadata and search functionality



Our Workflow

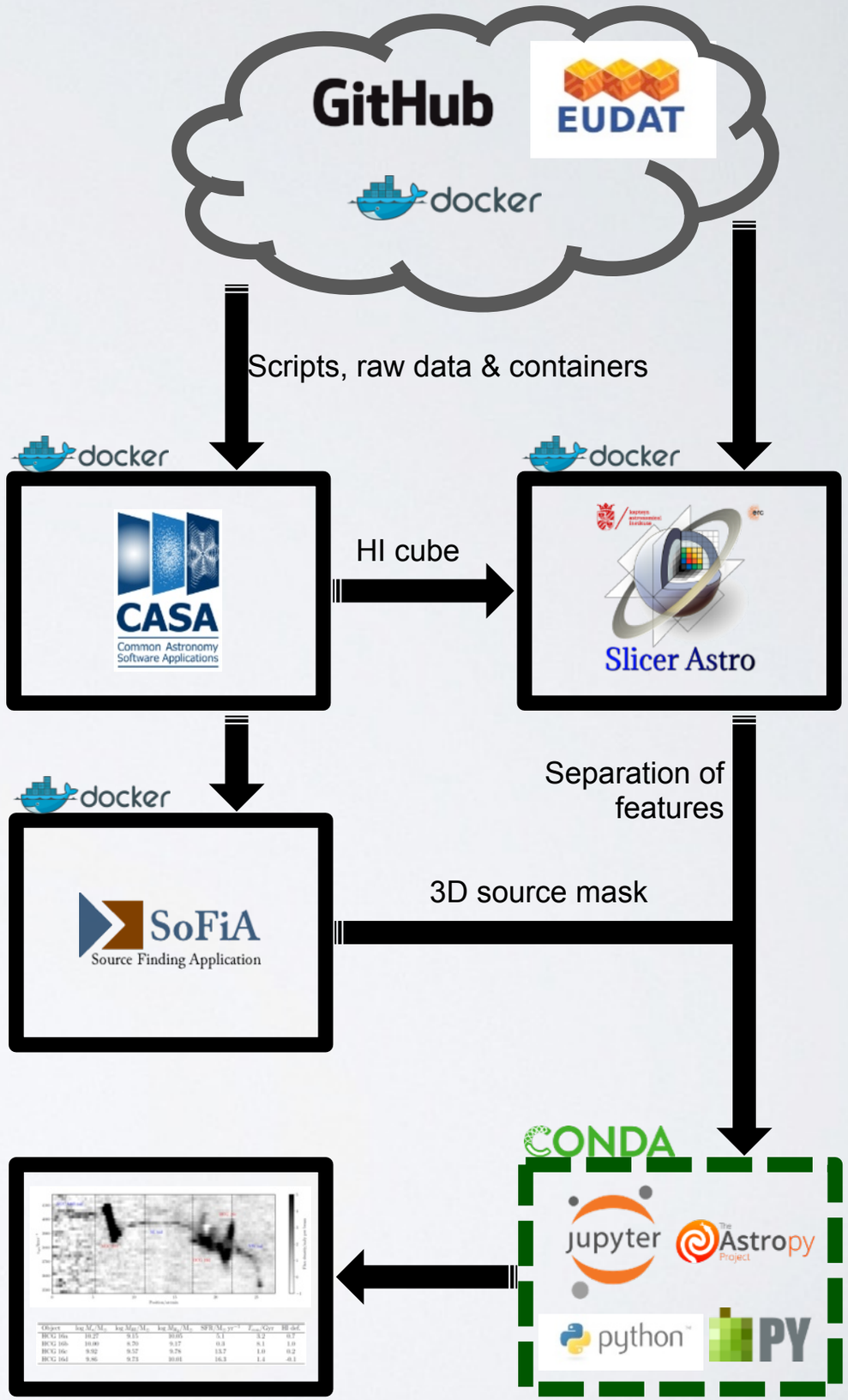
- All the code for the all of the workflow from raw data to final plots is stored in github and is openly accessible.



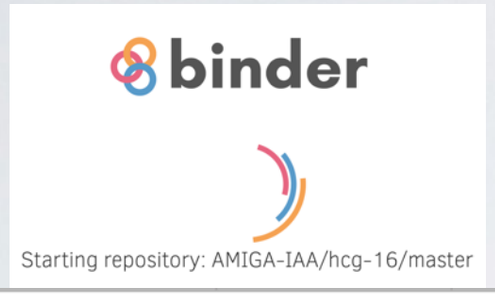
https://github.com/AMIGA-IAA/hcg-16

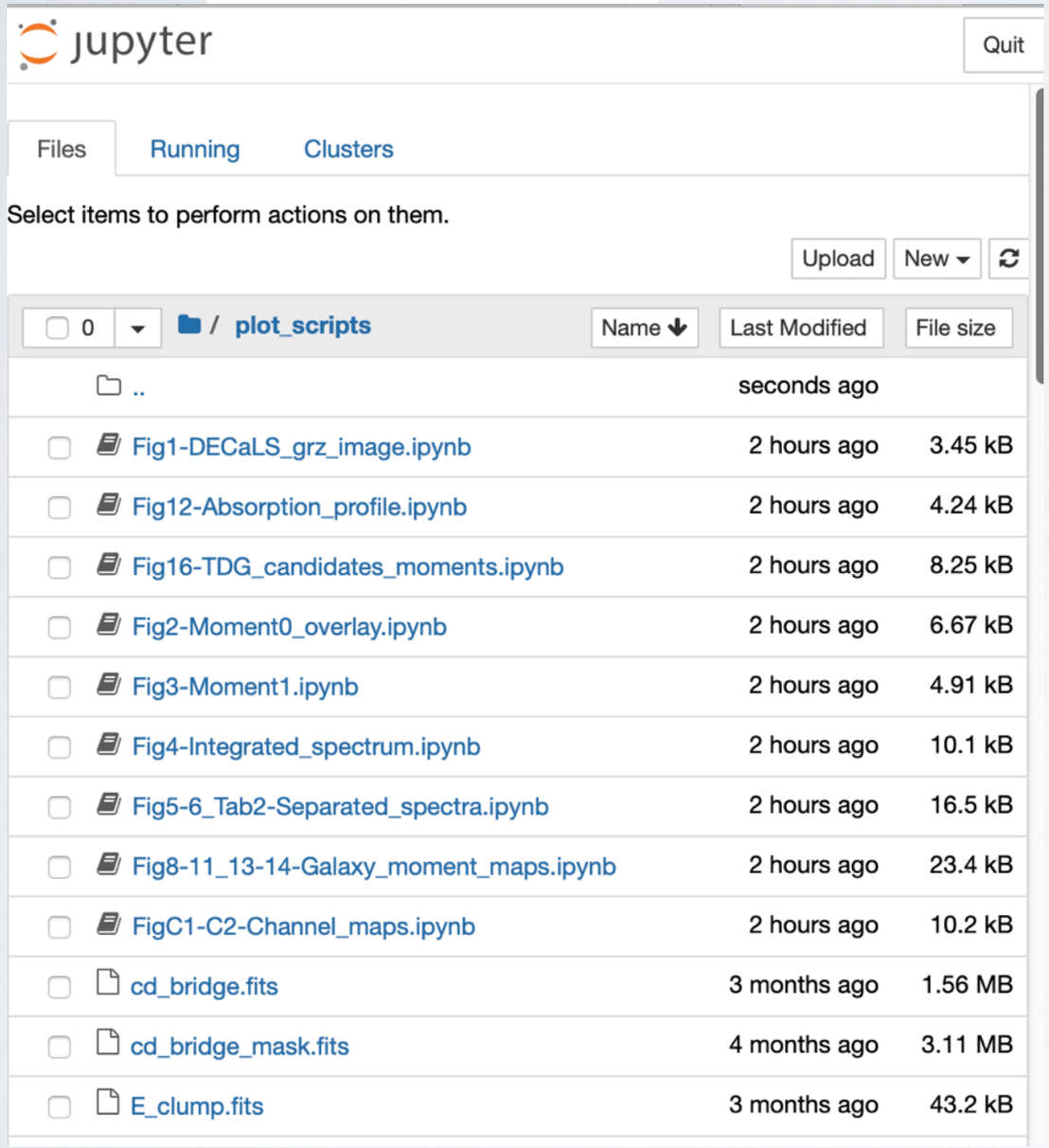
Our Workflow

- run.sh will do automatically the following steps:
  - download and install conda
  - download and install cgatcore, a workflow management system
  - construct a conda python environment with which to run the code
  - download the source code
  - download the input data
  - run the pipeline



REPRODUCIBLE FIGURES





https://mybinder.org/v2/gh/AMIGA-IAA/hcg-16/master

Figure 2. HCG 16 HI moment zero map and overlay

```
In [ ]: import matplotlib,aplpy
        from astropy.wcs import WCS
        from astropy.io import fits
        from general_functions import *
        import matplotlib.pyplot as plt
```

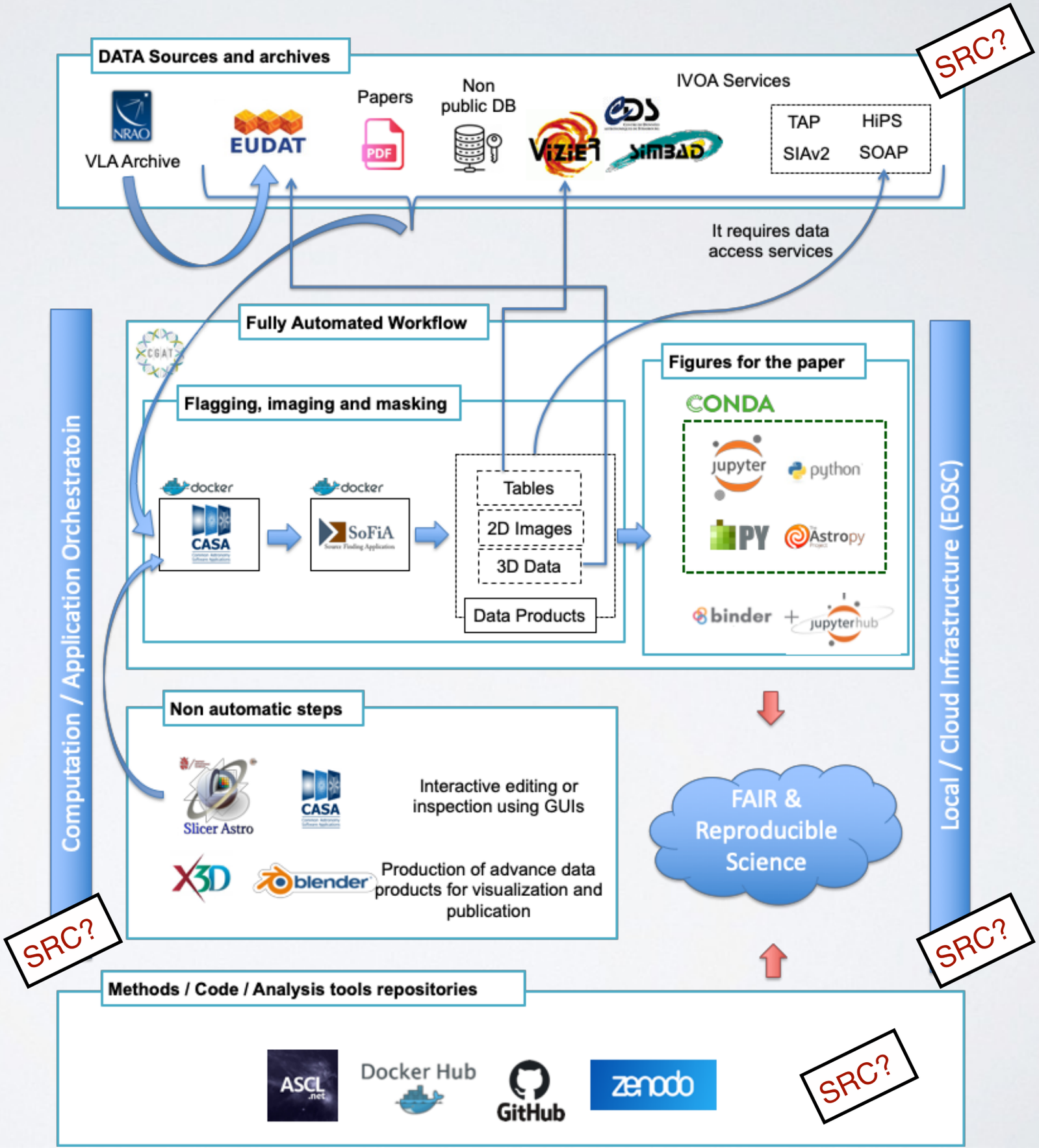
```
In [ ]: font = {'size'   : 14, 'family' : 'serif', 'serif' : 'cm'}
        plt.rc('font', **font)
        plt.rcParams['image.interpolation'] = 'nearest'
        plt.rcParams['lines.linewidth'] = 1
        plt.rcParams['axes.linewidth'] = 1

        #Set to true to save pdf versions of figures
        save_figs = True
```

The files used to make the following plot are:

```
In [ ]: moment0_casa = 'HCG16_CD_rob2_MS.mom0.pbcor.fits'
        moment0_sofia = 'HCG16_CD_rob2_MS_mom0.fits'
        r_image_decals = 'HCG16_DECaLS_r_cutout.fits'
```

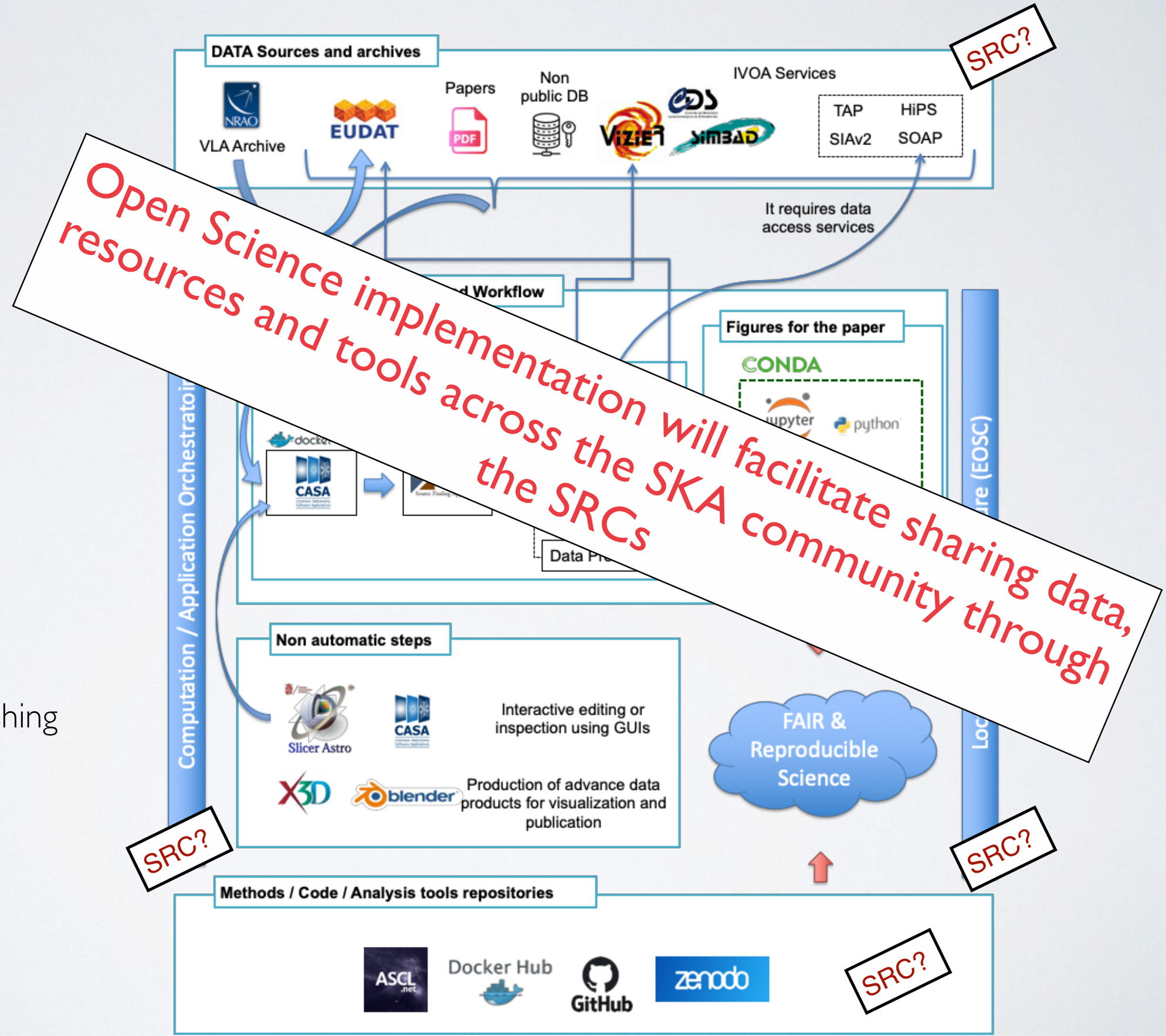
1. A moment 0 map of HCG 16 generated using a simple 3σ threshold in each channel (made with CASA). This file was generated in the *imaging* step of the workflow, which is described in the script imaging.py.
2. A moment 0 map of HCG 16 generated using 3.5σ mask made with SoFiA after smoothing over various kernel sizes. This file was generated in the *masking* step of the workflow. The SoFiA parameters file which makes this file is HCG16_CD_rob2_MS.3.5s.dil.session.
3. An *r*-band DECaLS fits image of HCG 16. This file was downloaded directly from the DECaLS public website. The exact parameters defining the region and pixel size of this images is contained in the pipeline.yml file.

TOWARDS FAIR & OPEN SCIENCE



- Open access publishing
- Open data
- Open Source
- Open peer review
- Open research
- Citizen science

TOWARDS FAIR & OPEN SCIENCE



- Open access publishing
- Open data
- Open Source
- Open peer review
- Open research
- Citizen science

If I have seen further it is by standing on the shoulders of Giants (Isaac Newton)