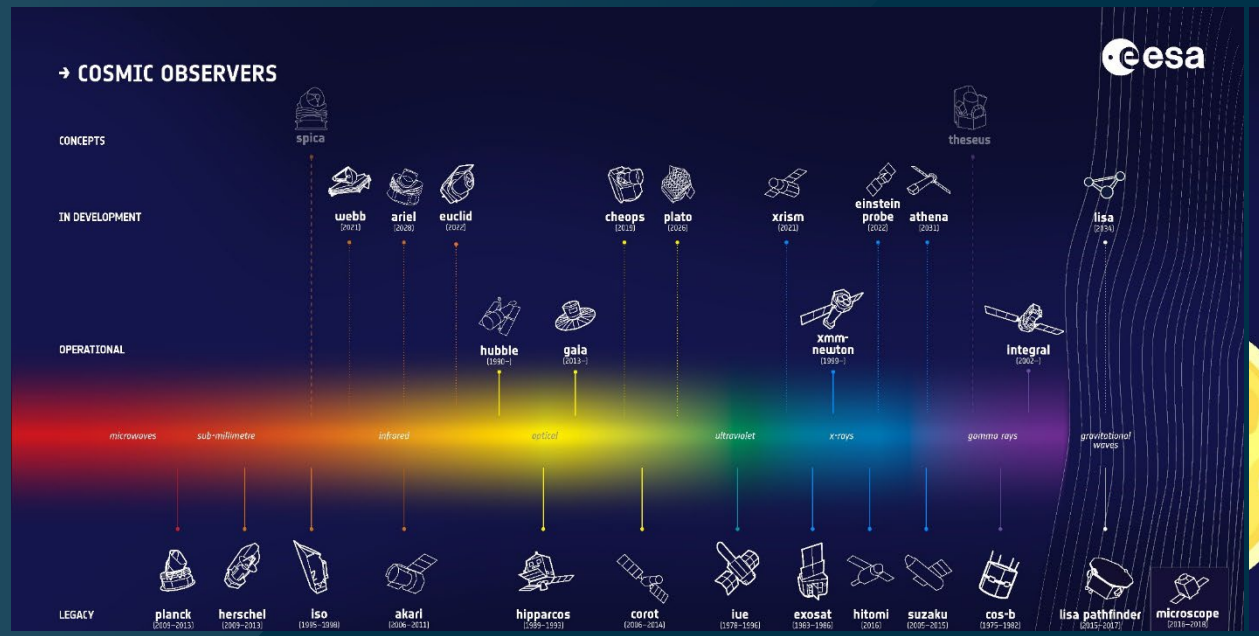
ESA Datalabs: Expanding Space Archives with Innovative Analysis and Collaboration Capabilities

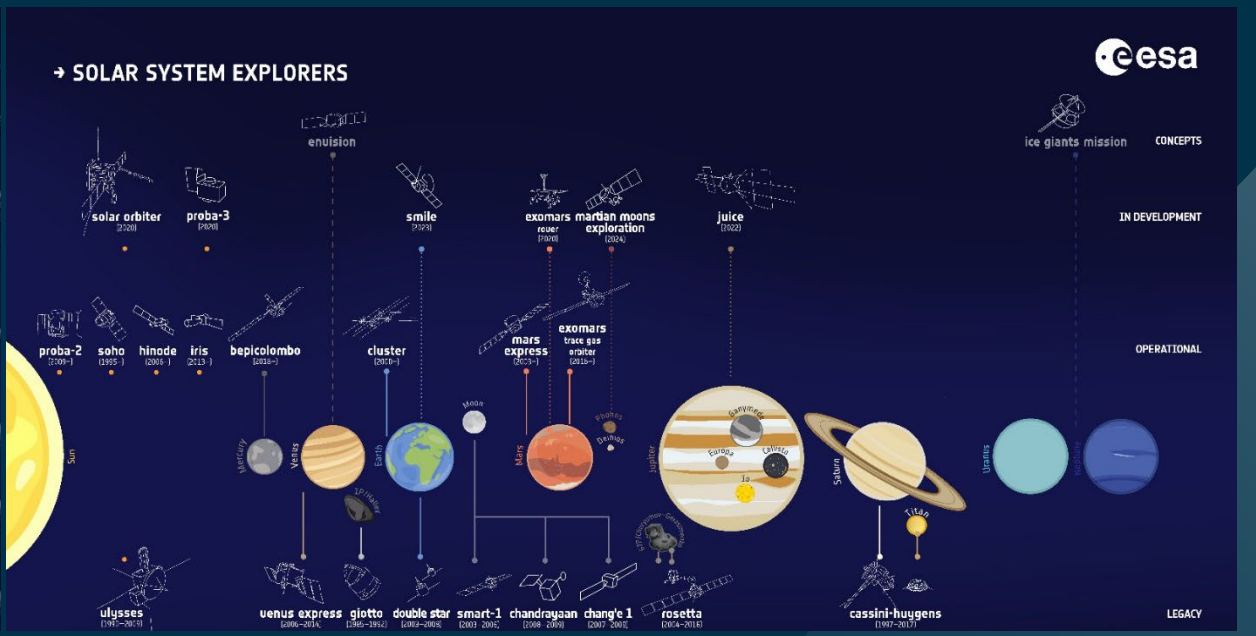
V. Navarro, F. Marinic, S. del Rio, M. A. Diego, M. Lopez-Caniego, A. Anku, C. Arviset

PSIDA 2022, 21.06.2021

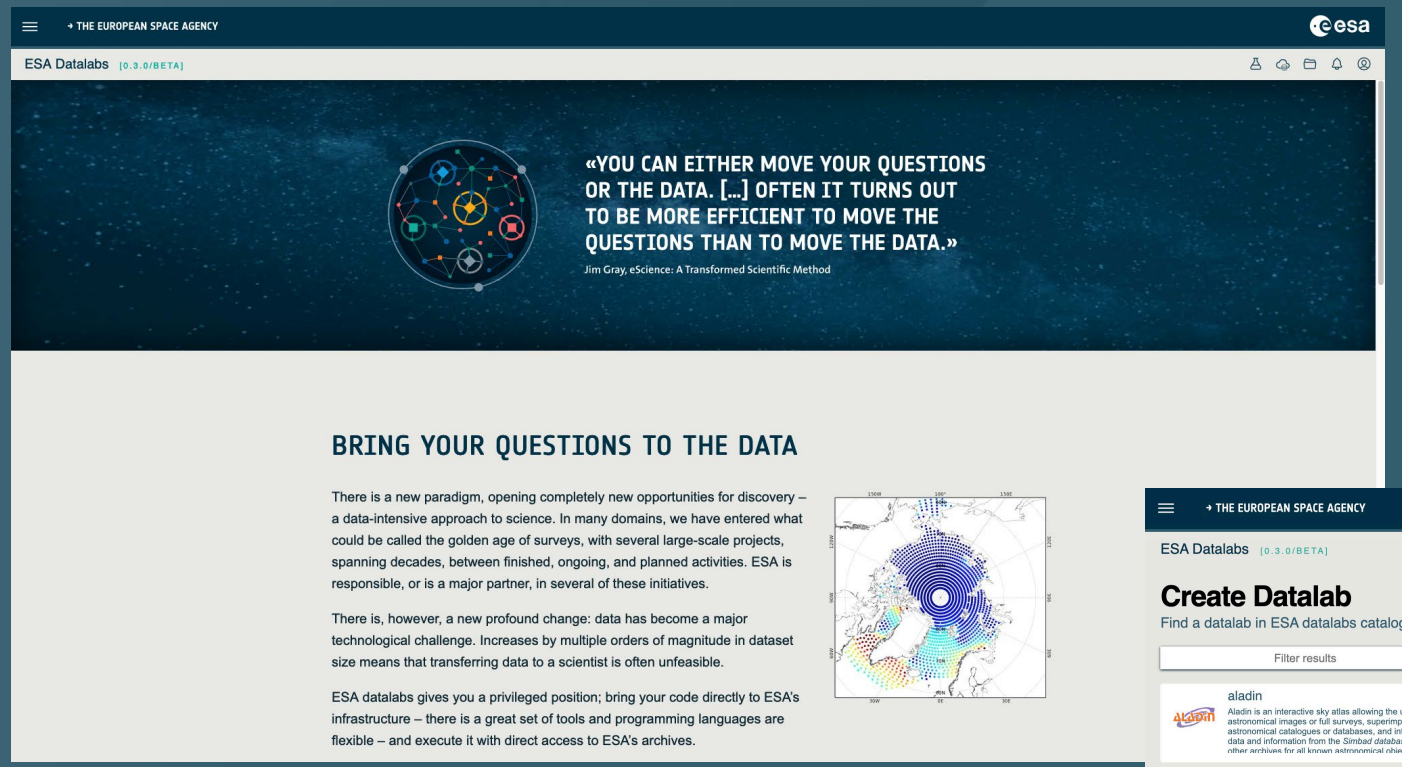
@ivicentenavarro

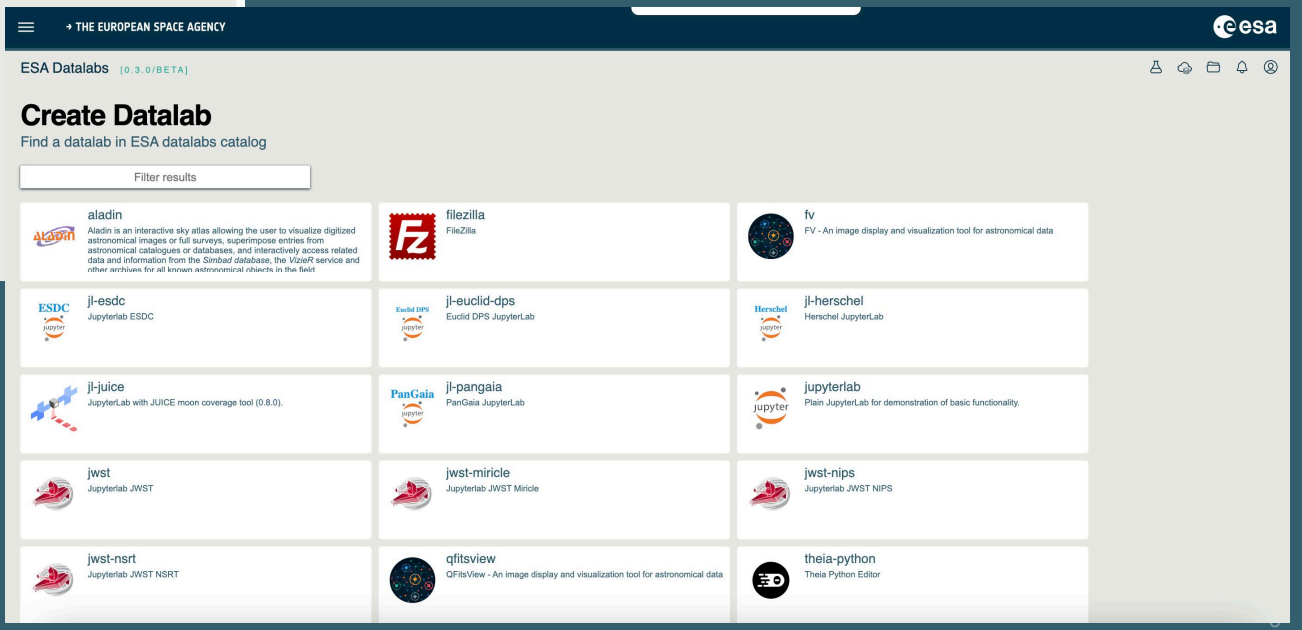
Space Science Context





ESA Datalabs – Cross Domain Data Exploitation Platform

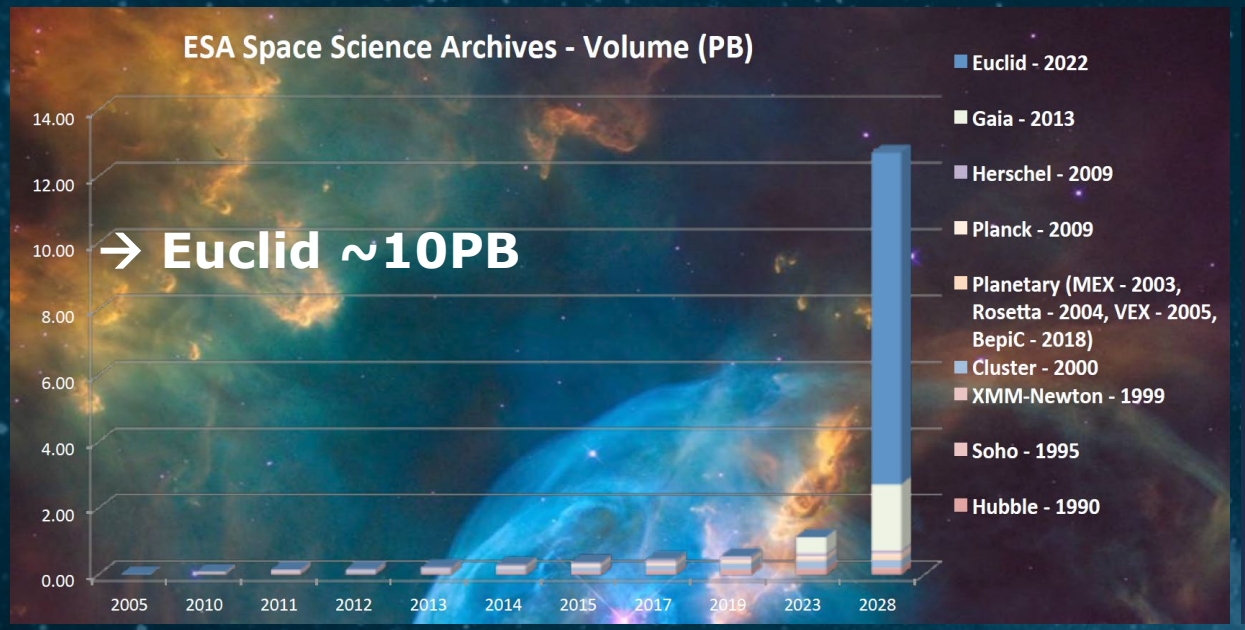


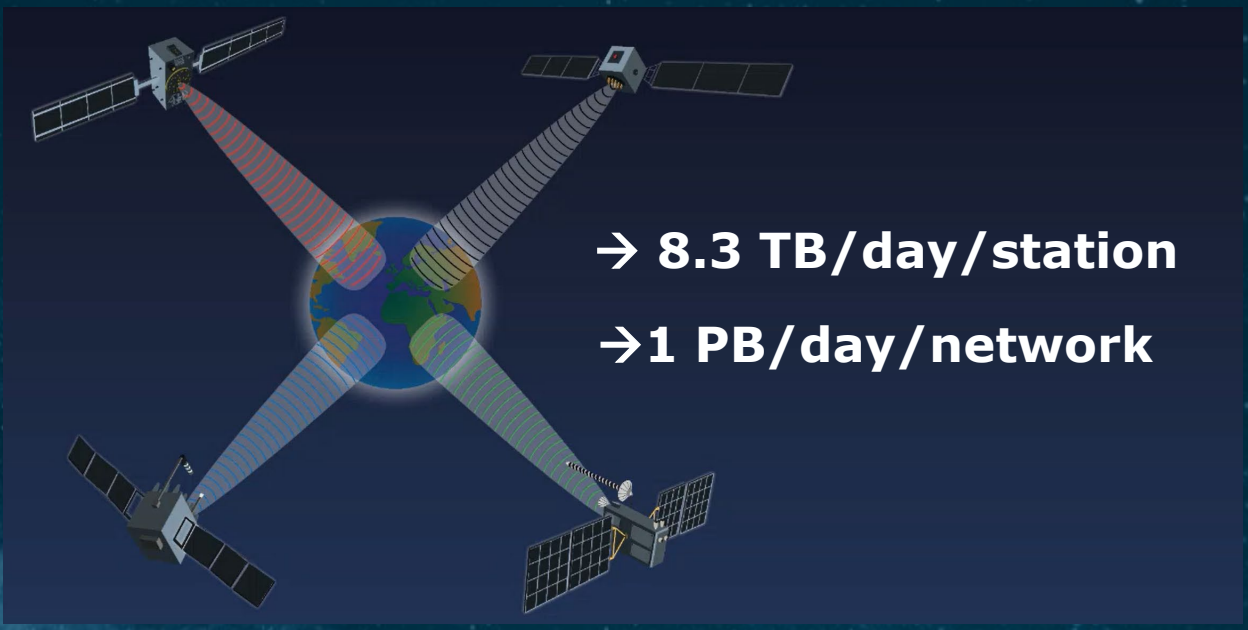


datalabs.esa.int


Data Access
Innovation

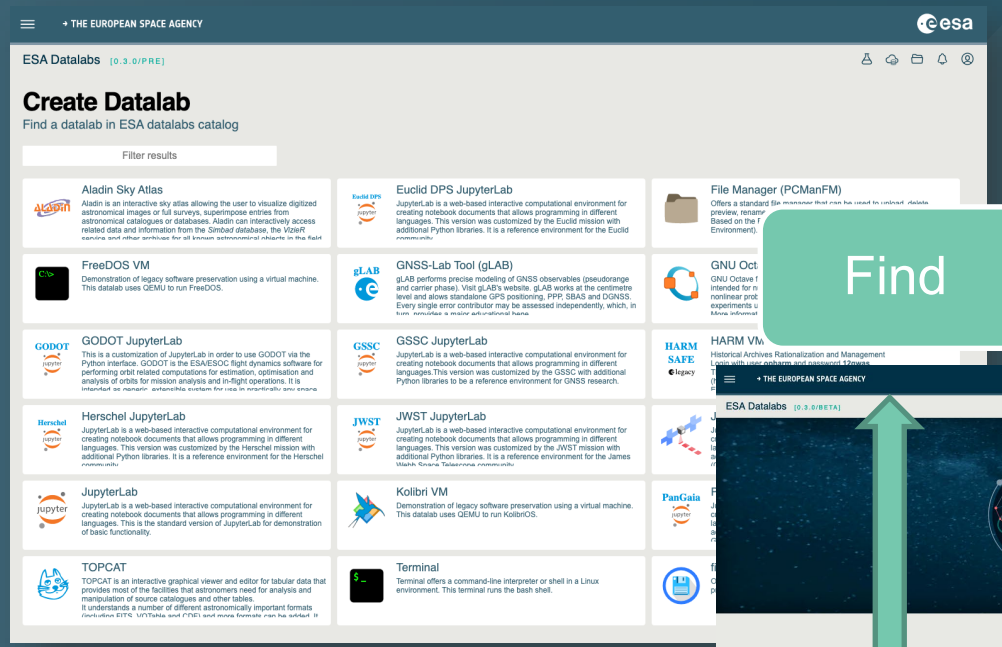
Data Exploitation Paradigm Shift



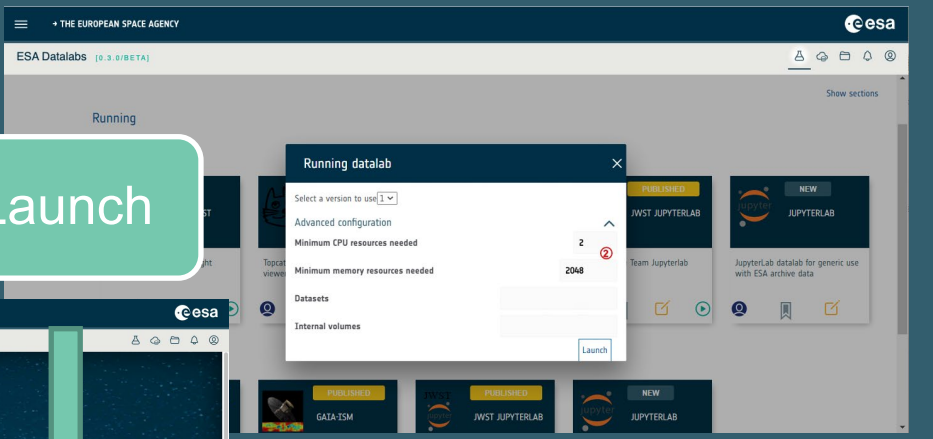


From *bring the data to the user*
To *bring the user to the data*

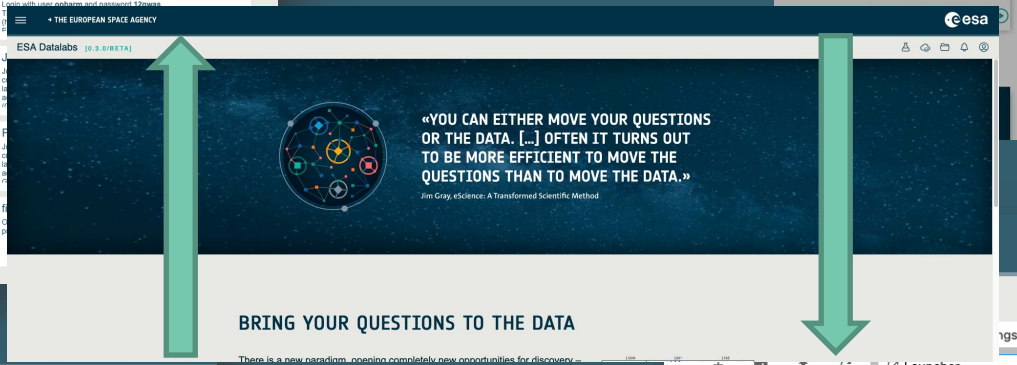
Datalabs – Standard Flow

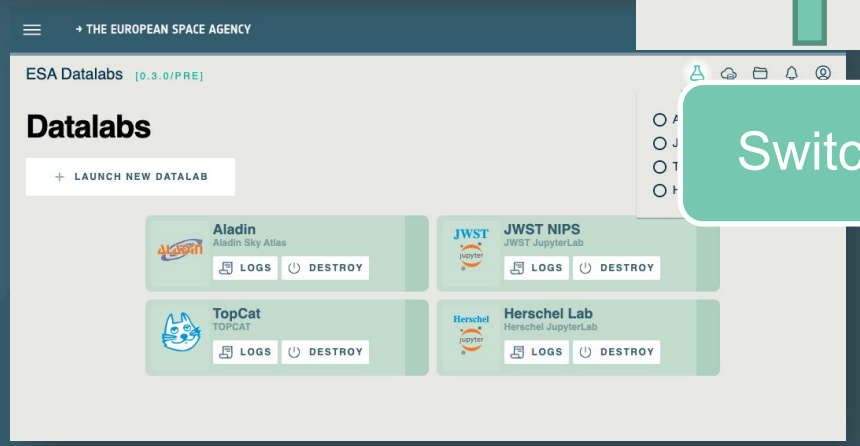


Find



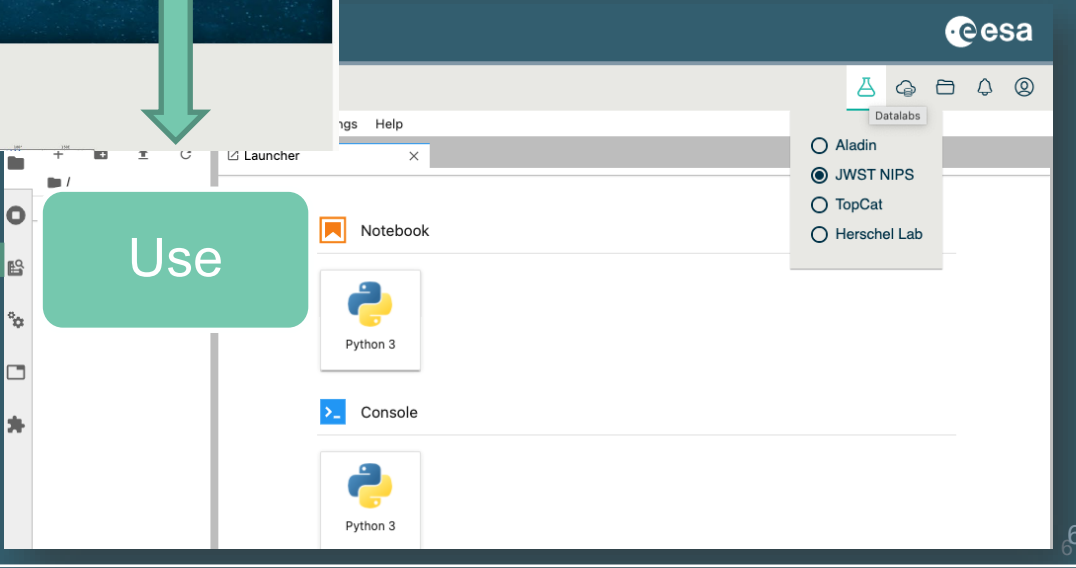
Launch



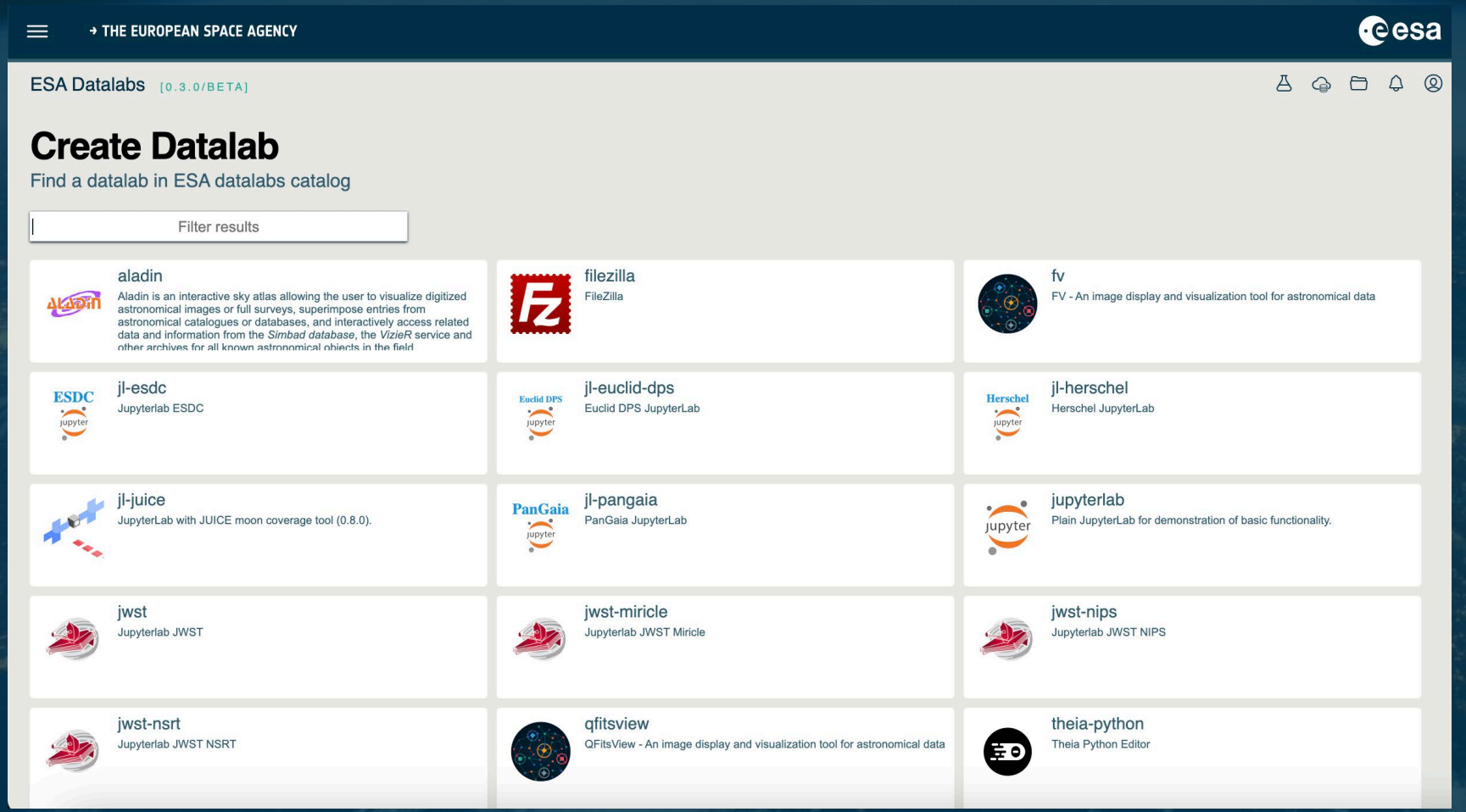


Switch

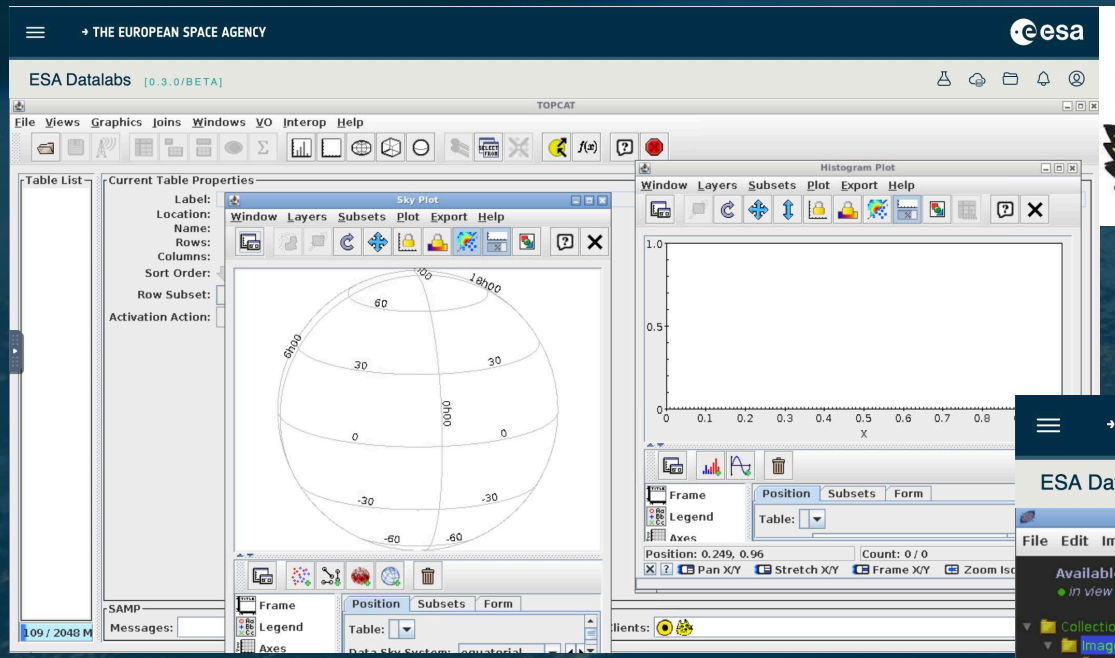
Use

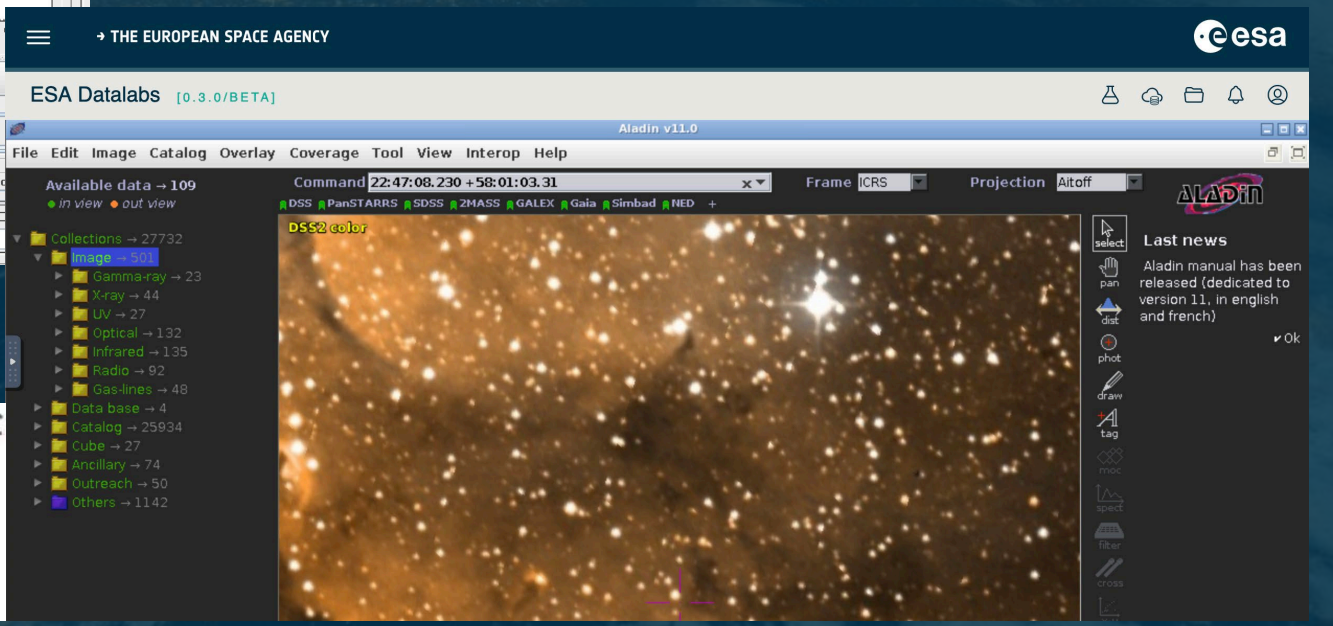


FIND

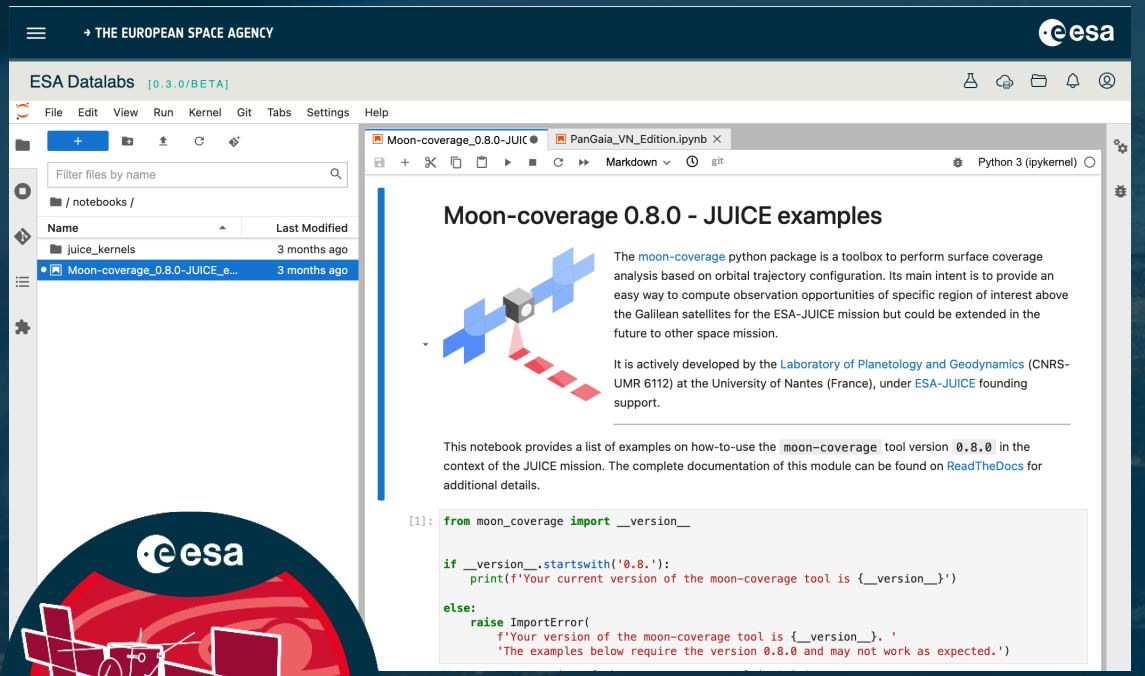

→ THE EUROPEAN SPACE AGENCY
esa
ESA Datalabs [0.3.0/BETA]
Create Datalab
Find a datalab in ESA datalabs catalog
Filter results
aladin
Aladin is an interactive sky atlas allowing the user to visualize digitized astronomical images or full surveys, superimpose entries from astronomical catalogues or databases, and interactively access related data and information from the Simbad database, the VizieR service and
filezilla
FileZilla
fv
FV - An image display and visualization tool for astronomical data
jl-esdc
Jupyterlab ESDC
jl-euclid-dps
Euclid DPS JupyterLab
jl-herschel
Herschel JupyterLab
jl-juice
JupyterLab with JUICE moon coverage tool (0.8.0).
jl-pangaia
PanGaia JupyterLab
jupyterlab
Plain JupyterLab for demonstration of basic functionality.
jwst
Jupyterlab JWST
jwst-miricle
Jupyterlab JWST Miricle
jwst-nips
Jupyterlab JWST NIPS
jwst-nsrt
Jupyterlab JWST NSRT
qfitsview
QFitsView - An image display and visualization tool for astronomical data
theia-python
Theia Python Editor

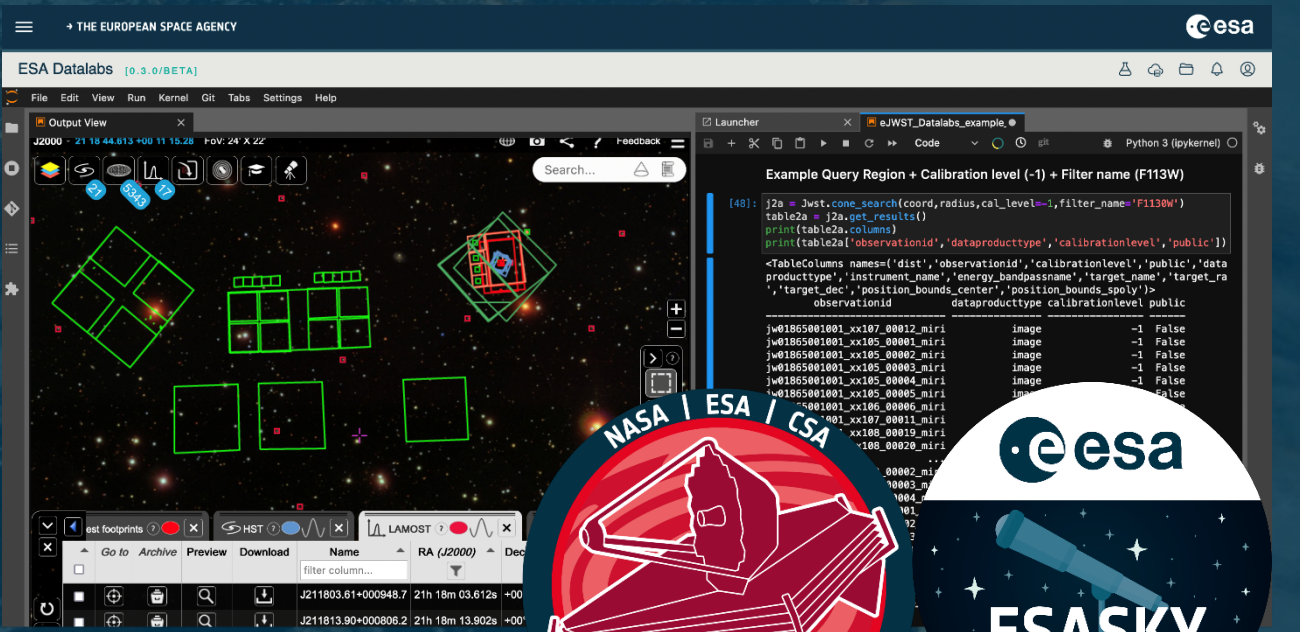
FIND – Desktop Apps



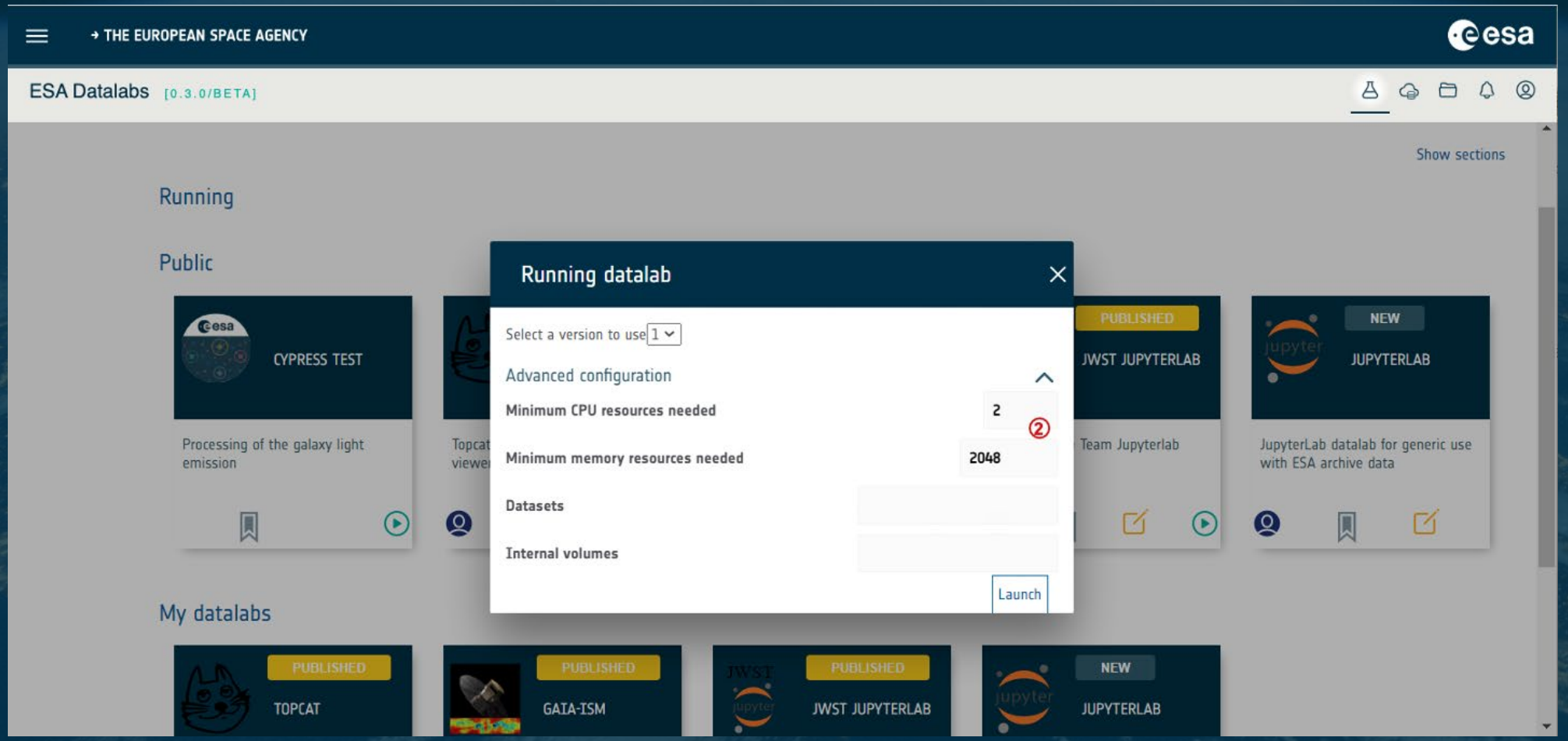


FIND – Web Apps (i.e. JupyterLab)





LAUNCH


→ THE EUROPEAN SPACE AGENCY
esa
ESA Datalabs [0.3.0/BETA]
Show sections
Running
Public
CYPRESS TEST
Processing of the galaxy light emission
PUBLISHED
JWST JUPYTERLAB
Team Jupyterlab
NEW
JUPYTERLAB
JupyterLab datalab for generic use with ESA archive data
Running datalab
Select a version to use 1
Advanced configuration
Minimum CPU resources needed
2
Minimum memory resources needed
2048
Datasets
Internal volumes
Launch
My datalabs
PUBLISHED
TOPCAT
PUBLISHED
GAIA-ISM
PUBLISHED
JWST JUPYTERLAB
NEW
JUPYTERLAB

USE

→ THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

File Edit View Run Kernel Git Tabs Settings Help

Launcher | Moon-coverage_0.8.0-JUIC | PanGaia_VN_Edition.ipynb

Markdown | git | Python 3 (ipykernel)

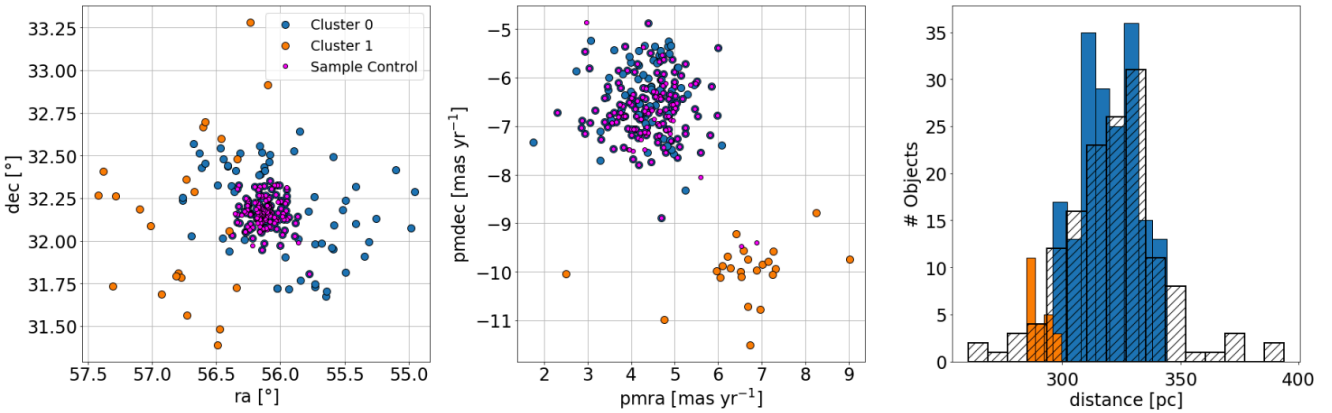
Filter files by name

/ my_workspace / My Notebooks /

Name	Last Modified
PanGaia_VN_Edition.ipy...	2 months ago
Untitled.ipynb	2 months ago

[11]:
```
# Show selected HDBSCAN results ========================
if step_4:
    data_clustering.run_hdbscan(probability = data_clustering.probability, min_samples = data_clustering.min_samples, verbose = True)
    if step_1:
        data_clustering.load_control_obj(control)
        data_clustering.compare_to_control()
    data_clustering.plot_clusters()
```

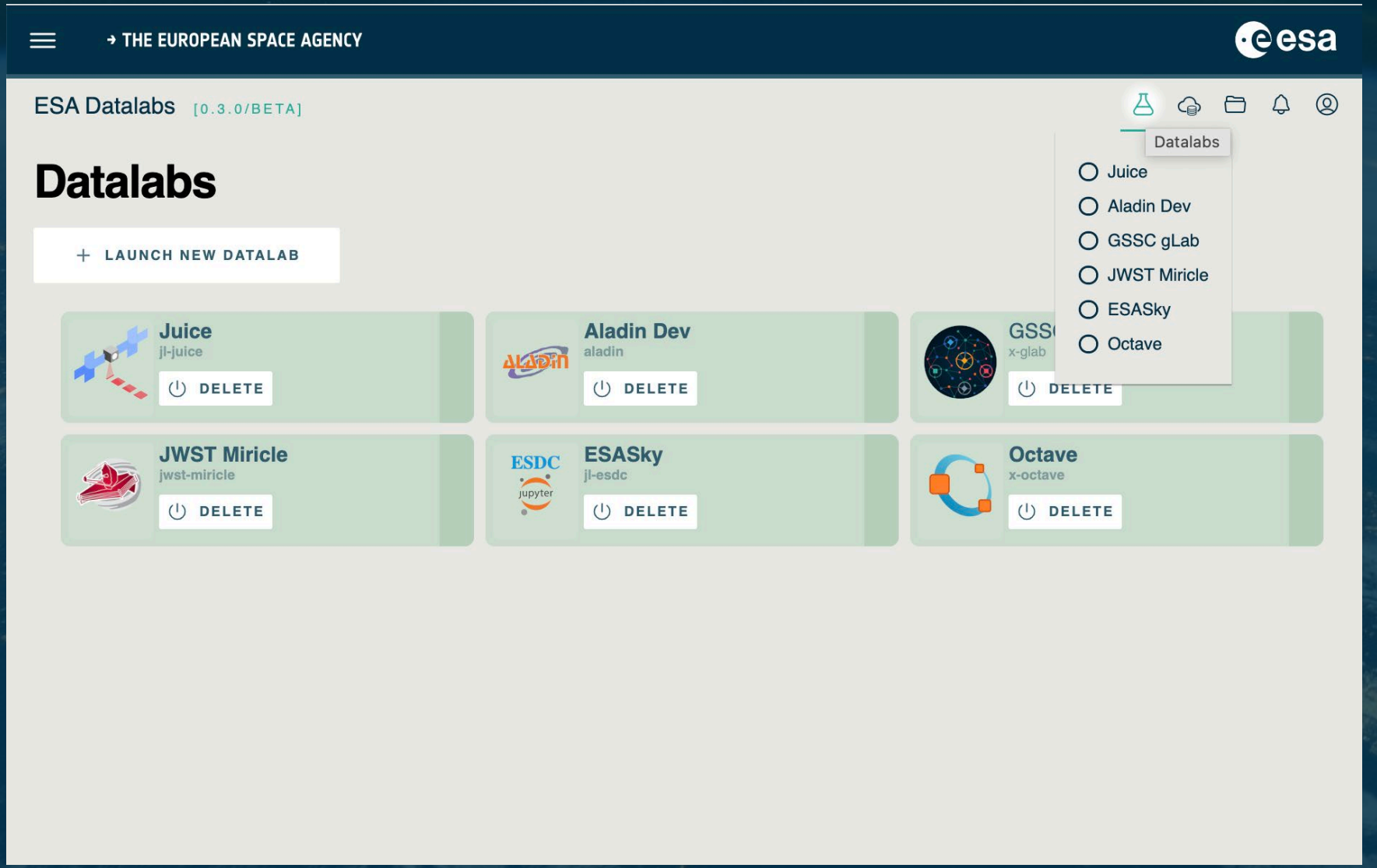
```
Insert min_Cluster_size: 20
mCls = 20; clusters = 2; N_members = [183, 23]

Cluster 0 contains   183 Elements, including 120 (83.3%) of the Control Sample
Cluster 1 contains    23 Elements, including 0 ( 0.0%) of the Control Sample
```



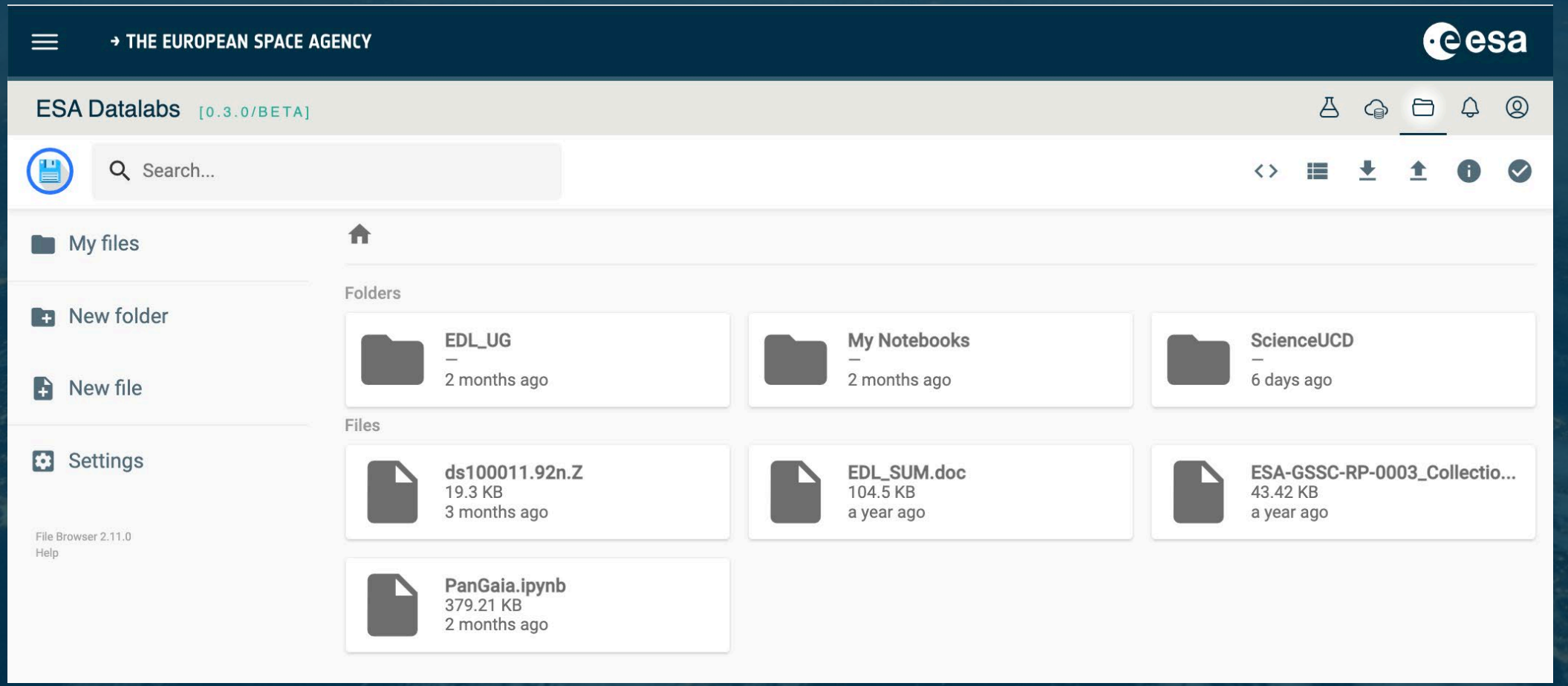

```
==================================================
PDF saved as: test_hdb_minsamp_1_prob_0.8_mCls_20.pdf
==================================================
```

Simple | 0 | 2 | Python 3 (ipykernel) | Idle | Mode: Command | Ln 1, Col 1 | PanGaia_VN_Edition.ipynb

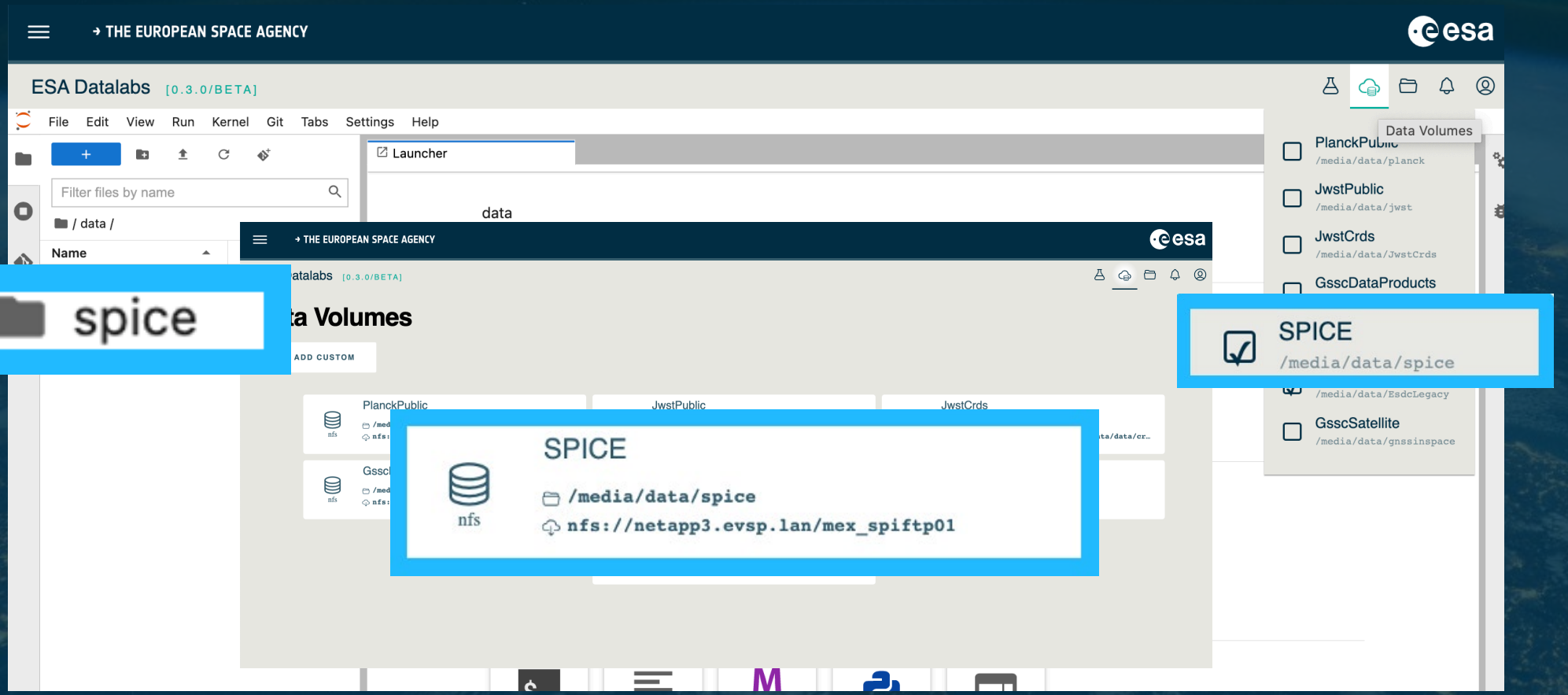
SWITCH


→ THE EUROPEAN SPACE AGENCY
esa
ESA Datalabs [0.3.0/BETA]
Datalabs
Datalabs
+ LAUNCH NEW DATALAB
Juice
Aladin Dev
GSSC gLab
JWST Miricle
ESASky
Octave
Juice
jl-juice
DELETE
Aladin Dev
aladin
DELETE
x-glab
DELETE
JWST Miricle
jwst-miricle
DELETE
ESASky
jl-esdc
DELETE
Octave
x-octave
DELETE

Personal Workspace

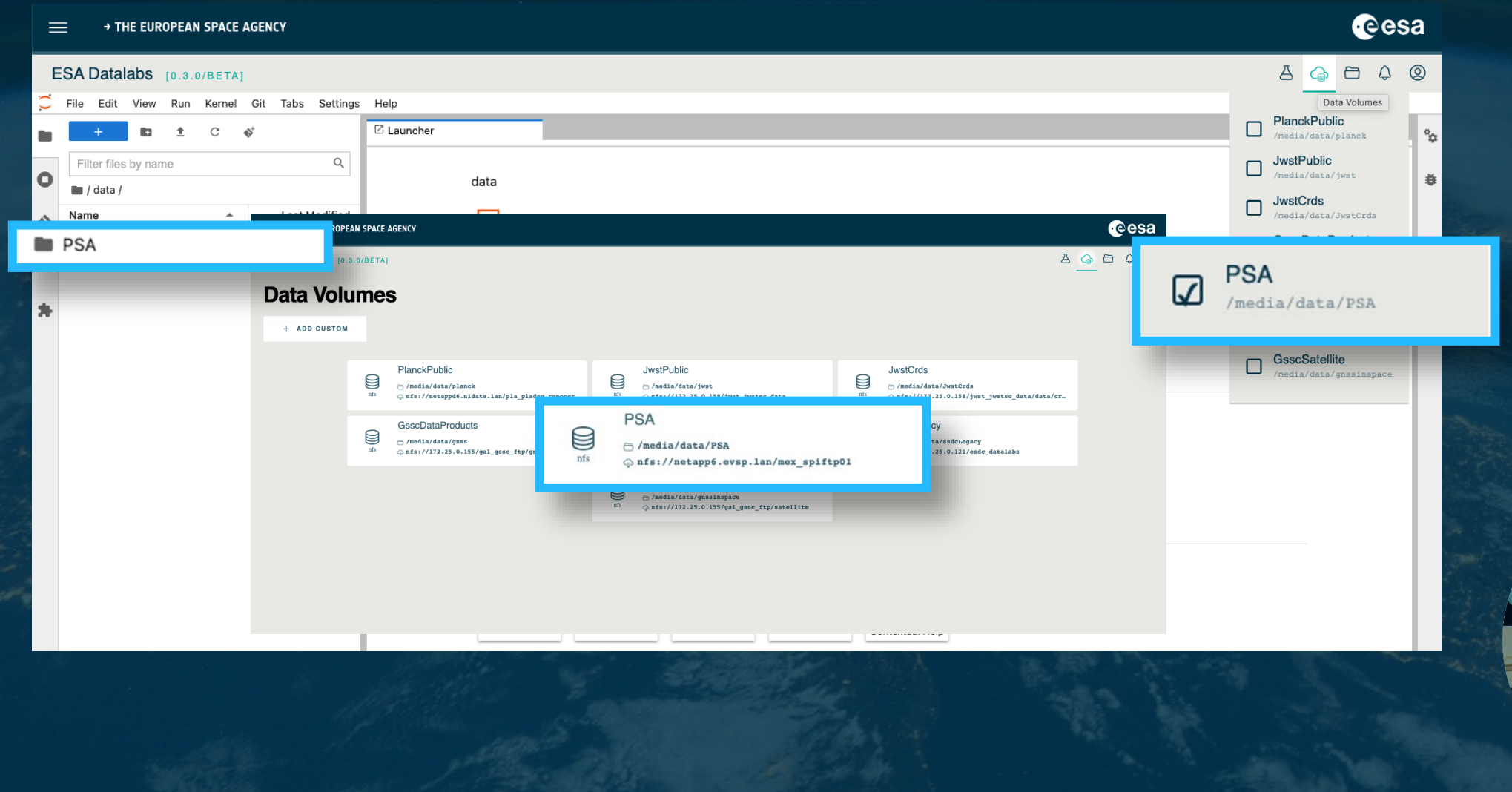


Data and Computing Colocation - Coarse Grain

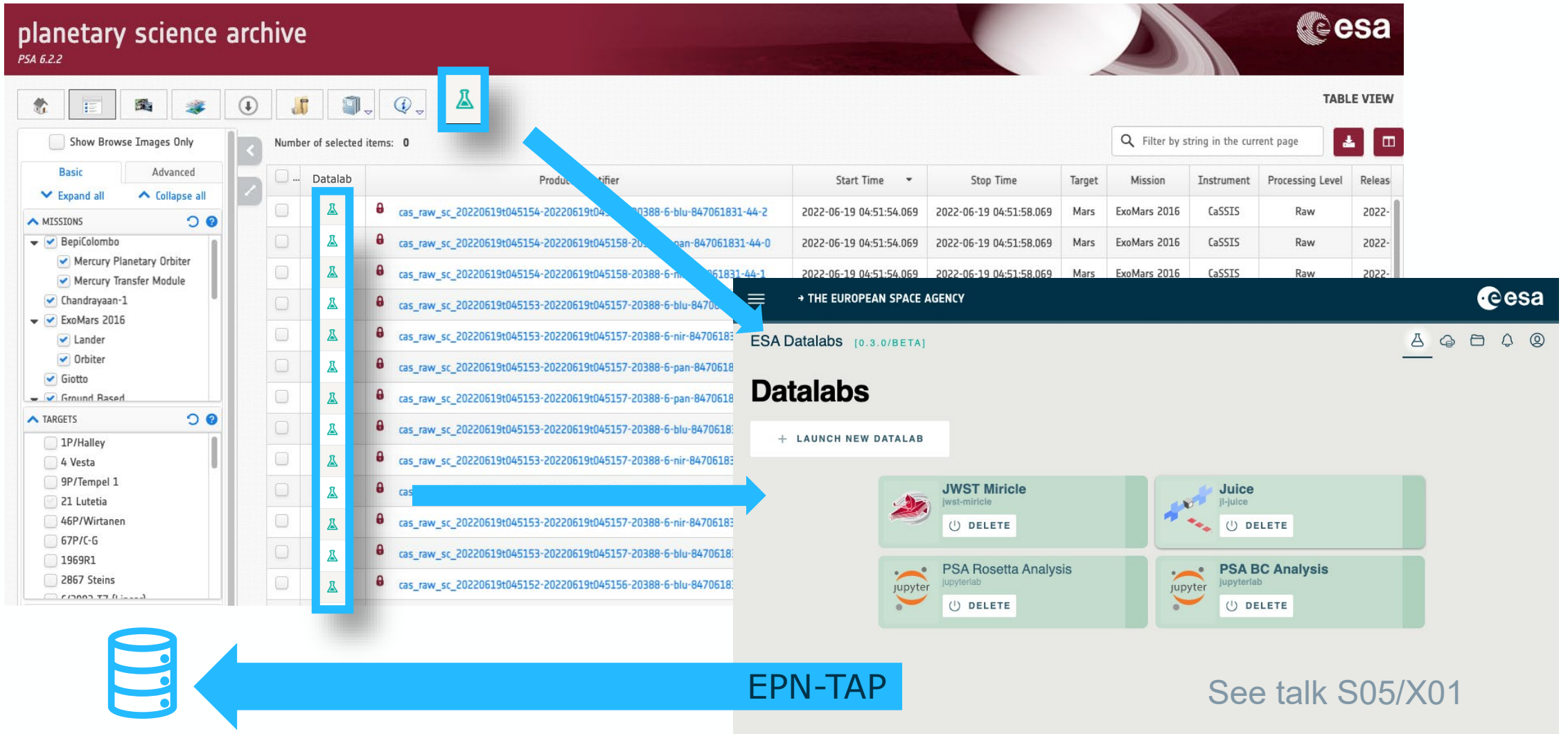


ESAC SCIENCE DATA CENTRE

Coarse Grain Integration with PSA (evolution opt)

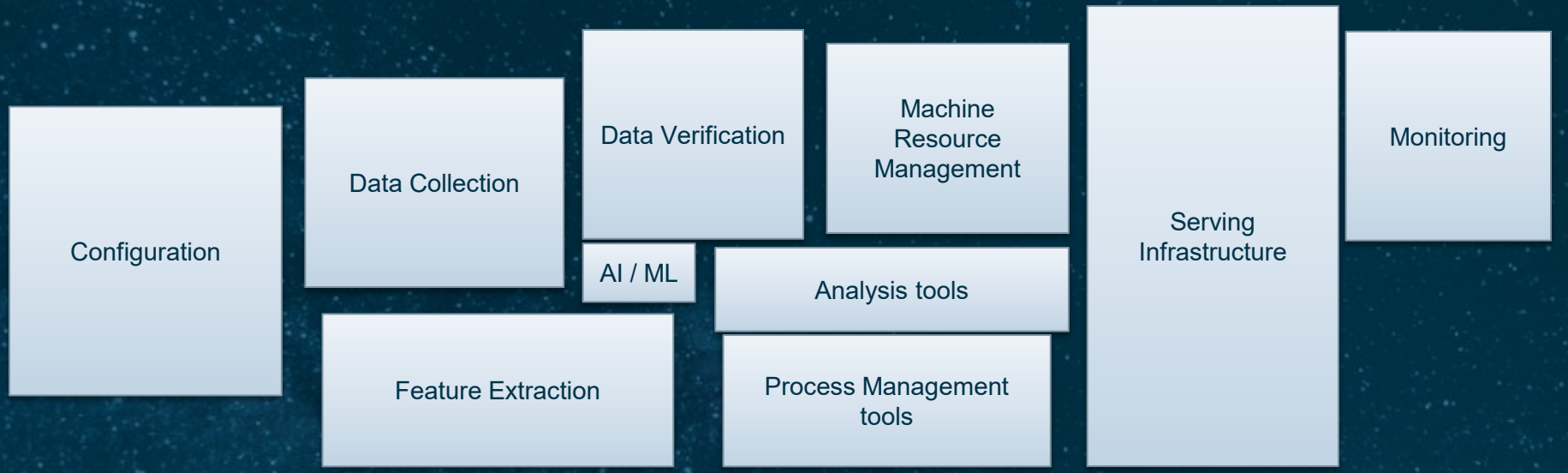


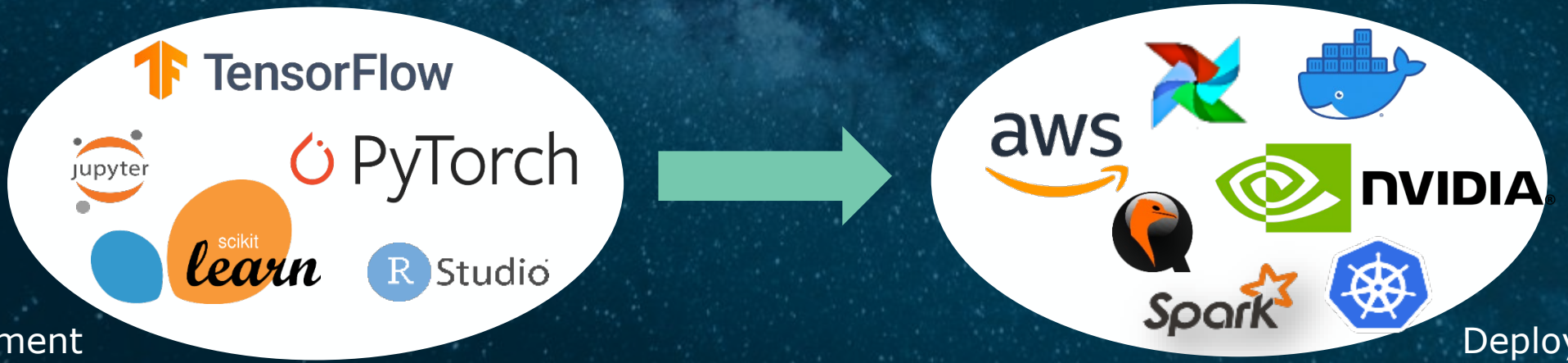
Fine Grain Integration with PSA (evolution opt)



Data Processing Innovation

MLOps



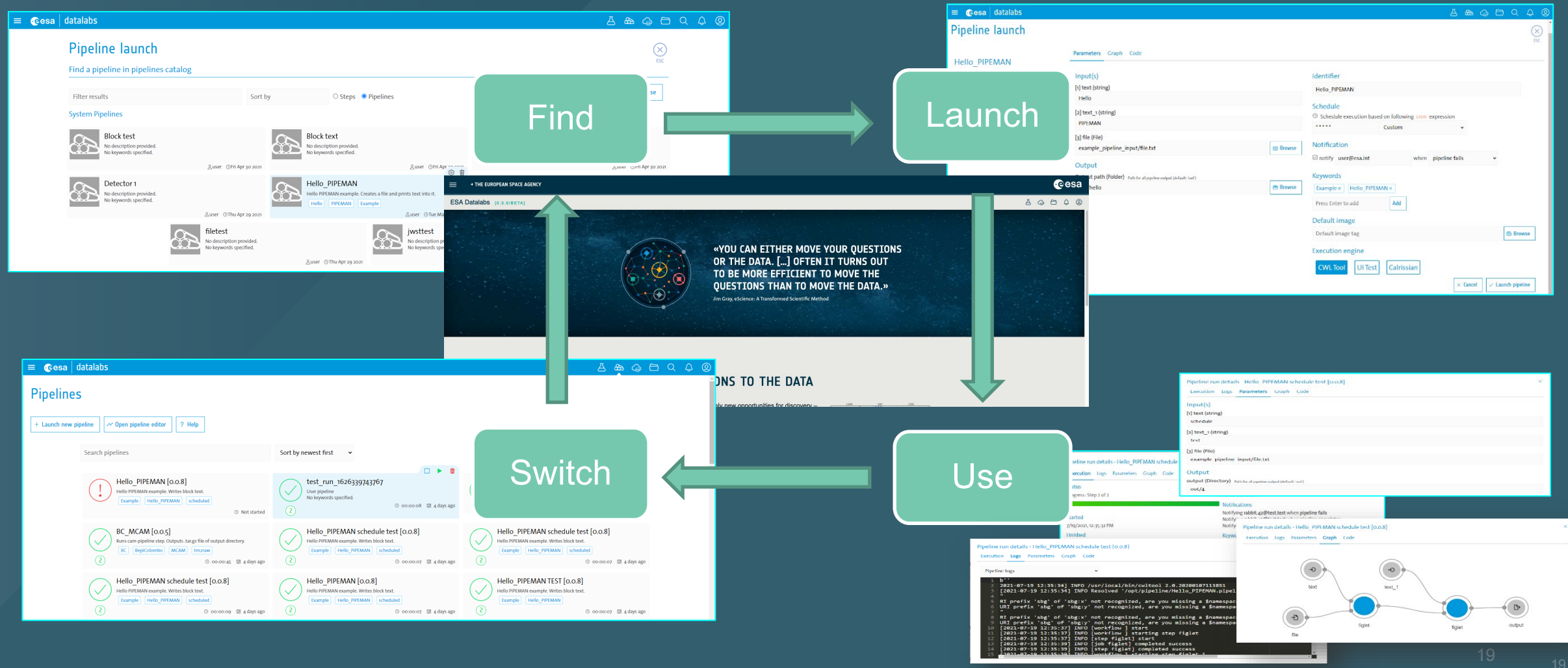


Model Development

Model Deployment

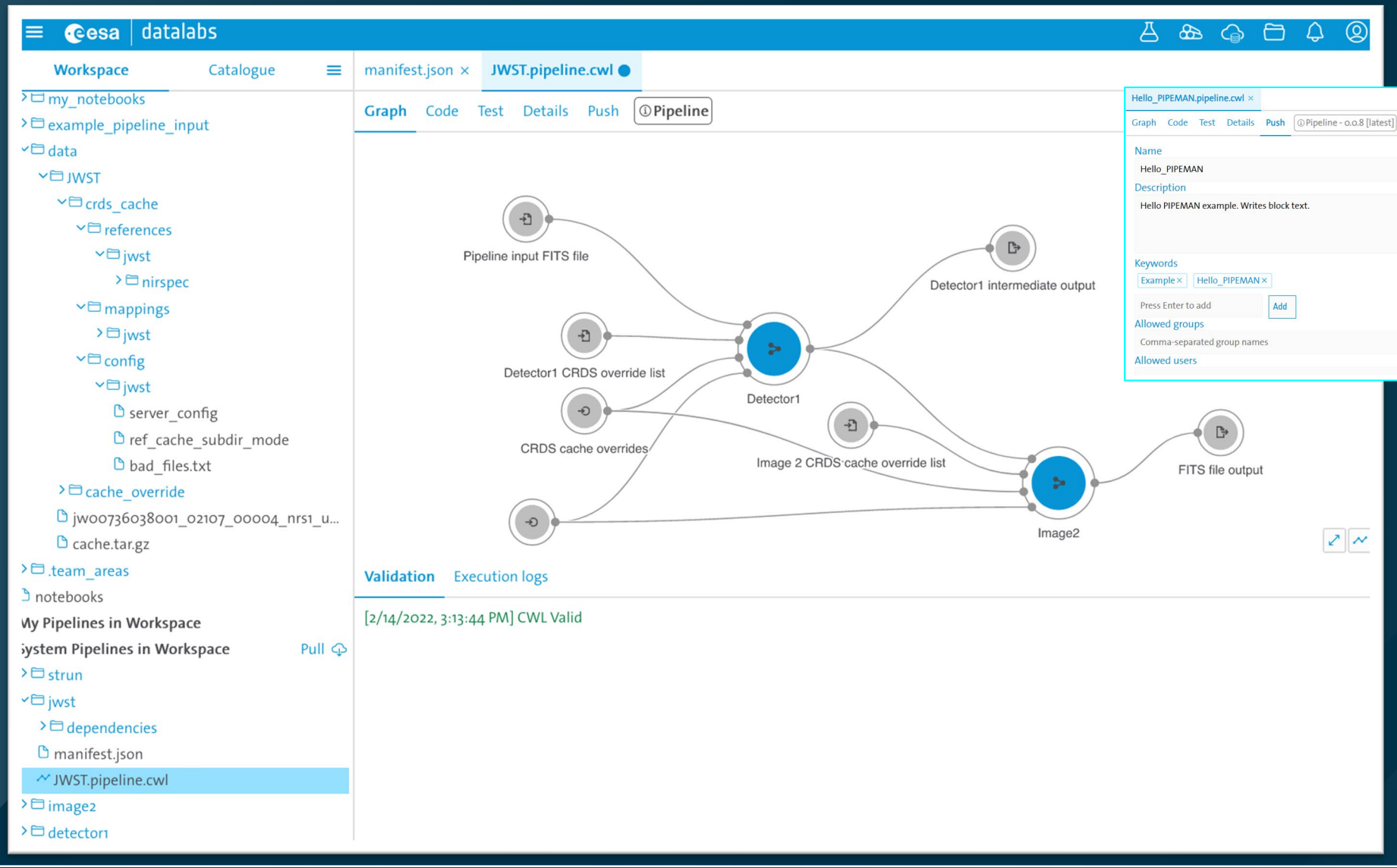
*Source: Hidden Technical Debt in Machine Learning Systems

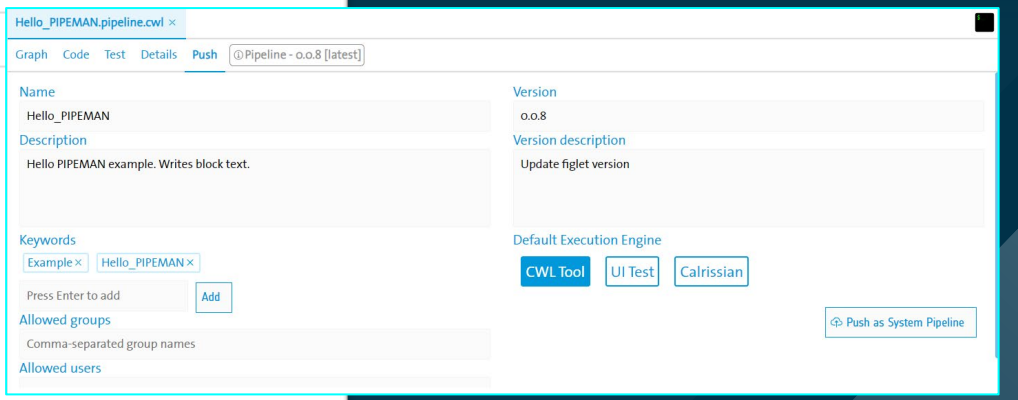
Pipelines – Standard Flow



19

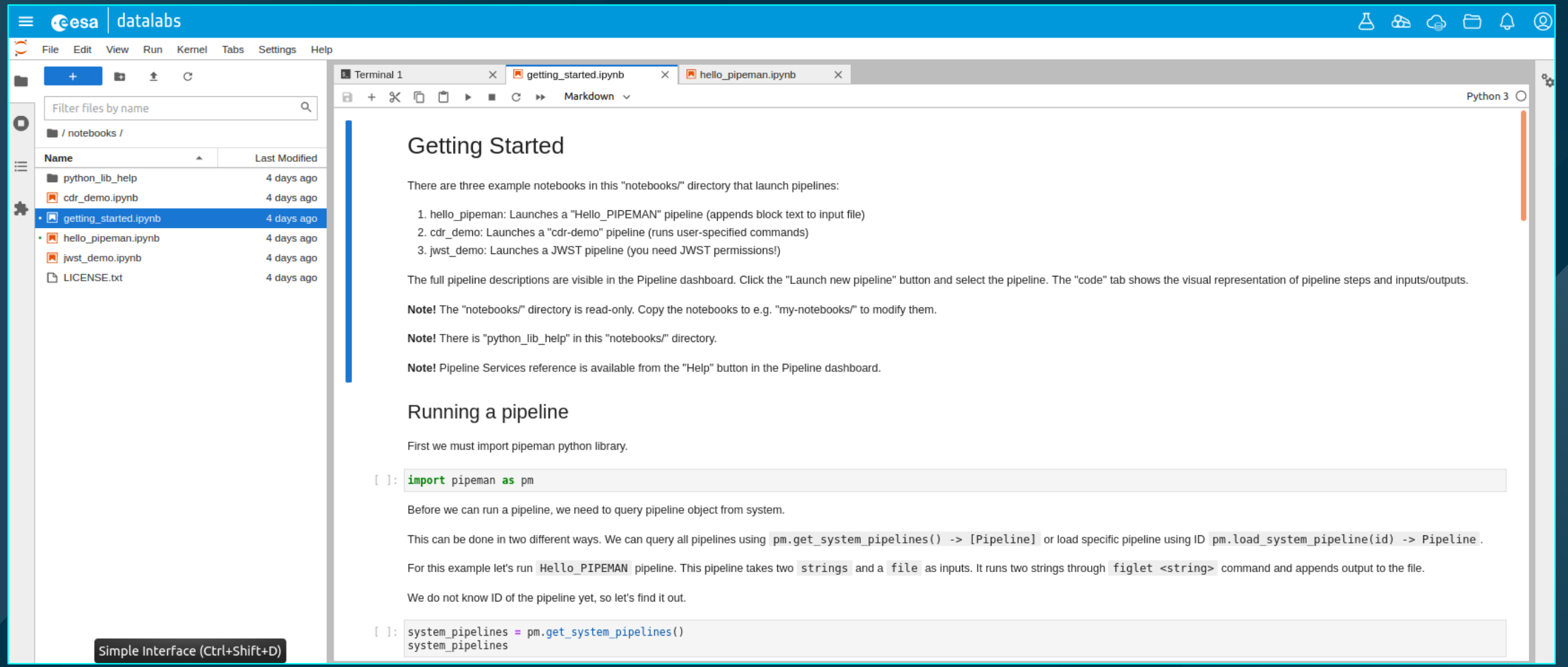
Pipeline Editor



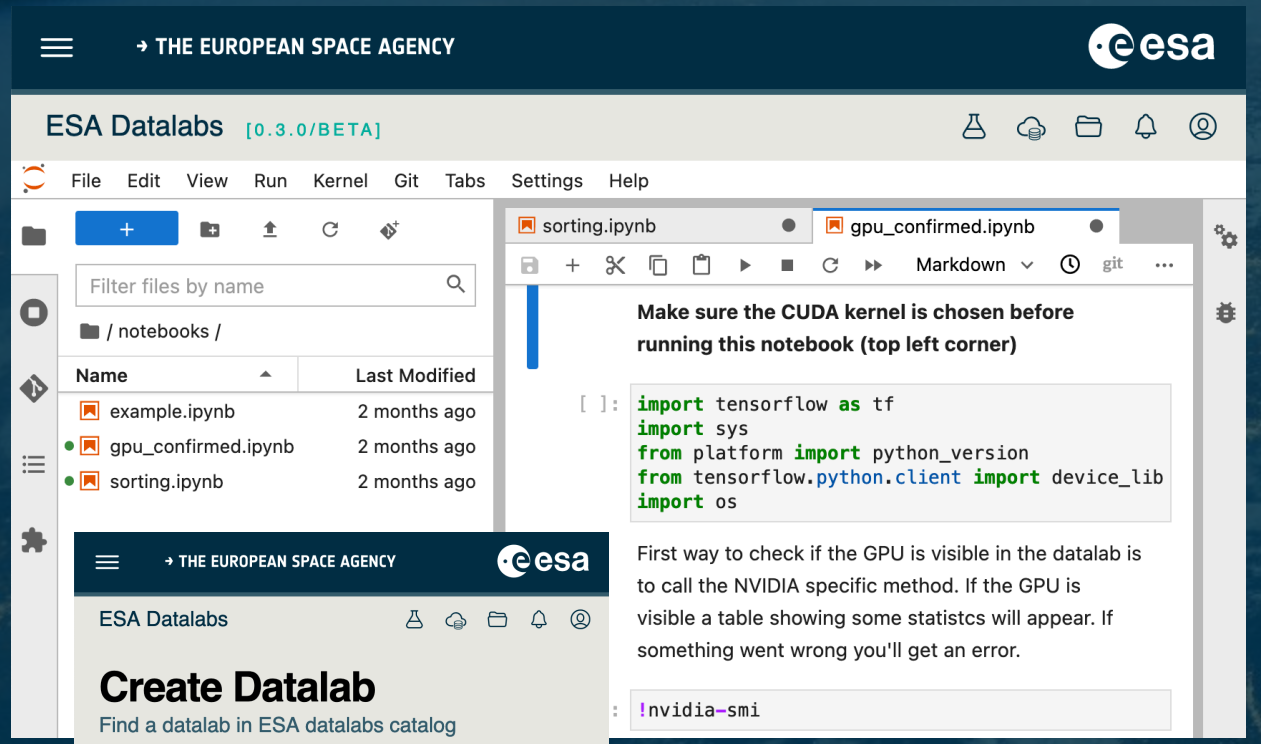


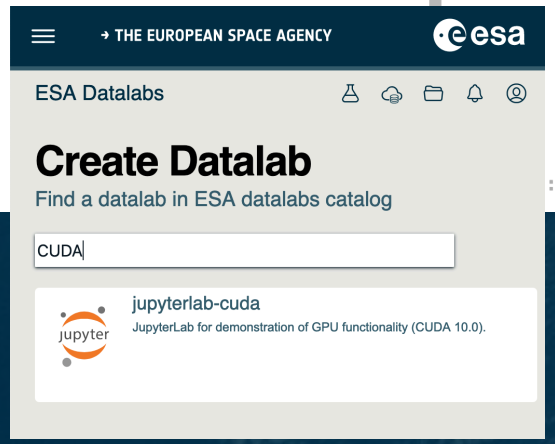
JWST @ ESA Datalabs

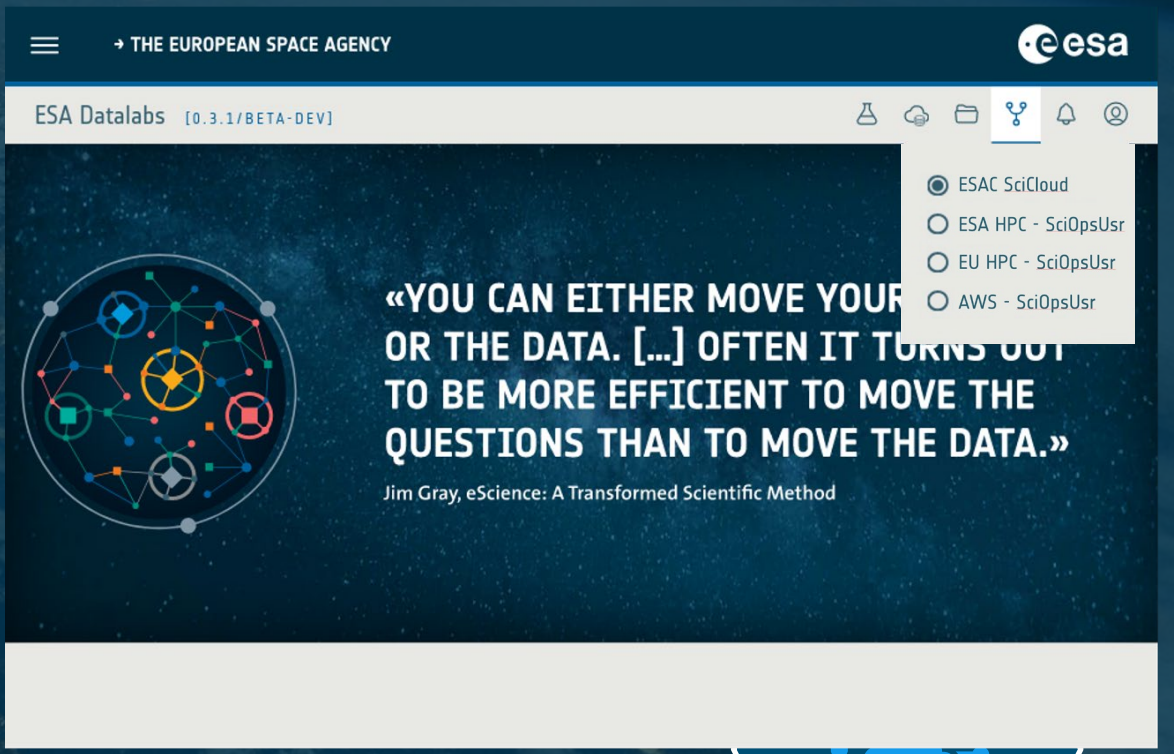
JupyterLab Pipeline Management



GPU Computing and Federated Scalability

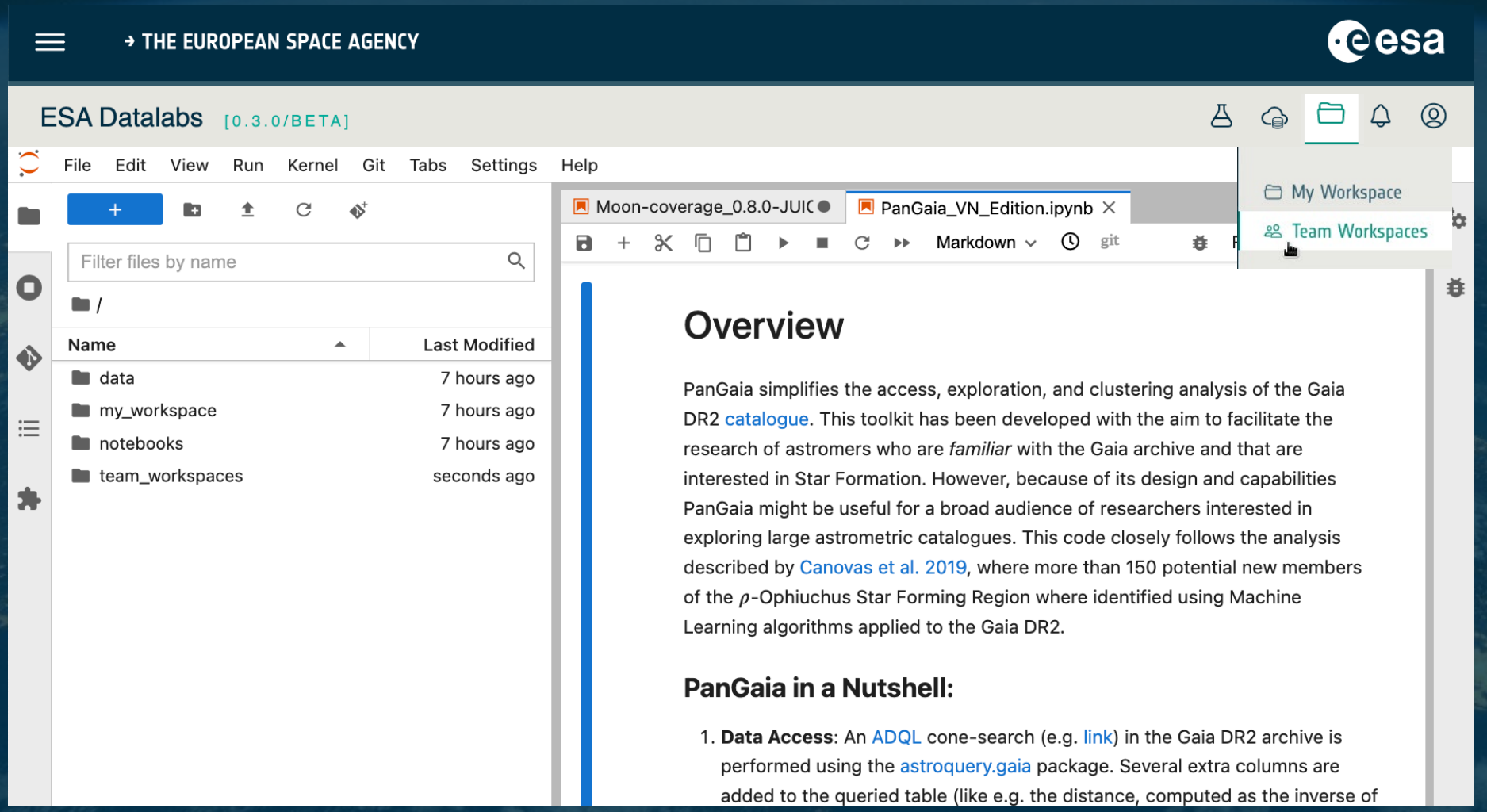




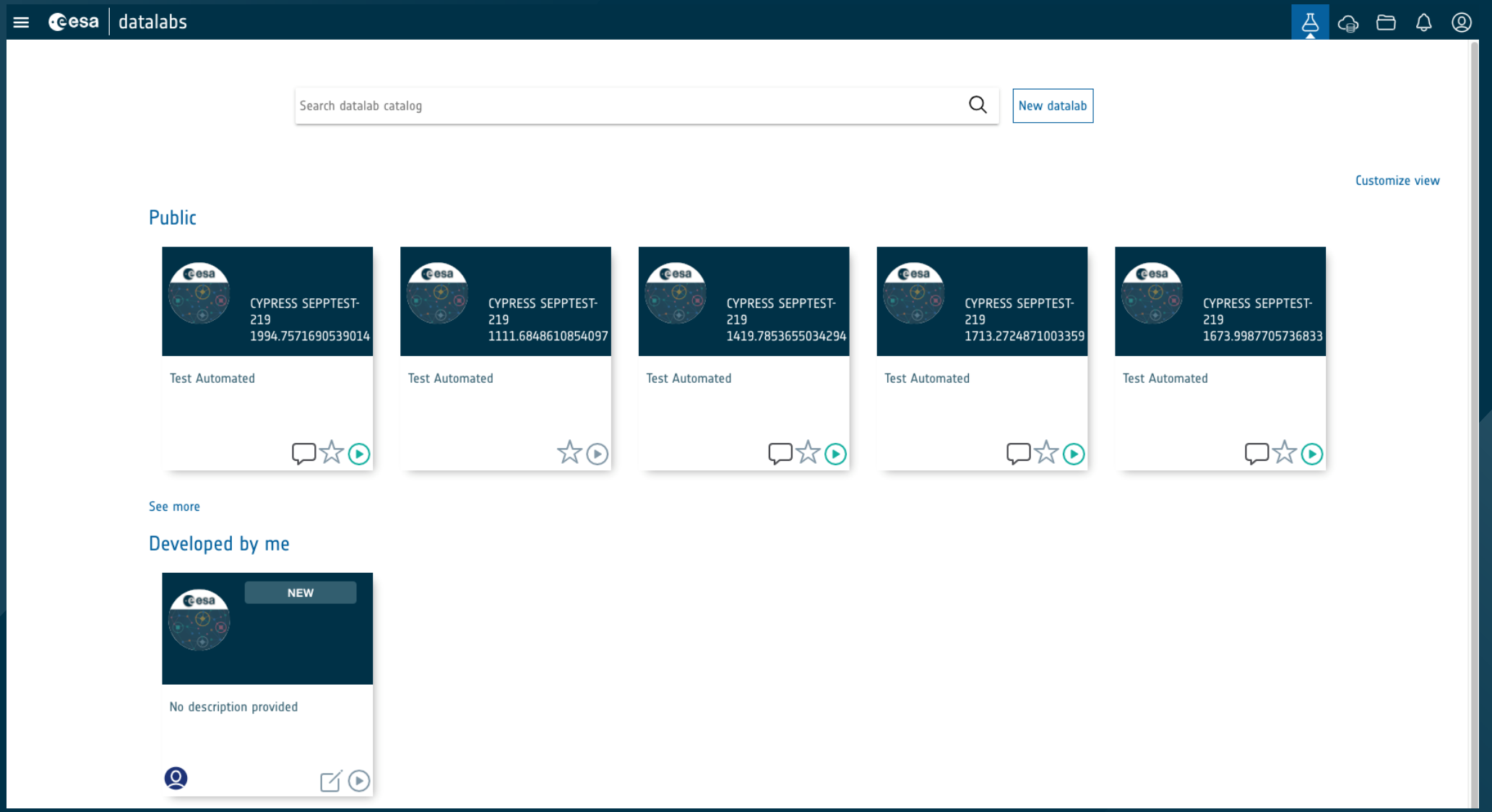


Collaborative Innovation

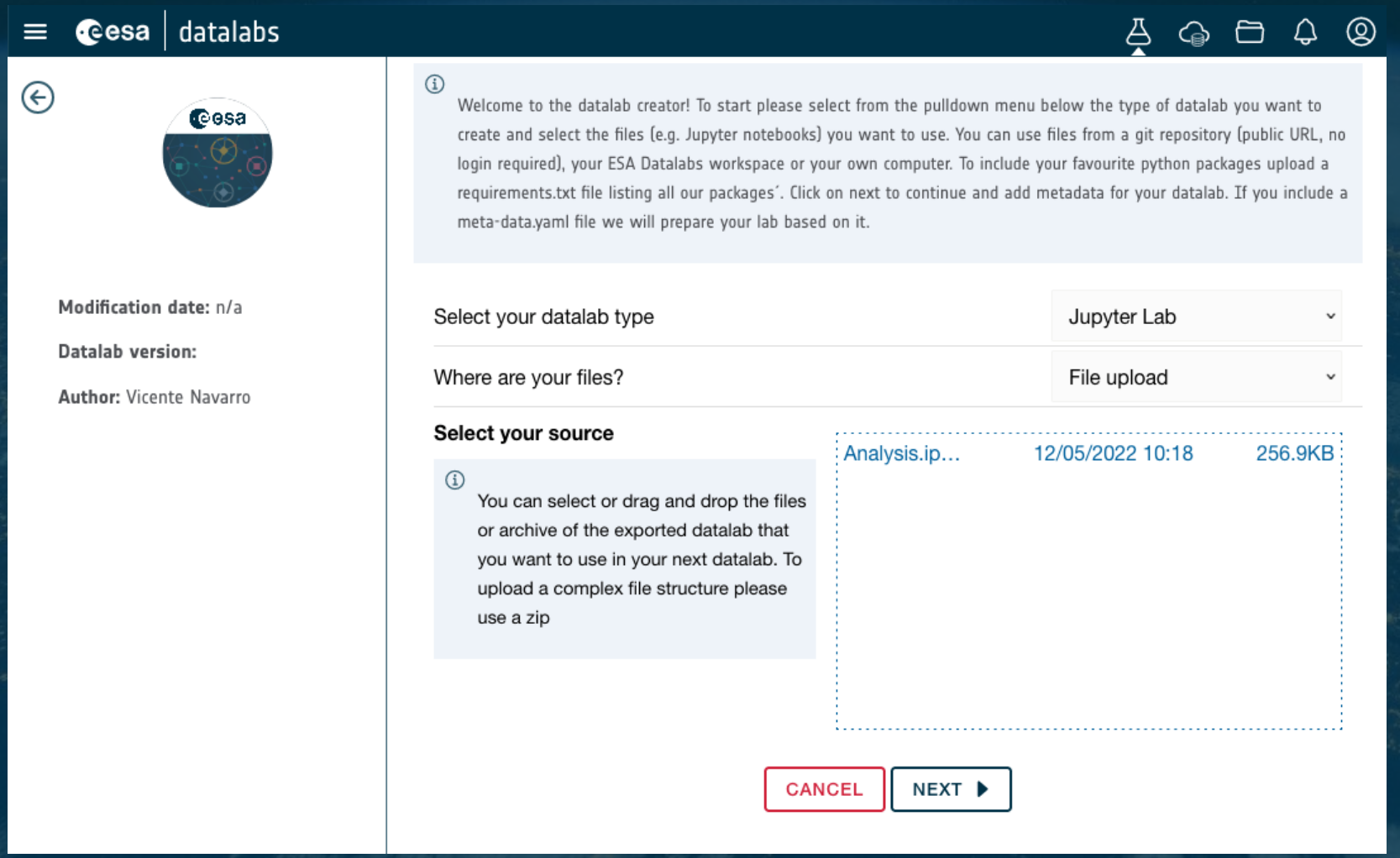
Data Centric Collaboration - Team Workspaces



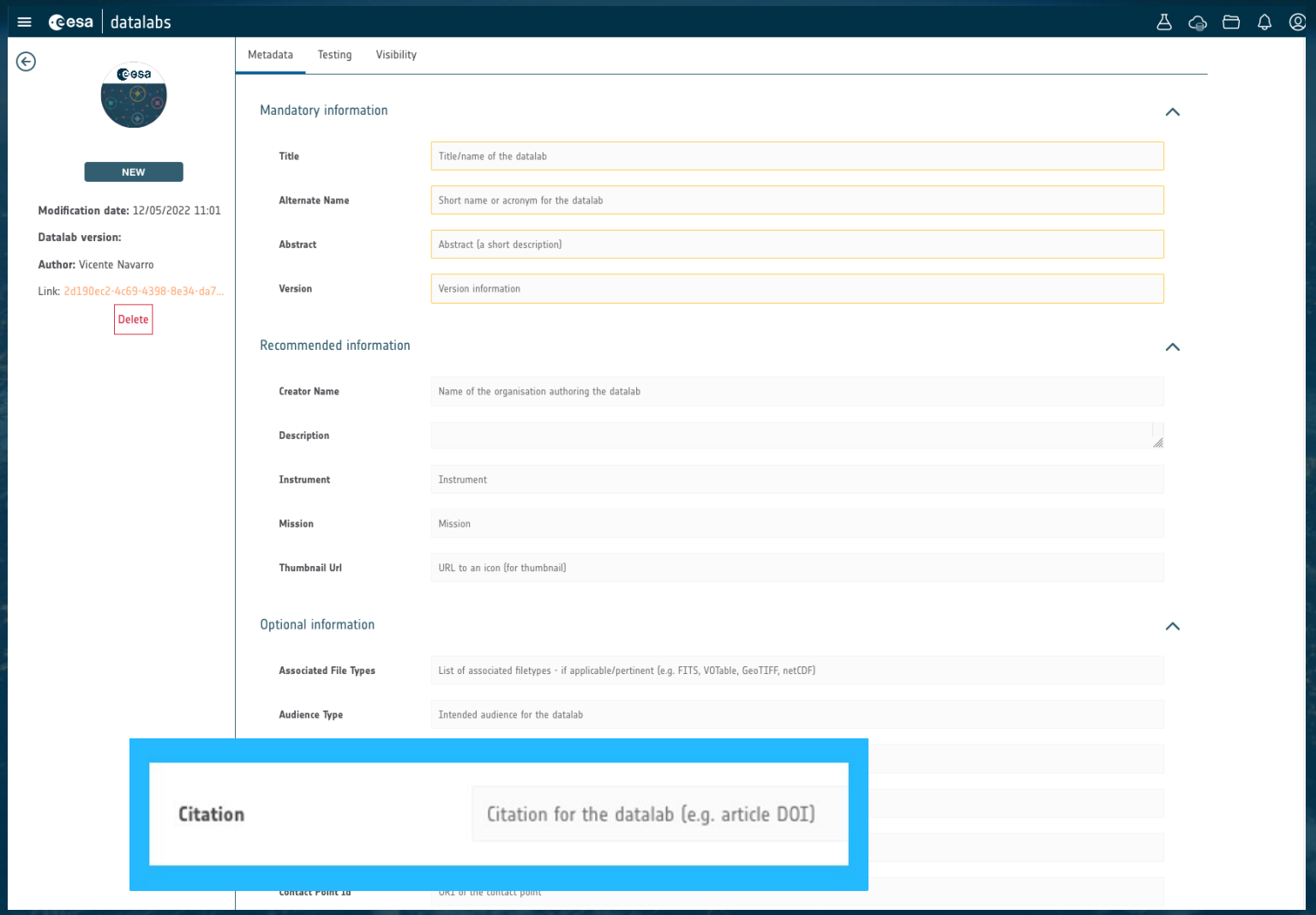
Application Centric Collaboration - AppStore



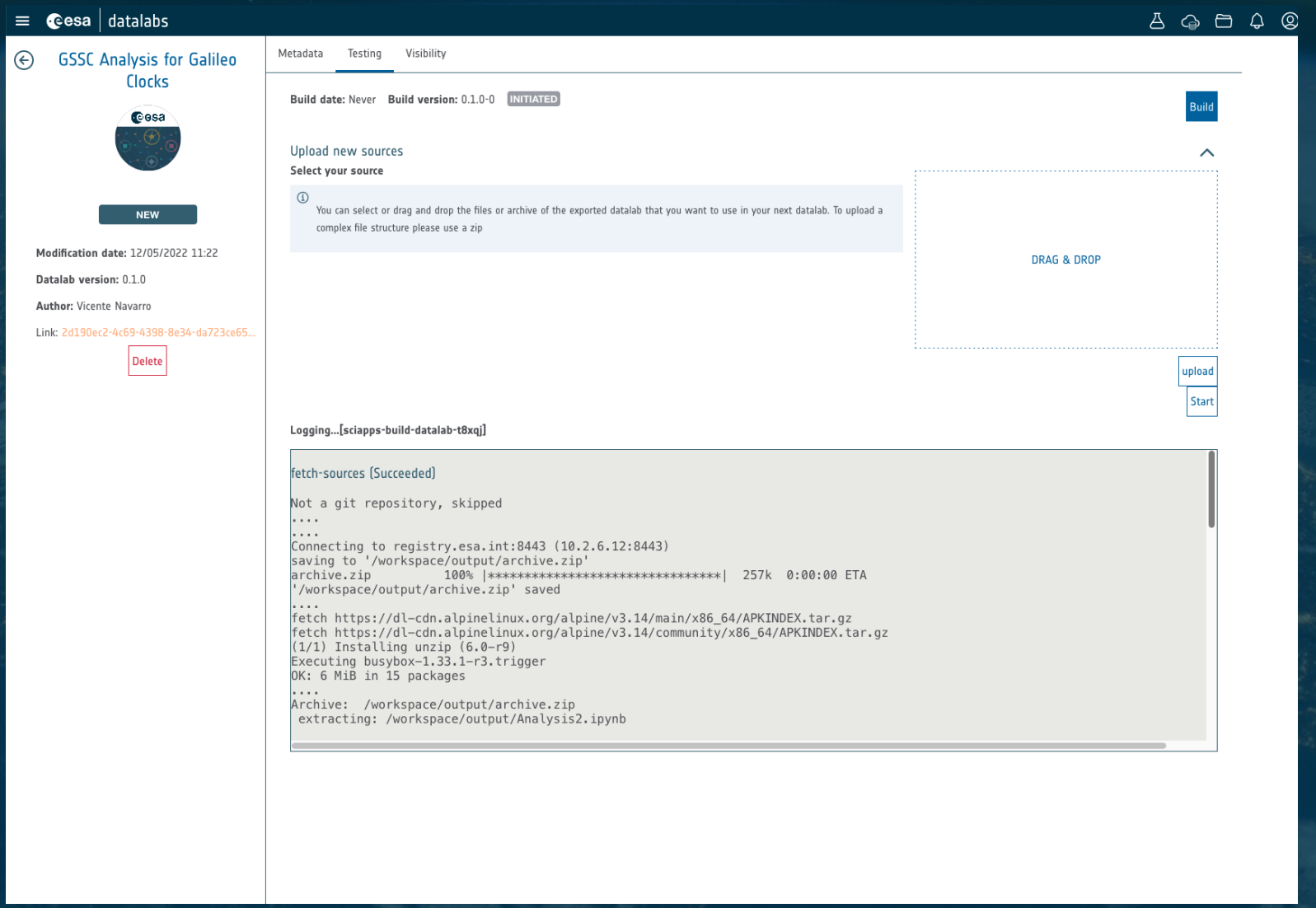
Datalabs Editor



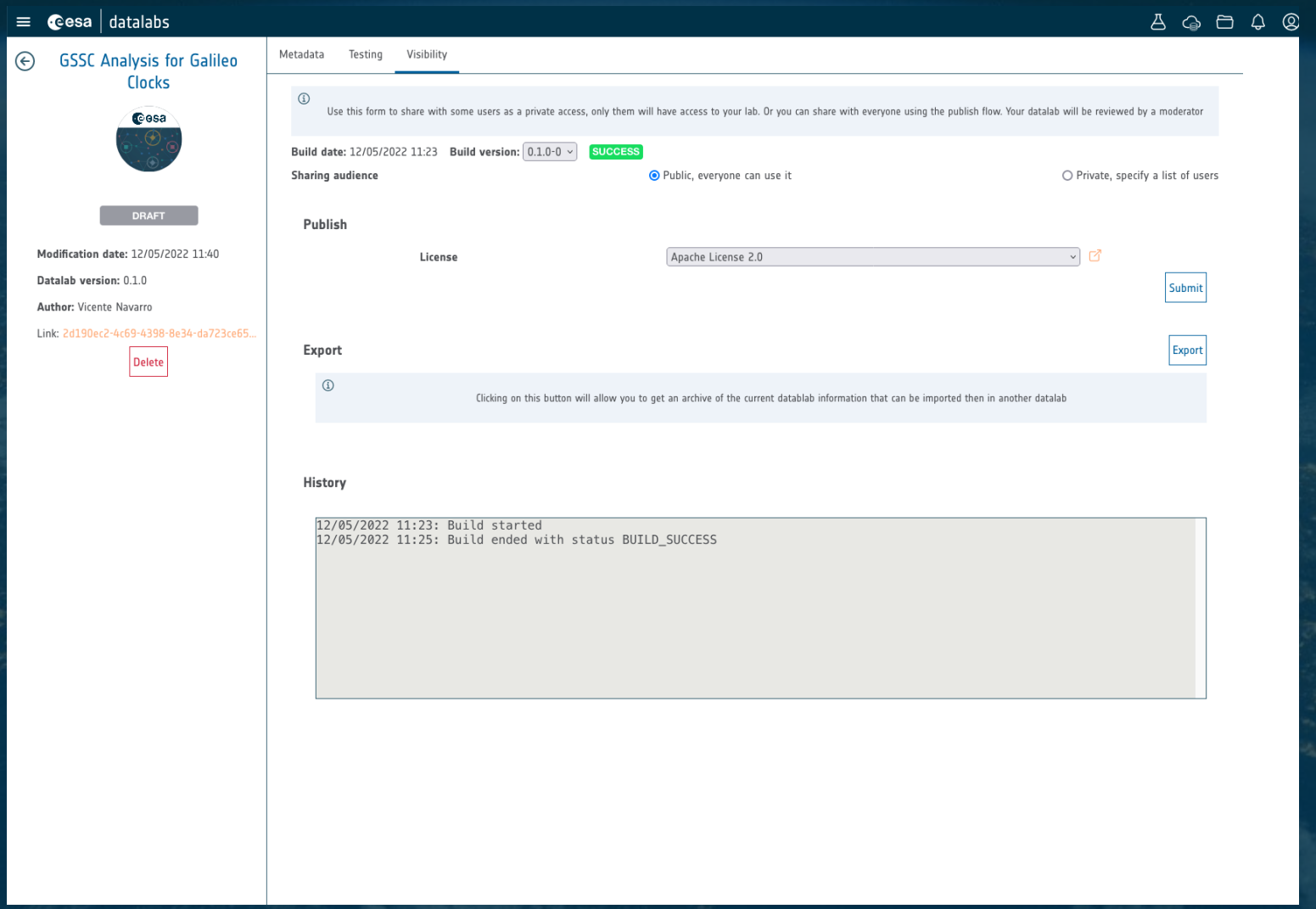
Datalabs Editor - Metadata



Datalabs Editor - Build



Datalabs Editor - Share



Take away messages

ESA Science strategy support

- Increased science return from its missions
- Increased science operations efficiency

Innovation traits

- Science data exploitation coupled with data
- Science pipelines for current and future needs
- Collaborative research & citizen science

Thank You!

datalabs.esa.int

… special thanks to ESA Datalabs Industry Teams and User's Group (+50 members)

Bonus Slides

IT Paradigms

Science Development

System Development

IT Development

SaaS

PaaS

IaaS

IT Monitoring

IT capacity

- Pre-production Environment: https://datalabs.esa.int
  - System Cluster:
    - 6 nodes
    - 56 vCPUs
    - 100 GB RAM
    - Currently, 70 containers
  - User Cluster:
    - 6 nodes
    - 48 vCPUs
    - 256 GB RAM
    - Currently, 216 containers
  - Storage:
    - 500GB for system persistence
    - 1TB for user persistence
- Development & Acceptance environments:
  - System cluster: running on shared cluster, where other systems are located
  - User cluster:
    - 4 nodes, one with GPU
    - 20 vCPUs
    - 74 GB RAM
    - Currently, 73 containers
    - Shared by DEV and E2E envs.
  - Storage: shared storage with other projects.