

# ESA Datalabs

## Multi Mission Science Exploitation and Preservation Platform

→ Astronomy Archives User Group

V. Navarro, R. Alvarez

ESA/ESAC

15/10/2019



# → COSMIC OBSERVERS



CONCEPTS



IN DEVELOPMENT



**webb**  
(2021)



**ariel**  
(2028)



**euclid**  
(2022)



**cheops**  
(2019)



**plato**  
(2026)



**xrism**  
(2021)



**einstein probe**  
(2022)



**athena**  
(2031)



**lisa**  
(2034)

OPERATIONAL



**hubble**  
(1990-)



**gaia**  
(2013-)



**xmm-newton**  
(1999-)



**integral**  
(2002-)

microwaves

sub-millimetre

infrared

optical

ultraviolet

x-rays

gamma rays

gravitational waves

LEGACY



**planck**  
(2009-2013)



**herschel**  
(2009-2013)



**iso**  
(1995-1998)



**akari**  
(2006-2011)



**hipparcos**  
(1989-1993)



**corot**  
(2006-2014)



**iue**  
(1978-1996)



**exosat**  
(1983-1986)



**hitomi**  
(2016)



**suzaku**  
(2005-2015)



**cos-b**  
(1975-1982)

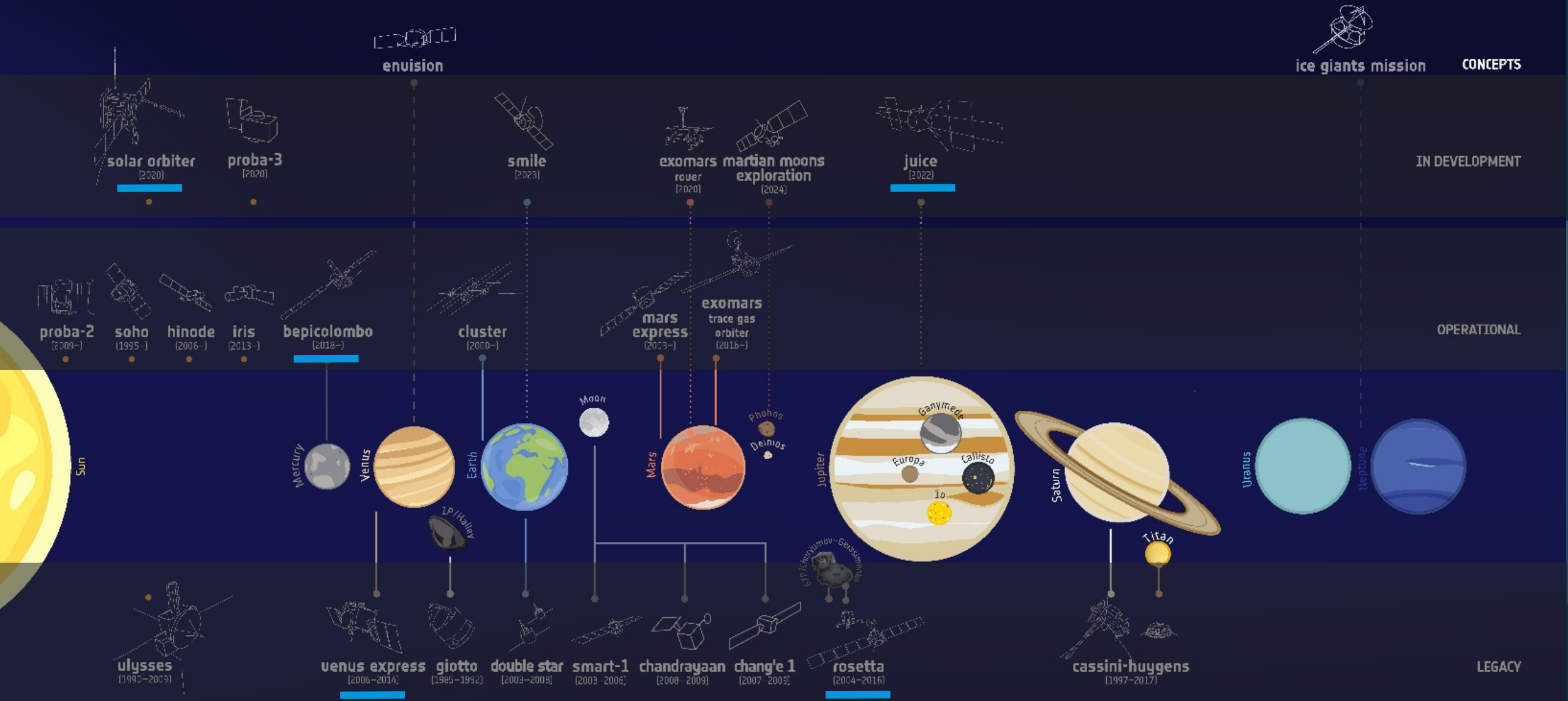


**lisa pathfinder**  
(2015-2017)



**microscope**  
(2016-2018)

# → SOLAR SYSTEM EXPLORERS

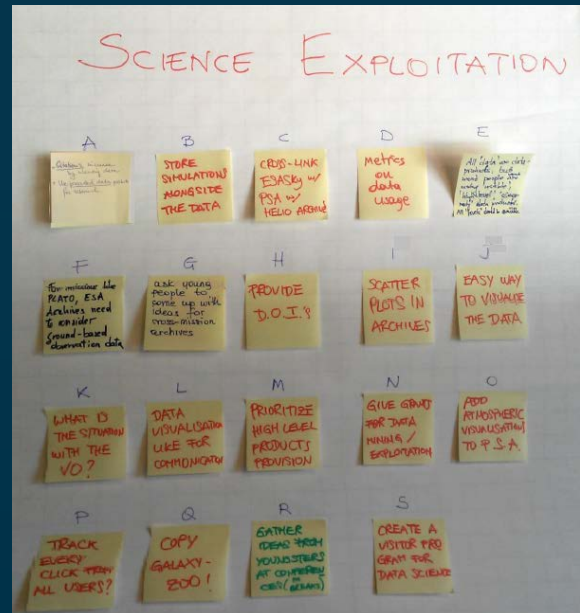


## SEPP-013: ESDC Collaborative Research Lab

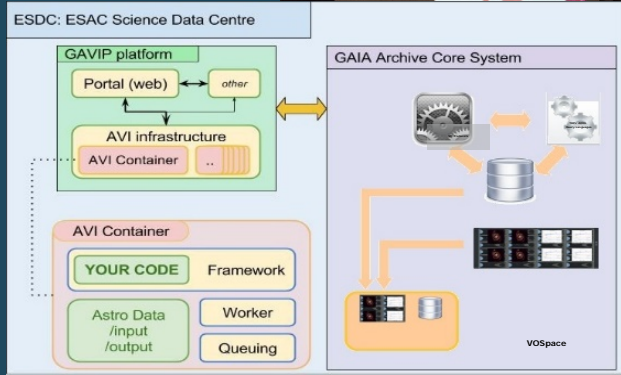
Primary actors: astronomers, planetary scientists and heliophysicists with knowledge of one or more of the data products in the ESDC science archives and wanting to data mine, visualise and/or analyse data from one or a combination of missions.

Secondary actors: members of the general public, educators, students, etc. wanting to explore and analyse archival data in the ESDC Science archives.

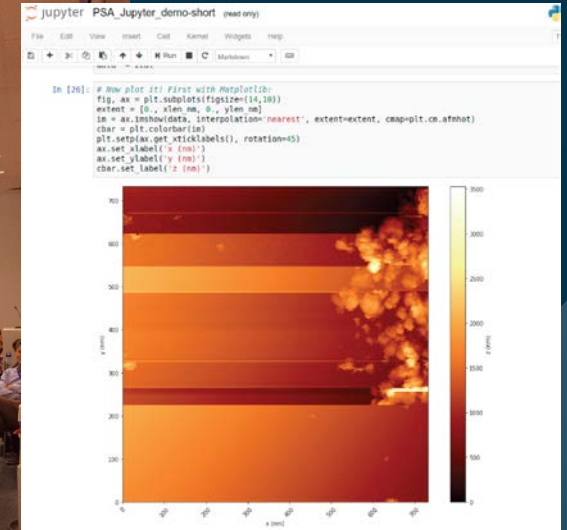
Efficient long-term preservation of data, software and knowledge



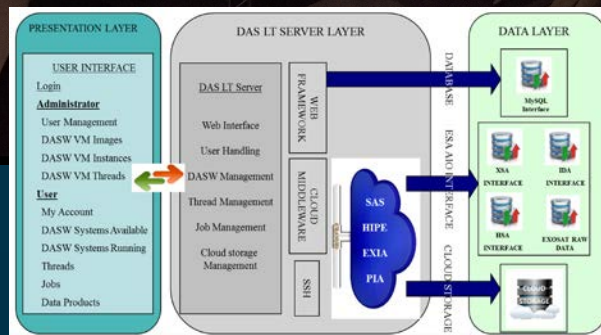
Enabling maximum scientific exploitation of datasets



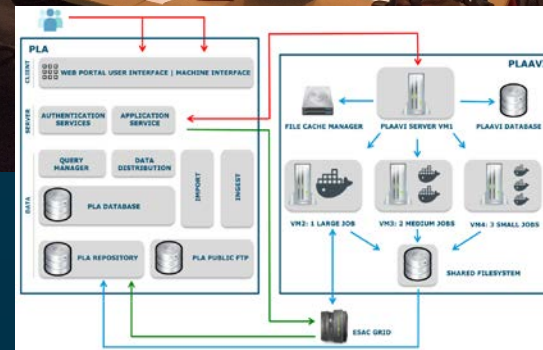
Gaia AVI



SCI-O JupyterHub



DAS-LT



Planck AVI



## Preservation



## Exploitation



## Pipeline Management



## Collaboration

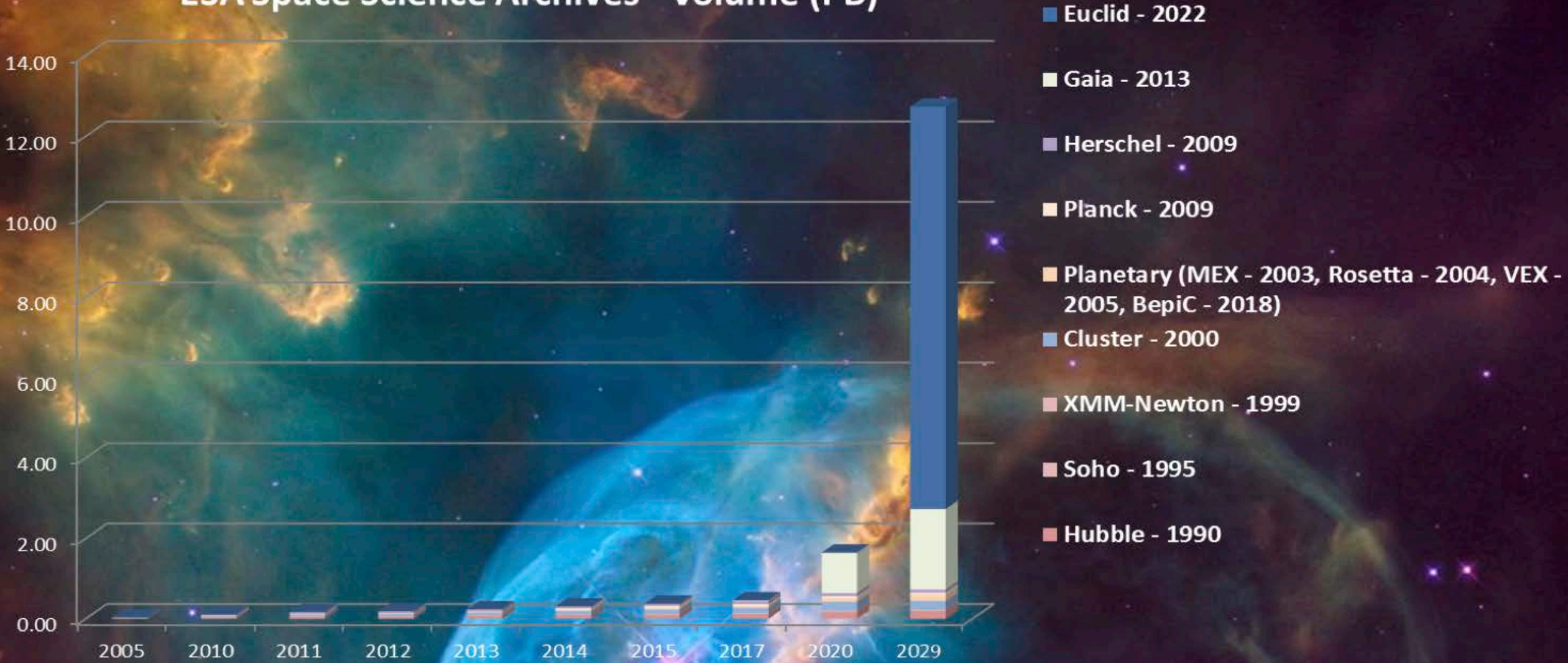


# Big Data



# The ESA Space Science Case

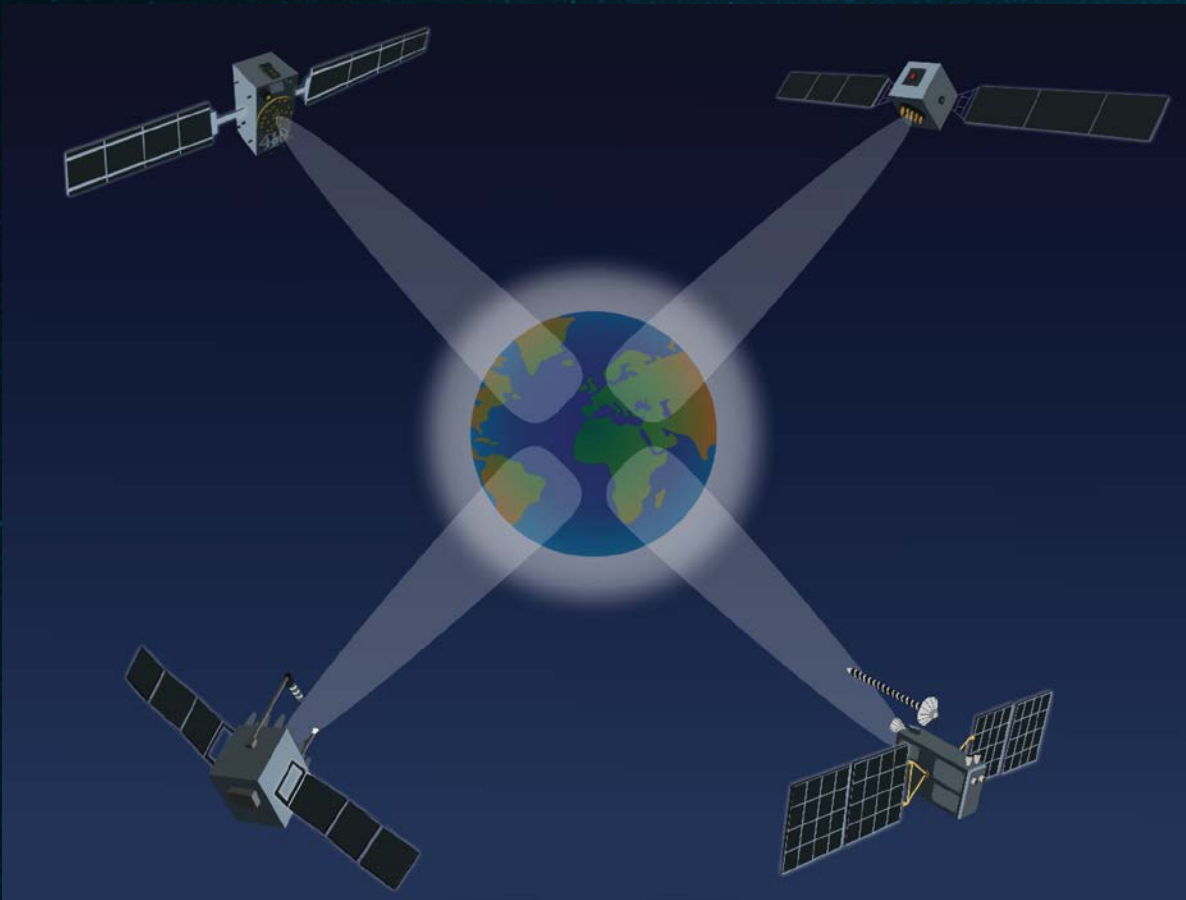
## ESA Space Science Archives - Volume (PB)





## IF Data

▶ High Potential for Discoveries and Innovations



→ Need to manage **8.3 TB of data** per day at each station

→ **1 PB of data** per day for 120 stations

**“Move compute to the data,  
instead of the data to the  
compute**

**... keeping the focus on end-users**

«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

→ THE ARCHIVES, A SCIENTIFIC TREASURE TROVE

The vast amounts of scientific data obtained during a space science mission have a much longer lifetime than the satellite mission itself. The data are archived and made freely accessible online to the global scientific community, and these archives are frequently a mine of unexpected discoveries. They allow researchers to study, for instance, the evolution of a certain celestial object with time, or its appearance at different wavelengths as observed by different telescopes.

The screenshot displays the ESA DataLabs JupyterLab environment. The top navigation bar includes the ESA logo and the text 'datalabs'. Below this, there are icons for a flask, a folder, a search magnifying glass, a bell, and a user profile. The main interface is divided into three sections:

- File Browser (Left):** Shows a directory structure under '/ notebooks /'. A table lists various notebooks and folders with their last modified dates.
- Code Editor (Center):** Contains several code cells. The first cell (cell [6]) imports various classes from the 'pyesasky' package. The second cell (cell [7]) creates an 'ESaskyWidget' instance. The third cell (cell [8]) displays the widget. The fourth cell (cell [4]) shows a comment and a call to 'esasky.setHiPS'.
- Top Bar:** Includes the ESA logo, 'datalabs' text, and utility icons.

Name	Last Modified
aces	3 months ago
astroquery	6 days ago
esasky	3 months ago
gaia	5 days ago
healpy	3 months ago
herschel	3 months ago
hst	3 months ago
psa	3 months ago
sunpy	3 months ago
pyESasky-Basic_MLC...	7 days ago

```
[6]: from pyesasky.pyesasky import ESaskyWidget
from pyesasky.cooFrame import CooFrame
from pyesasky.imgFormat import ImgFormat
from pyesasky.catalogueDescriptor import CatalogueDescriptor
from pyesasky.metadataDescriptor import MetadataDescriptor
from pyesasky.metadataType import MetadataType
from pyesasky.cooFrame import CooFrame
from pyesasky.catalogue import Catalogue
from pyesasky.pyesasky import FootprintSet
from pyesasky.pyesasky import FootprintSetDescriptor

[7]: esasky = ESaskyWidget()

[8]: esasky
ESaskyWidget(layout=Layout(align_self='stretch', height='400px'))

[4]: # change HiPS programmatically by using one of the preloaded HiPS in ESasky
esasky.setHiPS("DSS2 color", "http://cdn.skies.esac.esa.int/DSSColor/")
```

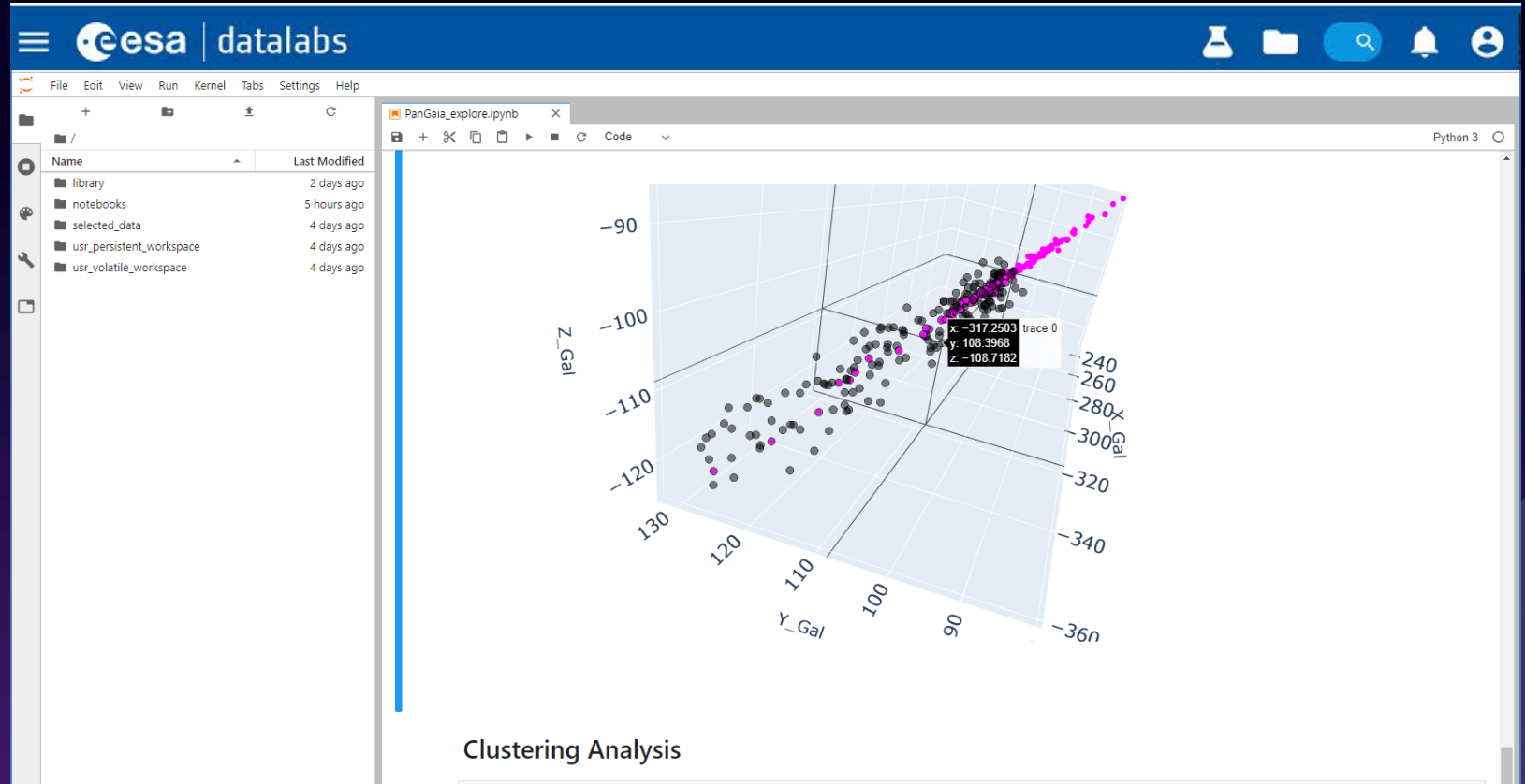
Customised, ready-to-use environment to maximise focus on scientific work

The screenshot shows the Octave desktop environment. The top bar features the 'esa | datalabs' logo and navigation icons. The main interface is divided into several panes:

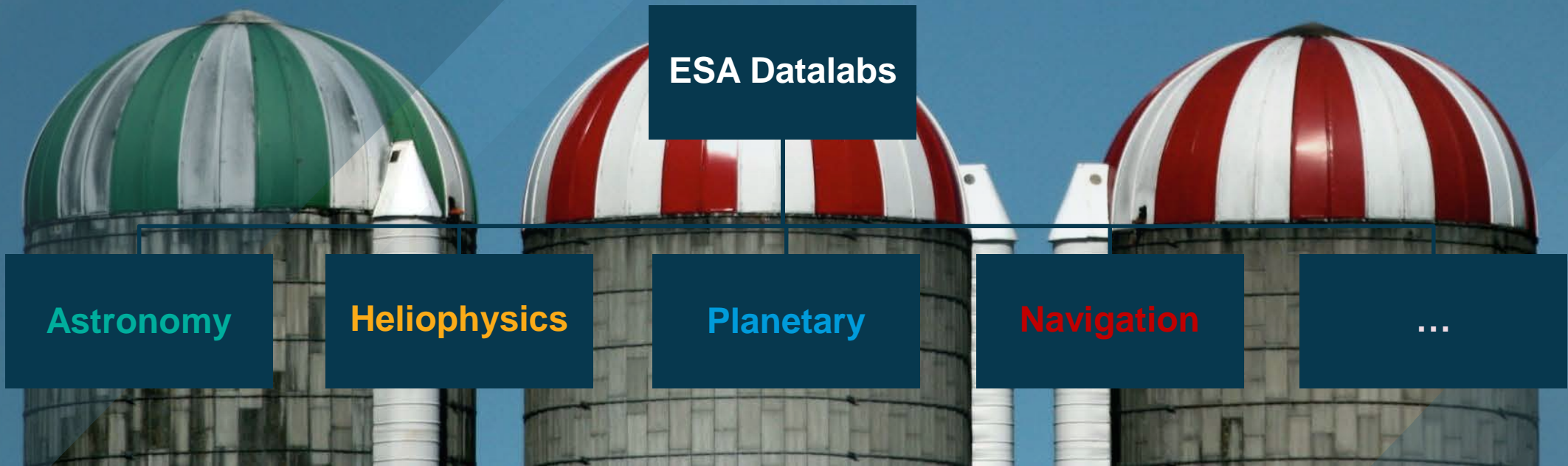
- File Browser:** Shows a directory tree with folders like bin, boot, dev, etc, home, lib, lib64, media, mnt, opt, and proc.
- Command Window:** Displays the GNU Octave version 4.2.2, copyright information, and the execution of the command `plot(sin(0))`, resulting in `ans = 2`.
- Figure 1:** A window showing a plot of a sine wave, with the y-axis ranging from 0 to 1.

Full access via web browser to desktop based applications

# HDBSCAN Clustering on Gaia DR2



Canovas et al. 2019b in preparation
















































































# Science Archives Integration: Planck

Planck Legacy Archive

RESULTS

FREQUENCY MAPS (32)

Map name	Size	Channels	Resolution	Bandwidth	Calibrated	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	Full	yes	no	N/A	HFI
<input type="checkbox"/>        HFI_SkyMap_217_2048_R3.00_full.fits	1.9 GB	217	Full	yes	no	N/A	HFI
<input type="checkbox"/>        HFI_SkyMap_353-psb_2048_R3.00_full.fits	1.9 GB	353	Full	yes	no	N/A	HFI
<input type="checkbox"/>        HFI_SkyMap_545_2048_R3.00_full.fits	576 MB	545	Full	yes	no	N/A	HFI
<input type="checkbox"/>        HFI_SkyMap_857_2048_R3.00_full.fits	576 MB	857	Full	yes	no	N/A	HFI
<input type="checkbox"/>        LFI_SkyMap_030-BPassCorrected-field-IQU_1024_R3.00_full.fits	144 MB	30	Full	yes	no	N/A	LFI
<input type="checkbox"/>        LFI_SkyMap_030_1024_R3.00_full.fits	480 MB	30	Full	no	no	N/A	LFI
<input type="checkbox"/>        LFI_SkyMap_030-BPassCorrected_1024_R3.00_full.fits	480 MB	30	Full	yes	no	N/A	LFI
<input type="checkbox"/>        LFI_SkyMap_030-field-IQU_1024_R3.00_full.fits	144 MB	30	Full	no	no	N/A	LFI
<input type="checkbox"/>        LFI_SkyMap_044-BPassCorrected-field-IQU_1024_R3.00_full.fits	144 MB	44	Full	yes	no	N/A	LFI

of 1 Page size: 10

Displaying 1-32 of 32

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

Open in a Notebook  
Add to Data Volume

COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED [V-3.4-oper]



## gnss science support centre



Filters | Labs | Collections | Logout

Search GNSS assets

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

Sort by Most relevant

- Information Data
- Information Product
- Information Data
- Information Product

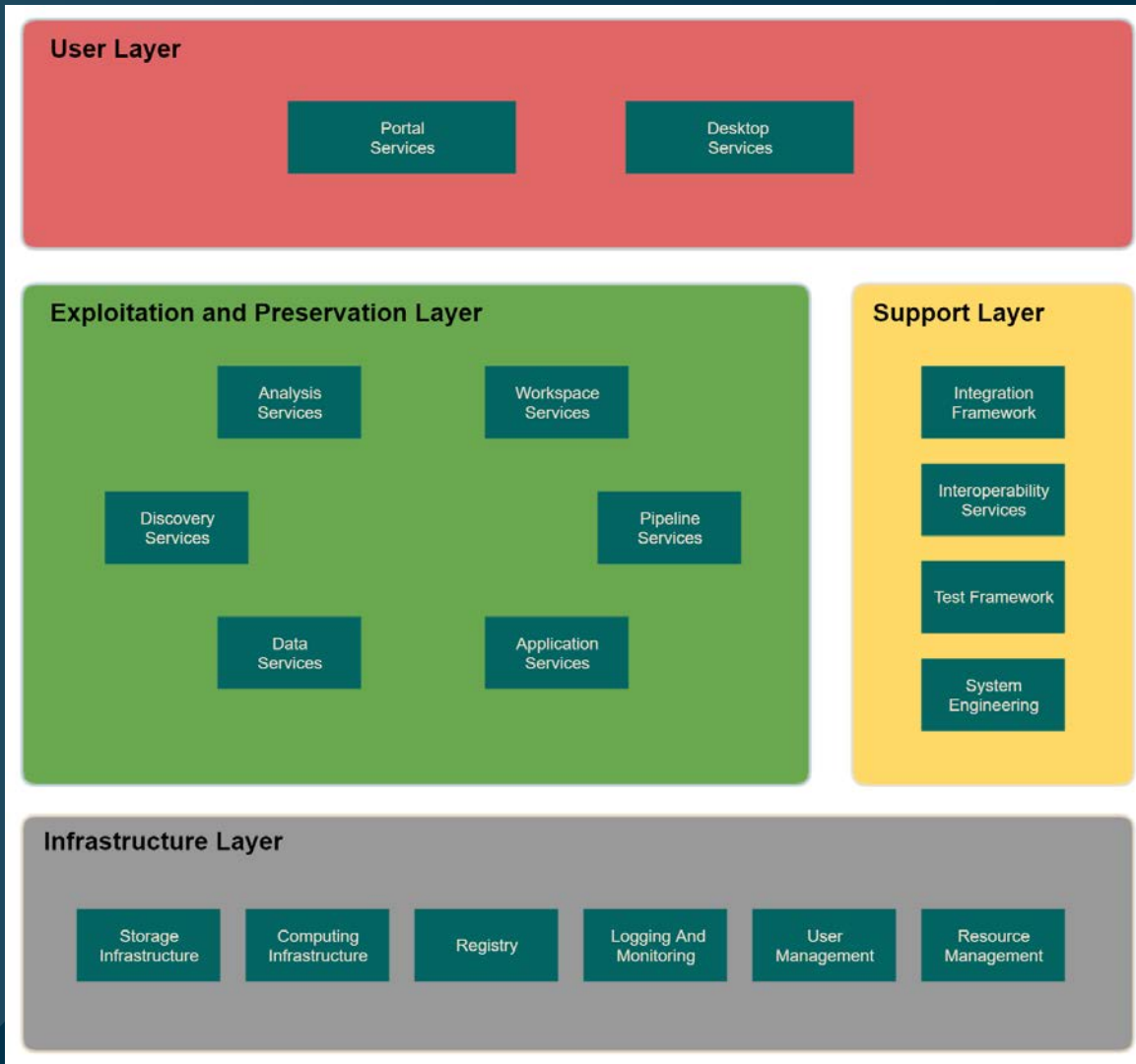
Description of this Data Asset  
Publisher:  
License:  
Created:

Tags: Atmosphere Experiment Test

- Open in a Notebook
- Add to Data Volume

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

# What's inside?



ESA Datalabs - Chromium

ESA Datalabs | datalabs

History x GraphiQL Prettify History

```
mutation { createDatalab(sap...
query { datalabs { id, base ...
```

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

```
{
  "data": {
    "createDatalab": {
      "datalab": {
        "id": "aae6c2bbb1eeaa3",
        "base": "sepp/jupyterlab:v1"
      }
    }
  }
}
```

Schema Mutations x

Search Mutations...

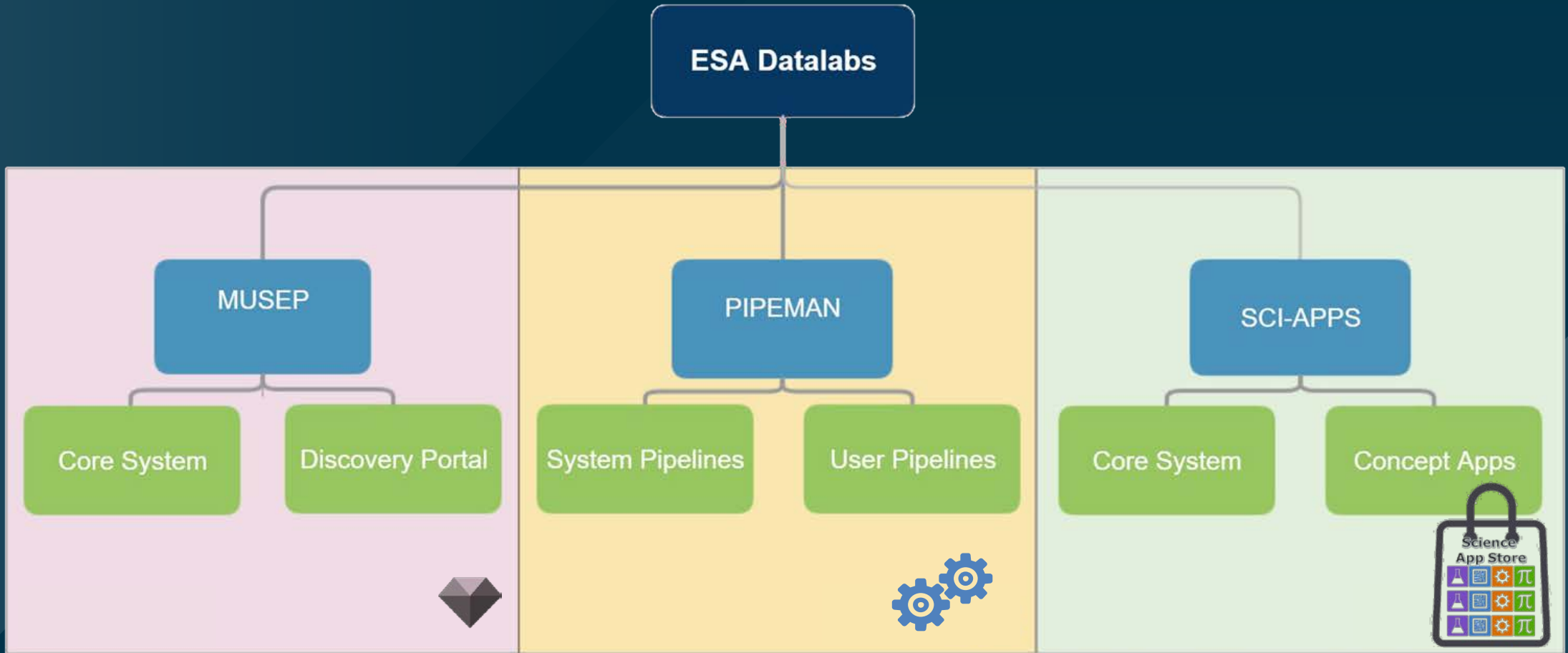
This is the root of all SEPP mutations.

FIELDS

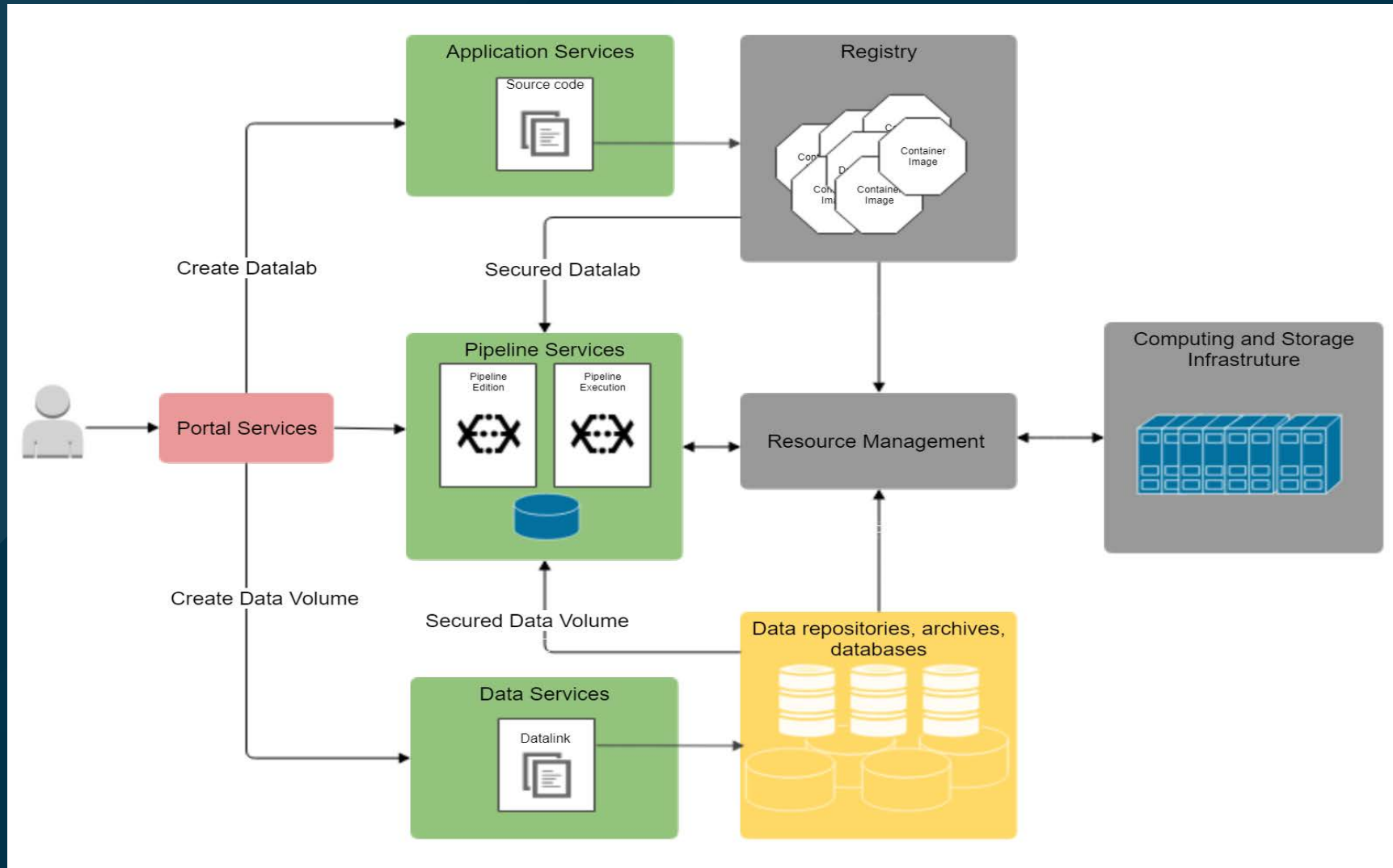
**createDatalab(sapId: String): CreateDatalab**  
Create a new datalab for the current user

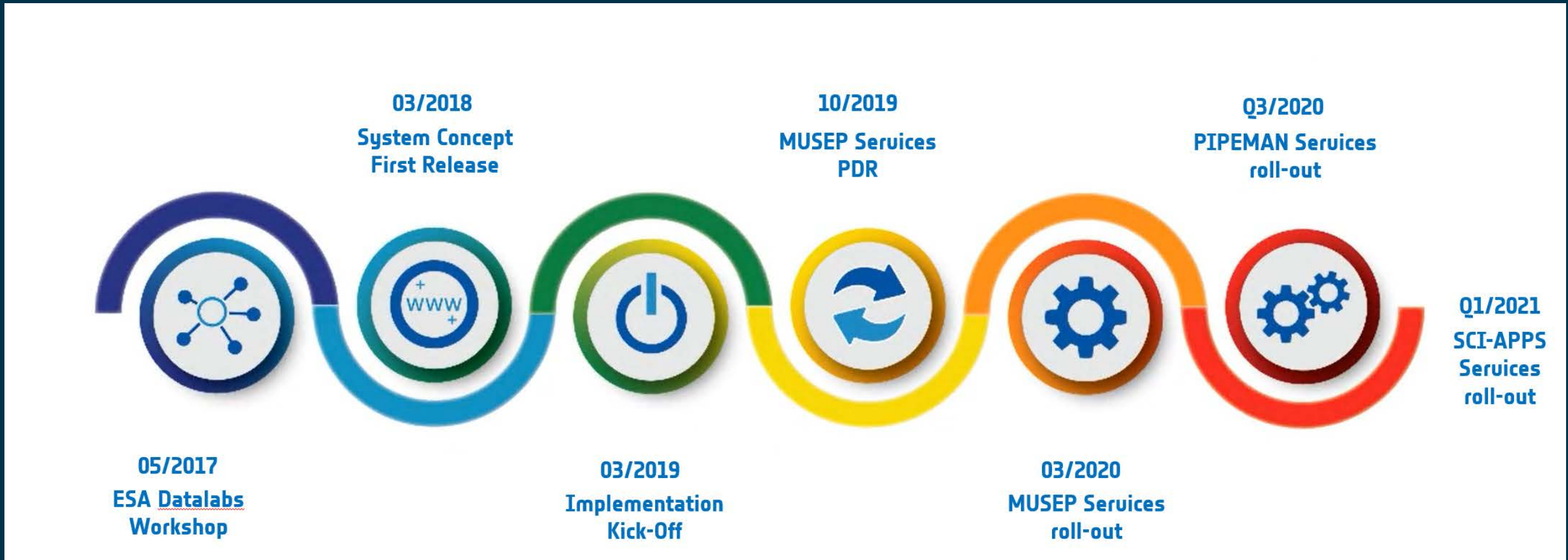
**destroyDatalab(cid: String): DestroyDatalab**  
Destroy the specified datalab for the current user

QUERY VARIABLES



# Bringing all together





## Security

Single-Sign-On, public, shared and restricted areas, code-injection ...

1

## Storage and Computing Dimensioning

Small evolution, quotas, reuse / evolution of existing infrastructure, pay-per-use hybrid models

2

## Fast evolution of the technology:

Jupyter compatibility, Containers vs VMs, Kubernetes vs Swarm ...

3

## IPR

Define clear boundaries and recognition for user property rights

4

## Open platform

Avoid vendor lock-in, develop and adopt OSS, standards and available ICT

5

## Community driven

Close involvement of the scientific community to maximise science return

6

**New missions call for:**

- **Paradigm shift from “bring the data to the user” to “bring the user to the data”**
- **Close interaction between archives and data processing services**

**Legacy missions call for data and software long term preservation**

**Scientists call for collaborative research environment**

**Leverage on existing VO standards, archives, mission and IT systems**



Thank You!



[datalabs.esa.int](http://datalabs.esa.int)

