

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



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



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

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



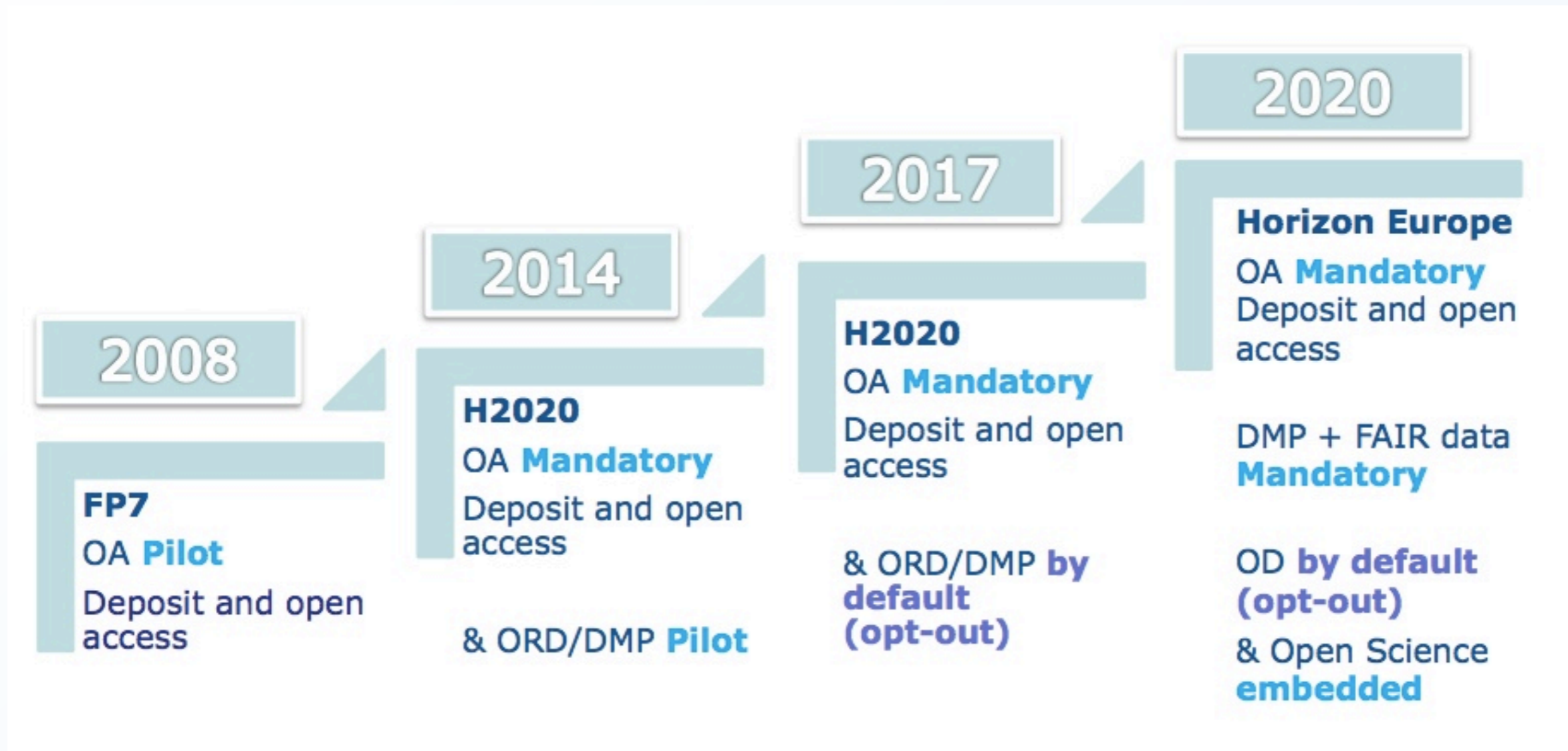
**A WORLD WHERE KNOWLEDGE CREATES
POWER FOR THE MANY, NOT THE FEW.
*THIS IS THE WORLD WE CHOOSE.***

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

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 ACCESS IN THE FPs

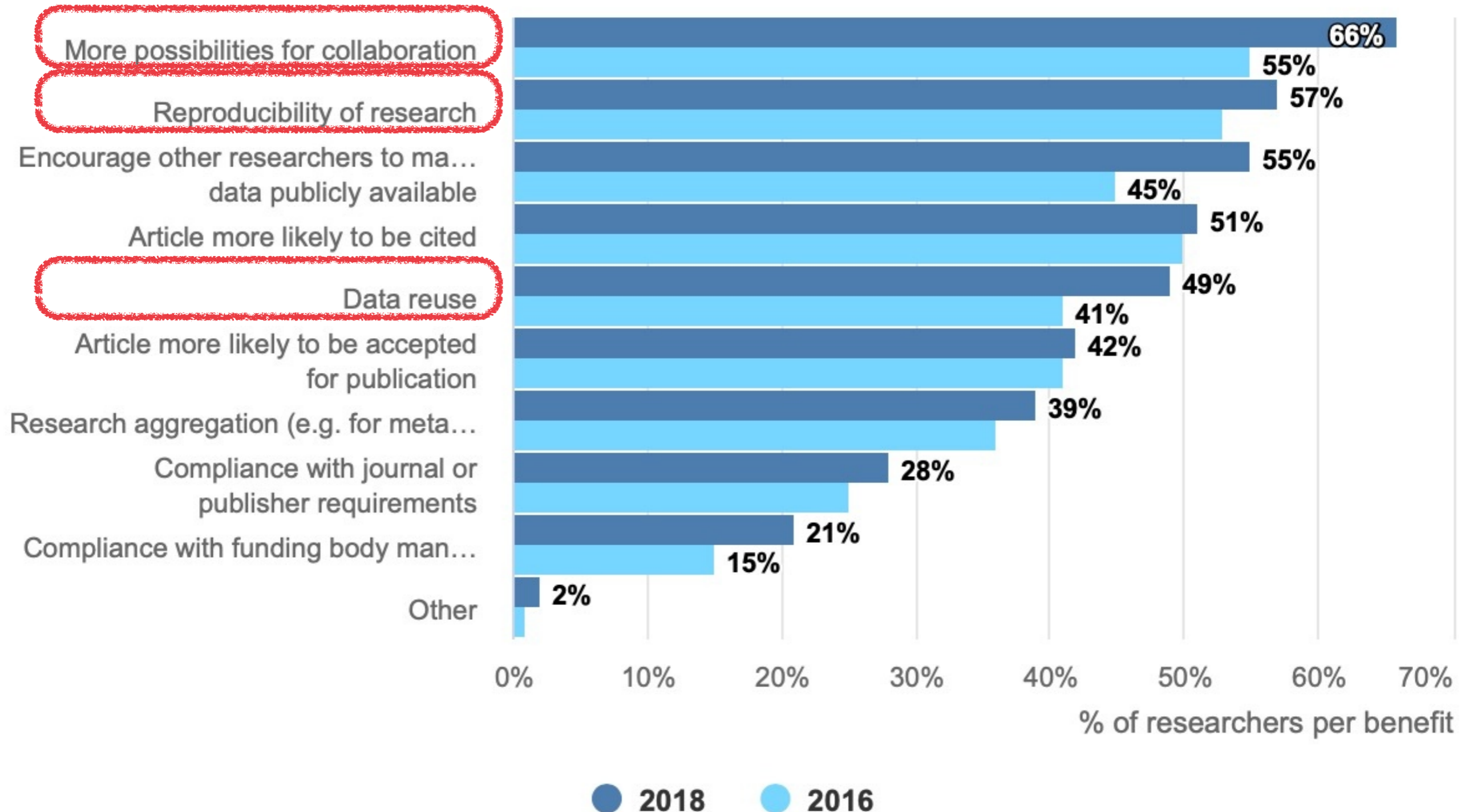


SRCs:
EU + SKA policies

ATTITUDE TOWARDS DATA SHARING

Benefits of sharing research data

Reference dates: 2016 and 2018

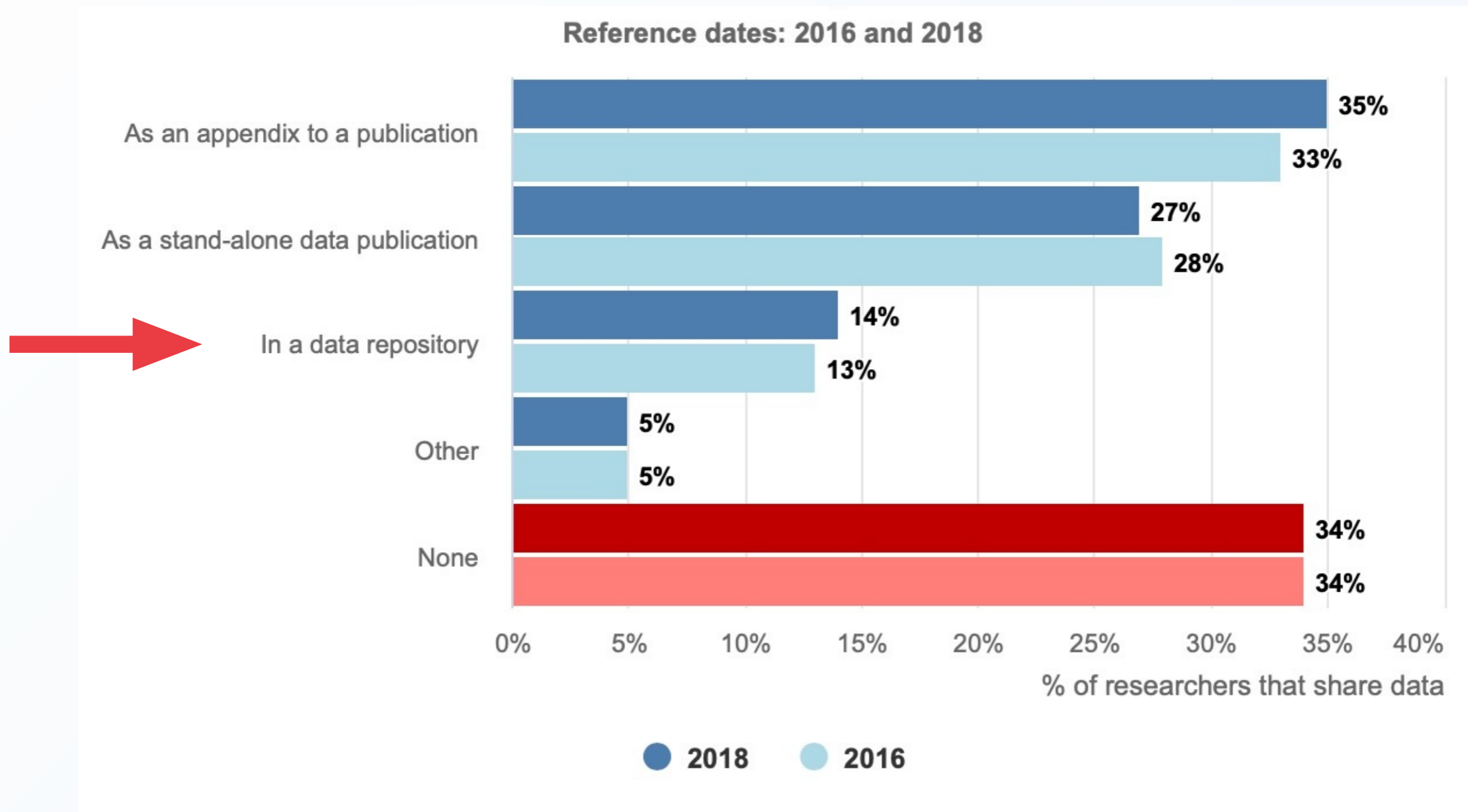


% of researchers per benefit

Source: Open science monitor

ATTITUDE TOWARDS DATA SHARING

% of researchers that share data



Open Science in a real scientific experiment



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.

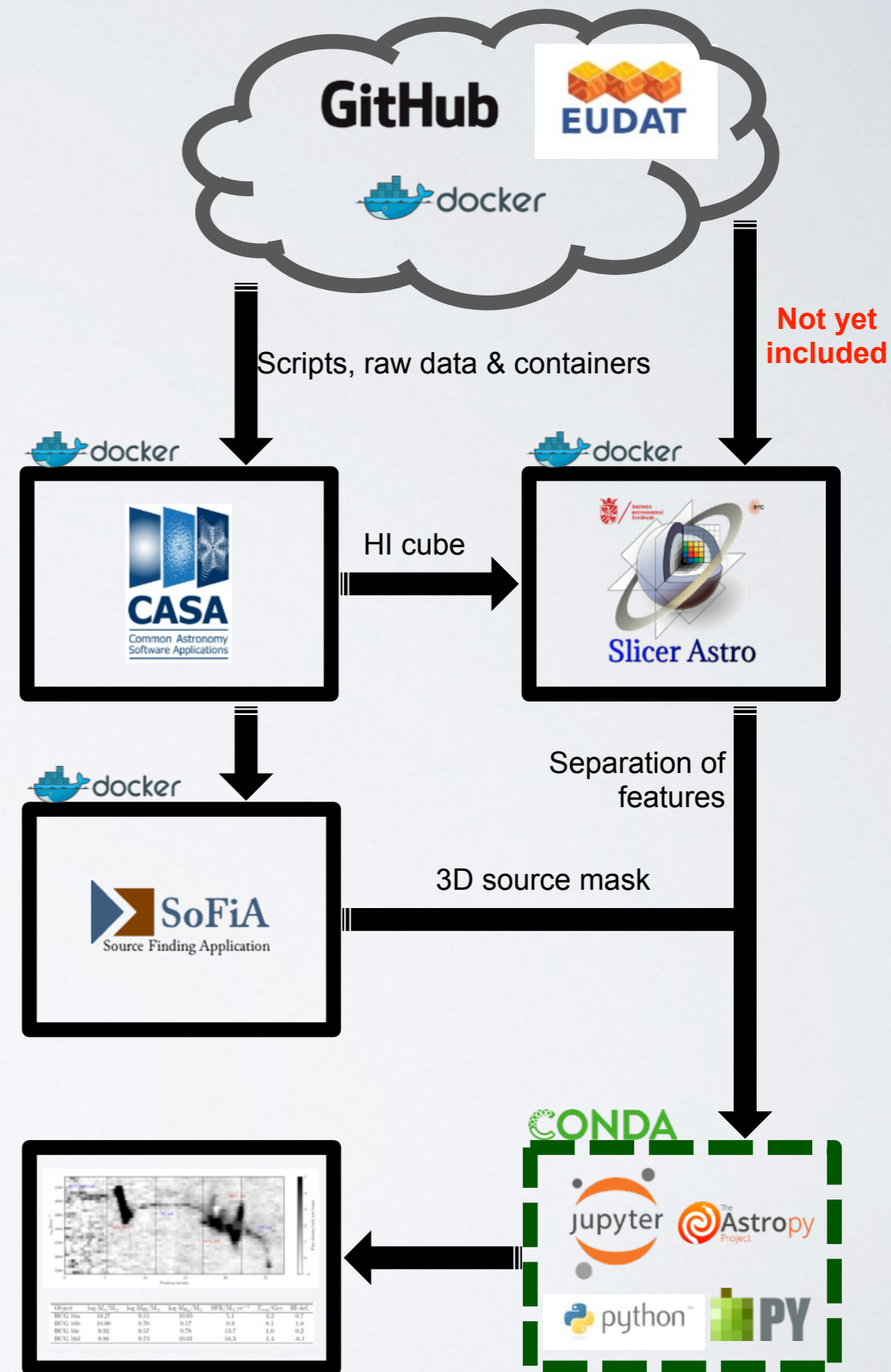
Common astronomy examples of un-FAIR practices:

- 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



hcg 16

HELP COMMUNITIES UPLOAD CONTACT

» RECORDS » AF679ED67B644432AE1A5F61B9654255

HCG16 L-band VLA C+D array data

by [Unknown]

Mar 5, 2019

TechnicalInfo: The VLA D and C array data of HCG 16 were collected by the Very Large Array (<http://www.vla.nrao.edu/>) in 1989 and 1999, under PI projects of Jacqueline van Gorkom and Marcus Verheijen. The project numbers are AW234 and AW500 respectively. The full original data of these projects are hosted by the VLA Archive (<https://science.nrao.edu/facilities/vla/archive/index>).

Disciplines: 3.5.2.1.1 → Observational astronomy → Radio astronomy;

DOI: [10.23728/b2share.af679ed67b644432ae1a5f61b9654255](https://doi.org/10.23728/b2share.af679ed67b644432ae1a5f61b9654255)

PID: [11304/16c0eb14-0bb0-4ec0-9ff4-11e000033c8](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0033-8)

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.

AMIGA-IAA / hcg-16

Unwatch 5 Star 2 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

HCG-16 Project Edit

Manage topics

130 commits 3 branches 0 releases 3 contributors MIT

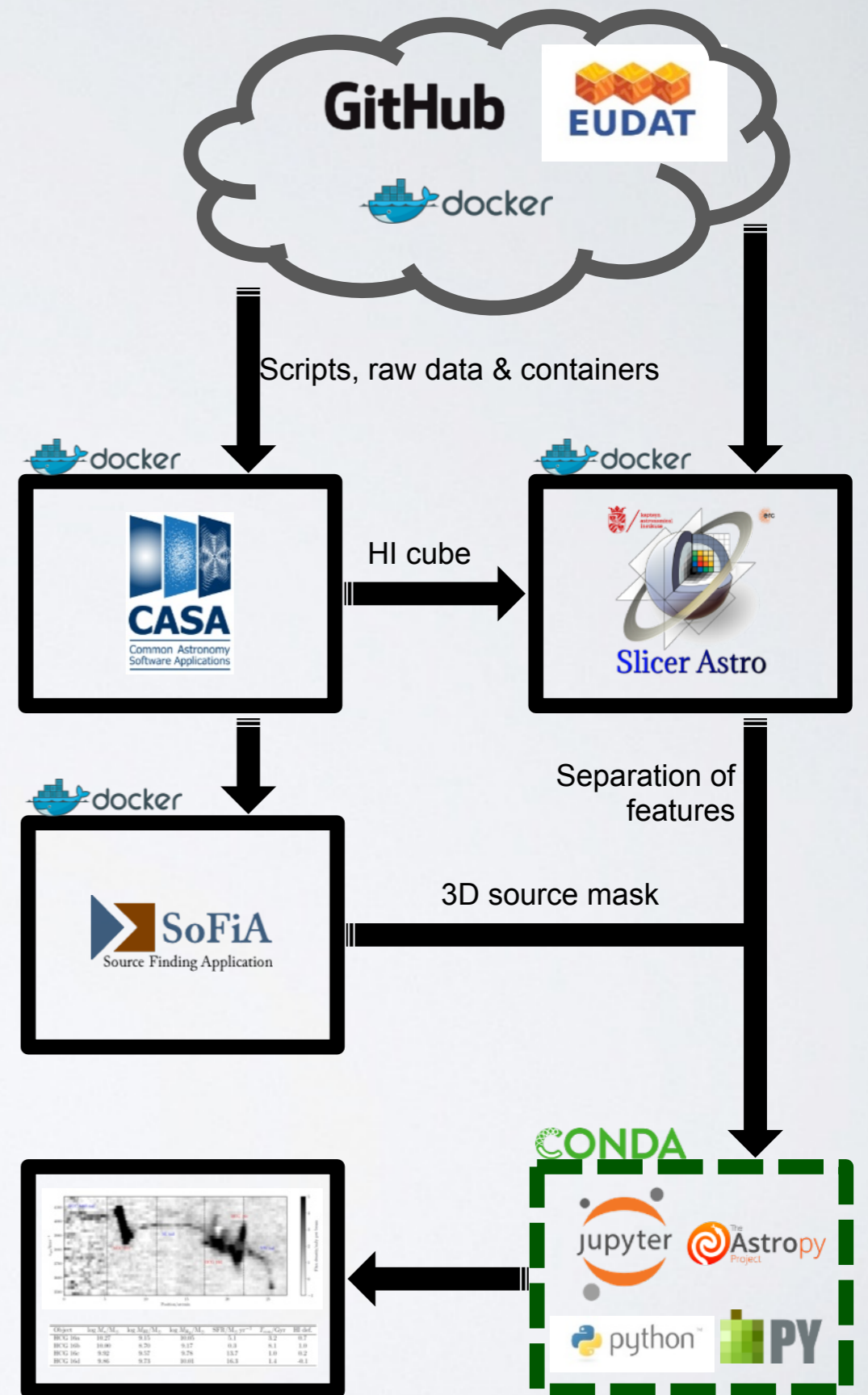
Branch: master New pull request Create new file Upload files Find File Clone or download

jonesmg Merge branch 'master' of github.com:AMIGA-IAA/hcg-16 Latest commit ef70764 6 days ago

casa	Added NW clump to moments generation task.	3 months ago
cgatcore	Update pipeline.py	10 days ago
docker	bugfix docker/Dockerfile.sofia	7 months ago
plot_scripts	Changed plot range to include NW clump.	6 days ago
sofia	Added HIPASS cube SoFiA step to pipeline.	12 days ago
LICENSE	Initial commit	7 months ago
README.md	edit README file to explain how to run ipynb files from local	6 days ago
environment.yml	change plotting task	4 months ago
postBuild	Added HIPASS mask to tar	10 days ago
run.sh	jupyter nbconvert --to python	2 months ago


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

Figure 2. HCG 16 HI moment zero map and overlay



Starting repository: AMIGA-IAA/hcg-16/master

jupyter Quit

Files Running Clusters

Select items to perform actions on them.

Upload New ↻

<input type="checkbox"/> 0	Name ↓	Last Modified	File size
	..	seconds ago	
<input type="checkbox"/>	Fig1-DECaLS_grz_image.ipynb	2 hours ago	3.45 kB
<input type="checkbox"/>	Fig12-Absorption_profile.ipynb	2 hours ago	4.24 kB
<input type="checkbox"/>	Fig16-TDG_candidates_moments.ipynb	2 hours ago	8.25 kB
<input type="checkbox"/>	Fig2-Moment0_overlay.ipynb	2 hours ago	6.67 kB
<input type="checkbox"/>	Fig3-Moment1.ipynb	2 hours ago	4.91 kB
<input type="checkbox"/>	Fig4-Integrated_spectrum.ipynb	2 hours ago	10.1 kB
<input type="checkbox"/>	Fig5-6_Tab2-Separated_spectra.ipynb	2 hours ago	16.5 kB
<input type="checkbox"/>	Fig8-11_13-14-Galaxy_moment_maps.ipynb	2 hours ago	23.4 kB
<input type="checkbox"/>	FigC1-C2-Channel_maps.ipynb	2 hours ago	10.2 kB
<input type="checkbox"/>	cd_bridge.fits	3 months ago	1.56 MB
<input type="checkbox"/>	cd_bridge_mask.fits	4 months ago	3.11 MB
<input type="checkbox"/>	E_clump.fits	3 months ago	43.2 kB

```
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

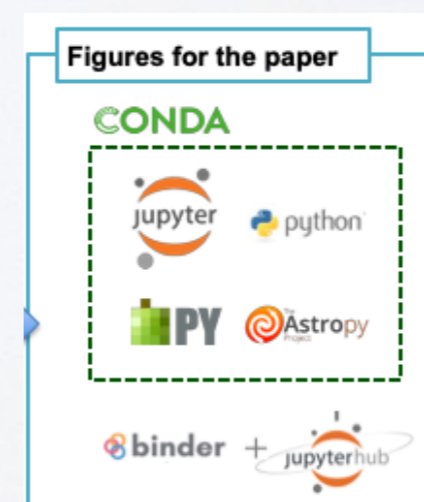
        #Set to true to save pdf versions of figures
        save_figs = True
```

The files used to make the following plot are:

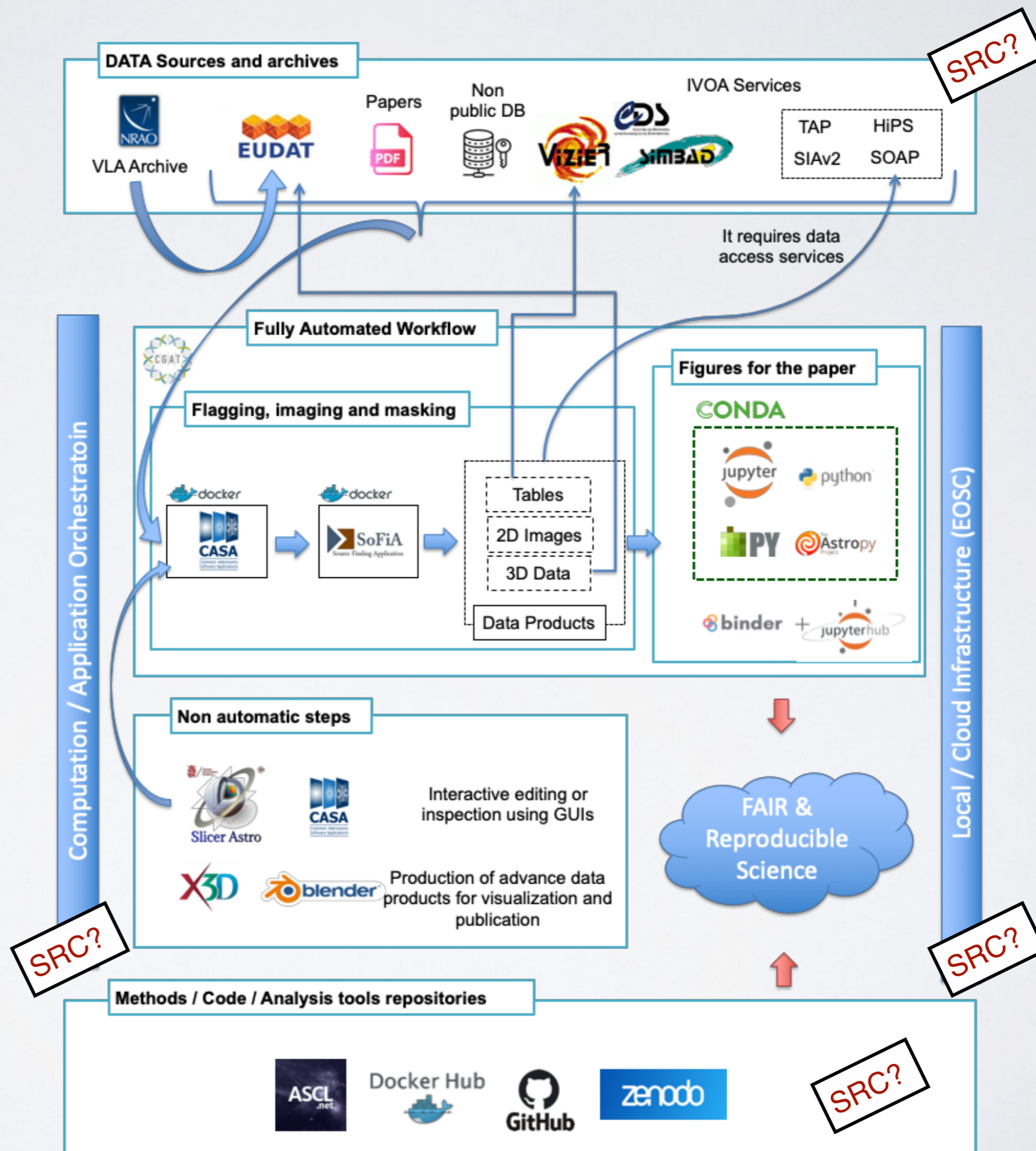
```
In [ ]: moment0_casa = 'HCG16_CD_rob2_MS.mom0.pbcor.fits'
        moment0_sofia = 'HCG16_CD_rob2_MS_mom0.fits'
        r_image_decals = 'HCG16_DECaLS_r_cutout.fits'
```

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.

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

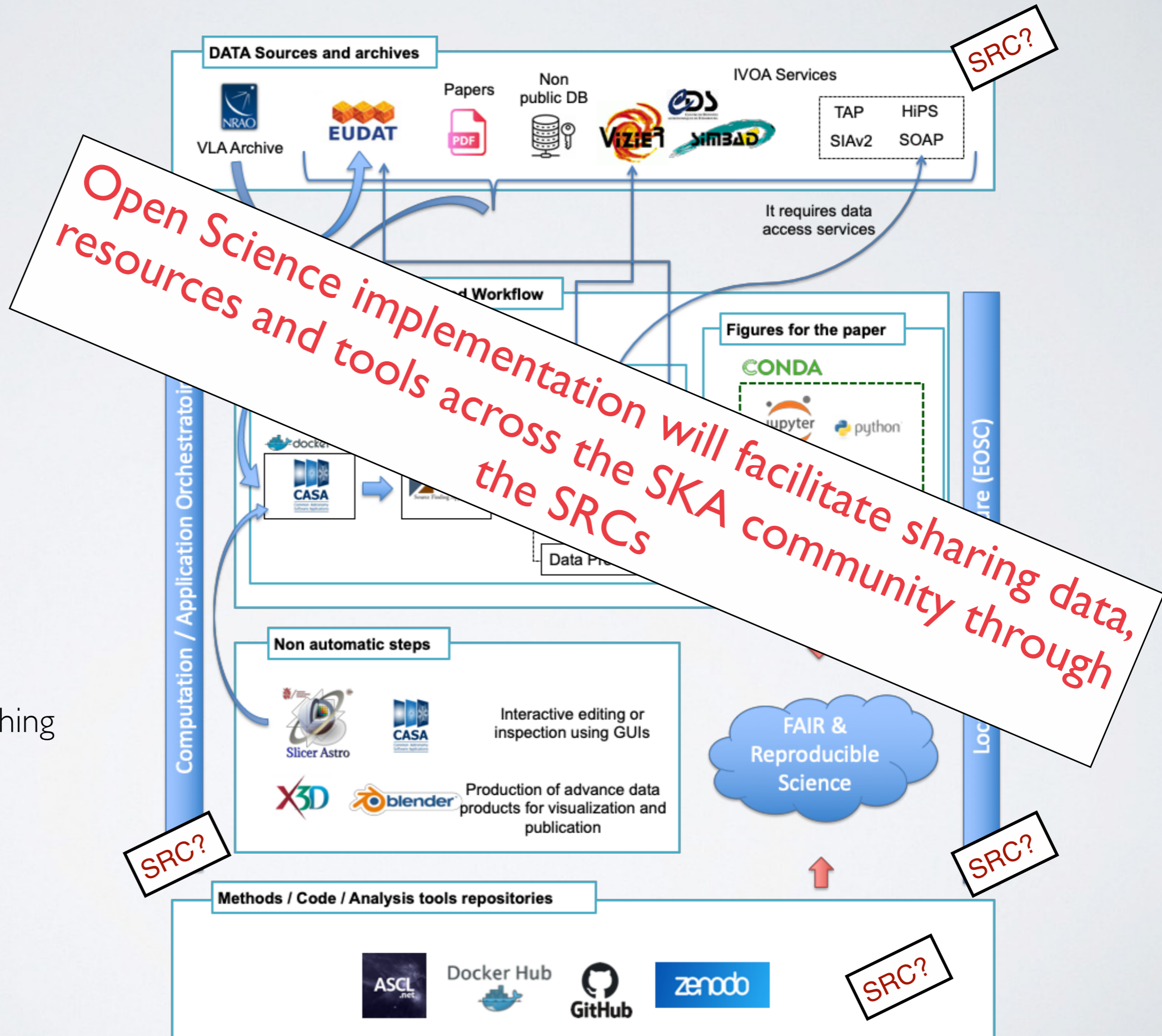


TOWARDS FAIR & OPEN SCIENCE



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

TOWARDS FAIR & OPEN SCIENCE



Open Science implementation will facilitate sharing data, resources and tools across the SKA community through the SRCs

- 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)