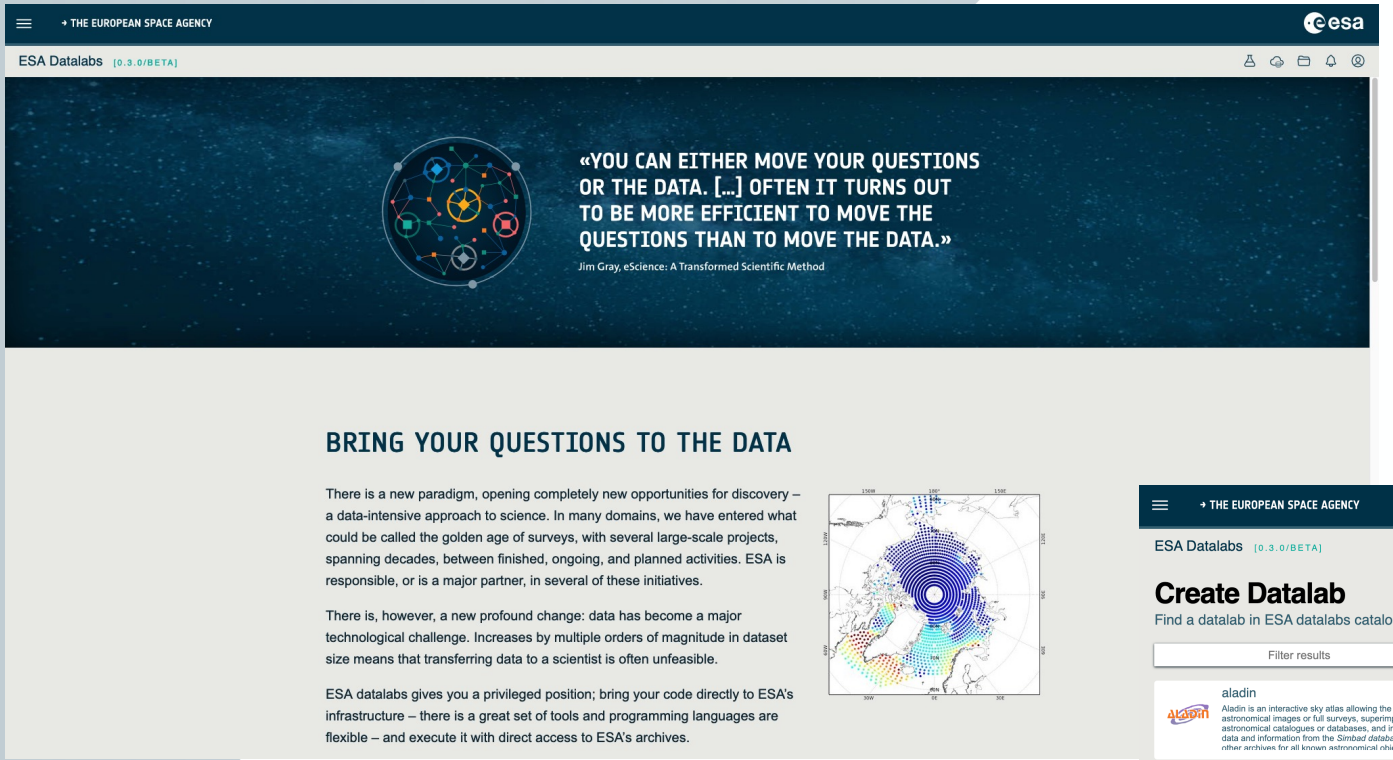
ESA Datalabs

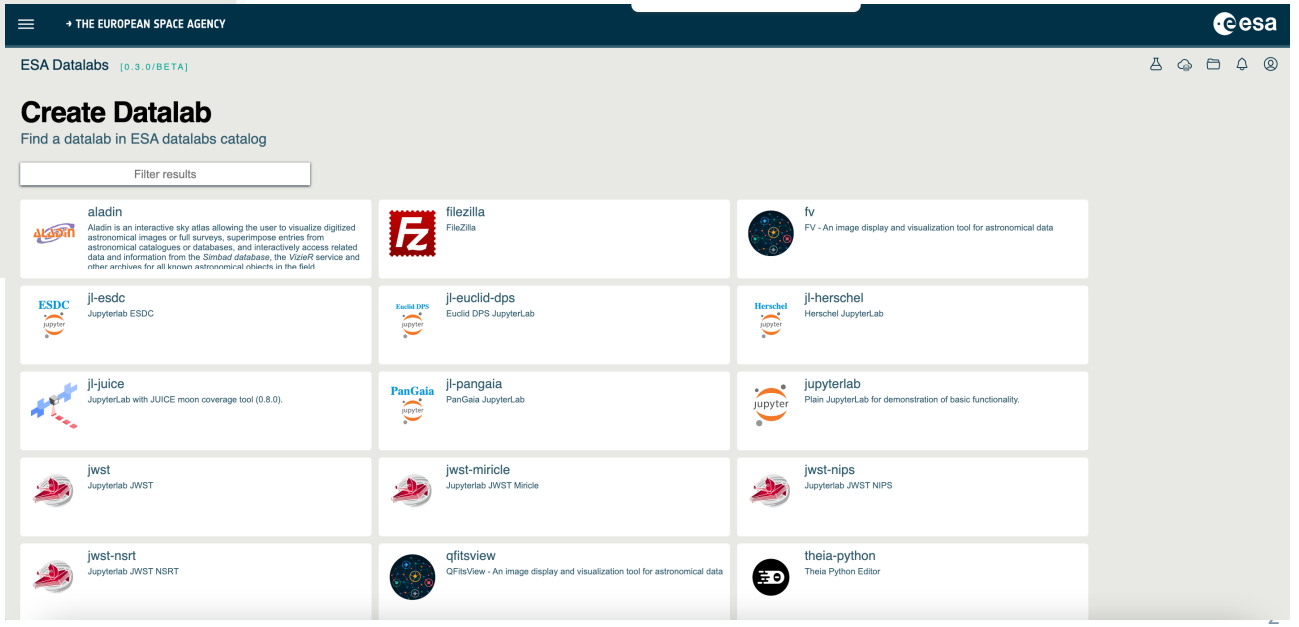
Vicente Navarro, on behalf of the ESA Datalabs Team

IVOA Interop

18.10.2022, Virtual

ESA Datalabs – Across Science Domains





ESA Datalabs – Across IT Domains

Science Development

System Development

IT Development

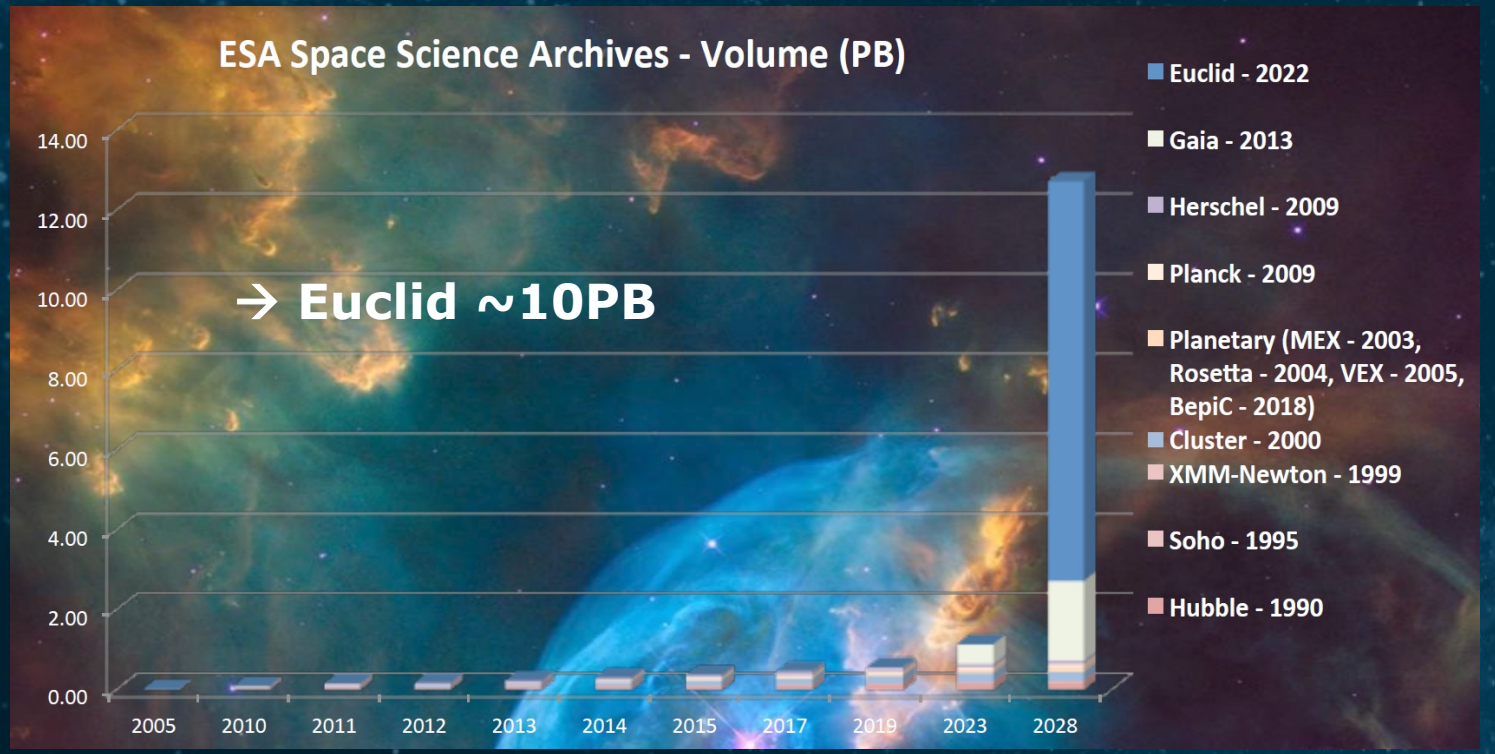
SaaS

PaaS

IaaS

Data Access
Innovation

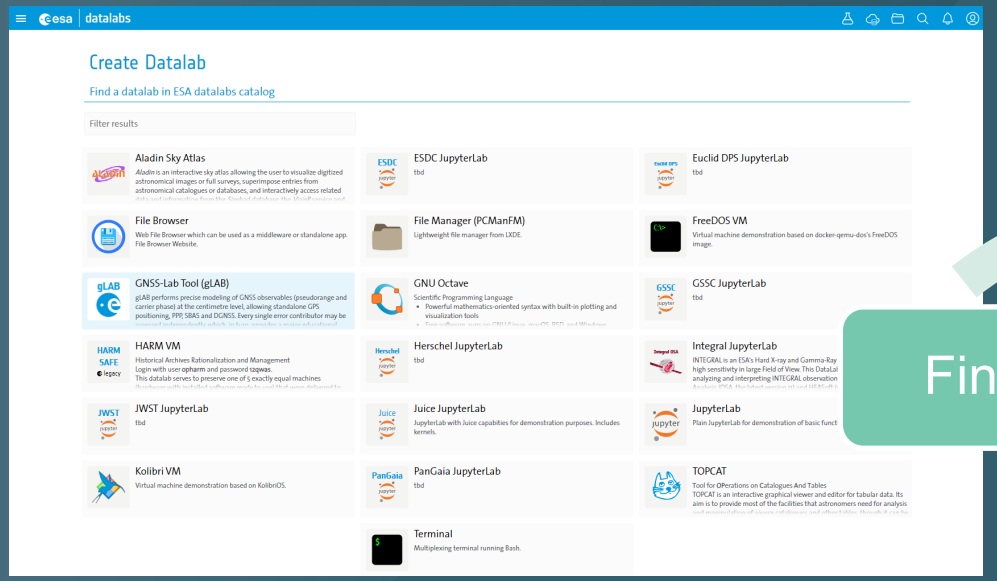
Data Exploitation Paradigm Shift



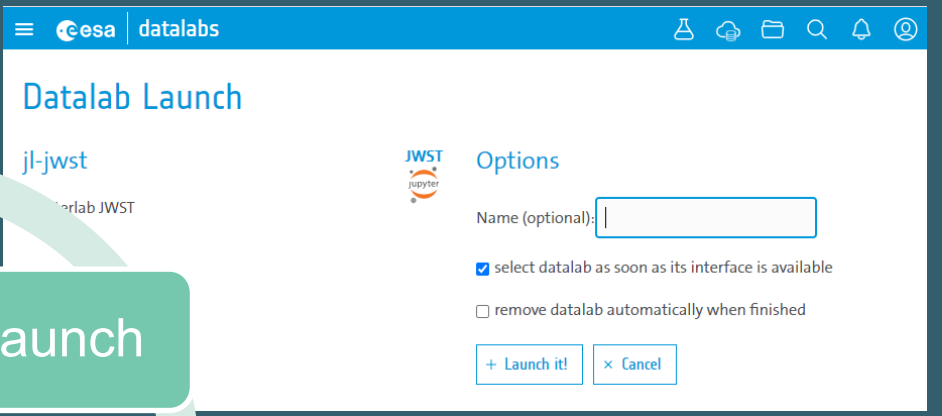
From *bring the data to the user*

To *bring the user to the data*

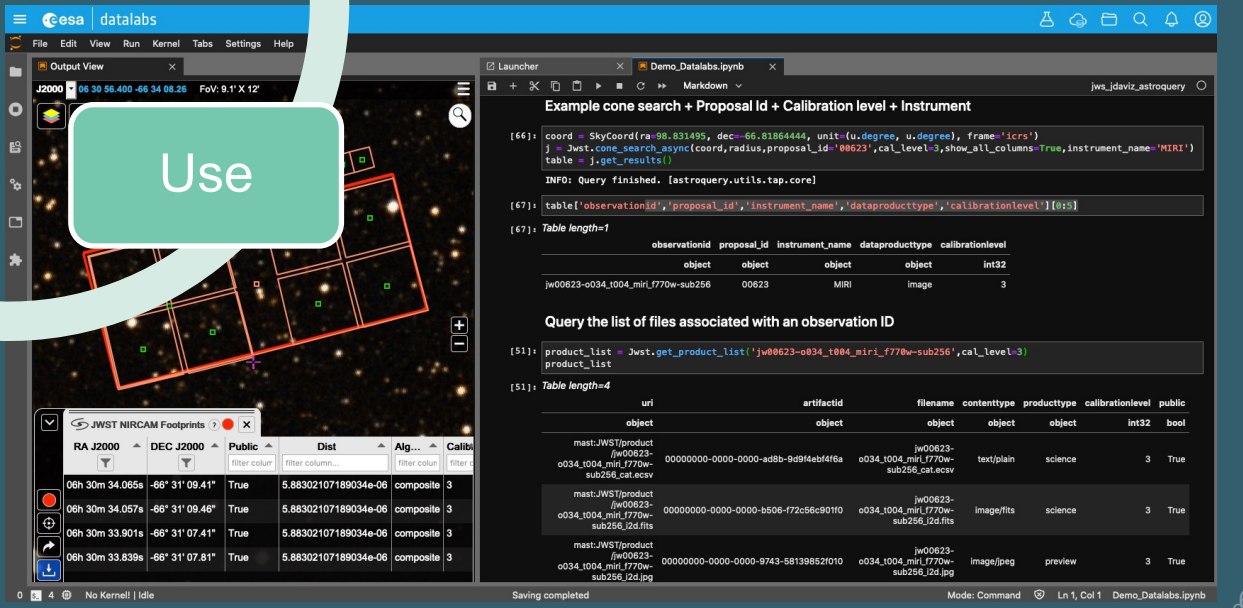
Main Scenario / Use Case



Find

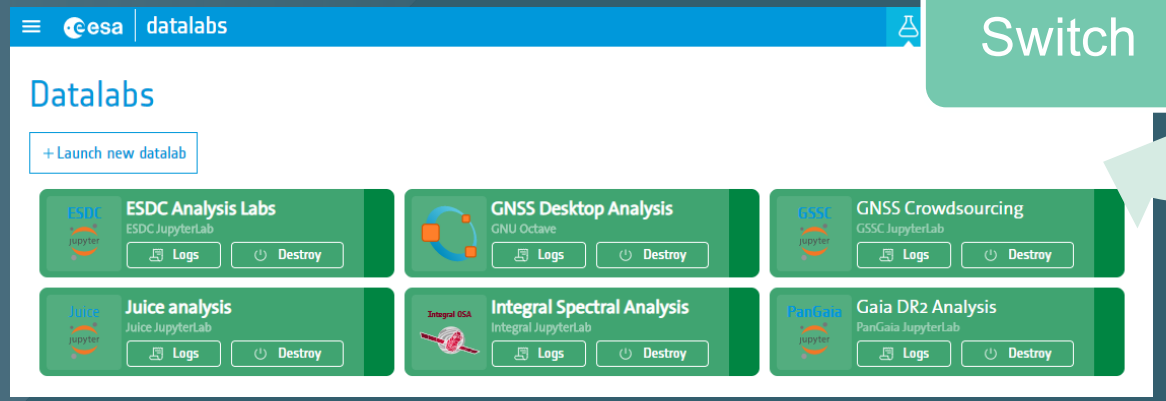


Launch

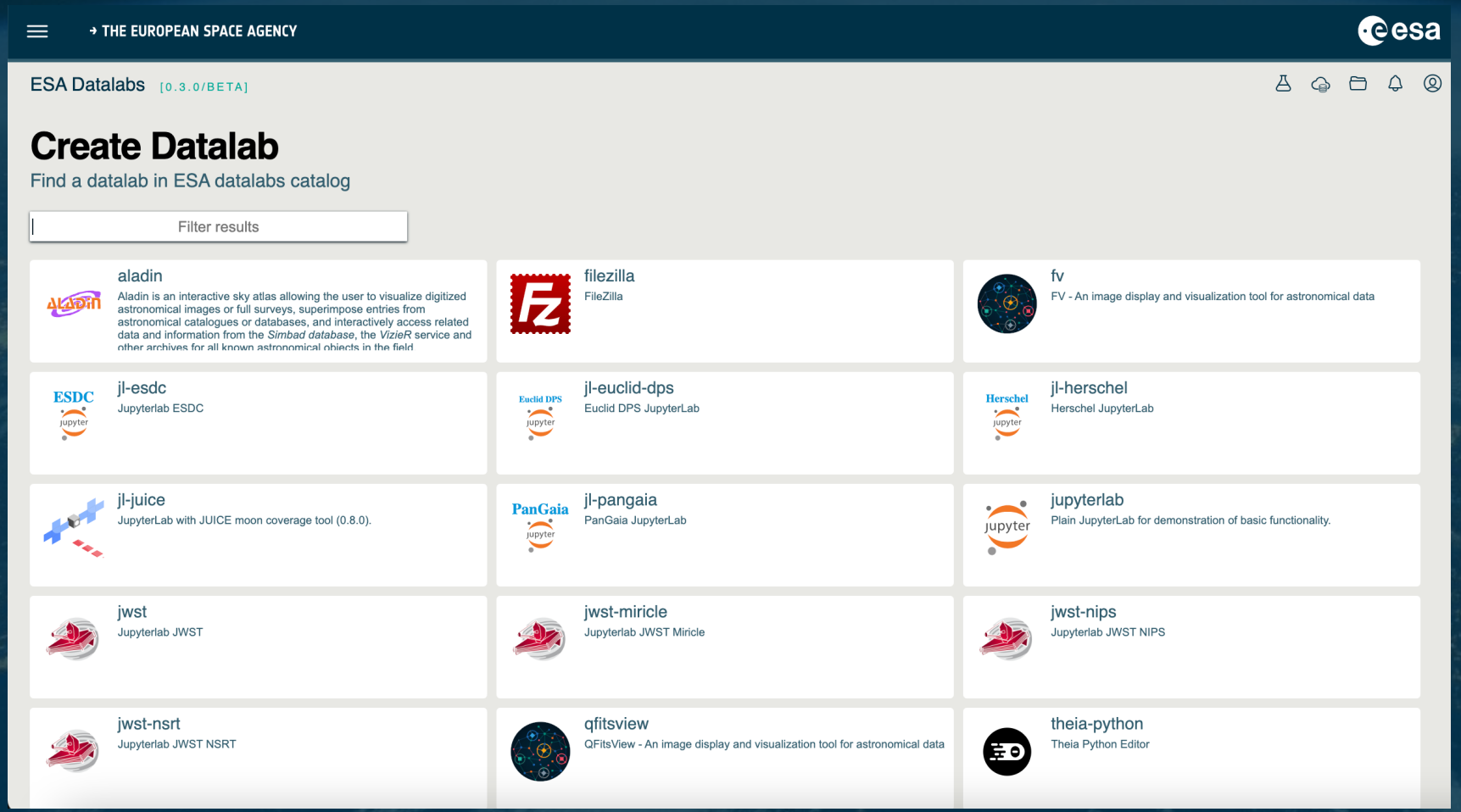


Use

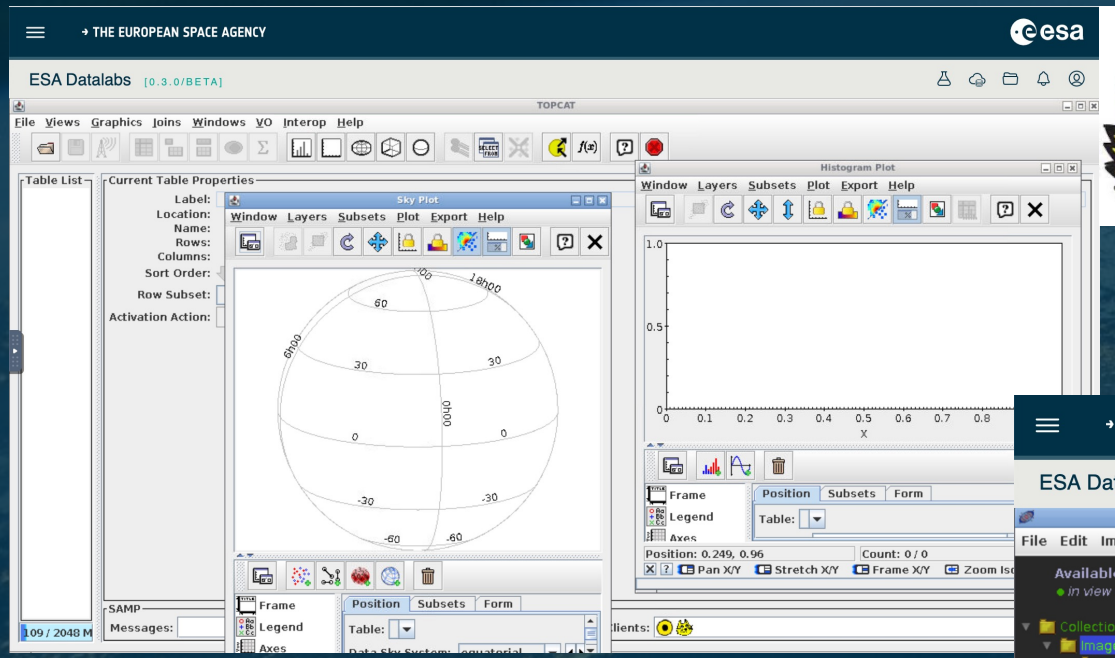
Switch

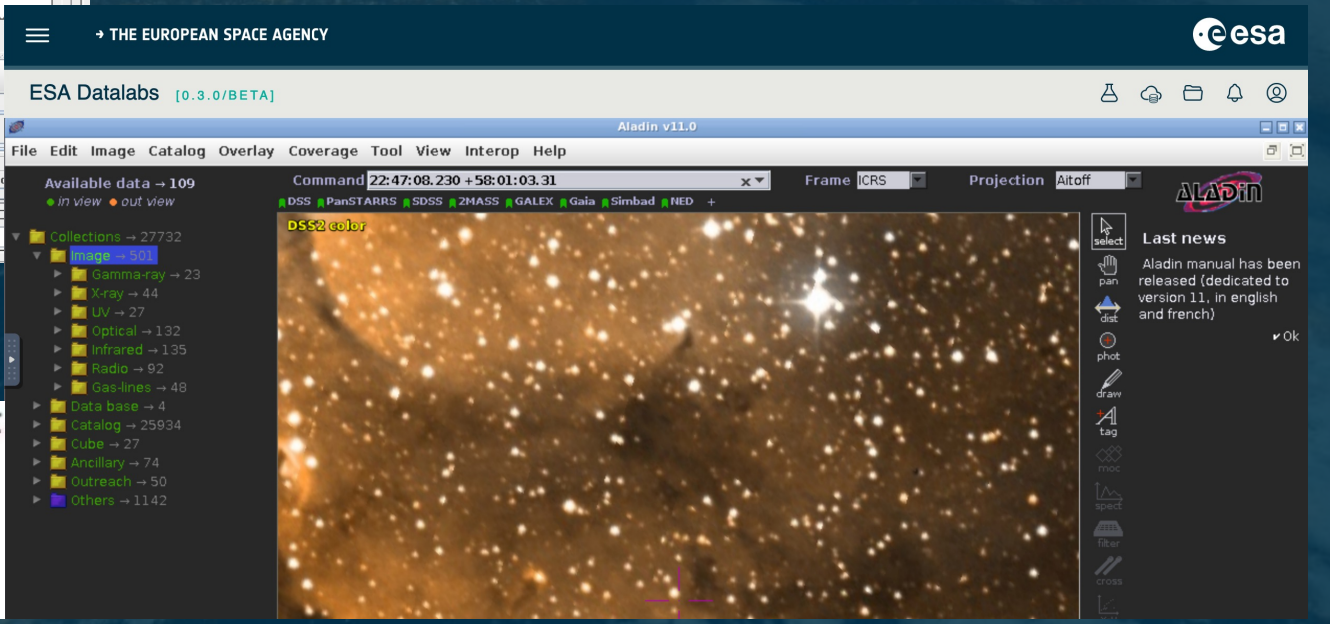


FIND

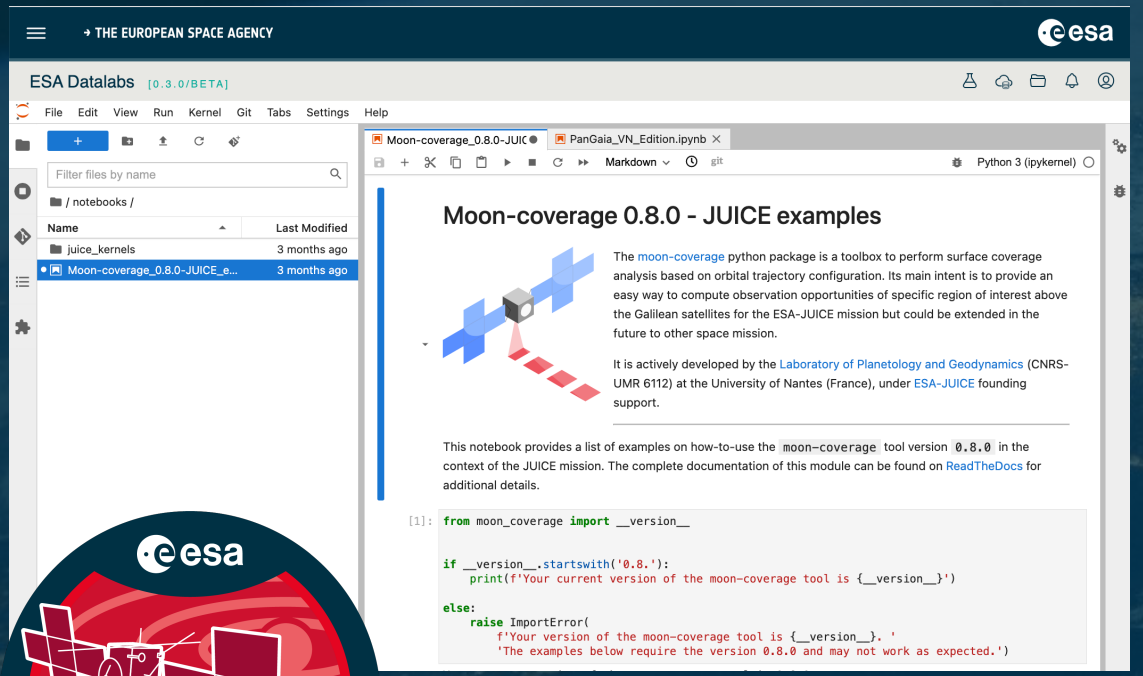

→ THE EUROPEAN SPACE AGENCY
esa
ESA Datalabs [0.3.0/BETA]
Create Datalab
Find a datalab in ESA datalabs catalog
Filter results
aladin
Aladin is an interactive sky atlas allowing the user to visualize digitized astronomical images or full surveys, superimpose entries from astronomical catalogues or databases, 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

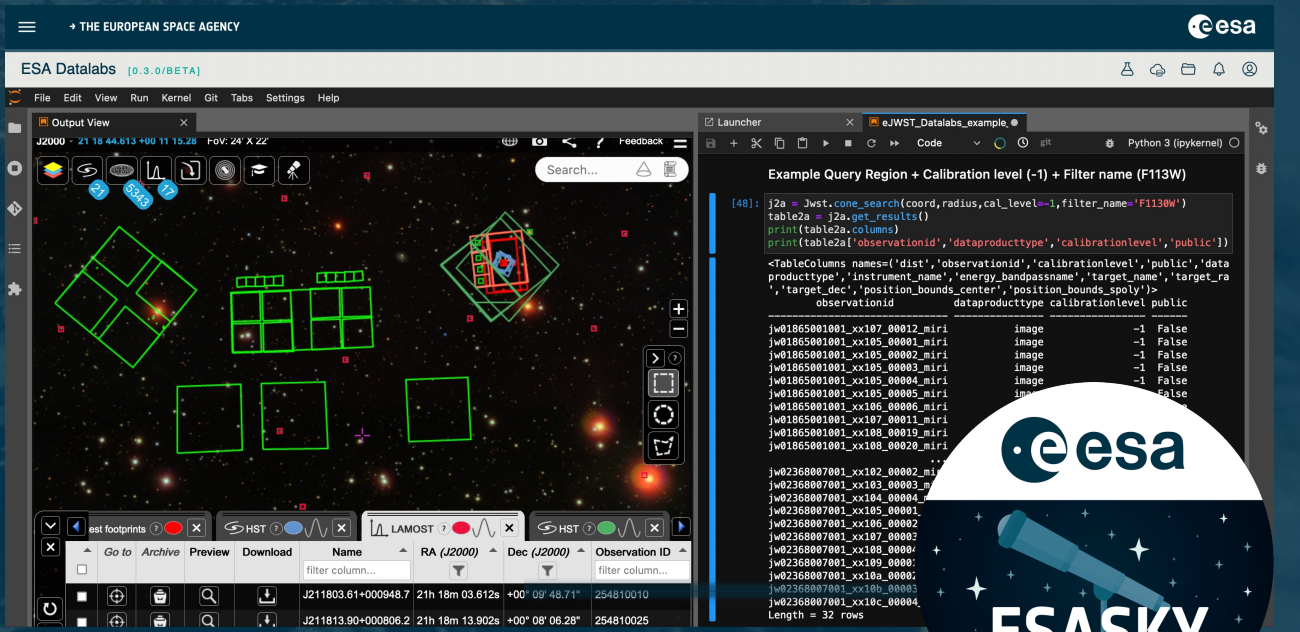
FIND – Desktop Apps



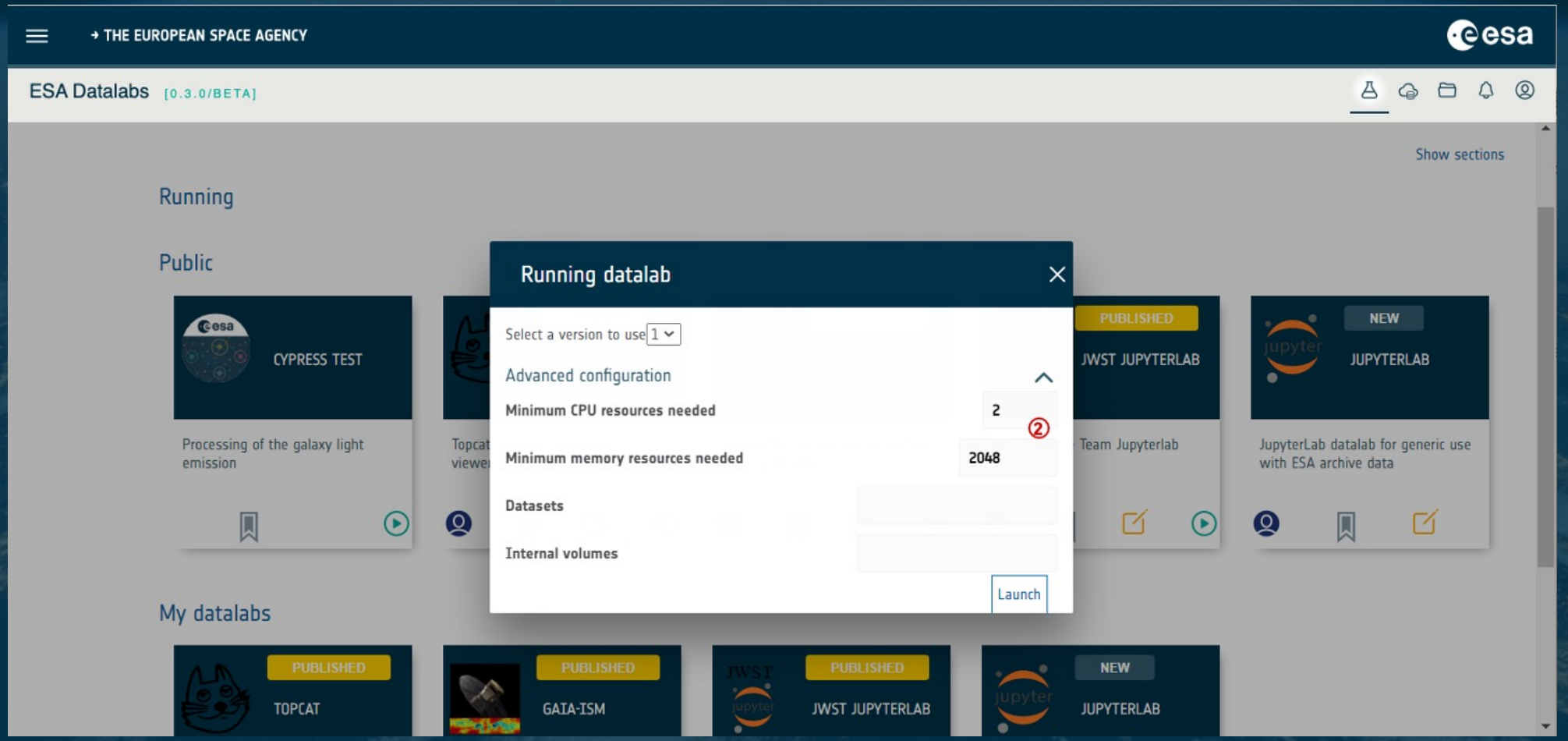


FIND – Web Apps (i.e. JupyterLab)





LAUNCH


→ THE EUROPEAN SPACE AGENCY
ESA 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)

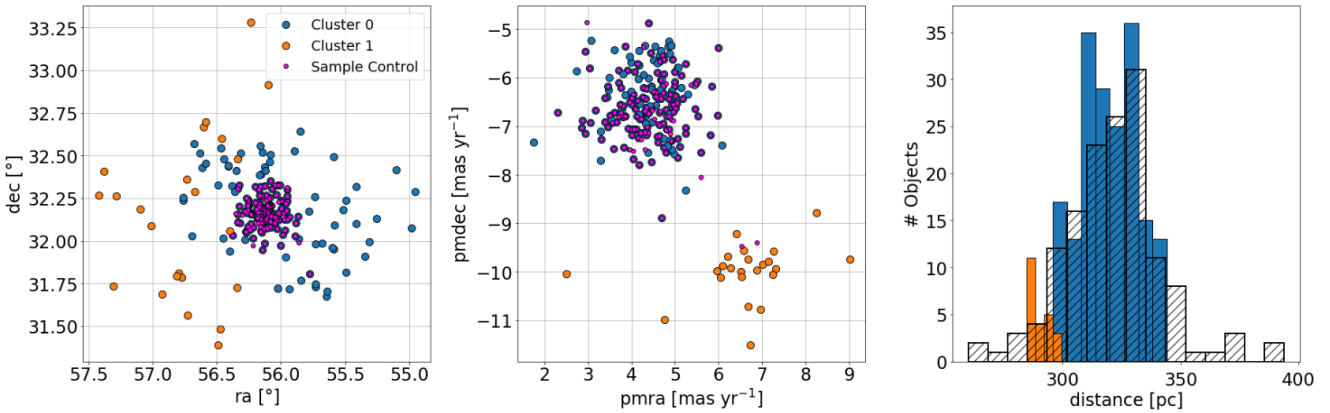
Filter files by name

/ 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()
```

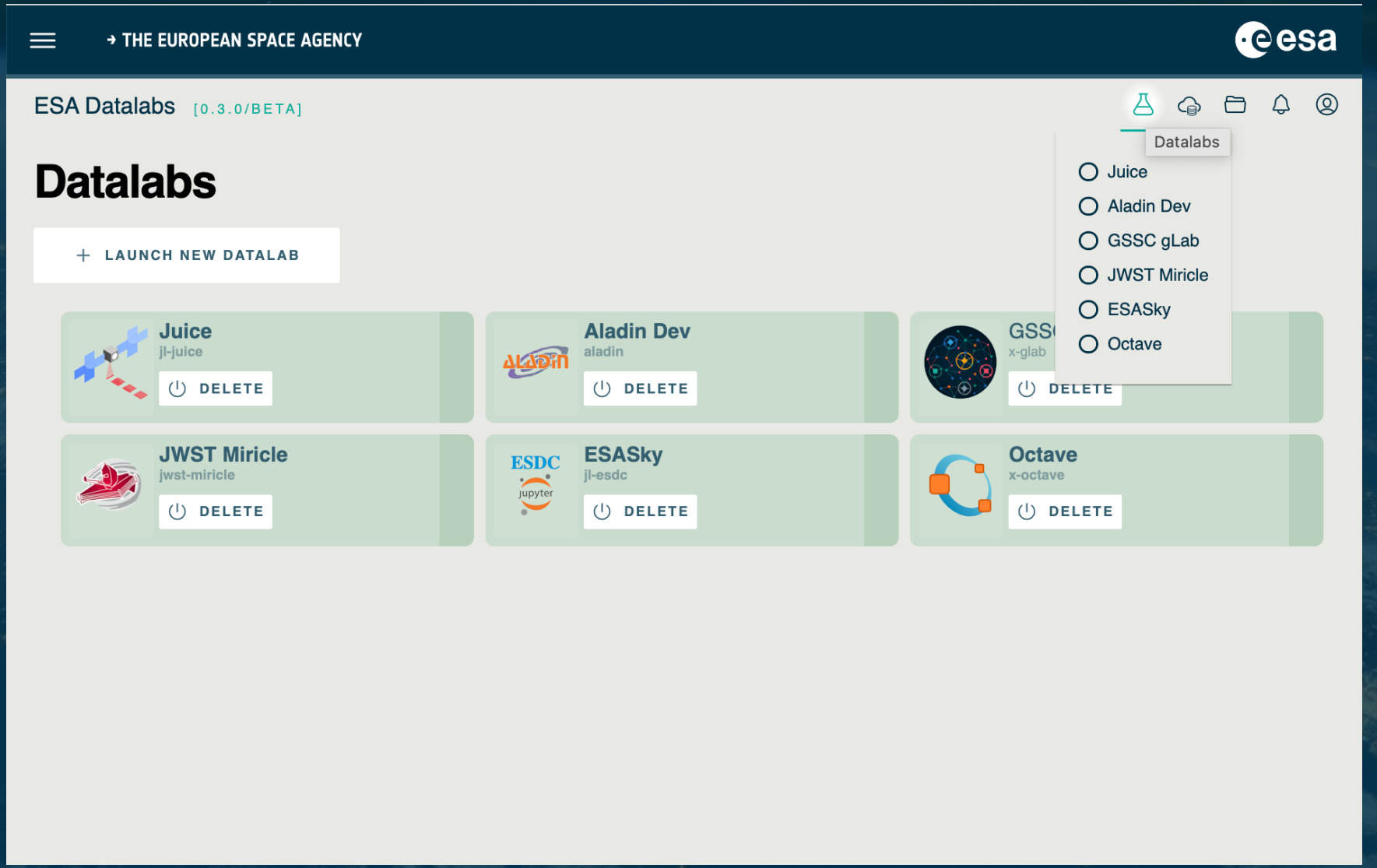
```
Insert min_Cluster_size: 20
mCls = 20; clusters = 2; N_members = [183, 23]

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
```



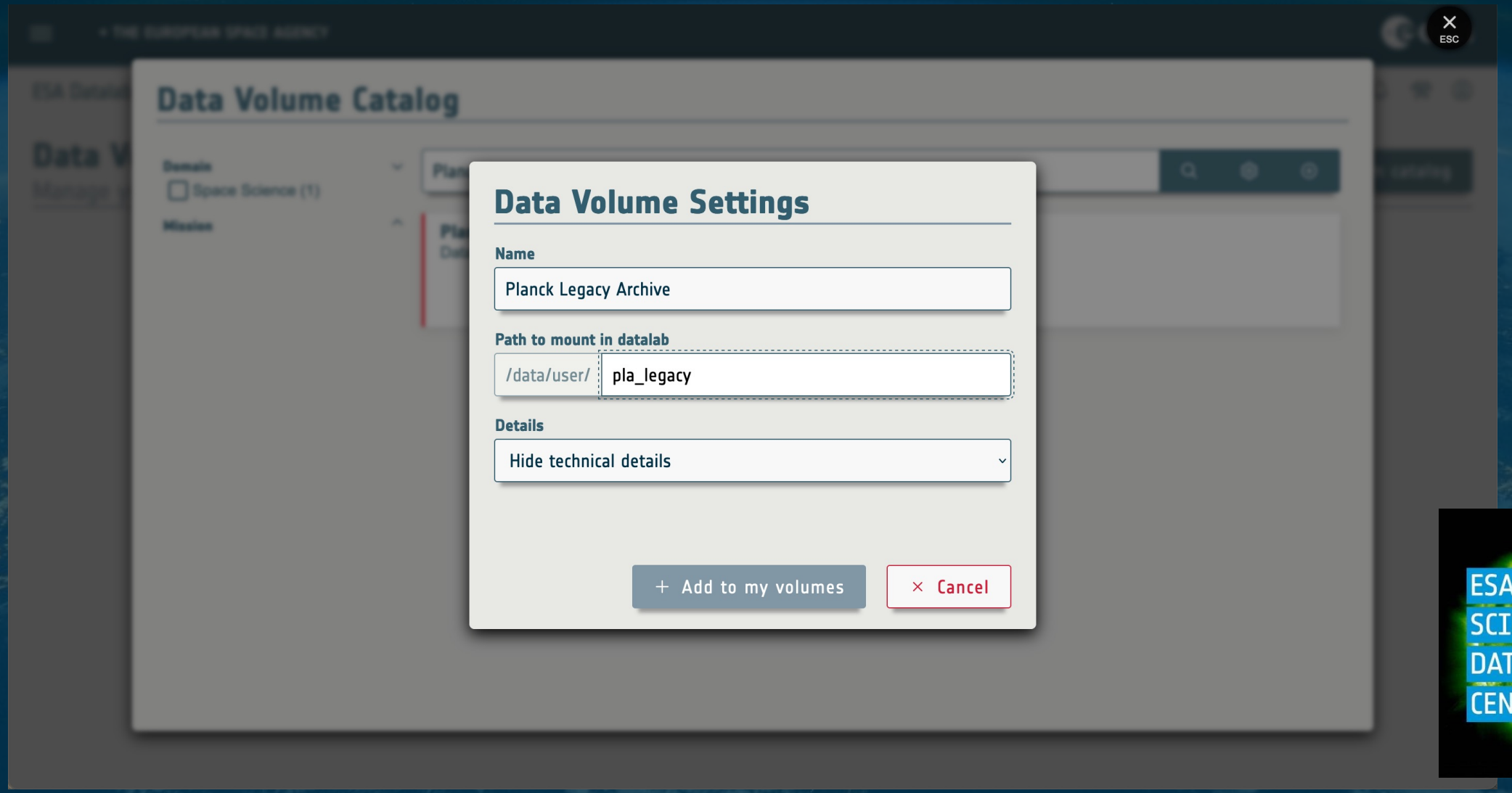

```
==================================================
PDF saved as: test_hdb_minsamp_1_prob_0.8_mCls_20.pdf
==================================================
```

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

SWITCH

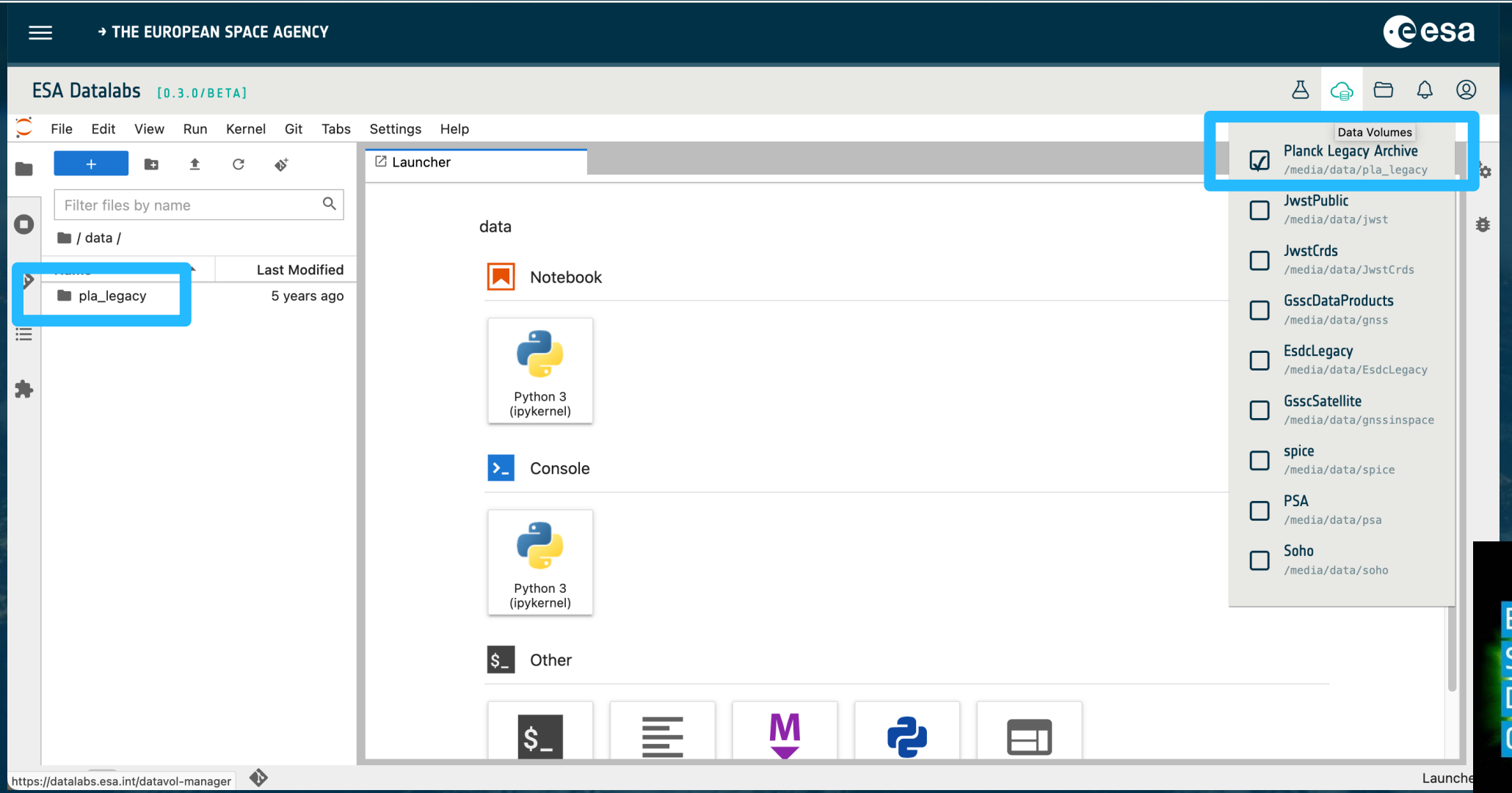

→ THE EUROPEAN SPACE AGENCY
esa
ESA Datalabs [0.3.0/BETA]
Datalabs
Datalabs
Juice
Aladin Dev
GSSC gLab
JWST Miricle
ESASky
Octave
+ LAUNCH NEW DATALAB
Juice
jl-juice
DELETE
Aladin Dev
aladin
DELETE
GSS
x-glab
DELETE
JWST Miricle
jwst-miricle
DELETE
ESASky
jl-esdc
DELETE
Octave
x-octave
DELETE

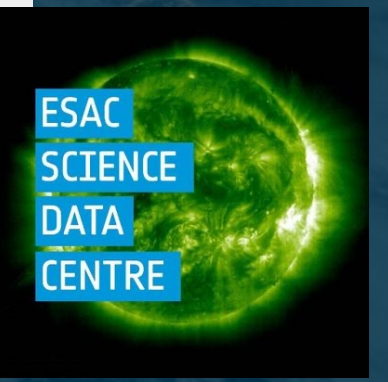
Data and Computing Colocation – Data Volumes



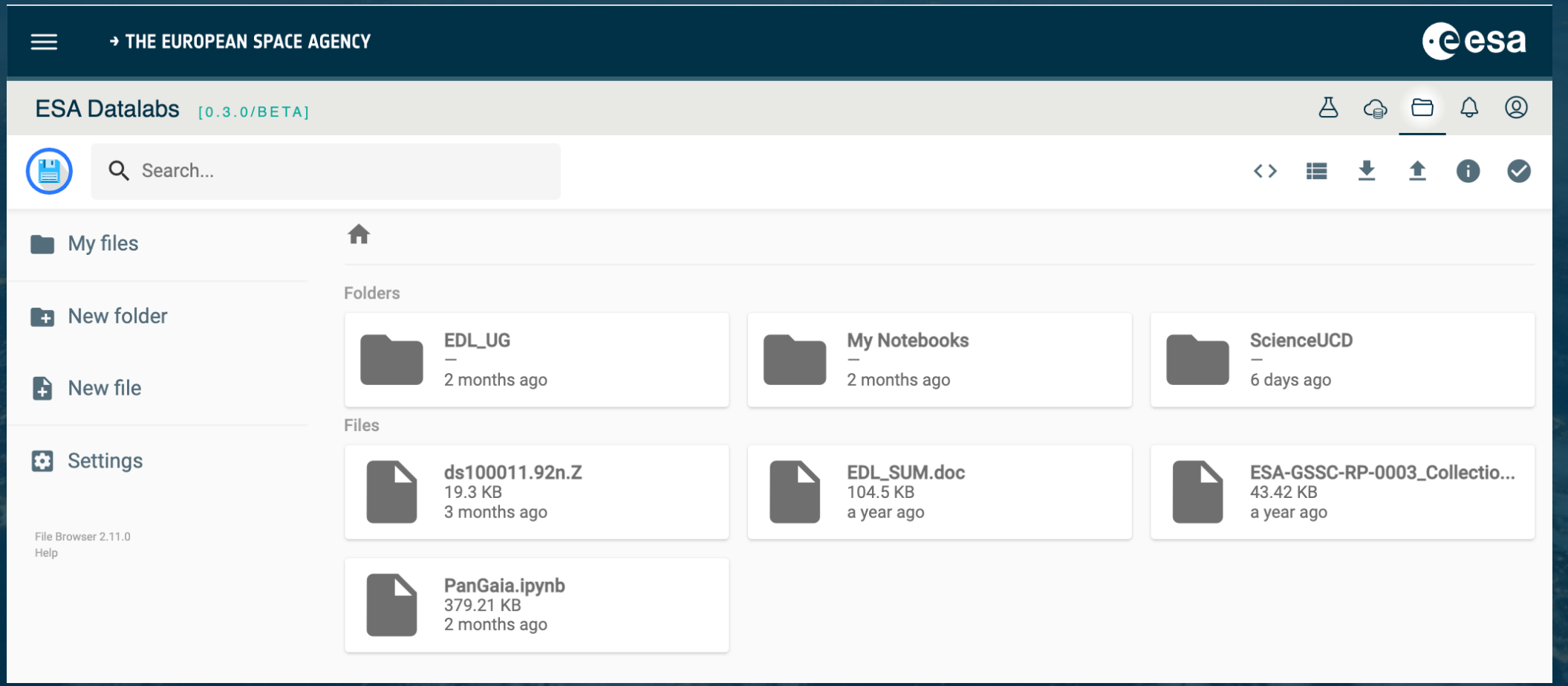
ESAC SCIENCE DATA CENTRE

Data and Computing Colocation – Data Volumes



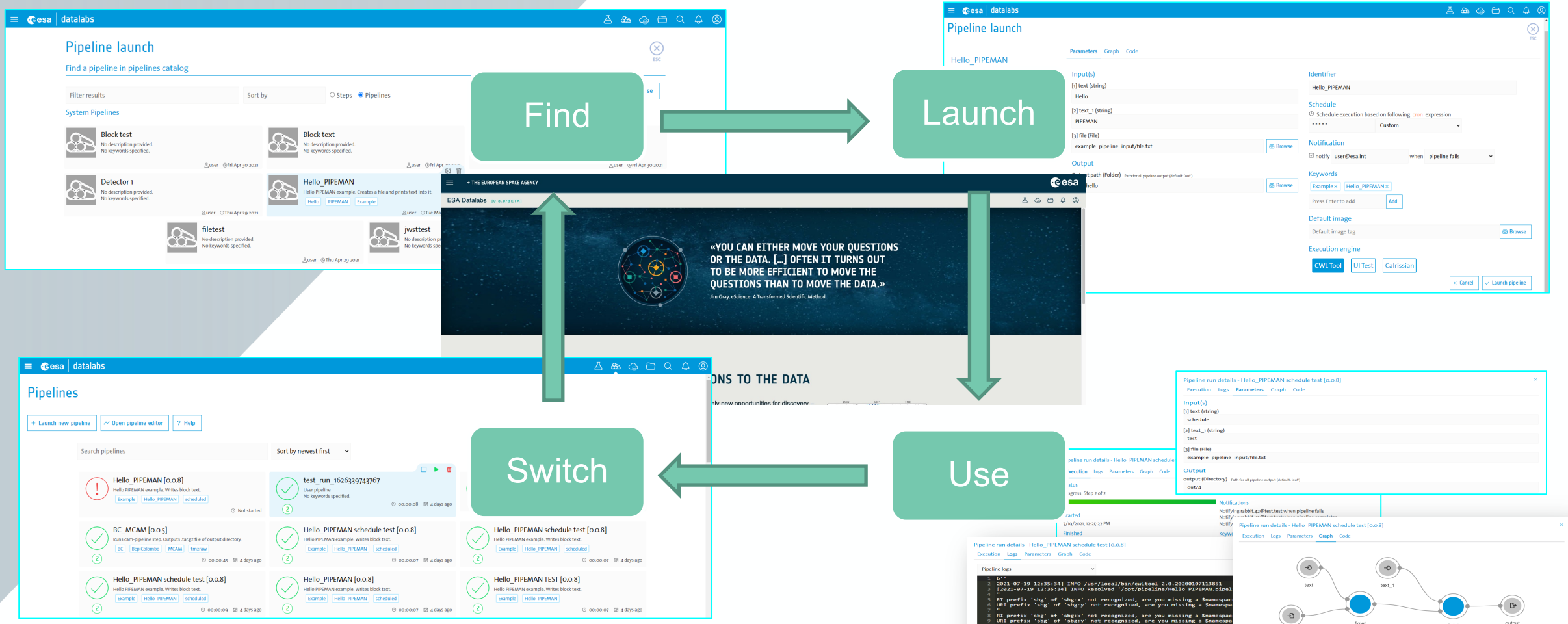


Data and Computing Colocation – Personal Workspace



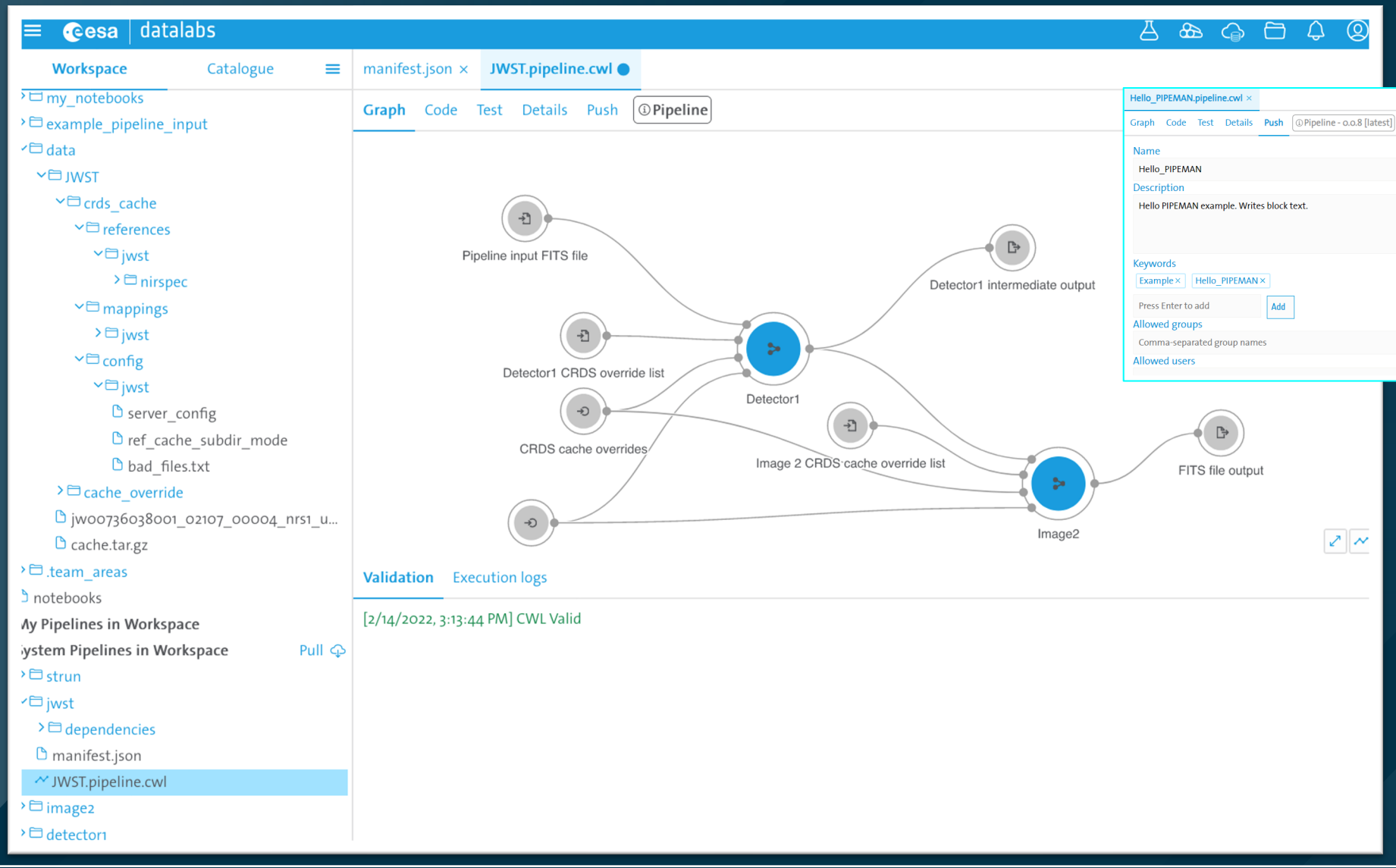
Data Processing Innovation

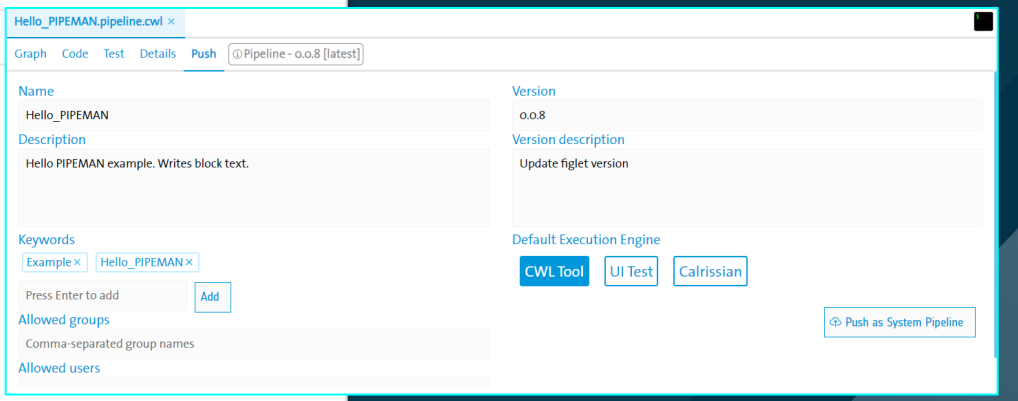
Pipelines – Standard Flow



17

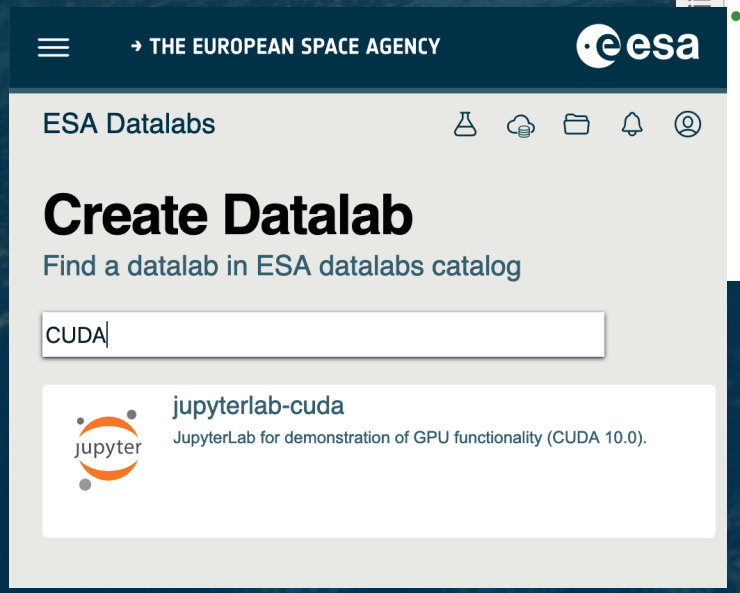
Pipeline Editor

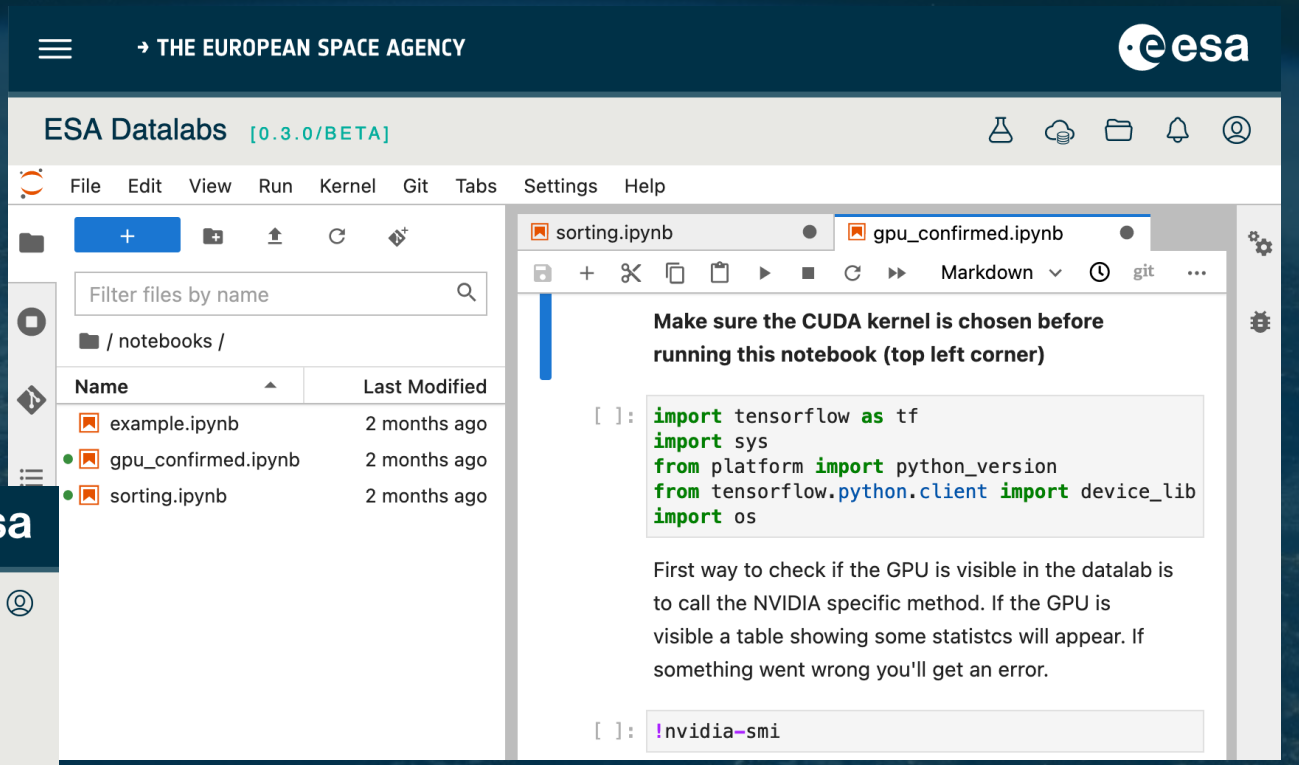




JWST @ ESA Datalabs

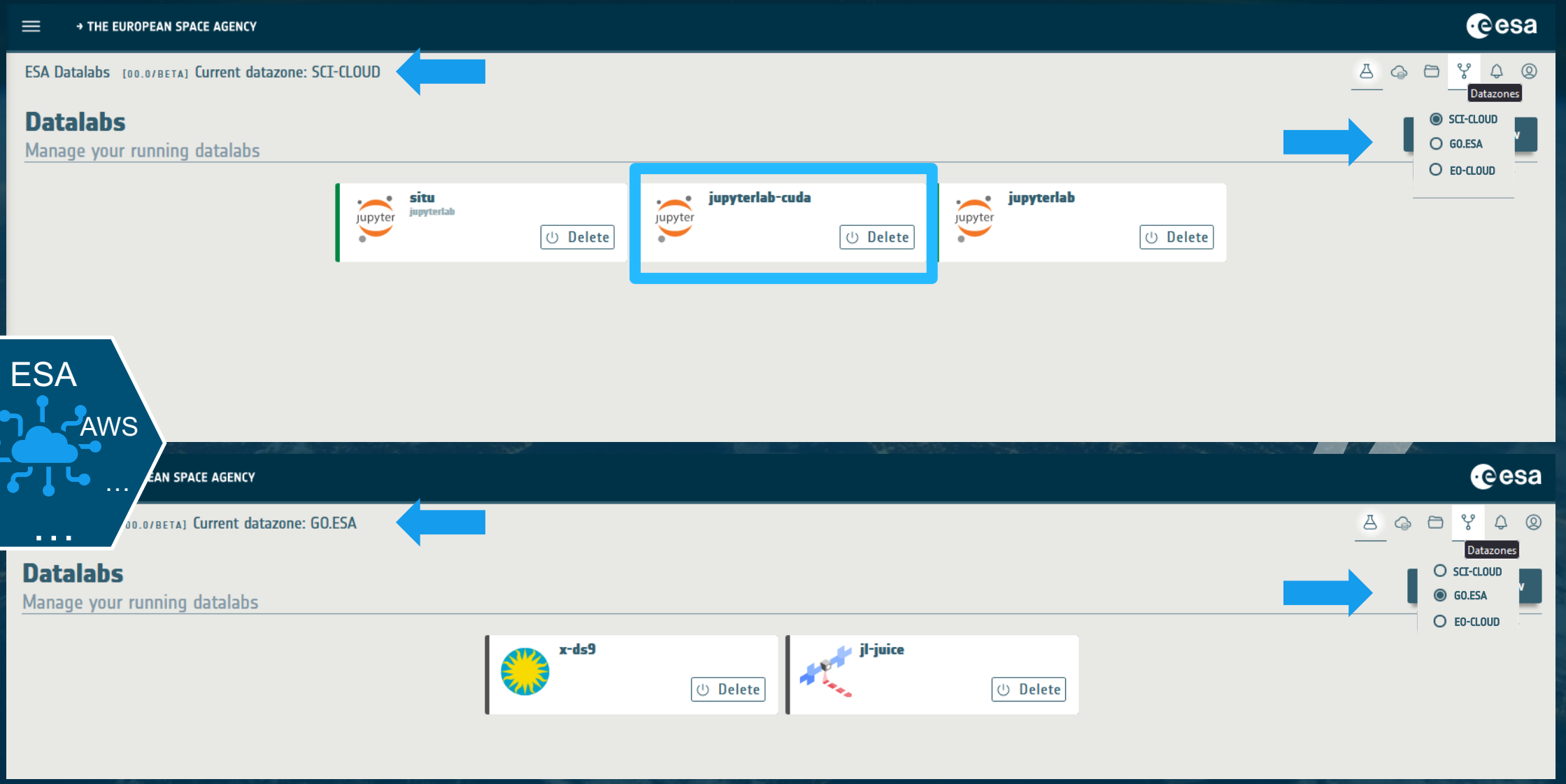
GPU Computing





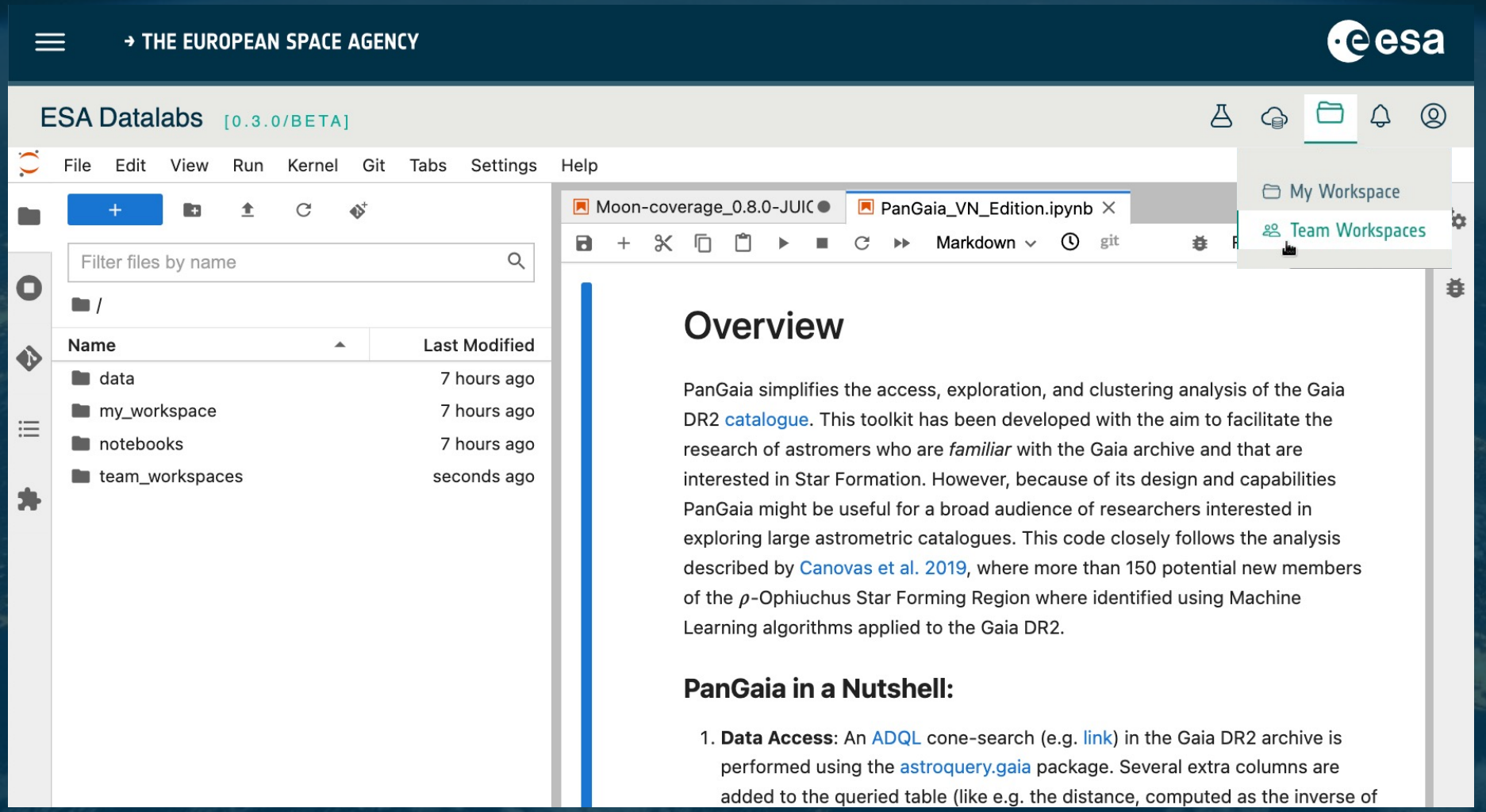
Time measured	numpy + no gpu	numpy + gpu	cupy + gpu
Total execution time	≈ 44m47s	≈ 14min26s	≈ 3min2s

Federated Scalability

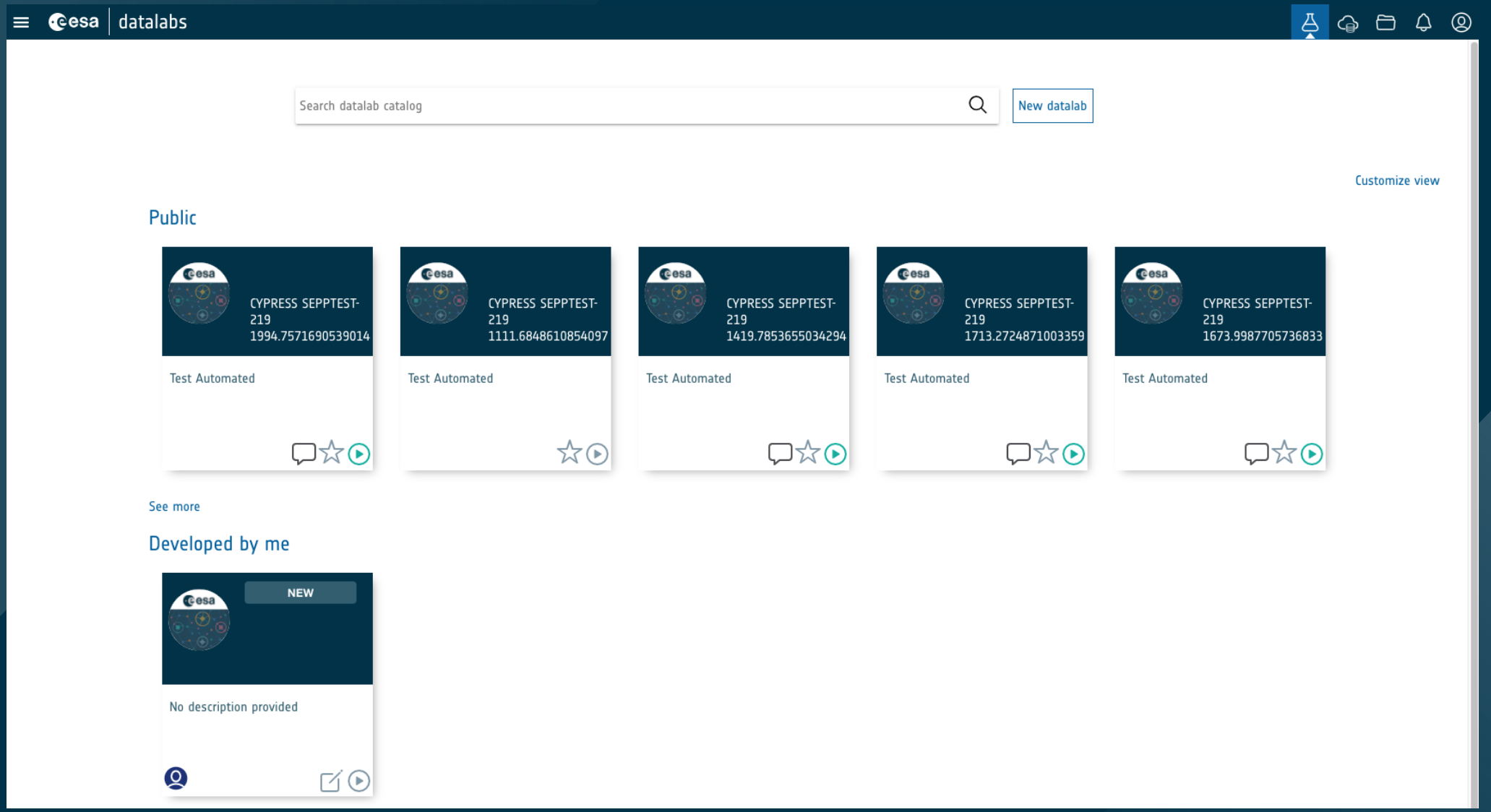


Collaborative Innovation

Data Centric Collaboration - Team Workspaces



Application Centric Collaboration - AppStore



Interoperability

Datalab Metadata Specification

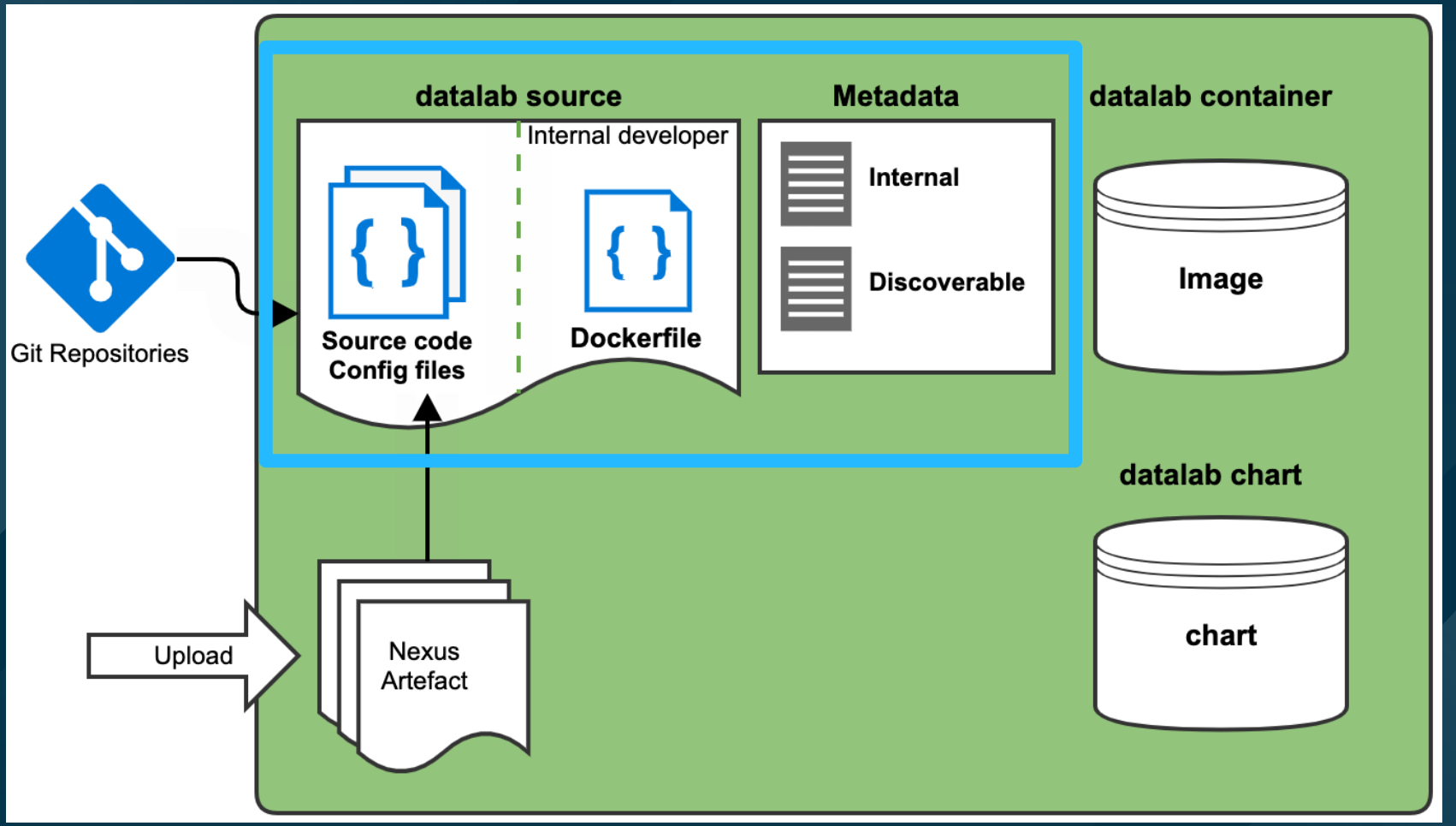
ESA DDP Term	Namespace:Term	Description	Visibility (private / public)	Write access (developer / moderator / sciapps)	Data type
abstract	sdo:abstract	Abstract (a short description)	PUBLIC	DEVELOPER	Text
alternateName	sdo:alternateName (pipeman: displayName)	Short name or acronym for the datalab	PUBLIC	DEVELOPER	Text
associatedFileType	esado:associatedFileType (new)	List of associated filetypes - if applicable/pertinent (e.g. FITS, VOTable, GeoTIFF, netCDF)	PUBLIC	DEVELOPER	Text
ESAOfficial	esado:ESAOfficial	flag indicating that the datalab was created by ESA	PUBLIC	DEVELOPER	Boolean
audienceType	sdo:audienceType	Intended audience for the datalab	PUBLIC	DEVELOPER	Text
citation	sdo:citation	Citation for the datalab (e.g. article DOI)	PUBLIC	DEVELOPER	Text
datalabImage.platform	sdo:version	ESA datalabs platform version	PRIVATE	APPLICATIO...	Text
datalabImage.id	sdo:identifier	Identifier of the datalab image	PRIVATE	APPLICATIO...	Text
datalabImage.keyword	dcat:keyword	Keyword tag of the current datalab image	PRIVATE	DEVELOPER	Text
dateCreated	sdo:dateCreated	Date the datalab was created	PRIVATE	APPLICATIO...	Date

PUBLIC

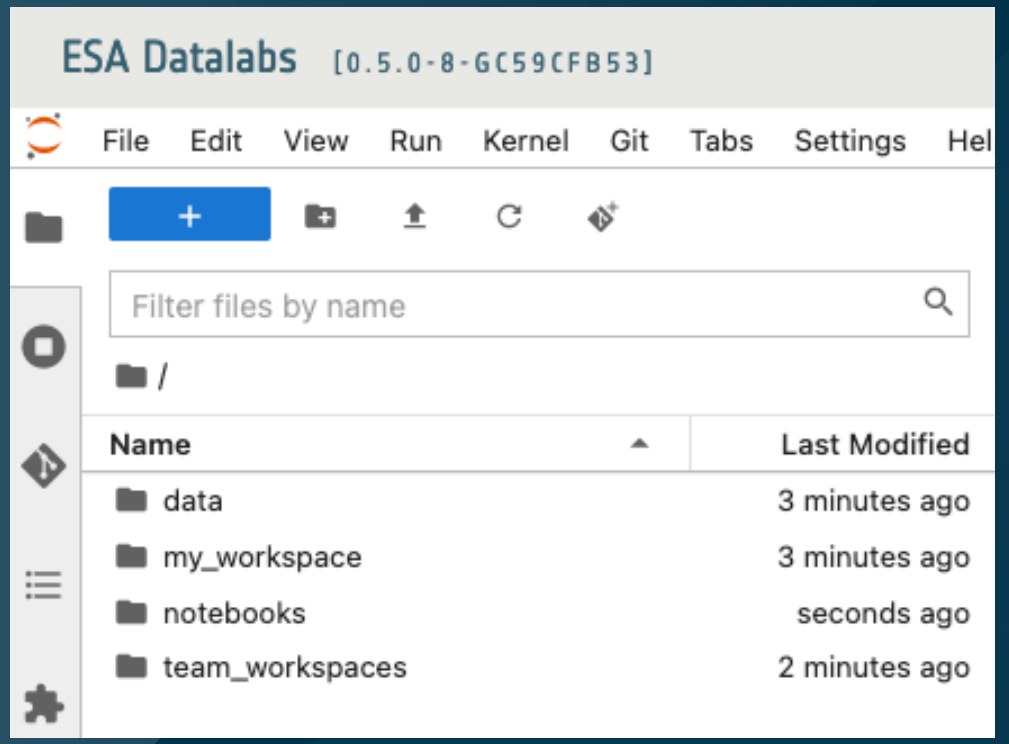
PRIVATE

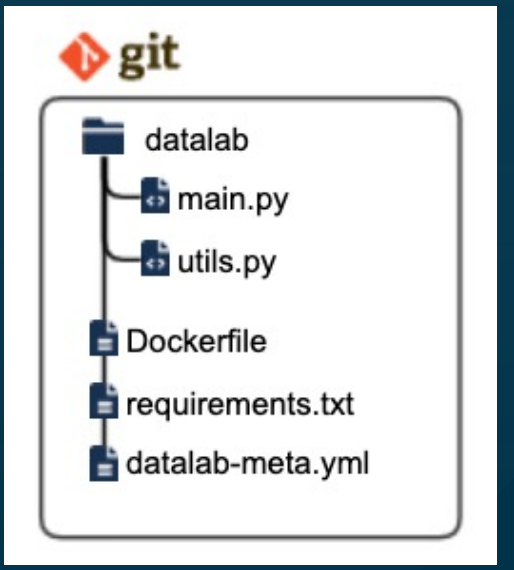
Leveraging on existing vocabularies and ontologies: sdo, dcat, skos, foaf … esado

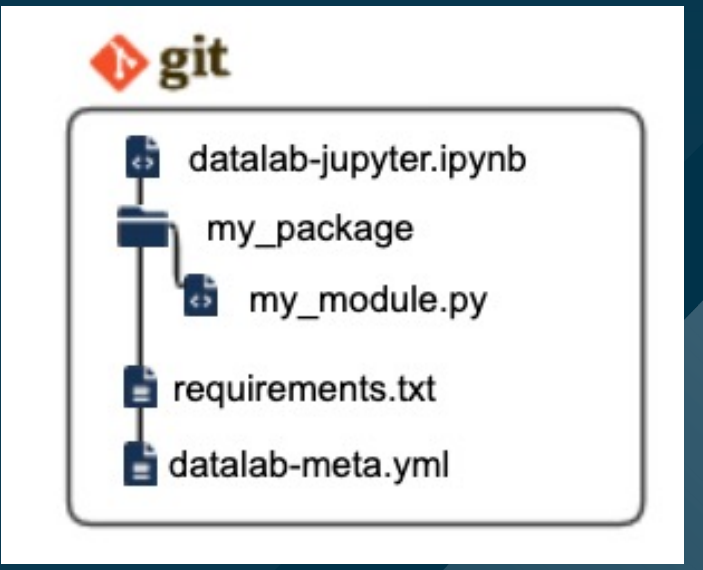
Datalab Package Specification



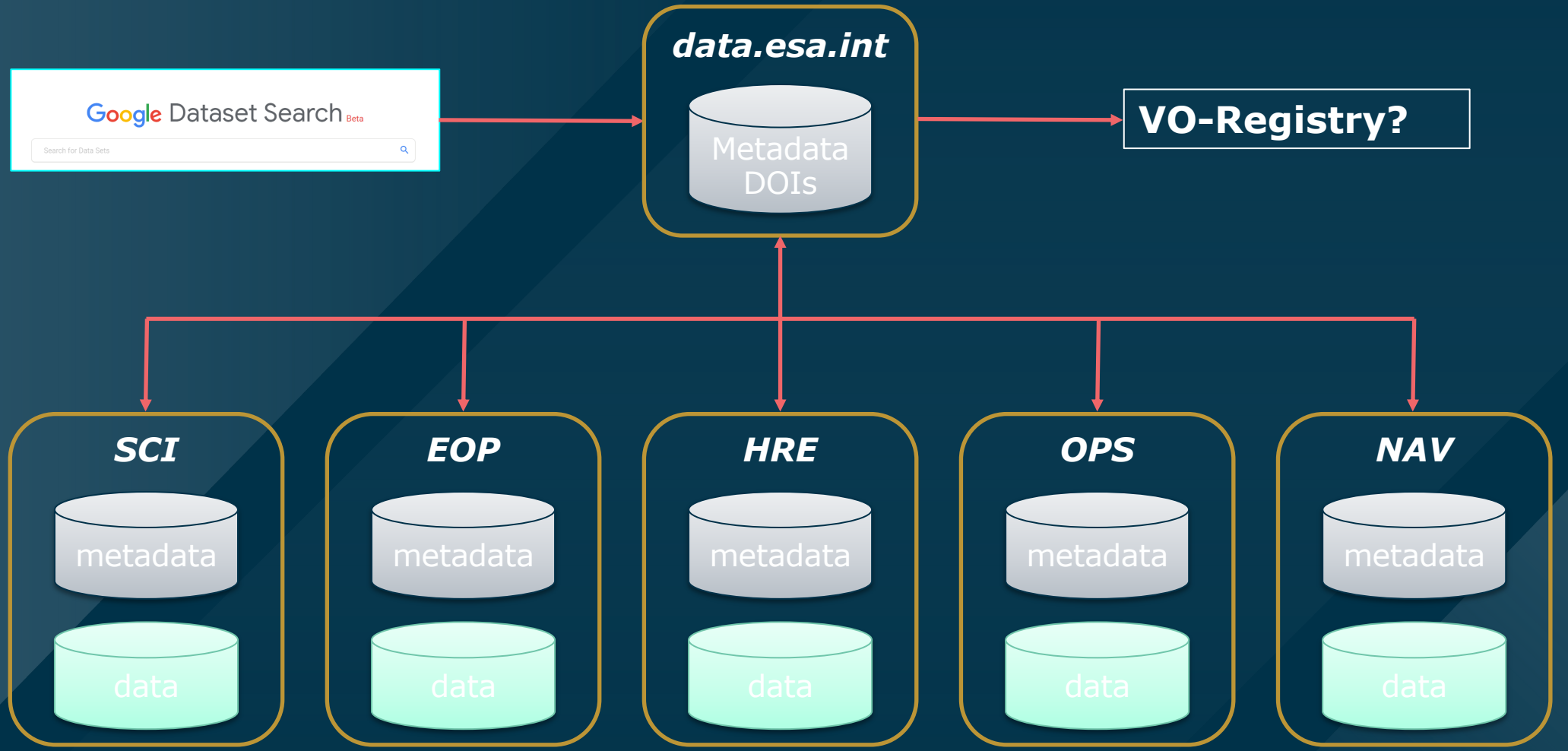
Datalab Structure Specification (development & runtime)



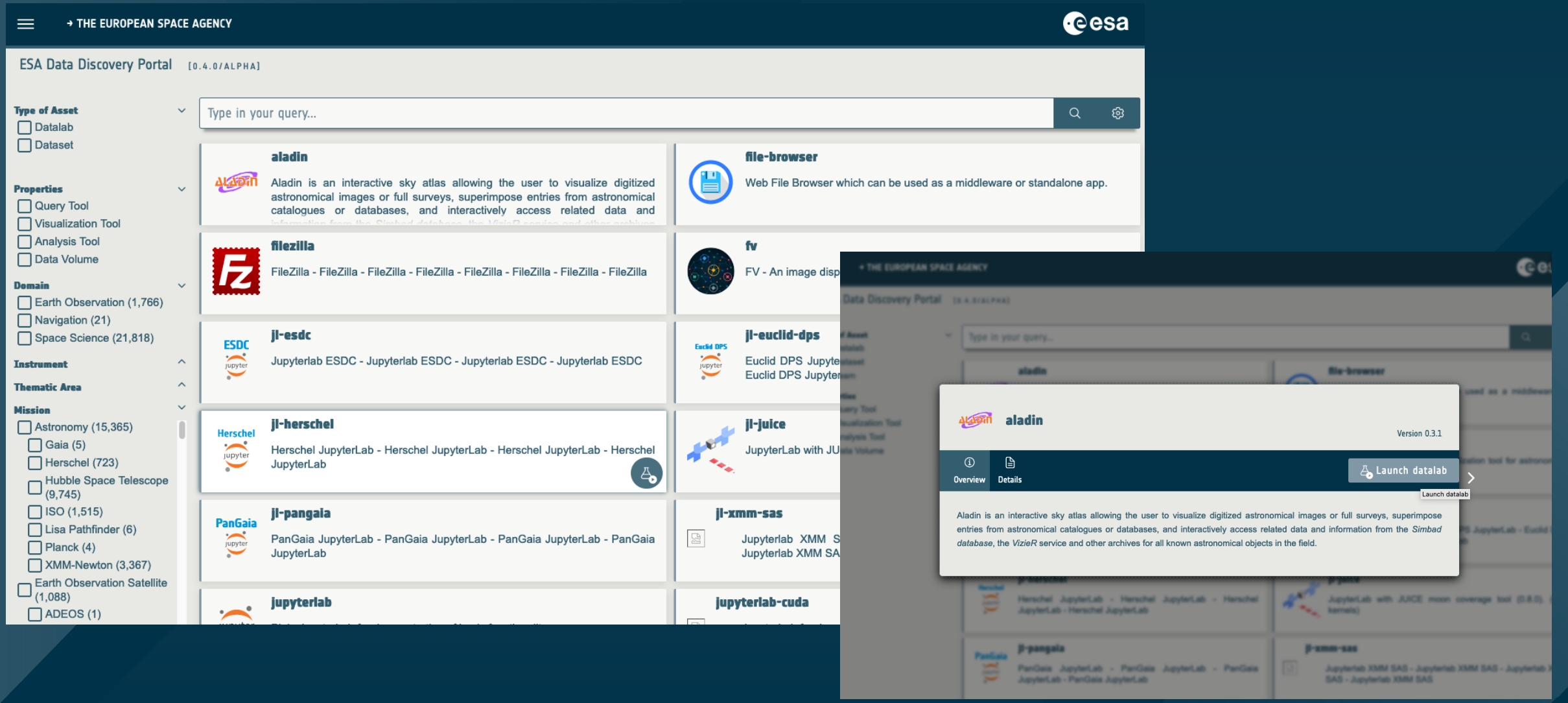




Metadata environment



Exploitation environment



Take away messages

ESA Science 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