

ESA Datalabs

Vicente Navarro, on behalf of the ESA Datalabs Team

IVOA Interop

18.10.2022, Virtual



ESA Datalabs – Across Science Domains



THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

«YOU CAN EITHER MOVE YOUR QUESTIONS OR THE DATA. [...] OFTEN IT TURNS OUT TO BE MORE EFFICIENT TO MOVE THE QUESTIONS THAN TO MOVE THE DATA.»
Jim Gray, eScience: A Transformed Scientific Method

BRING YOUR QUESTIONS TO THE DATA

There is a new paradigm, opening completely new opportunities for discovery – a data-intensive approach to science. In many domains, we have entered what could be called the golden age of surveys, with several large-scale projects, spanning decades, between finished, ongoing, and planned activities. ESA is responsible, or is a major partner, in several of these initiatives.

There is, however, a new profound change: data has become a major technological challenge. Increases by multiple orders of magnitude in dataset size means that transferring data to a scientist is often unfeasible.

ESA datalabs gives you a privileged position; bring your code directly to ESA's infrastructure – there is a great set of tools and programming languages are flexible – and execute it with direct access to ESA's archives.

THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

Create Datalab

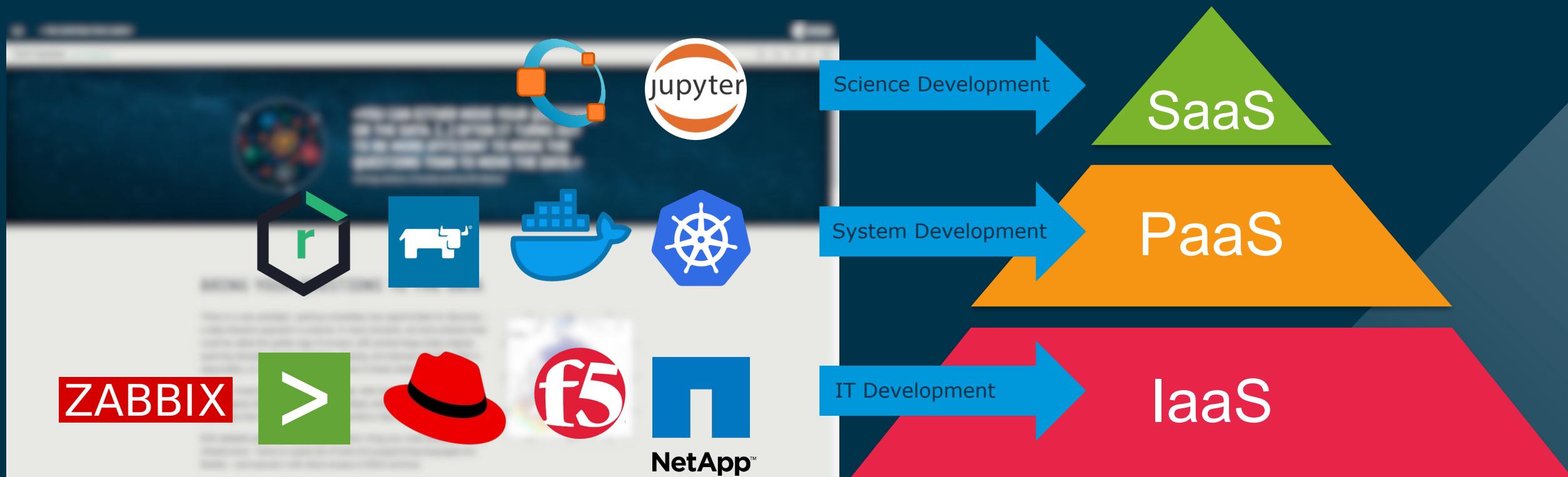
Find a datalab in ESA datalabs catalog

Filter results

| Project Name | Description | Tool |
|---------------|--|------|
| 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 other archives for all known astronomical objects in the field. | |
| 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 | |
| jupyterlab | Plain JupyterLab for demonstration of basic functionality. | |
| jl-juice | JupyterLab with JUICE moon coverage tool (0.8.0). | |
| jl-pangaia | PanGaea JupyterLab | |
| jwst | Jupyterlab JWST | |
| jwst-miricle | Jupyterlab JWST Miricle | |
| jwst-nsrc | Jupyterlab JWST NSRT | |
| qfitsview | QFitsView - An image display and visualization tool for astronomical data | |
| theia-python | Theia Python Editor | |



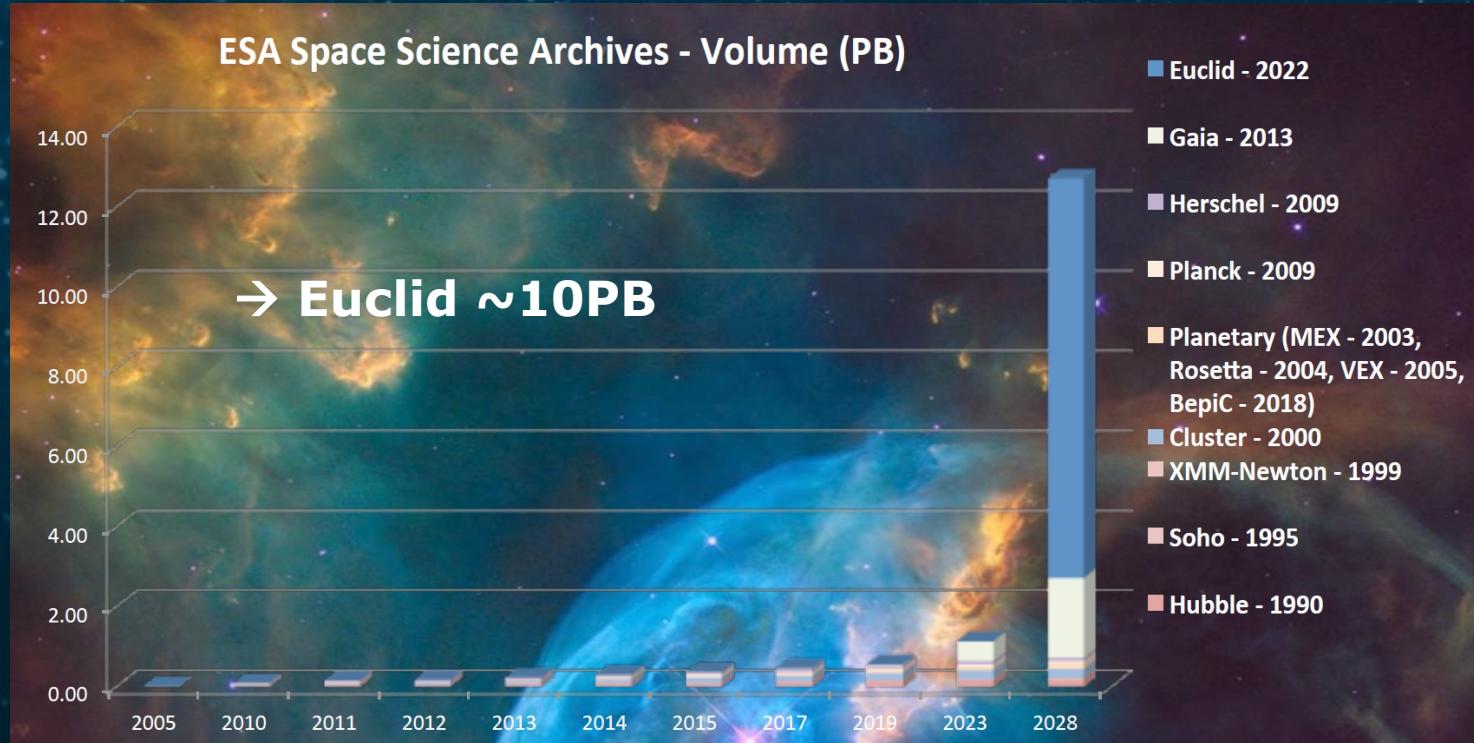
ESA Datalabs – Across IT Domains



Data Access Innovation

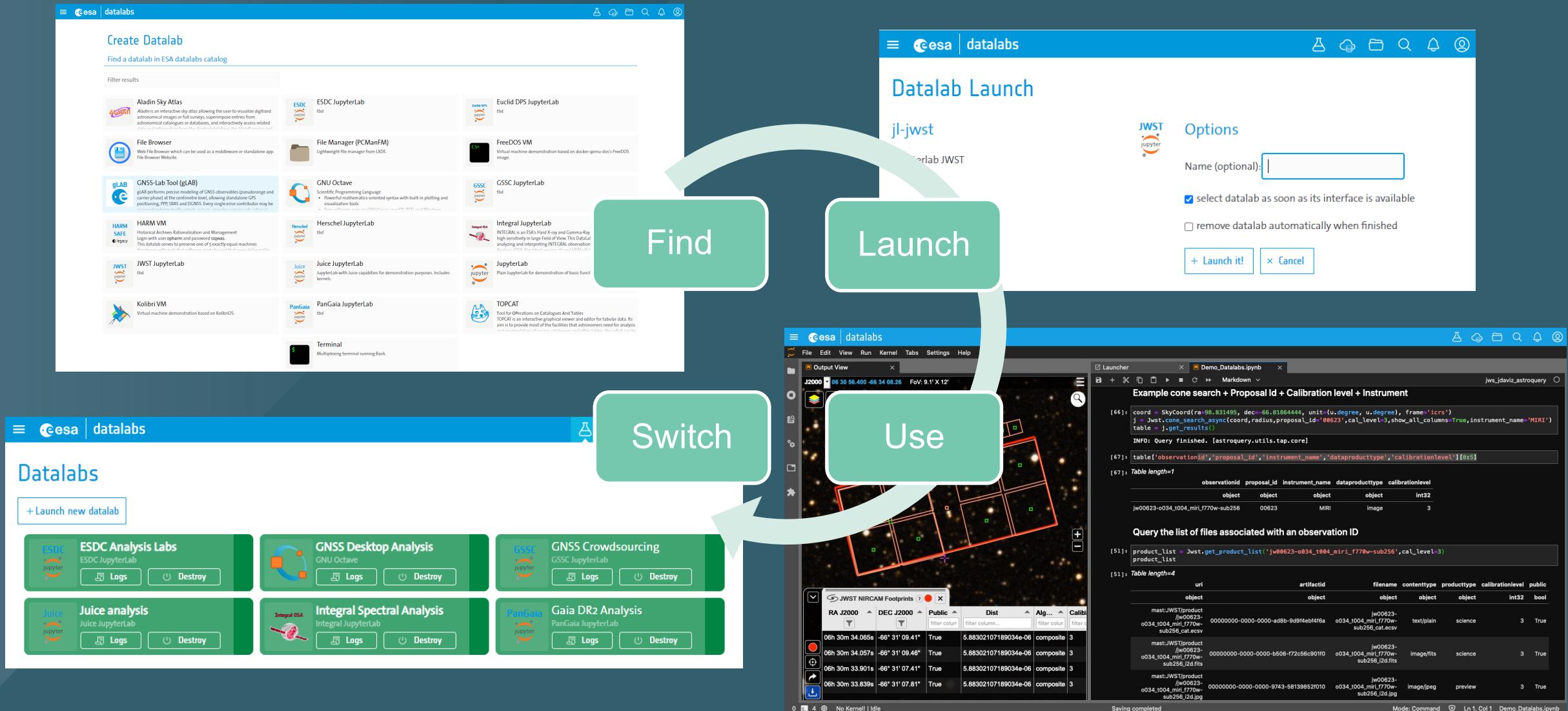


Data Exploitation Paradigm Shift



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

Main Scenario / Use Case



THE EUROPEAN SPACE AGENCY

esa

ESA Datalabs [0.3.0/BETA]

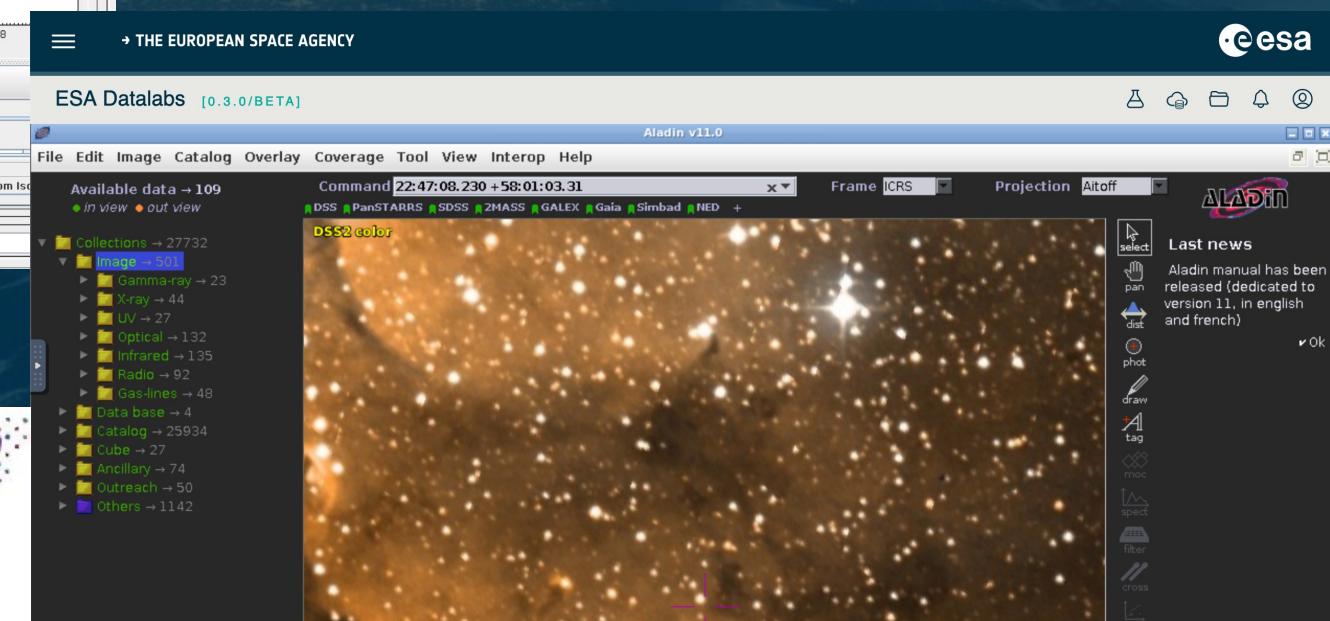
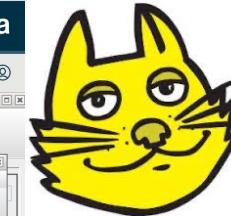
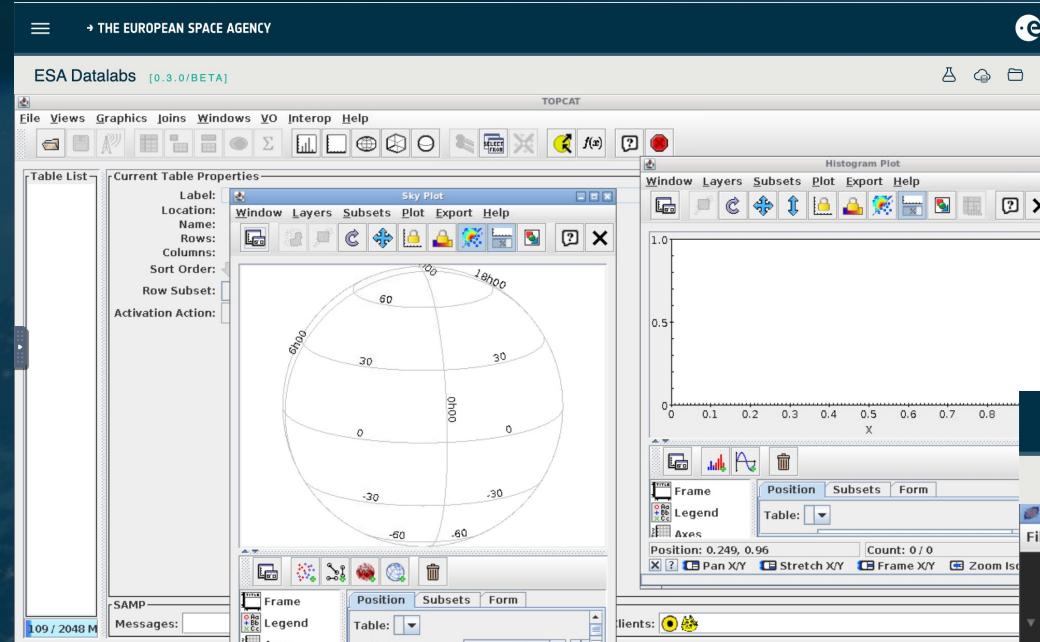
Filter results

Create Datalab

Find a datalab in ESA datalabs catalog

| | | |
|---|---|---|
|  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 <i>Simbad database</i> , the <i>VizieR</i> service and other archives for all known astronomical objects in the field |  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)



The screenshot shows the ESA Datalabs interface. On the left, a sidebar displays a file tree with notebooks, including 'juice_kernels' and 'Moon-coverage_0.8.0-JUICE_examples'. The main area shows a Jupyter Notebook titled 'Moon-coverage 0.8.0 - JUICE examples'. The notebook content includes a diagram of a satellite in space, text about the moon-coverage package, and a code cell that prints the current version of the tool.

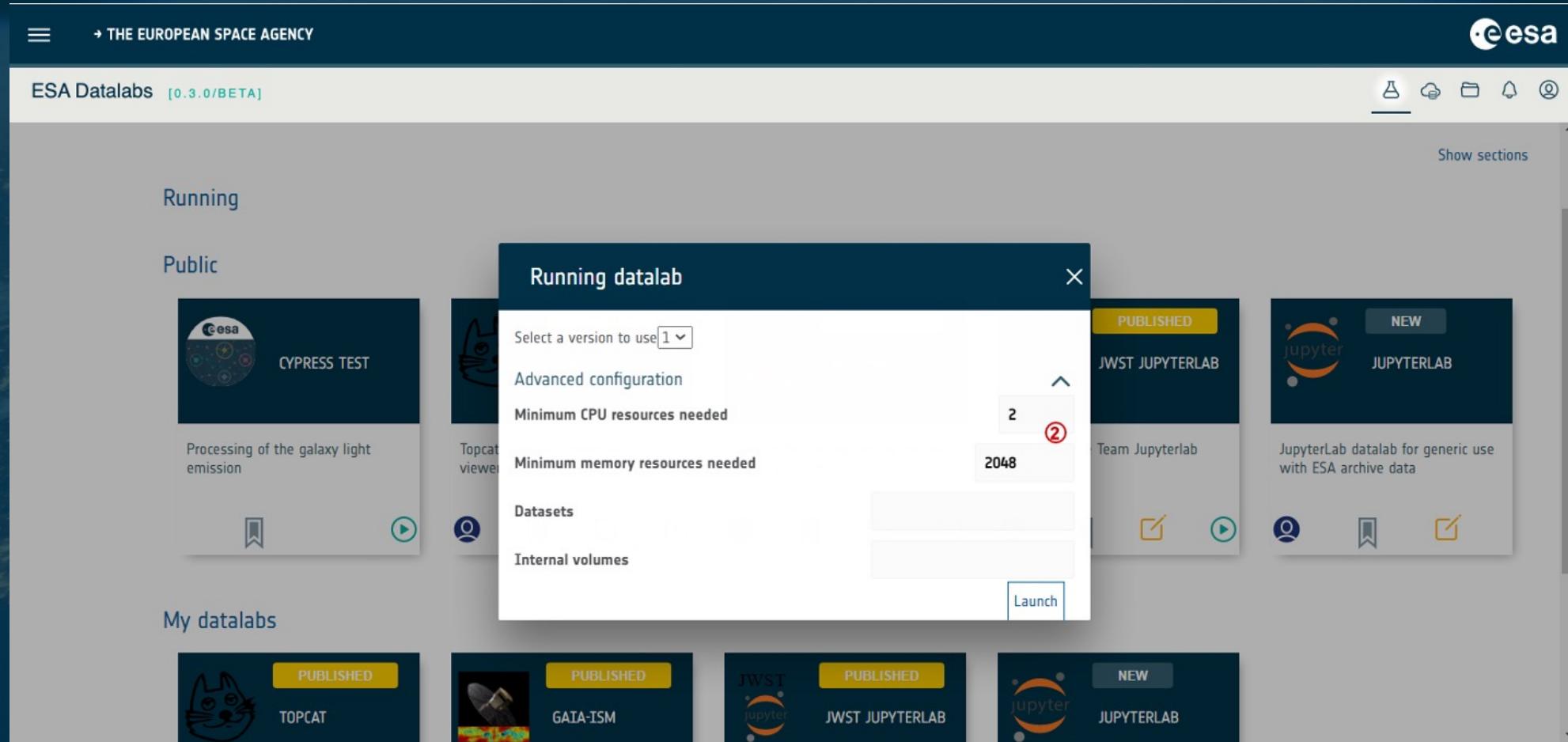
```
[1]: from moon_coverage import __version__

if __version__.startswith('0.8.'):
    print(f'Your current version of the moon-coverage tool is {__version__}.')
else:
    raise ImportError(
        f'Your version of the moon-coverage tool is {__version__}. '
        'The examples below require the version 0.8.0 and may not work as expected.' )
```

The screenshot shows the ESA Datalabs interface with an 'Output View' tab selected. It displays a map of the sky with various green rectangular regions representing observation footprints. Below the map, there are several tabs for different instruments: 'test footprints', 'HST', 'LAMOST', and 'HST'. A code cell in the bottom right shows an example query for JWST data, specifically for the F1130W filter.

```
[48]: j2a = Jwst.cone_search(coord, radius, cal_level=-1, filter_name='F1130W')
table2a = j2a.get_results()
print(table2a.columns)

<TableColumns names='dist', 'observationid', 'calibrationlevel', 'public', 'dataproducttype', 'instrument_name', 'energy_bandpassname', 'target_name', 'target_ra', 'target_dec', 'position_bounds_center', 'position_bounds_spoly'>
observationid      dataproducttype calibrationlevel public
jw01865001001_xx107_00012_miri      image      -1 False
jw01865001001_xx105_00001_miri      image      -1 False
jw01865001001_xx105_00002_miri      image      -1 False
jw01865001001_xx105_00003_miri      image      -1 False
jw01865001001_xx105_00004_miri      image      -1 False
jw01865001001_xx105_00005_miri      image      -1 False
jw01865001001_xx105_00006_miri      image      -1 False
jw01865001001_xx107_00011_miri      image      -1 False
jw01865001001_xx108_00019_miri      image      -1 False
jw01865001001_xx108_00020_miri      image      -1 False
jw02368007001_xx102_00002_miri      image      -1 False
jw02368007001_xx103_00003_miri      image      -1 False
jw02368007001_xx104_00004_miri      image      -1 False
jw02368007001_xx105_00001_miri      image      -1 False
jw02368007001_xx105_00002_miri      image      -1 False
jw02368007001_xx107_00003_miri      image      -1 False
jw02368007001_xx108_00009_miri      image      -1 False
jw02368007001_xx109_00008_miri      image      -1 False
jw02368007001_xx108_00007_miri      image      -1 False
jw02368007001_xx109_00003_miri      image      -1 False
jw02368007001_xx109_00004_miri      image      -1 False
Length = 32 rows
```



The screenshot shows the ESA Datalabs interface. At the top, there's a header with the ESA logo and navigation links. Below the header, the main area is divided into sections: "Running" (which is currently active), "Public", and "Mydatalabs".

The "Running" section displays a "CYPRESS TEST" datalab card. The "Public" section shows cards for "TOPCAT", "GAIA-ISM", and "JWST JUPYTERLAB". The "Mydatalabs" section shows cards for "JUPYTERLAB" and "JUPYTERLAB" (marked as NEW).

A modal window titled "Running datalab" is open in the center. It contains fields for "Select a version to use" (set to 1), "Advanced configuration", "Minimum CPU resources needed" (set to 2048), and "Minimum memory resources needed" (set to 2048). It also has sections for "Datasets" and "Internal volumes". A "Launch" button is at the bottom right of the modal.

THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

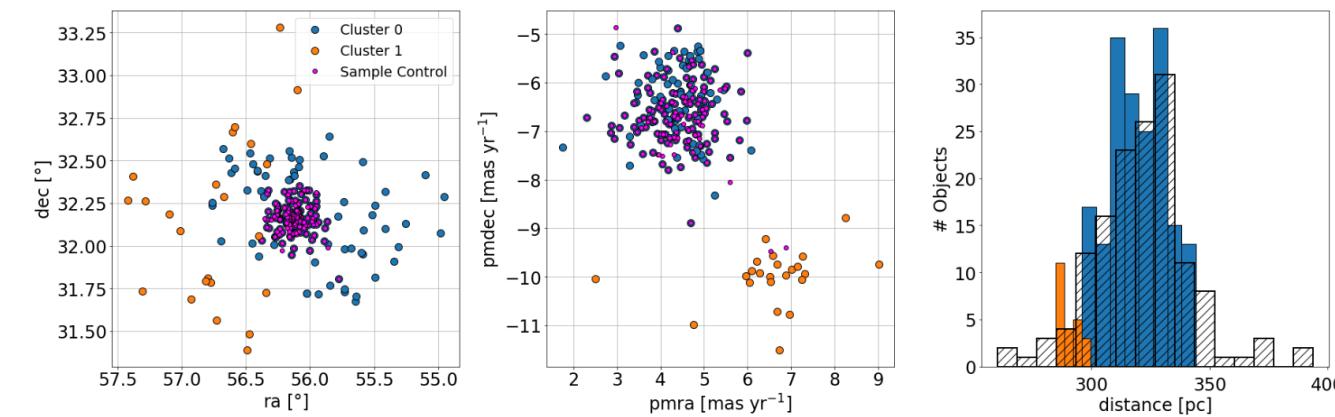
File Edit View Run Kernel Git Tabs Settings Help

Launcher Moon-coverage_0.8.0-JUIC PanGaia_VN_Edition.ipynb Python 3 (ipykernel)

[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
=====

Mode: Command Ln 1, Col 1 PanGaia_VN_Edition.ipynb

☰ → THE EUROPEAN SPACE AGENCY

esa

ESA Datalabs [0.3.0/BETA]

Datalabs

+ LAUNCH NEW DATALAB

 **Juice**
jl-juice


 **Aladin Dev**
aladin


 **GSSC gLab**
x-glab


 **JWST Miricle**
jwst-miricle


 **ESDC jupyter**
jl-esdc


 **Octave**
x-octave


☰ Datalabs

- Juice
- Aladin Dev
- GSSC gLab
- JWST Miricle
- ESASky
- Octave

12

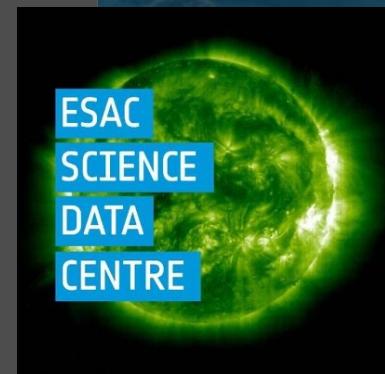
Data and Computing Colocation – Data Volumes



The screenshot shows a modal dialog titled "Data Volume Settings" overlaid on a "Data Volume Catalog" interface. The modal contains the following fields:

- Name:** Planck Legacy Archive
- Path to mount in databab:** /data/user/ pla_legacy
- Details:** A dropdown menu showing "Hide technical details".

At the bottom of the modal are two buttons: "+ Add to my volumes" (blue) and "Cancel" (red).



Data and Computing Colocation – Data Volumes



THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

File Edit View Run Kernel Git Tabs Settings Help

+ Filter files by name / data /

pla_legacy 5 years ago

Launcher

data

Notebook

Python 3 (ipykernel)

Console

Python 3 (ipykernel)

Other

\$ _ M Python 3 (ipykernel) \$ _

Data Volumes

- Planck Legacy Archive /media/data/pla_legacy
- JwstPublic /media/data/jwst
- JwstCrdz /media/data/JwstCrdz
- GsscDataProducts /media/data/gnss
- EsdcLegacy /media/data/EsdcLegacy
- GsscSatellite /media/data/gnssinspace
- spice /media/data/spice
- PSA /media/data/psa
- Soho /media/data/soho

https://datalabs.esa.int/datavol-manager

Launcher

The screenshot shows the ESA Datalabs interface version 0.3.0/BETA. On the left, there's a file browser with a sidebar for filtering files by name. A specific file named 'pla_legacy' is highlighted with a blue box. In the center, there's a launcher with sections for Notebook, Console, and Other, each containing a Python 3 (ipykernel) icon. On the right, a sidebar titled 'Data Volumes' lists several data volumes with checkboxes: Planck Legacy Archive (checked), JwstPublic, JwstCrdz, GsscDataProducts, EsdcLegacy, GsscSatellite, spice, PSA, and Soho. A blue box highlights the 'Planck Legacy Archive' entry. At the bottom, there's a URL 'https://datalabs.esa.int/datavol-manager' and a 'Launcher' button.



Data and Computing Colocation – Personal Workspace



THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

My files

New folder

New file

Settings

File Browser 2.11.0

Help

Search...

EDL_UG
2 months ago

My Notebooks
2 months ago

ScienceUCD
6 days ago

ds100011.92n.Z
19.3 KB
3 months ago

EDL_SUM.doc
104.5 KB
a year ago

ESA-GSSC-RP-0003_Collectio...
43.42 KB
a year ago

PanGaia.ipynb
379.21 KB
2 months ago

Cloud

Folder

Bell

User

Home

Search

Back

Forward

Grid

Download

Upload

Info

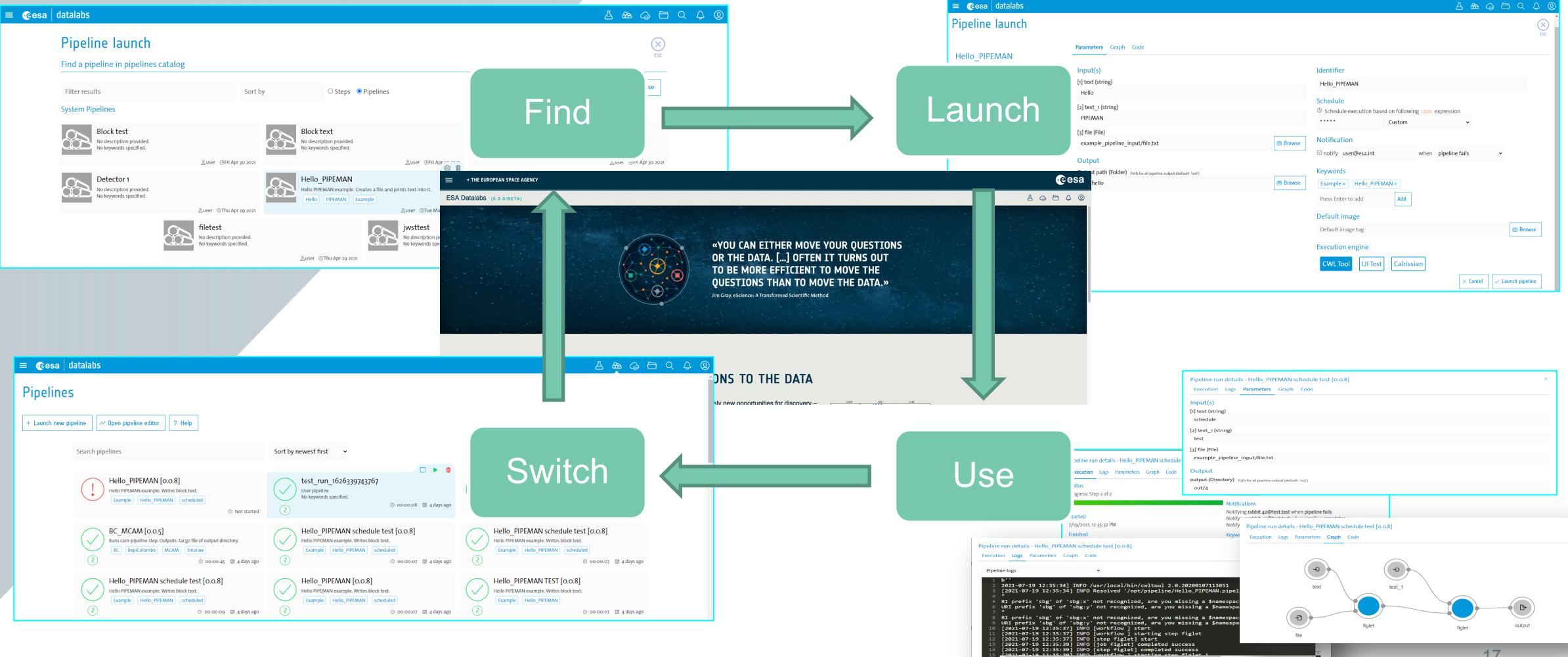
Checkmark



→ THE EUROPEAN SPACE AGENCY

Data Processing Innovation

Pipelines – Standard Flow



Pipeline Editor



esa | datalabs

Workspace Catalogue

manifest.json x JWST.pipeline.cwl ●

Graph Code Test Details Push Pipeline

Pipeline input FITS file

Detector1 intermediate output

Detector1 CRDS override list

CRDS cache overrides

Image 2 CRDS cache override list

FITS file output

Validation Execution logs

[2/14/2022, 3:13:44 PM] CWL Valid

my Pipelines in Workspace

system Pipelines in Workspace

JWST.pipeline.cwl

image2

detector1

Pipeline input FITS file

Detector1 intermediate output

Detector1 CRDS override list

CRDS cache overrides

Image 2 CRDS cache override list

FITS file output

Name: Hello_PIPEMAN

Description: Hello PIPEMAN example. Writes block text.

Version: 0.0.8

Default Execution Engine: CWL Tool

Allowed groups: Comma-separated group names

Allowed users:

Push as System Pipeline

The screenshot shows the Pipeline Editor interface with a manifest.json file selected. The main area displays a graph of nodes: Pipeline input FITS file, Detector1, Image2, and various override lists. The right panel shows the configuration for the 'Hello_PIPEMAN' pipeline, including fields for Name, Description, Version, Default Execution Engine (CWL Tool), and Allowed groups. A 'Push as System Pipeline' button is also visible. The bottom left shows a list of pipelines in the workspace, with 'JWST.pipeline.cwl' highlighted. The bottom right corner features the 'JWST @ ESA Datalabs' logo.

GPU Computing



The screenshot displays the ESA Datalabs interface, version 0.3.0/BETA, running on a JupyterLab environment. The top navigation bar includes the ESA logo and links to File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. The left sidebar shows a file tree under /notebooks/ with three files: example.ipynb, gpu_confirmed.ipynb (selected), and sorting.ipynb, all modified 2 months ago. The main area contains a code cell with Python code to check GPU visibility:

```
import tensorflow as tf
import sys
from platform import python_version
from tensorflow.python.client import device_lib
import os
```

A note in the cell states: "Make sure the CUDA kernel is chosen before running this notebook (top left corner)". Below the cell, a message explains how to check GPU visibility using the !nvidia-smi command.

On the right, a table compares execution times for different scenarios:

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

The bottom left shows a search bar for "CUDA" and a card for "jupyterlab-cuda" which is a JupyterLab for demonstration GPU functionality (CUDA 10.0). The bottom right shows the ESA footer with flags of member states and links to the European Space Agency.

Federated Scalability



THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.0.0/BETA] Current datazone: SCI-CLOUD

Datalabs Manage your runningdatalabs

jupyterlab-cuda

jupyterlab

Delete Delete

Datazones

- SCI-CLOUD
- GO.ESA
- EO-CLOUD



THE EUROPEAN SPACE AGENCY

0.0.0/BETA Current datazone: GO.ESA

Datalabs Manage your runningdatalabs

x-dsg

jl-juice

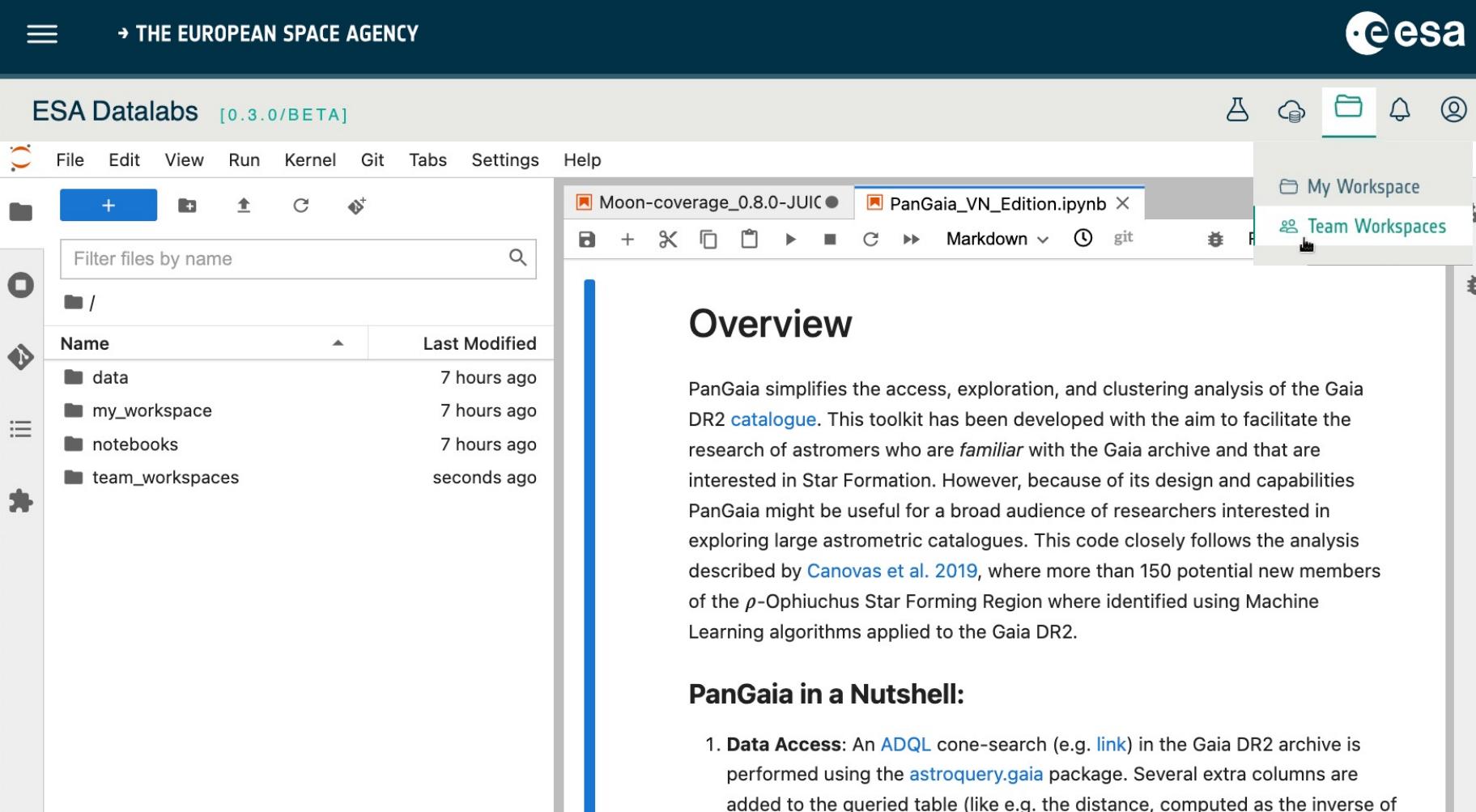
Delete Delete

Datazones

- SCI-CLOUD
- GO.ESA
- EO-CLOUD

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et ad officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum hic tenetur a sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat.

Collaborative Innovation



The screenshot shows the ESA Datalabs interface version 0.3.0/BETA. The top navigation bar includes 'File', 'Edit', 'View', 'Run', 'Kernel', 'Git', 'Tabs', 'Settings', and 'Help'. The toolbar features icons for file operations like '+', up, down, and search, along with a 'Filter files by name' input field. The main area displays a file tree on the left with entries: 'data' (7 hours ago), 'my_workspace' (7 hours ago), 'notebooks' (7 hours ago), and 'team_workspaces' (seconds ago). On the right, two Jupyter notebook tabs are open: 'Moon-coverage_0.8.0-JUIC.ipynb' and 'PanGaia_VN_Edition.ipynb'. The 'Team Workspaces' tab is currently selected. Below the tabs, there's a toolbar with icons for file operations, a 'Markdown' dropdown, and a 'git' button. The central content area contains an 'Overview' section with text about PanGaia, followed by a 'PanGaia in a Nutshell:' section and a numbered list.

Overview

PanGaia simplifies the access, exploration, and clustering analysis of the Gaia DR2 catalogue. This toolkit has been developed with the aim to facilitate the research of astromers who are *familiar* with the Gaia archive and that are interested in Star Formation. However, because of its design and capabilities PanGaia might be useful for a broad audience of researchers interested in exploring large astrometric catalogues. This code closely follows the analysis described by [Canovas et al. 2019](#), where more than 150 potential new members of the ρ -Ophiuchus Star Forming Region were identified using Machine Learning algorithms applied to the Gaia DR2.

PanGaia in a Nutshell:

1. **Data Access:** An ADQL cone-search (e.g. [link](#)) in the Gaia DR2 archive is performed using the `astroquery.gaia` package. Several extra columns are added to the queried table (like e.g. the distance, computed as the inverse of

Application Centric Collaboration - AppStore



≡ e esa | datalabs

Search datalab catalog 🔍 New datalab Customize view

Public

CYPRESS SEPPTEST-219
1994.7571690539014

Test Automated

💬 ⭐ ▶

CYPRESS SEPPTEST-219
1111.6848610854097

Test Automated

⭐ ▶

CYPRESS SEPPTEST-219
1419.7853655034294

Test Automated

💬 ⭐ ▶

CYPRESS SEPPTEST-219
1713.2724871003359

Test Automated

💬 ⭐ ▶

CYPRESS SEPPTEST-219
1673.9987705736833

Test Automated

💬 ⭐ ▶

[See more](#)

Developed by me

NEW

No description provided

👤 📝 ▶



Interoperability

Datalab Metadata Specification

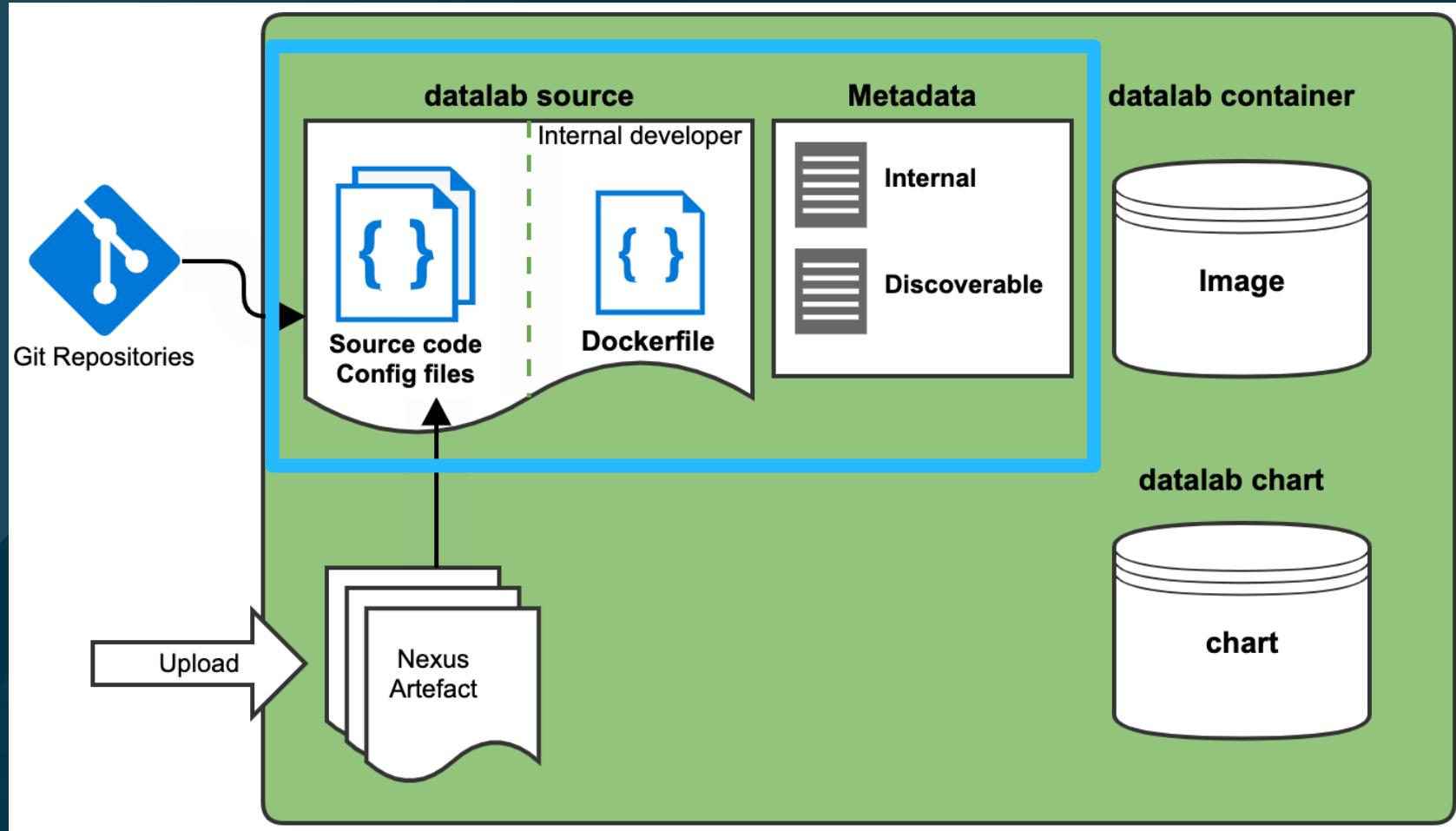
| ESA DDP Term | Namespace:Term | Description | Visibility (private / public) | Write access (developer / moderator / sciapps) | Data type |
|-----------------------|---|--|----------------------------------|---|-----------|
| abstract | sdo:abstract | Abstract (a short description) | PUBLIC | DEVELOPER | 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 | APPLICATION... | Text |
| datalabImage.id | sdo:identifier | Identifier of the datalab image | PRIVATE | APPLICATION... | Text |
| datalabImage.keyword | dcat:keyword | Keyword tag of the current datalab image | PRIVATE | DEVELOPER | Text |
| dateCreated | sdo:dateCreated | Date the datalab was created | PRIVATE | APPLICATION... | Date |

PUBLIC

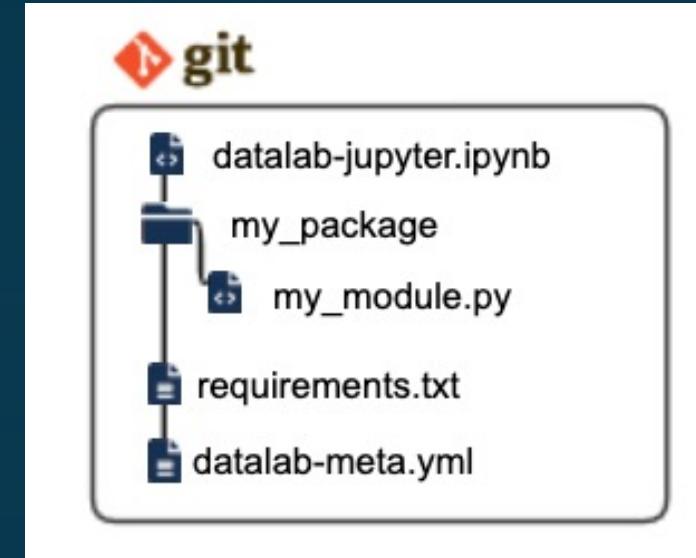
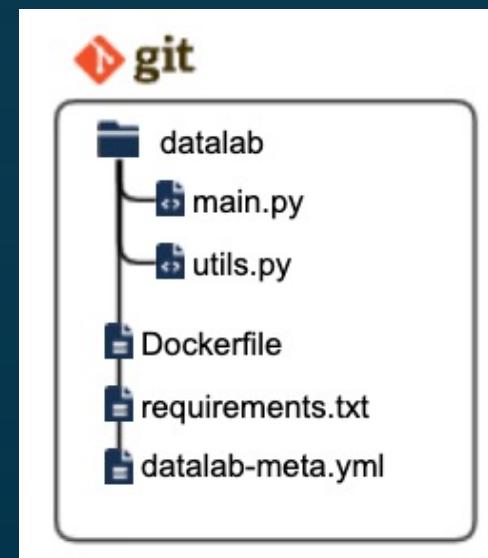
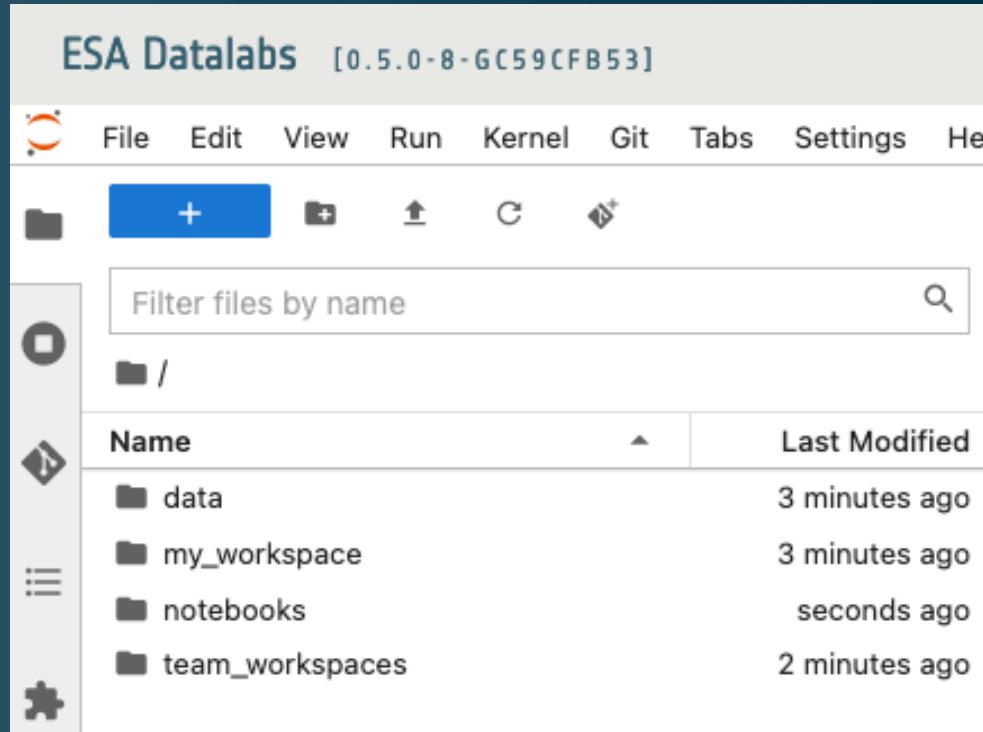
PRIVATE

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

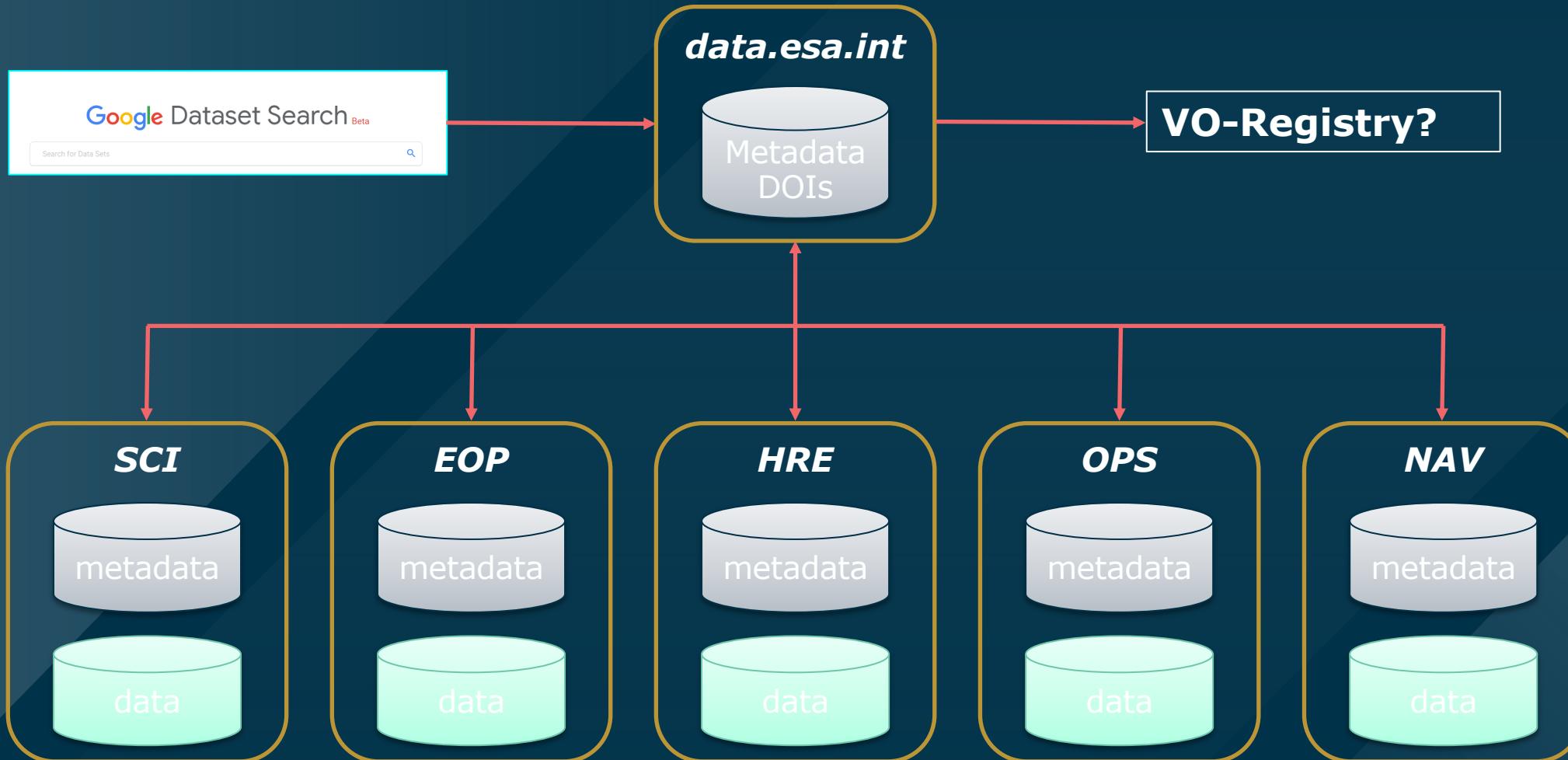
Datalab Package Specification



Datalab Structure Specification (development & runtime)



Metadata environment



Exploitation environment



≡ THE EUROPEAN SPACE AGENCY

ESA Data Discovery Portal [0.4.0/ALPHA]

Type of Asset

- Databab
- Dataset

Properties

- Query Tool
- Visualization Tool
- Analysis Tool
- Data Volume

Domain

- Earth Observation (1,766)
- Navigation (21)
- Space Science (21,818)

Instrument

Thematic Area

Mission

- Astronomy (15,365)
 - Gaia (5)
 - Herschel (723)
 - Hubble Space Telescope (9,745)
 - ISO (1,515)
 - Lisa Pathfinder (6)
 - Planck (4)
 - XMM-Newton (3,367)
- Earth Observation Satellite (1,088)
- ADEOS (1)

Type in your query...

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 other archives for all known astronomical objects in the field.

file-browser

Web File Browser which can be used as a middleware or standalone app.

filezilla

FileZilla - FileZilla

fv

FV - An image display tool.

ji-esdc

Jupyterlab ESDC - Jupyterlab ESDC - Jupyterlab ESDC - Jupyterlab ESDC

ji-euclid-dps

Euclid DPS Jupyterlab - Euclid DPS Jupyterlab

ji-herschel

Herschel JupyterLab - Herschel JupyterLab - Herschel JupyterLab - Herschel JupyterLab

ji-juice

JupyterLab with JUICE

ji-pangala

PanGaia JupyterLab - PanGaia JupyterLab - PanGaia JupyterLab - PanGaia JupyterLab

ji-xmm-sas

Jupyterlab XMM SAS - Jupyterlab XMM SAS

jupyterlab

jupyterlab-cuda

The European Space Agency Data Discovery Portal [0.4.0/ALPHA]

Type in your query...

aladin

Version 0.3.1

[Launch databab](#)

[Launch databab](#)

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 other archives for all known astronomical objects in the field.

ji-herschel

Herschel JupyterLab - Herschel JupyterLab - Herschel JupyterLab

ji-juice

JupyterLab with JUICE moon coverage tool (0.8.0) - JUICE kernels

ji-pangala

PanGaia JupyterLab - PanGaia JupyterLab - PanGaia JupyterLab

ji-xmm-sas

Jupyterlab XMM SAS - Jupyterlab XMM SAS - Jupyterlab XMM SAS



→ THE EUROPEAN SPACE AGENCY

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