

ESA Datalabs for Space Science and GNSS

→ 2020 EIROforum Workshop: Big Data - From Acquisition to Data Mining

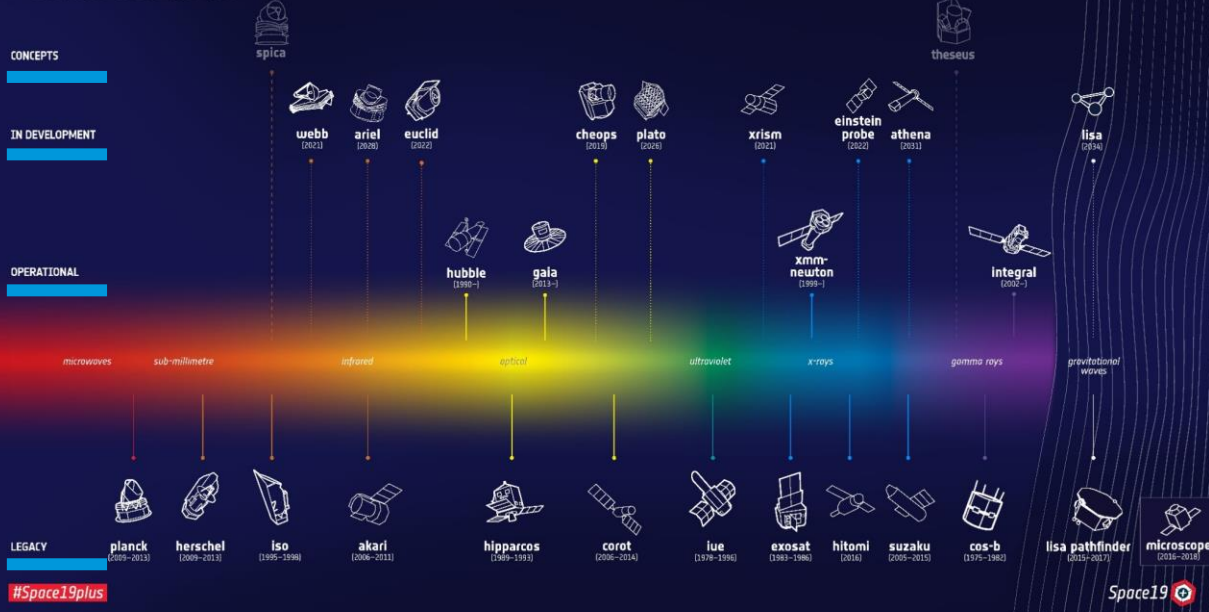
Vicente Navarro

ESA/ESAC

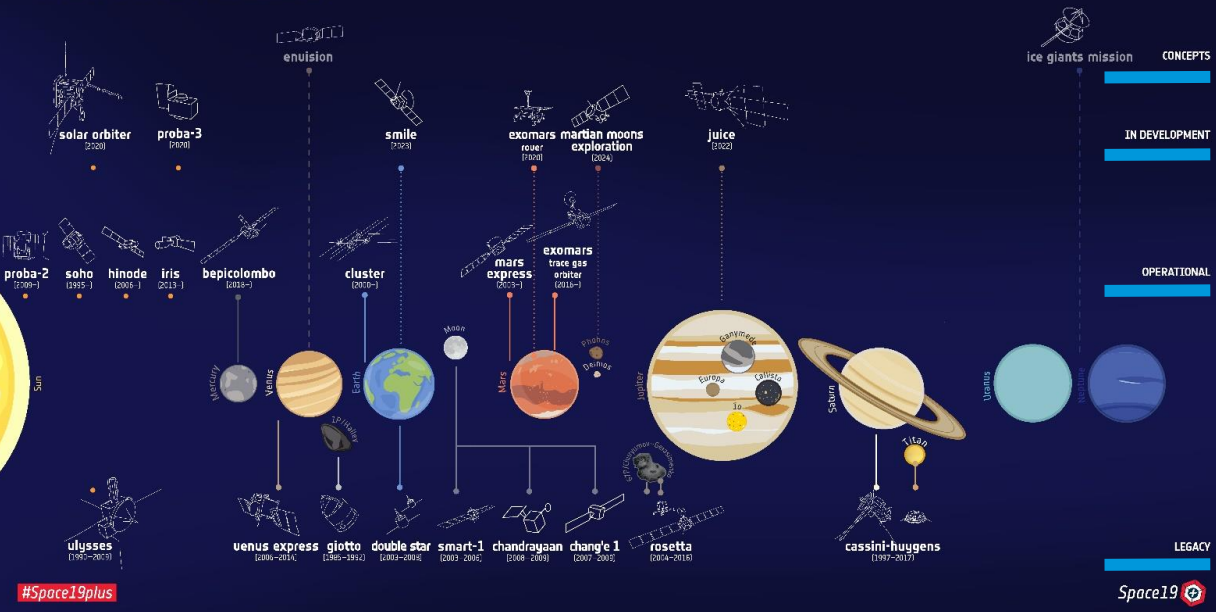
26/10/2020



→ COSMIC OBSERVERS



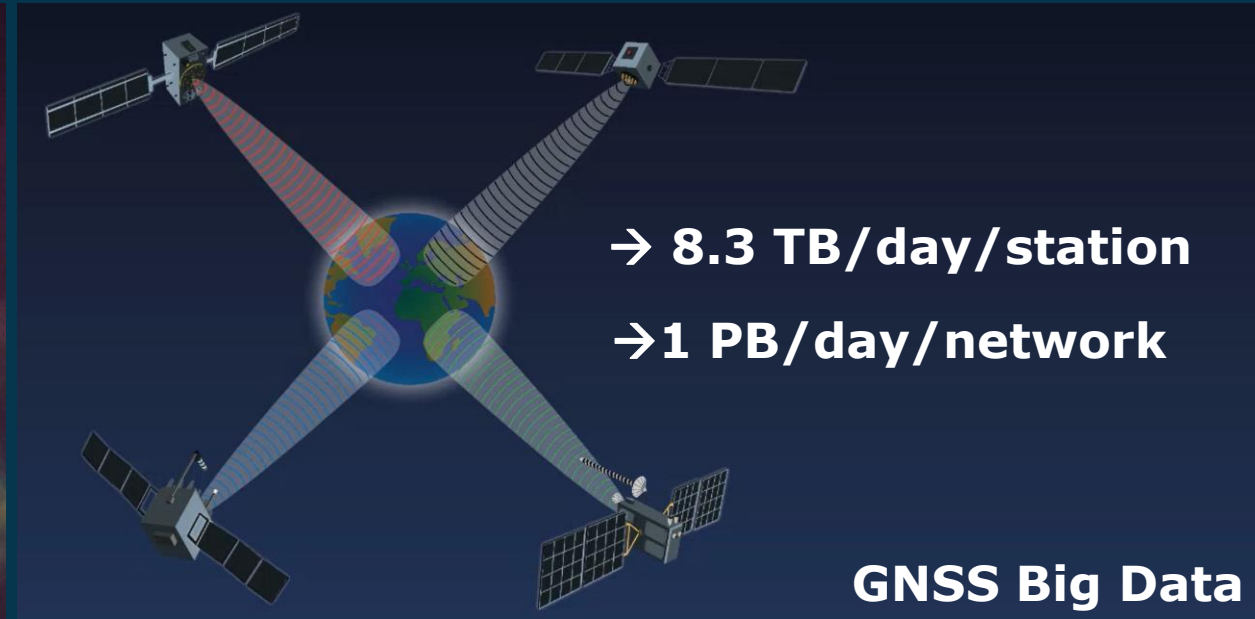
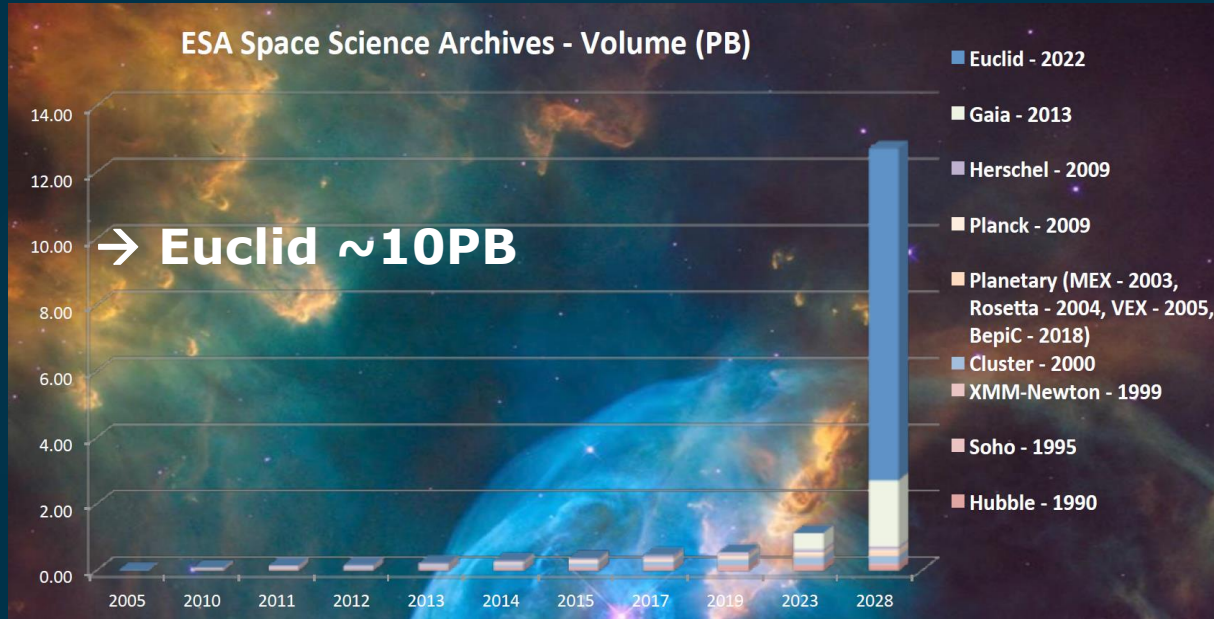
→ SOLAR SYSTEM EXPLORERS



Need to combine **exploitative** and **explorative** mind sets

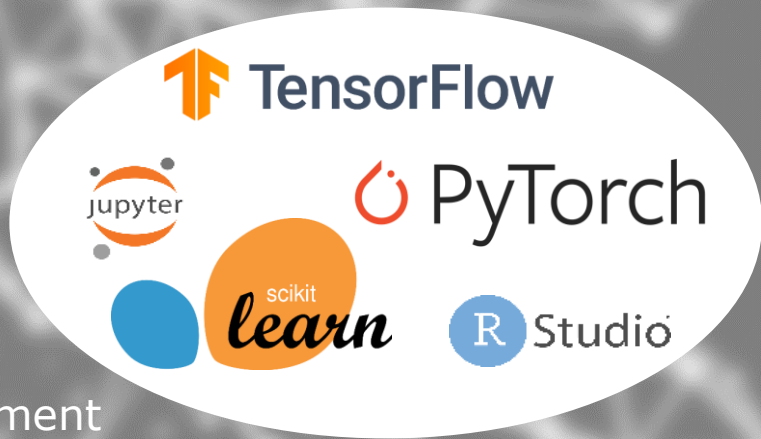
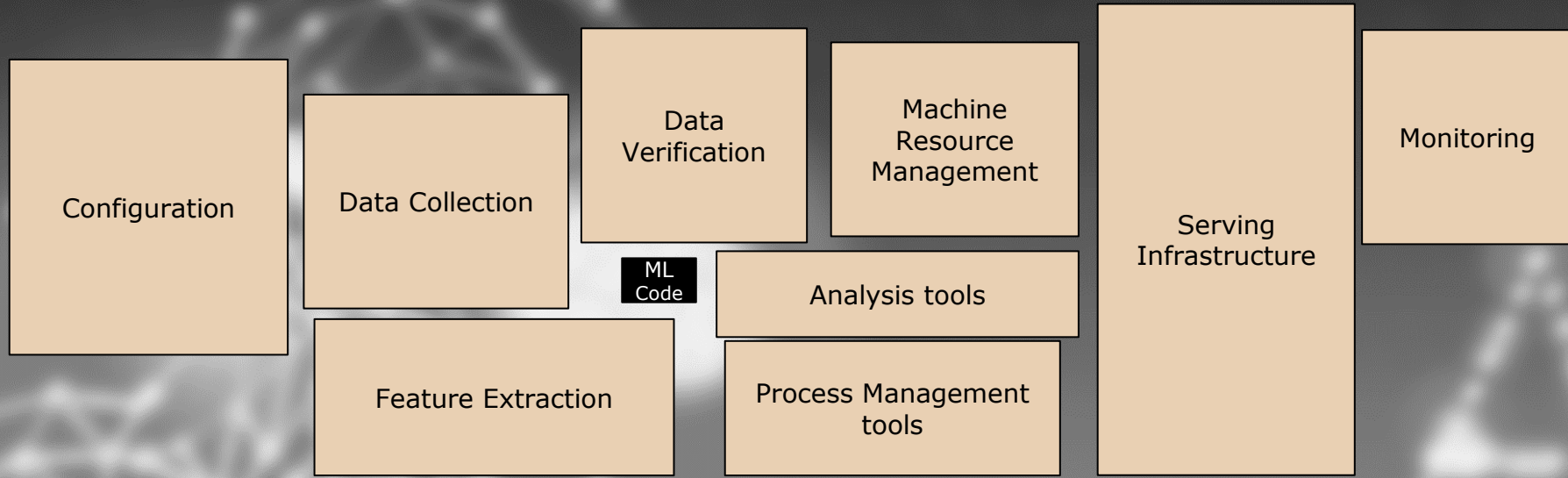
Big Data





From Data transfer to Computation to
Computation transfer to Data

Innovative Data Analysis

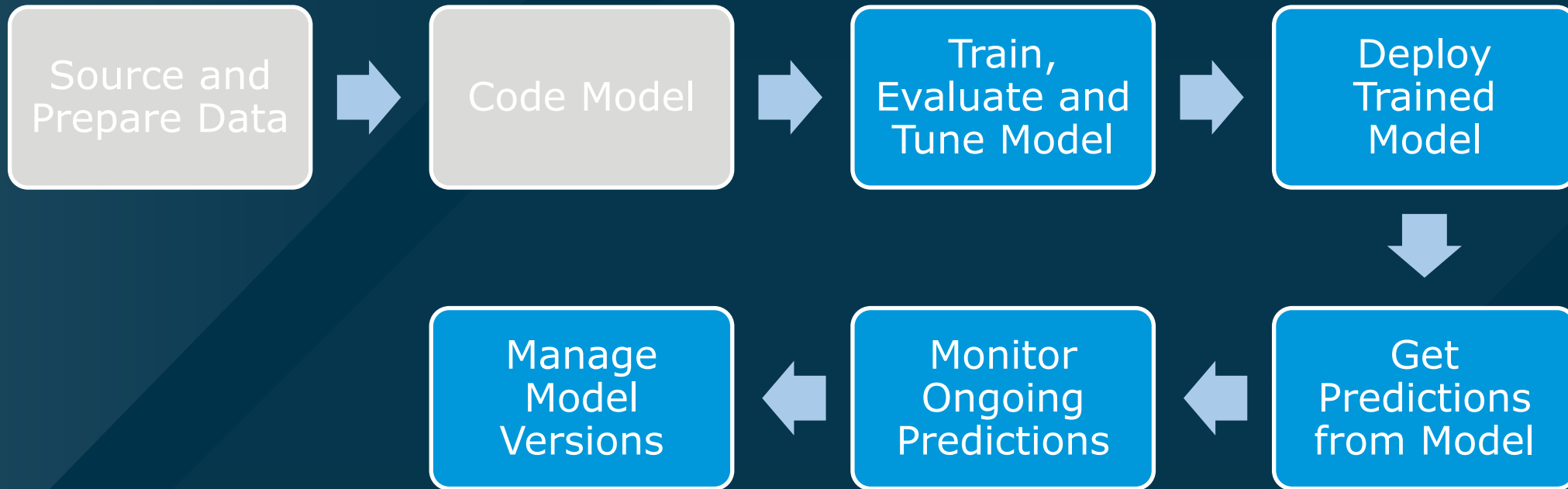


Model Development



Model Deployment

*Source: Hidden Technical Debt in Machine Learning Systems



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, omni voluptas assumenda est, omnis dolor repellendus. Itaque earum rerum hic tenetur a sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat.

Collaboration

INNOVATION
SCIENTIFIC
BRANDING
IDEAS
MARKETING
SUCCESS
MANAGEMENT
ANALYSIS



The diagram features a background of three large, cylindrical structures with domes. The left dome is green and white striped, the middle is red and white striped, and the right is red and white striped. A dark blue box labeled 'ESA Datalabs' is centered at the top. Below it, five dark blue boxes are arranged horizontally, each containing a label: 'SCI' (teal), 'NAV' (yellow), 'OPS' (light blue), 'HRE' (red), and '...' (white). A thin blue line connects the top of these boxes to the 'ESA Datalabs' box.

ESA Datalabs

SCI

NAV

OPS

HRE

...

esa | datalabs

Beta testing for early adopters scheduled for Q4 2020

esa
datalabs

«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.»

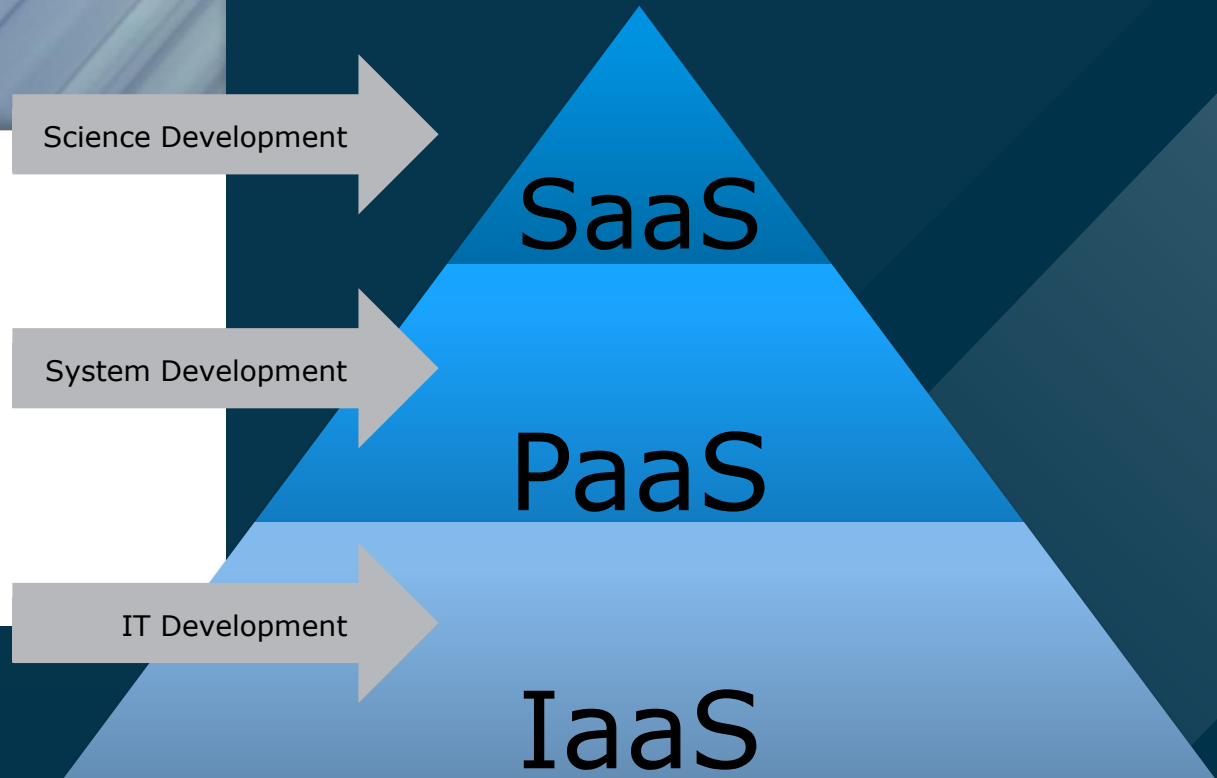
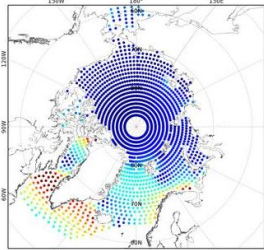
Jim Gray, eScience: A Transformed Scientific Method

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.



☰ datalabs

Create Datalab

Find a datalab in ESA datalabs catalog

Filter results

| | | |
|--|--|---|
| <p>Aladin Sky Atlas 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</p> | <p>ESDC JupyterLab tbd</p> | <p>Euclid DPS JupyterLab tbd</p> |
| <p>File Browser Web File Browser which can be used as a middleware or standalone app. File Browser Website.</p> | <p>File Manager (PCManFM) Lightweight file manager from LXDE.</p> | <p>FreeDOS VM Virtual machine demonstration based on docker-qemu-dos's FreeDOS image.</p> |
| <p>GNSS-Lab Tool (gLAB) gLAB performs precise modeling of GNSS observables (pseudorange and carrier phase) at the centimetre level, allowing standalone GPS positioning, PPP, SBAS and DGNSS. Every single error contributor may be accessed independently which, in turn, provides a major educational</p> | <p>GNU Octave Scientific Programming Language</p> <ul style="list-style-type: none"> • Powerful mathematics-oriented syntax with built-in plotting and visualization tools • Free software, runs on GNU/Linux, macOS, BSD, and Windows | <p>GSSC JupyterLab tbd</p> |
| <p>HARM VM Historical Archives Rationalization and Management Login with user opharm and password 12qwaz. This datalab serves to preserve one of 5 exactly equal machines (hardware with installed software ready to use) that were delivered to</p> | <p>Herschel JupyterLab tbd</p> | <p>Integral JupyterLab INTEGRAL is an ESA's Hard X-ray and Gamma-Ray observatory, featuring high sensitivity in large Field of View. This DataLab includes software for analyzing and interpreting INTEGRAL observations: Offline Scientific Analysis (OSA, the latest version 11) and HEASOFT (version 6.27.1). This</p> |
| <p>JWST JupyterLab tbd</p> | <p>Juice JupyterLab JupyterLab with Juice capabilities for demonstration purposes. Includes kernels.</p> | <p>JupyterLab Plain JupyterLab for demonstration of basic functionality.</p> |
| <p>Kolibri VM Virtual machine demonstration based on KolibriOS.</p> | <p>PanGaia JupyterLab tbd</p> | <p>TOPCAT Tool for OPerations on Catalogues And Tables TOPCAT is an interactive graphical viewer and editor for tabular data. Its aim is to provide most of the facilities that astronomers need for analysis and manipulation of source catalogues and other tables. Though it can be</p> |
| <p>Terminal Multiplexing terminal running Bash.</p> | | |

Navigation bar: datalabs

Create Datalab

Find a datalab in ESA datalabs catalog

| | | |
|-------------------------------|---|--|
| ESDC JupyterLab tbd | Euclid DPS JupyterLab tbd | GNSS-Lab Tool (gLAB) gLAB performs precise modeling of GNSS observables (pseudorange and carrier phase) at the centimetre level, allowing standalone GPS positioning, PPP, SBAS and DGNSS. Every single error contributor may be accessed independently which, in turn, provides a major educational |
| GSSC JupyterLab tbd | Herschel JupyterLab tbd | Integral JupyterLab INTEGRAL is an ESA's Hard X-ray and Gamma-Ray observatory, featuring high sensitivity in large Field of View. This DataLab includes software for analyzing and interpreting INTEGRAL observations: Offline Scientific Analysis (OSA, the latest version 11) and HEASoft (version 6.27.1). This |
| JWST JupyterLab tbd | Juice JupyterLab JupyterLab with Juice capabilities for demonstration purposes. Includes kernels. | JupyterLab Plain JupyterLab for demonstration of basic functionality. |
| | PanGaia JupyterLab tbd | |

Launching the Datalab

esa | datalabs

Datalab Launch

Juice JupyterLab

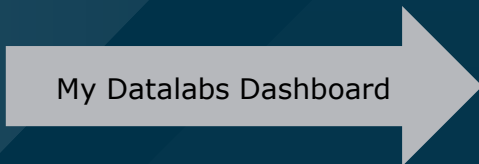
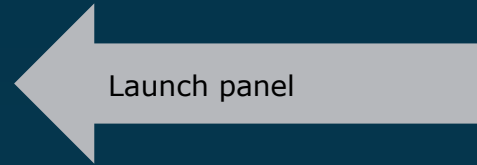
JupyterLab with Juice capabilities for demonstration purposes. Includes kernels.

Options

Name (optional):

select datalab as soon as its interface is available

remove datalab automatically when finished



esa | datalabs

Datalabs

| | | | | | |
|--|--|--|--|--|--|
| | ESDC Analysis Labs ESDC JupyterLab <input type="button" value="Logs"/> <input type="button" value="Destroy"/> | | GNSS Desktop Analysis GNU Octave <input type="button" value="Logs"/> <input type="button" value="Destroy"/> | | GNSS Crowdsourcing GSSC JupyterLab <input type="button" value="Logs"/> <input type="button" value="Destroy"/> |
| | Juice analysis Juice JupyterLab <input type="button" value="Logs"/> <input type="button" value="Destroy"/> | | Integral Spectral Analysis Integral JupyterLab <input type="button" value="Logs"/> <input type="button" value="Destroy"/> | | Gaia DR2 Analysis PanGaia JupyterLab <input type="button" value="Logs"/> <input type="button" value="Destroy"/> |

Data: ESA Archives, private space and more



esa | datalabs

Data Volumes

+ add from catalog + add custom

- My data selection from Planck**
ftp /media/data/planck1
ftp://guest:*****@planck.esa.int/user_sel/3...
- My Dropbox account**
dropbox /media/data/dropbox
dropbox://myusr:*****@dropbox
- Gaia data release v7**
nfs /media/data/gaia_data
nfs://nfs-server2.esa.int/gaia/v7
- Some random webdav data**
webdav /media/data/webdav_test
webdav://user:****@xyz.esa.int/some/path
- My Google Drive account**
gdrive /media/data/gdrive
gdrive://user:****@gdrive
- My home server**
sshfs /media/data/myhome
sshfs://user:****@home.dyndns.example/home/...

esa | datalabs

Search...

My files > persistent_area

New folder
New file
Settings

File Browser 2.1.0
Help

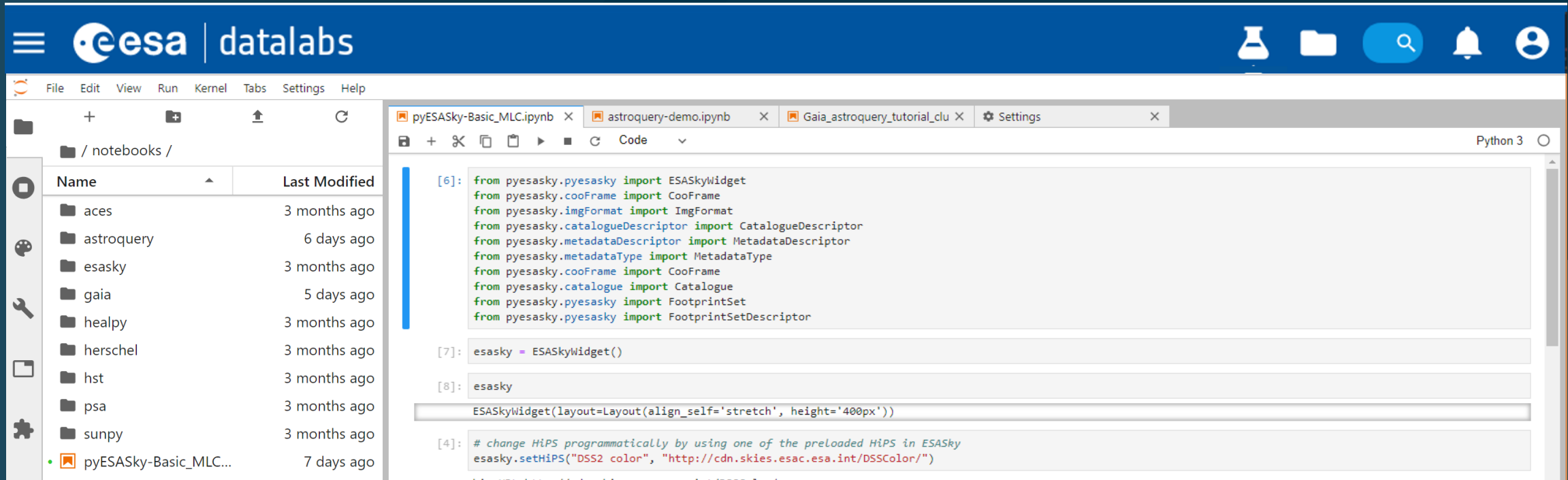
Folders

- .ipynb_checkpoints
3 months ago

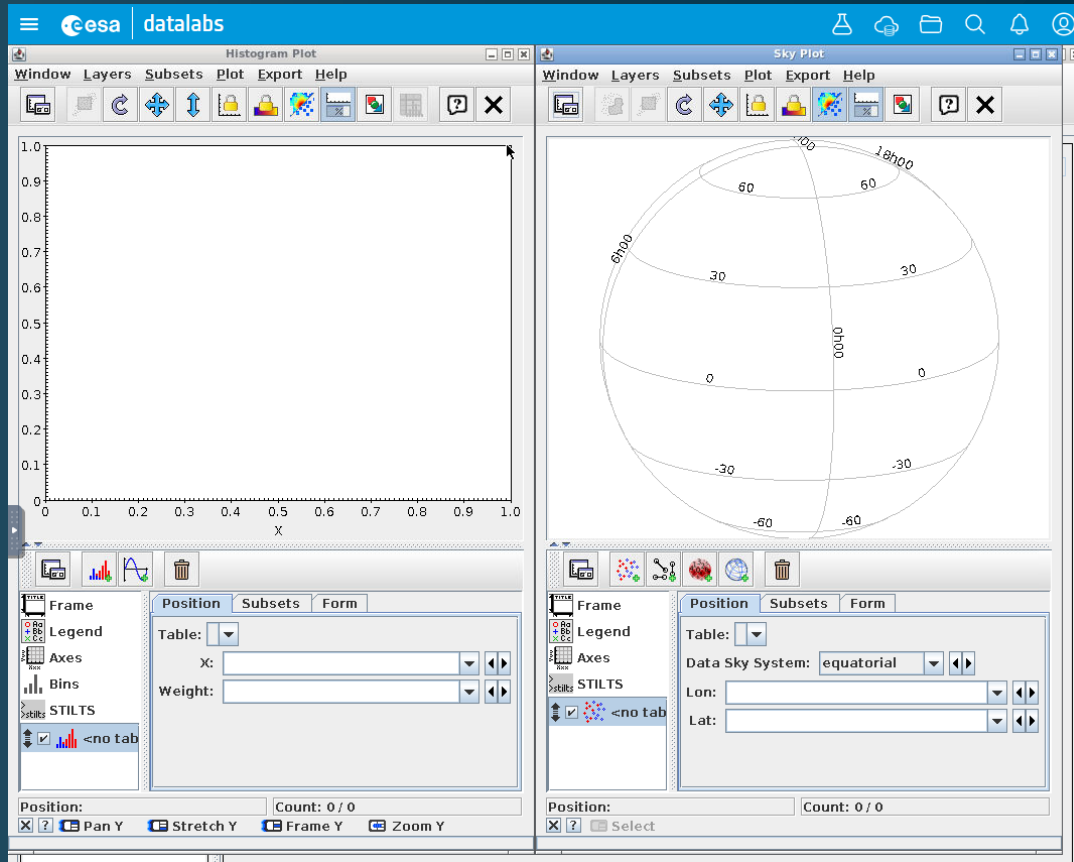
Files

- FeatureFinderNotebook_v1...
52.14 KB
2 months ago





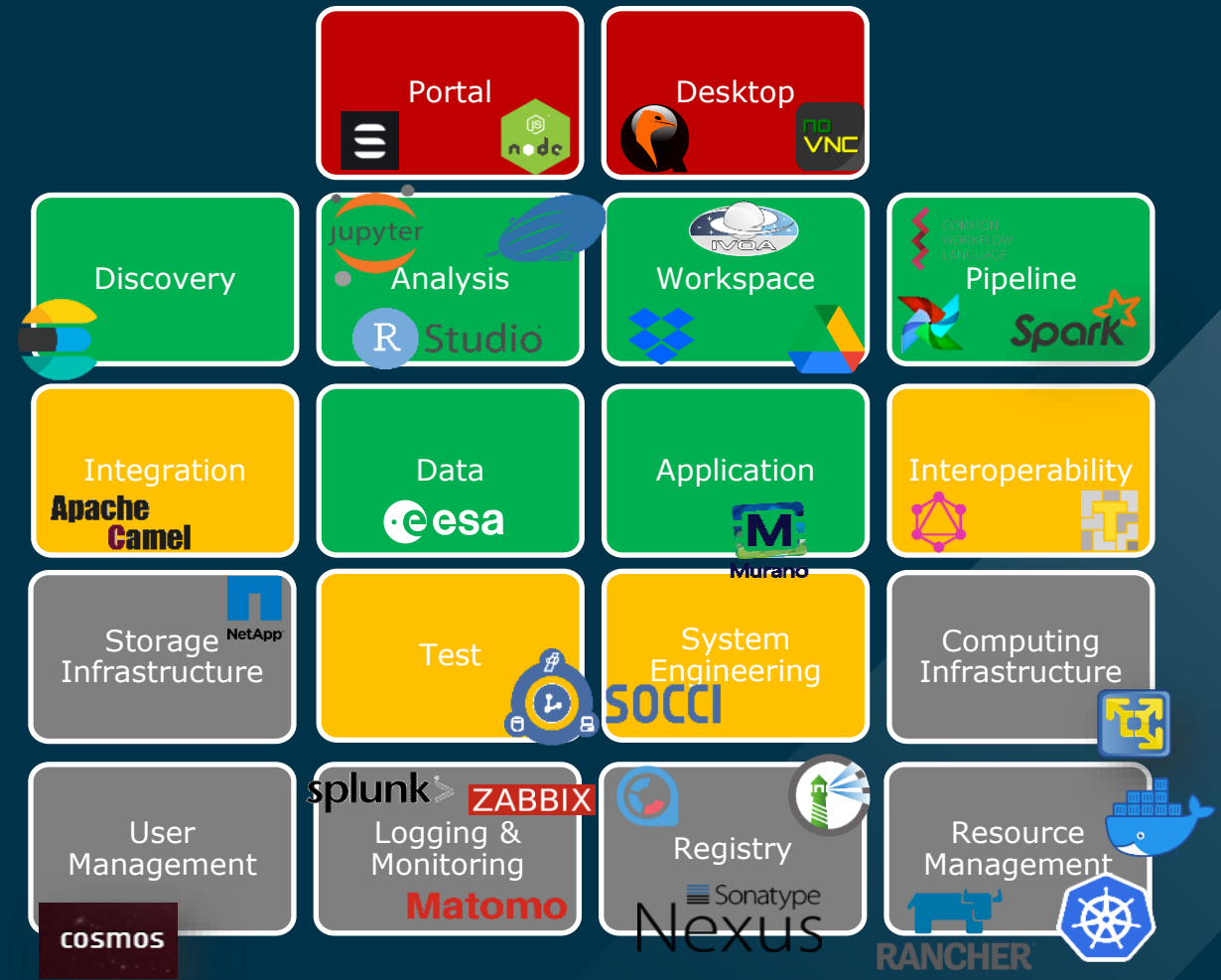
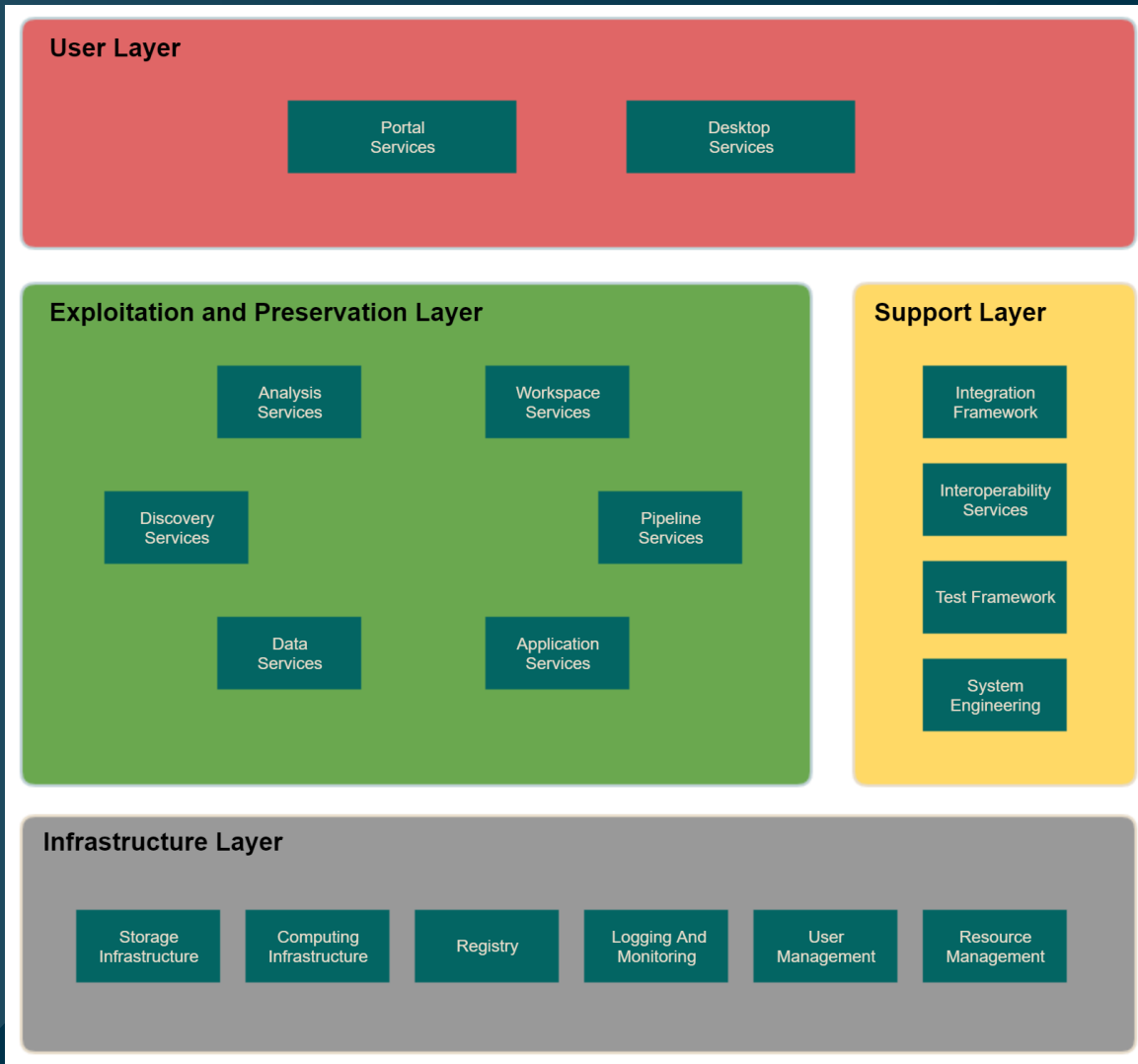
Notebooks pre-configured with AI packages and domain specific analysis tools

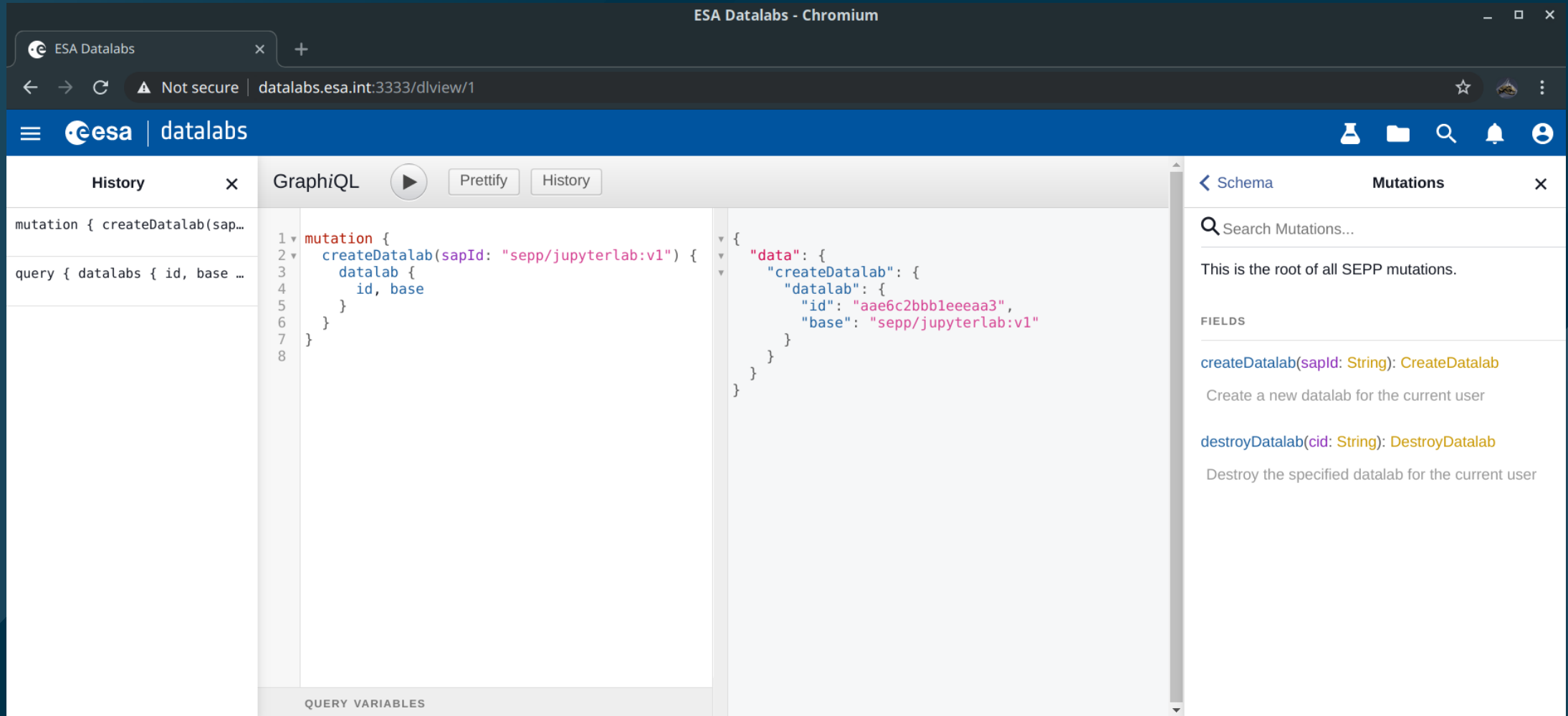


The screenshot shows the gLAB v5.4.4 web interface. At the top, there is a menu bar with 'Mode', 'Templates', 'Configuration', 'Preferences', and 'Help'. Below the header, there is a 'Templates' section with a grid of buttons for different plot types: 'NEU Positioning Error', 'Horizontal Positioning Error', 'Dilution Of Precision', 'Satellite Skyplot', 'Model Components', 'Pfit Residuals', 'Posfit Residuals', 'Measurement Multipath/Noise', 'Zenith Tropospheric Delay', 'Ionospheric Combinations', 'Carrier Phase Ambiguities', and 'Orbit and Clock Comparison'. Below this is the 'Global Graphic Parameters' section, which includes input fields for 'Title', 'X-label', and 'Y-label', a 'Clear' button, and a 'Label Position' dropdown set to 'Top Right'. There are also checkboxes for 'Automatic Limits' and 'Automatic Ticks'. The 'Individual Plot(s) Configuration' section at the bottom allows selecting a plot number (Plot Nr. 1, 2, 3, or 4) and configuring 'Source File', 'Condition(s)', 'X Column', 'Y Column', and 'Legend-label'.

VMs on demand with pre-packaged Domain specific Desktop Apps

Architecture – Technology Stack





The screenshot shows the ESA Datalabs web interface in a Chromium browser. The browser address bar shows the URL `datalabs.esa.int:3333/dlview/1`. The interface has a blue header with the ESA logo and the text "datalabs". Below the header, there are navigation icons for a home page, a folder, search, notifications, and a user profile.

The main content area is divided into three panels:

- History:** Shows a list of recent queries and mutations. The first entry is a mutation: `mutation { createDatalab(sapId: "sepp/jupyterlab:v1") }`. The second entry is a query: `query { datalabs { id, base ... }`.
- GraphiQL:** A central editor for GraphQL queries. It contains a query:

```
1 mutation {
2   createDatalab(sapId: "sepp/jupyterlab:v1") {
3     datalab {
4       id, base
5     }
6   }
7 }
8
```

 To the right of the editor are buttons for "Prettify" and "History".
- Mutations:** A panel on the right showing the JSON response of the query:

```
{
  "data": {
    "createDatalab": {
      "datalab": {
        "id": "aae6c2bbb1e3aa3",
        "base": "sepp/jupyterlab:v1"
      }
    }
  }
}
```

At the bottom of the interface, there is a "QUERY VARIABLES" section which is currently empty.

What's next?

Workflow editor

General
Name: test_dag_5

Datalabs selection

| |
|----------|
| Datalab1 |
| Datalab2 |
| Datalab3 |
| Datalab4 |
| Datalab5 |
| Datalab6 |
| Datalab7 |
| Datalab8 |

```
graph LR; datalab1((datalab1)) --> datalab2((datalab2)); datalab1 --> datalab3((datalab3)); datalab2 --> datalab4((datalab4)); datalab3 --> datalab4
```

Data Volumes

My data selection from Planck

Volume path: /user_sel/31241

Datalab path: /media/data/planck1

Configuration

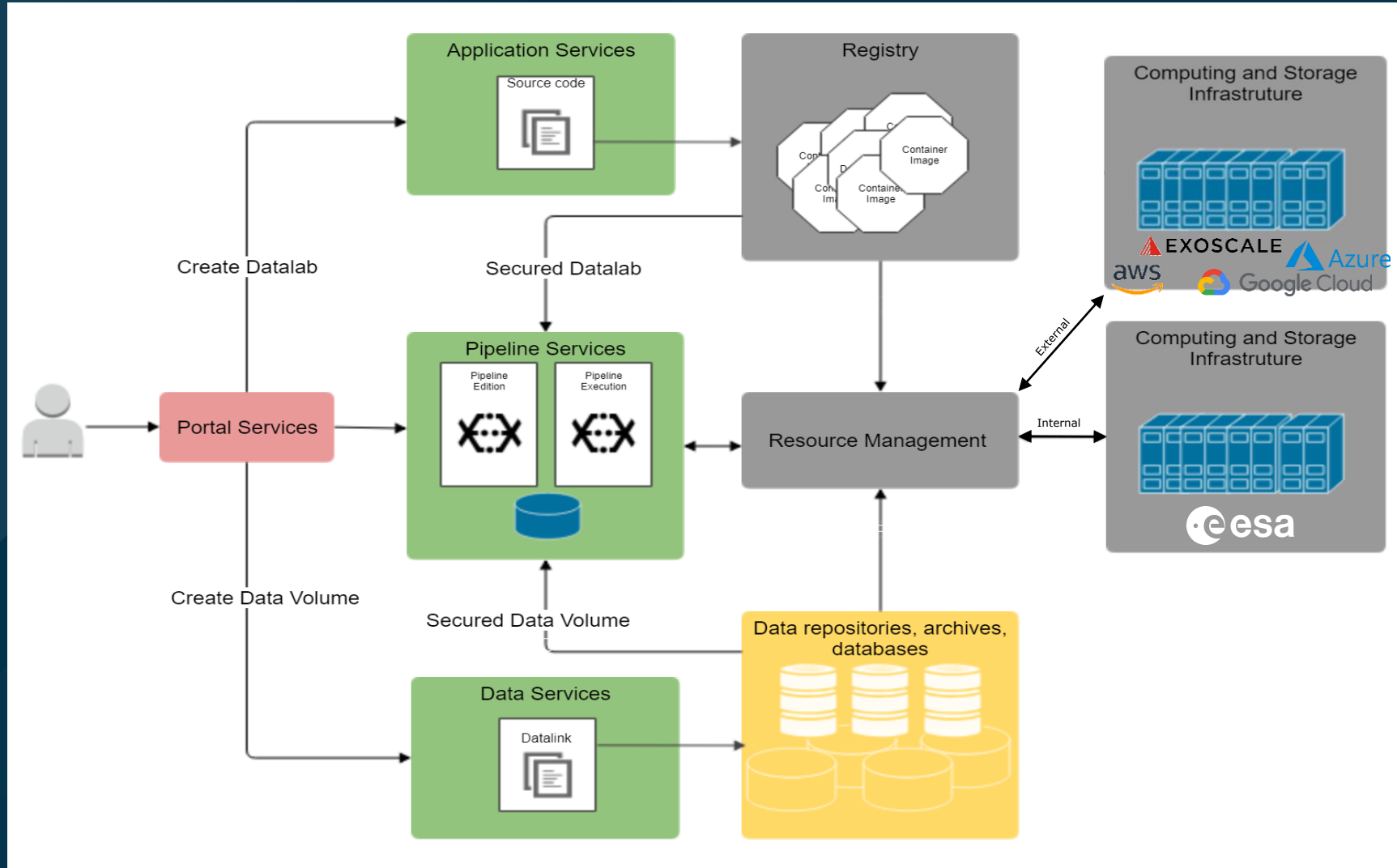
Environment variables

Workflow parameter 1

Workflow parameter 2

Workflow parameter 3

Confirm Cancel



Future Space and GNSS Science Archives Integration

EUROPEAN SPACE AGENCY SCIENCE & TECHNOLOGY MLOPEZCA

Planck Legacy Archive

RESULTS Close All << < Maps #1 > >>

FREQUENCY MAPS (32) ✕

| | Map name | Size | Frequency | Period | BPassCorrected | Zodiccorrected | Ringhalf | Instrument |
|-------------------------------------|--|--------|-----------|--------|----------------|----------------|----------|------------|
| <input checked="" type="checkbox"/> | HFI_SkyMap_100_2048_R3.00_full.fits | 1.9 GB | 100 | Full | yes | no | N/A | HFI |
| <input type="checkbox"/> | HFI_SkyMap_143_2048_R3.00_full.fits | 1.9 GB | 143 | F | | | | |
| <input type="checkbox"/> | HFI_SkyMap_217_2048_R3.00_full.fits | 1.9 GB | 217 | F | | | | |
| <input type="checkbox"/> | HFI_SkyMap_353-psb_2048_R3.00_full.fits | 1.9 GB | 353 | F | | | | |
| <input type="checkbox"/> | HFI_SkyMap_545_2048_R3.00_full.fits | 576 MB | 545 | F | | | | |
| <input type="checkbox"/> | HFI_SkyMap_857_2048_R3.00_full.fits | 576 MB | 857 | F | | | | |
| <input type="checkbox"/> | LFI_SkyMap_030-BPassCorrected-field-IQU_1024_R3.00_full.fits | 144 MB | 30 | F | | | | |
| <input type="checkbox"/> | LFI_SkyMap_030_1024_R3.00_full.fits | 480 MB | 30 | F | | | | |
| <input type="checkbox"/> | LFI_SkyMap_030-BPassCorrected_1024_R3.00_full.fits | 480 MB | 30 | F | | | | |
| <input type="checkbox"/> | LFI_SkyMap_030-field-IQU_1024_R3.00_full.fits | 144 MB | 30 | F | | | | |
| <input type="checkbox"/> | LFI_SkyMap_044-BPassCorrected-field-IQU_1024_R3.00_full.fits | 144 MB | 44 | F | | | | |

Open "All" the selected in a Notebook
Add "All" the selected to Data Volume

Open in a Notebook
Add to Data Volume

gnss science support centre

Filters Labs Collections Logout

Search GNSS assets

3 results filtered by Missions > SWARM Characteristics > Software Clear All

Sort by Most relevant

- Information Data 222
- Information Product 96
- Processing software 57

Organization

- IGS 178
- ESOC 256

Domains

- Earth Science 223

Information Data

Description of this Data Asset

Publisher:

License:

Created:

Tags: Atmosphere Experiment Test

Information Product

Description of this Data Asset

Publisher:

License:

Created:

Tags: Atmosphere Experiment Test

Information Data

Description of this Data Asset

Publisher:

License:

Created:

Tags: Atmosphere Experiment Test

Information Product

Description of this Data Asset

Publisher:

License:

Created:

Tags: Atmosphere Experiment Test

- ✓ **Big Data** Ubiquitous Wave calls for a compute to data paradigm shift
- ✓ **Innovative Data Analysis** calls for platforms hiding infrastructure complexities from scientists
- ✓ **Collaboration** across domains as open science enabler

Thank You!



datalabs.esa.int

Vicente Navarro

@ivicientenavarro

