

ESA Datalabs

Multi Mission Science Exploitation and Preservation Platform

→ ADASS 2019

V. Navarro, R. Alvarez, S. del Rio, R. Guerra, V. Martín, M.A. Diego, N. Alexandre, A. Lousa, X. Rodrigues, A. Pereira, D. Monteiro, A. Héliou, M. Marseille

Groningen, the Netherlands

08/10/2019



→ COSMIC OBSERVERS

CONCEPTS



spica

IN DEVELOPMENT



webb
(2021)



ariel
(2028)



euclid
(2022)



cheops
(2019)



plato
(2026)



xrism
(2021)



einstein
probe
(2022)



athena
(2031)



theseus



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)



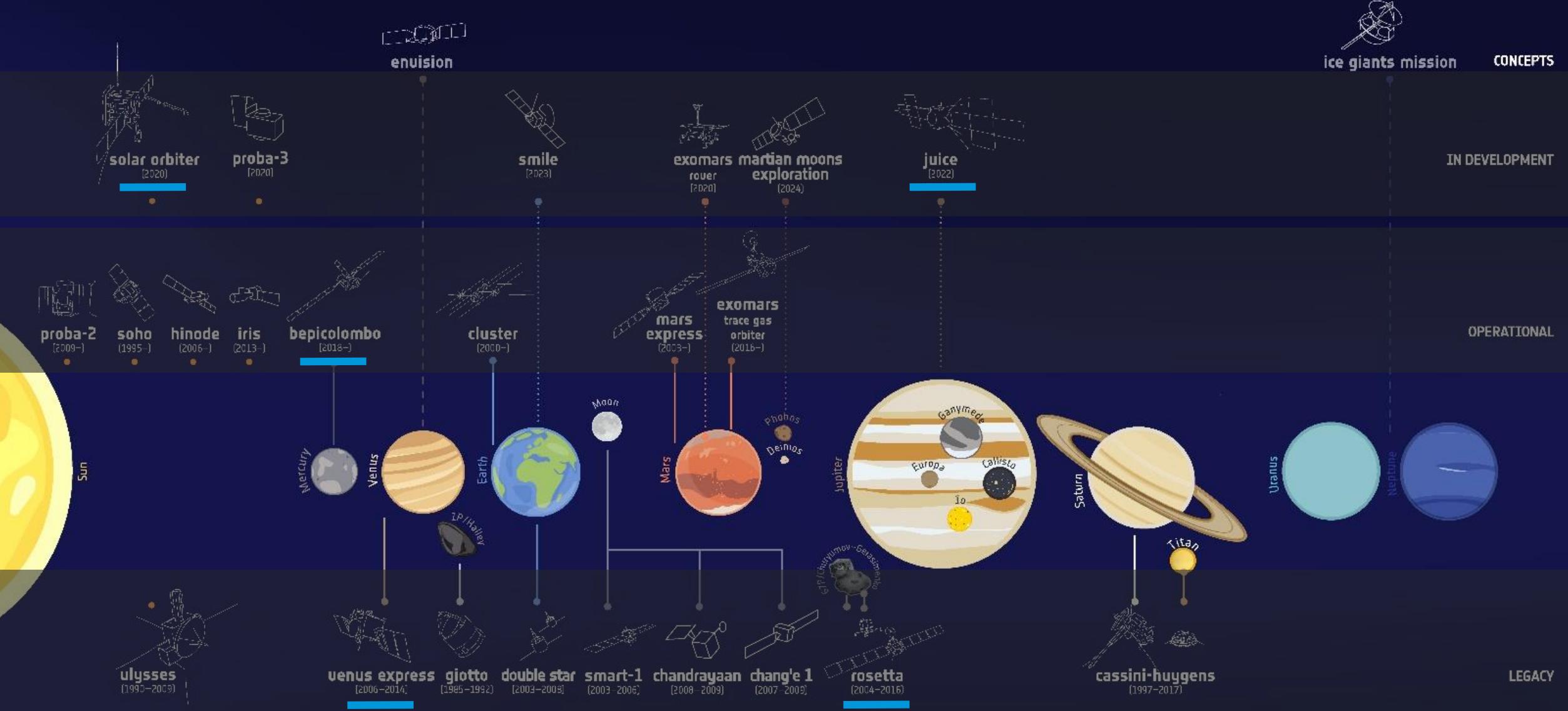
cos-b
(1975–1982)



microscope
(2016–2018)



→ SOLAR SYSTEM EXPLORERS



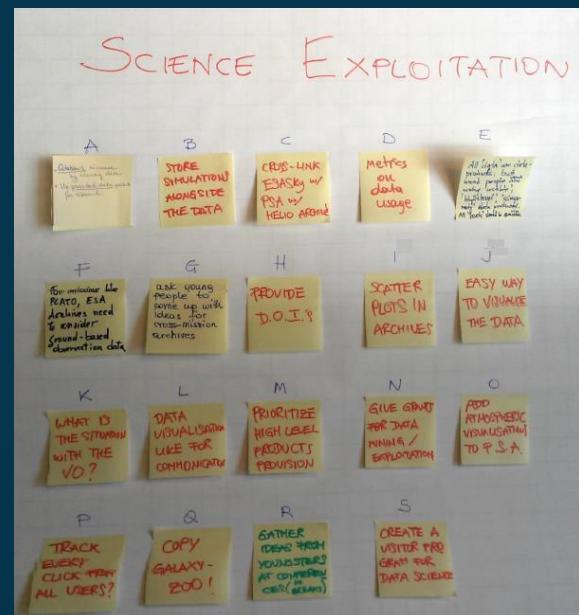
Strategy Definition

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.

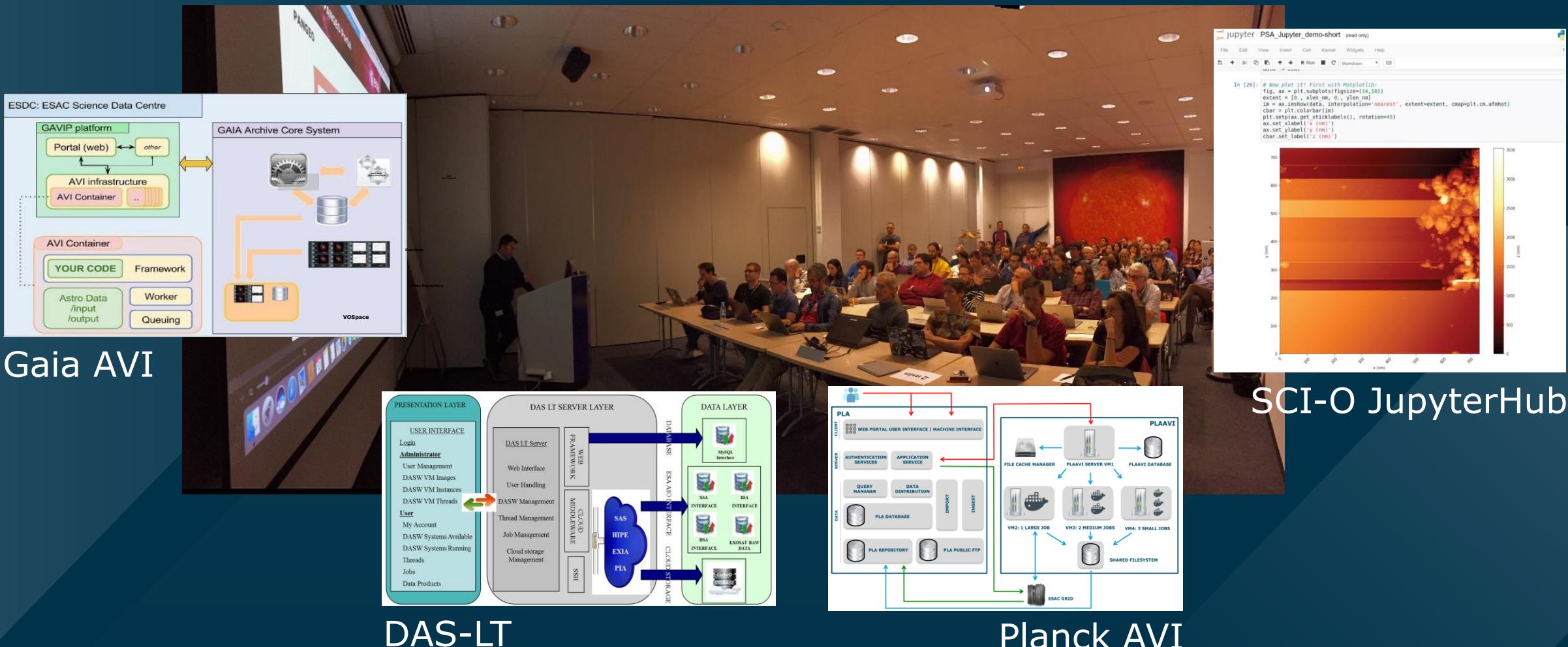
Secondary actors: members of the general public, educators, students, etc. wanting to explore and analyse archival data in the ESDC Science archives.

Efficient long-term preservation of data, software and knowledge

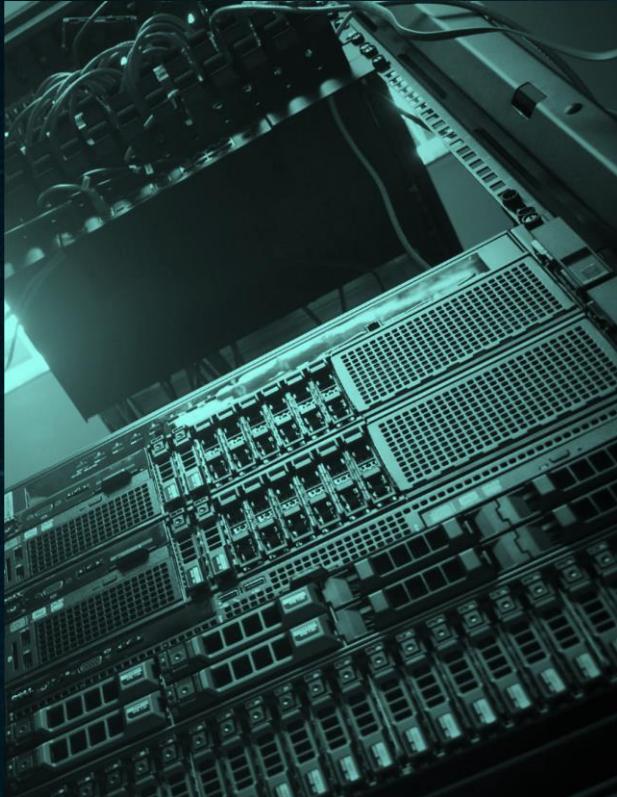


Enabling maximum scientific exploitation of datasets

SCI Related Activities



Preservation



Exploitation



Pipeline Management



Collaboration



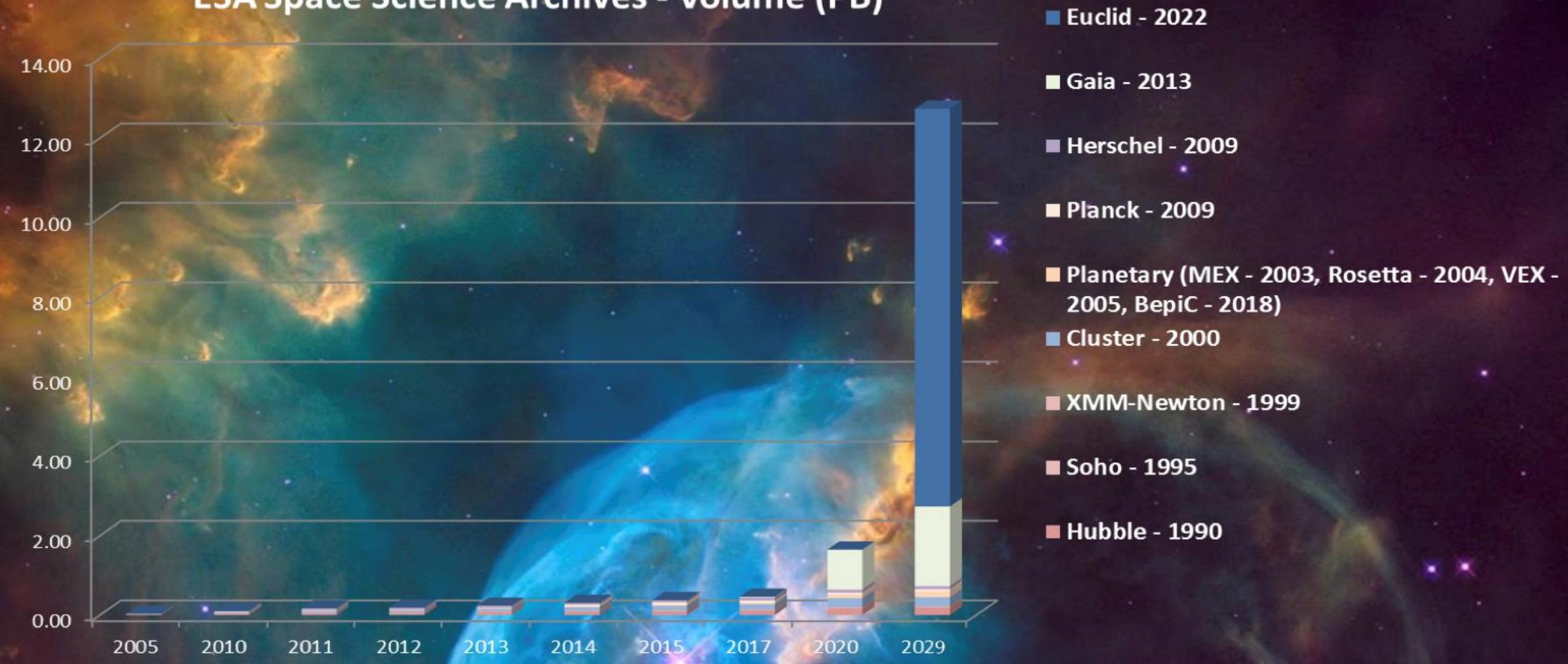
Big Data



The ESA Space Science Case

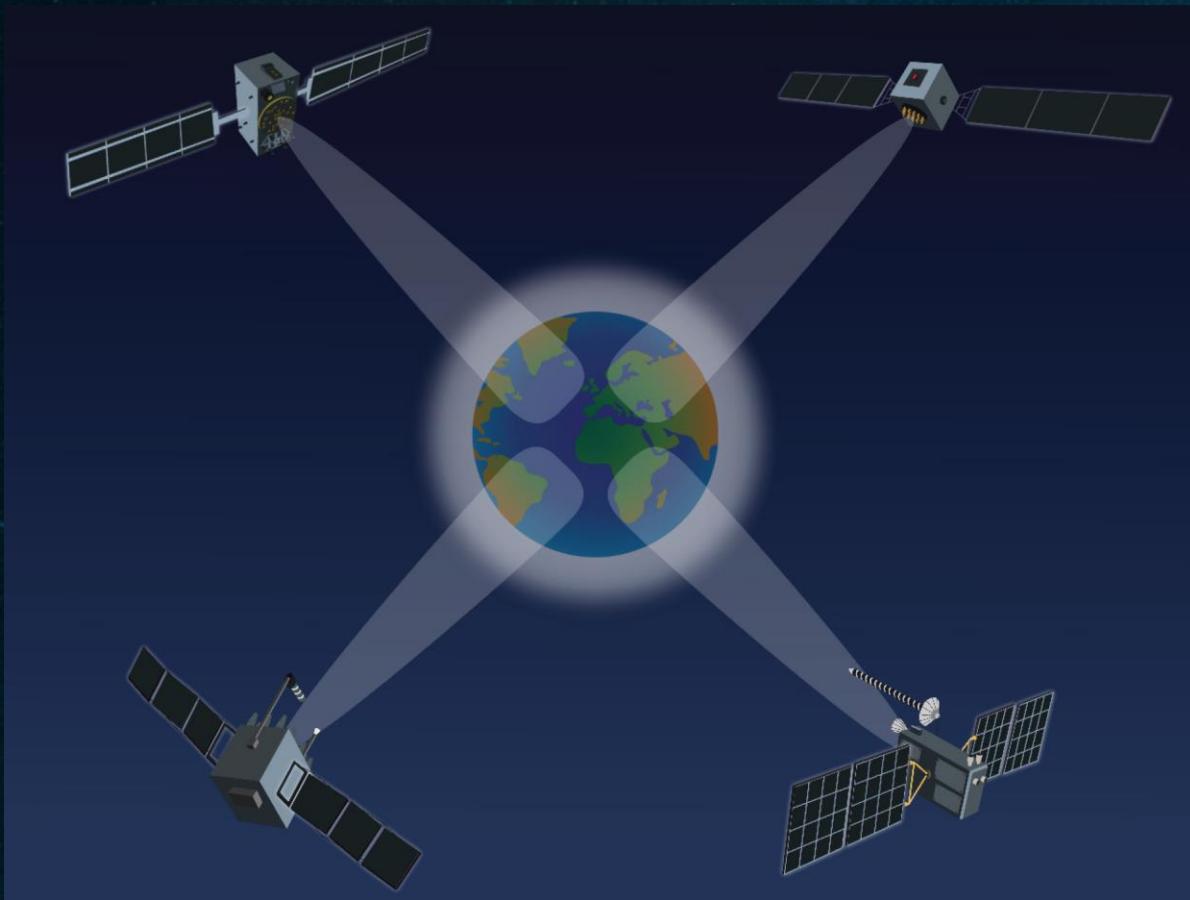


ESA Space Science Archives - Volume (PB)



IF Data

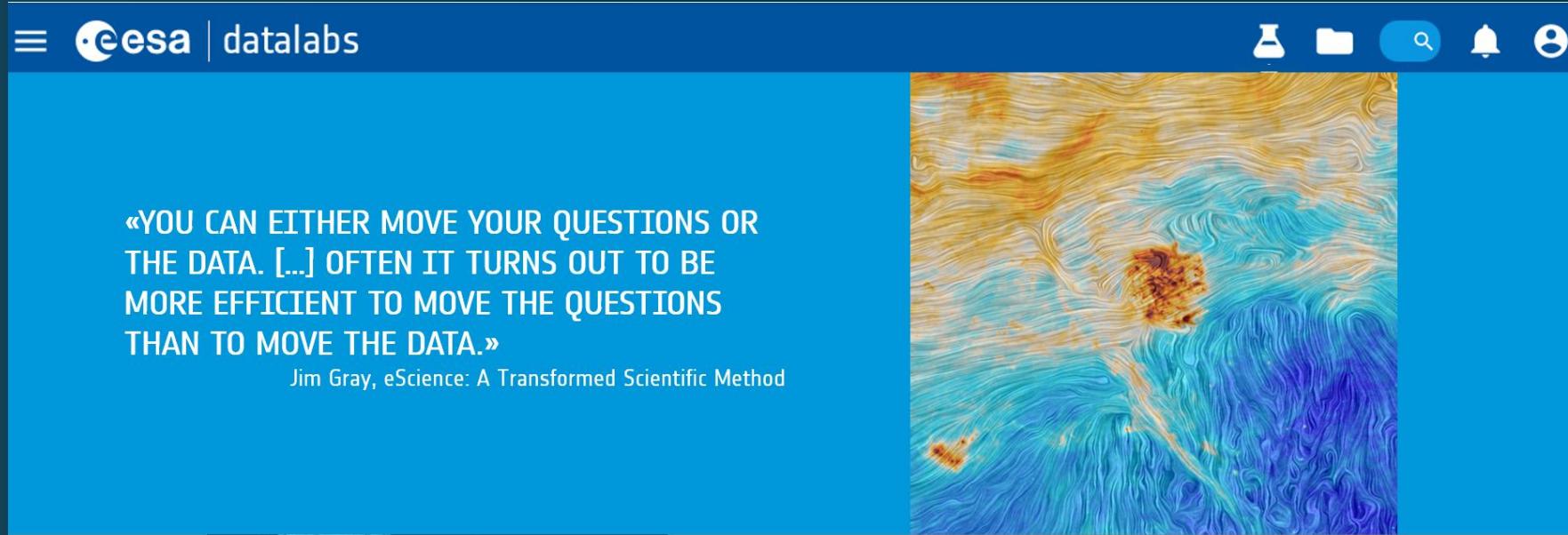
► High Potential for Discoveries and Innovations



- 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



≡  | 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



→ 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.

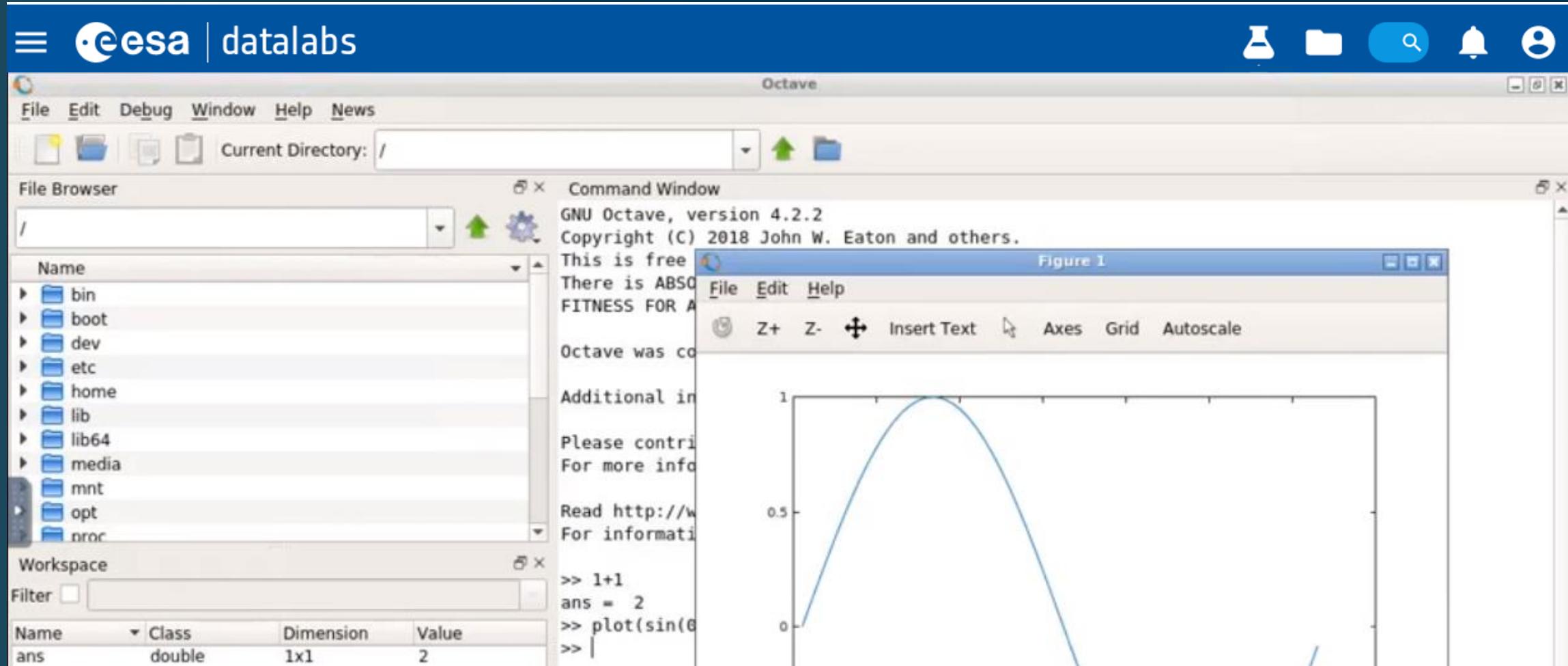
Analysis Services: JupyterLab



The screenshot shows the e esa JupyterLab environment. On the left, there is a sidebar with icons for navigation, a file browser showing a list of notebooks in the 'notebooks/' directory, and a search bar. The main area has a toolbar with icons for file operations like new, open, save, and run. There are three tabs open in the notebook editor: 'pyESASky-Basic_MLC.ipynb' (active), 'astroquery-demo.ipynb', and 'Gaia_astroquery_tutorial_clu'. The code editor displays Python code related to the pyESASky library, including imports for ESASkyWidget, CooFrame, ImgFormat, CatalogueDescriptor, MetadataDescriptor, and MetadataType, and definitions for 'esasky' and its configuration.

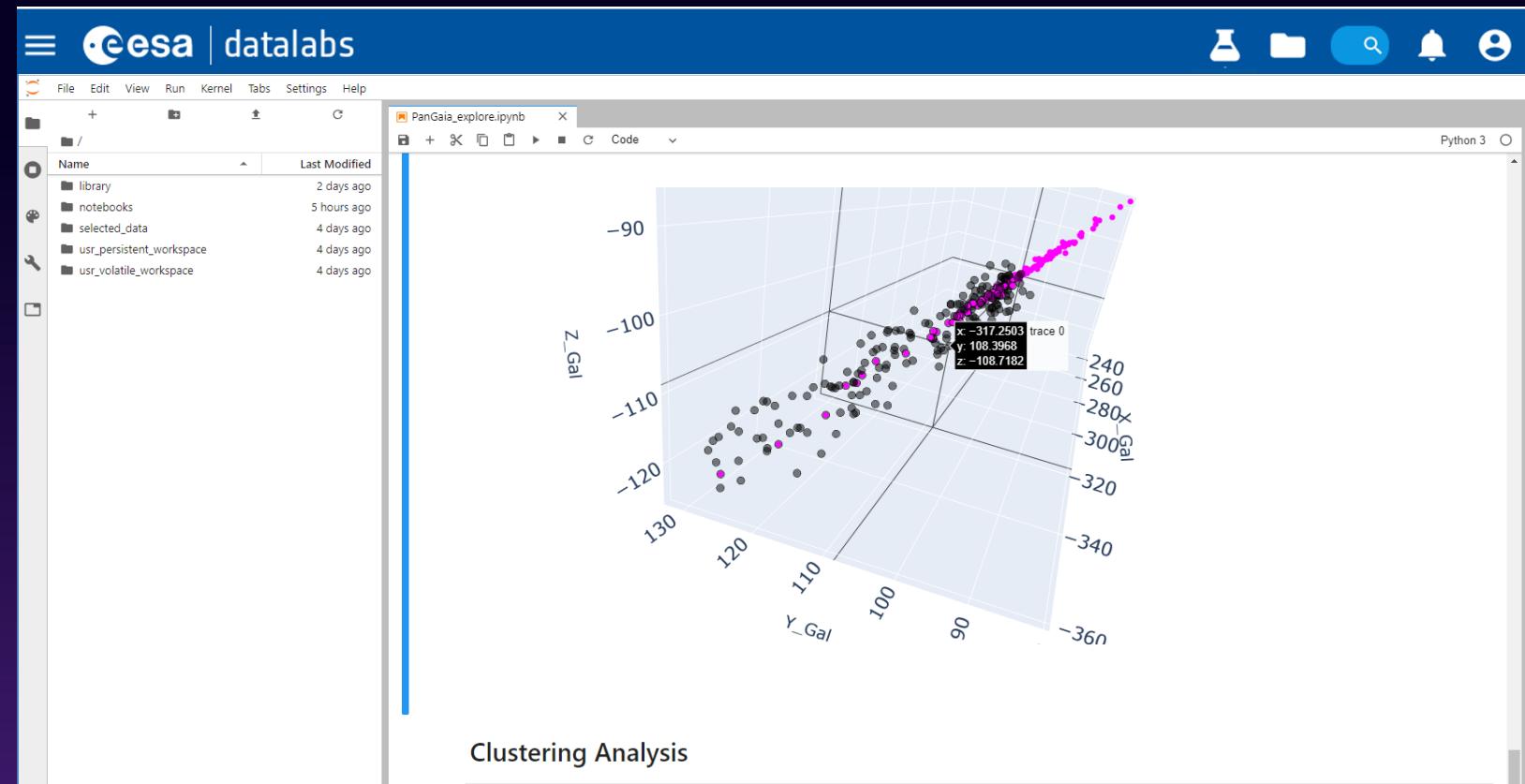
Customised, ready-to-use environment to maximise focus on scientific work

Desktop Services: Octave

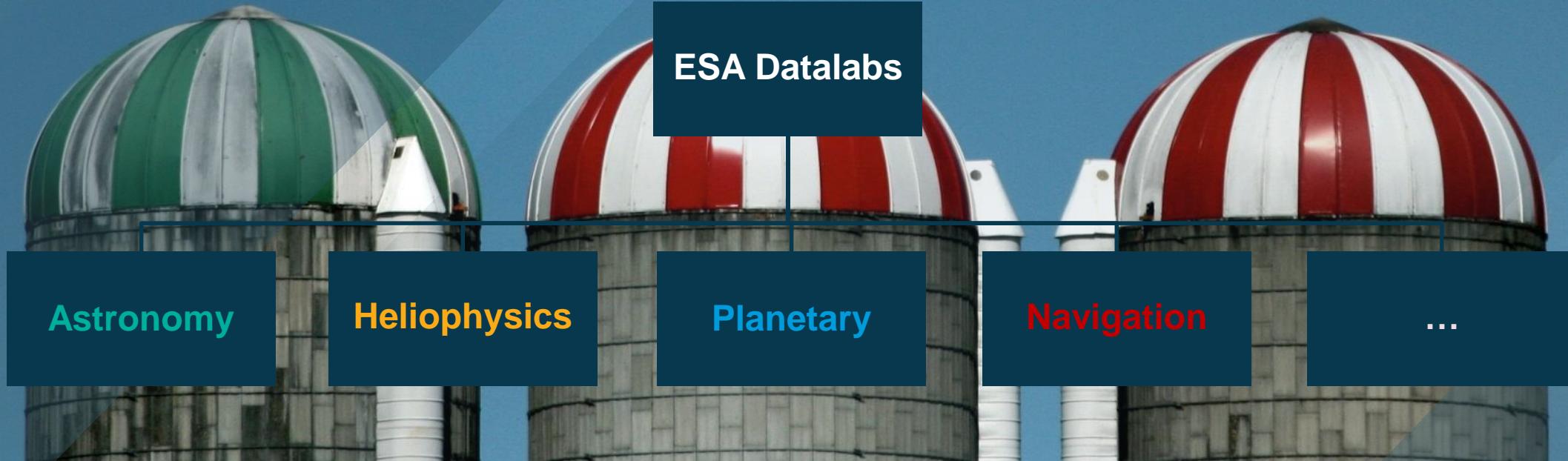


Full access via web browser to desktop based applications

HDBSCAN Clustering on Gaia DR2



Canovas et al. 2019b in preparation
P 3.3 Merin, Canovas et al.



Science Archives Integration: Planck



The screenshot shows a web-based archive interface for the Planck Legacy Archive. The top navigation bar includes links for 'EUROPEAN SPACE AGENCY' and 'SCIENCE & TECHNOLOGY'. The main title 'Planck Legacy Archive' is displayed above a red header banner featuring the Planck satellite and the 'e esa' logo.

The central area is titled 'RESULTS' and displays a list of 'FREQUENCY MAPS (32)'. Each entry in the table includes a checkbox, a preview icon, and a file name. The columns represent various metadata fields: Map name, Size, Resolution, Full, Yes/No, No, N/A, and Instrument (HFI or LFI). The table shows a mix of HFI and LFI map types, with sizes ranging from 144 MB to 1.9 GB.

Two callout boxes provide instructions for interacting with the data:

- A box on the right side of the table lists:
 - Open "All" the selected in a Notebook
 - Add "All" the selected to Data Volume
- A box at the bottom left of the interface lists:
 - Open in a Notebook
 - Add to Data Volume

At the bottom of the page, there is a footer bar with the text 'COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED (V-3.4-oper)' and a row of international flags.

Science Archives Integration: GNSS



gnss science support centre

esa

Filters | Labs | Collections

Logout

Collections

- GREAT
- GOCE RINEX
- SWARM RINEX
- More

Categories

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

Organization

- IGS 178
- ESOC 256

Domains

- Earth Science 223

Search GNSS assets

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

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

Sort by Most relevant

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:

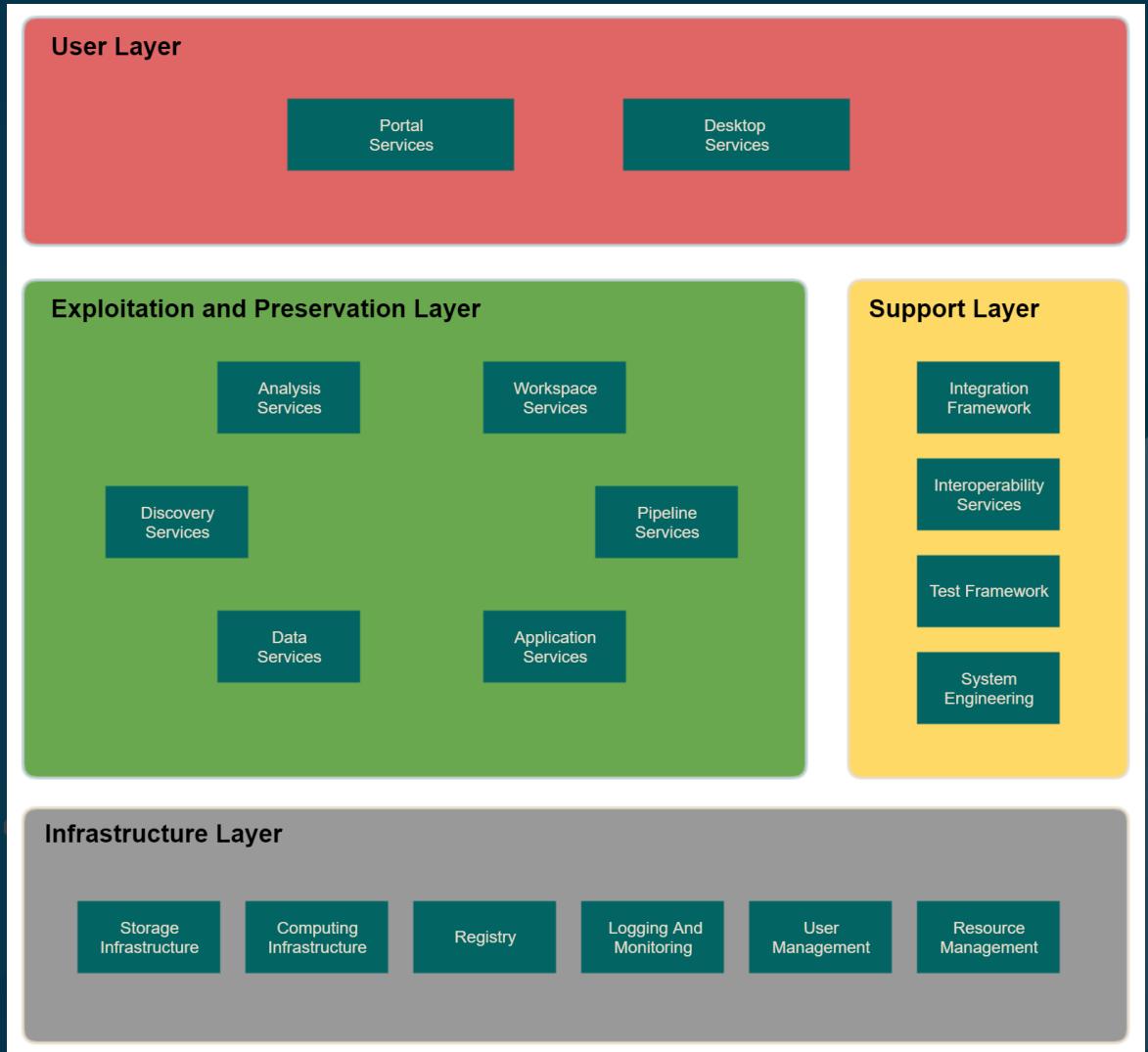
Tags: Atmosphere Experiment Test

Tools: Filter, Refresh, Download, CSV, Grid

- Open in a Notebook
- Add to Data Volume

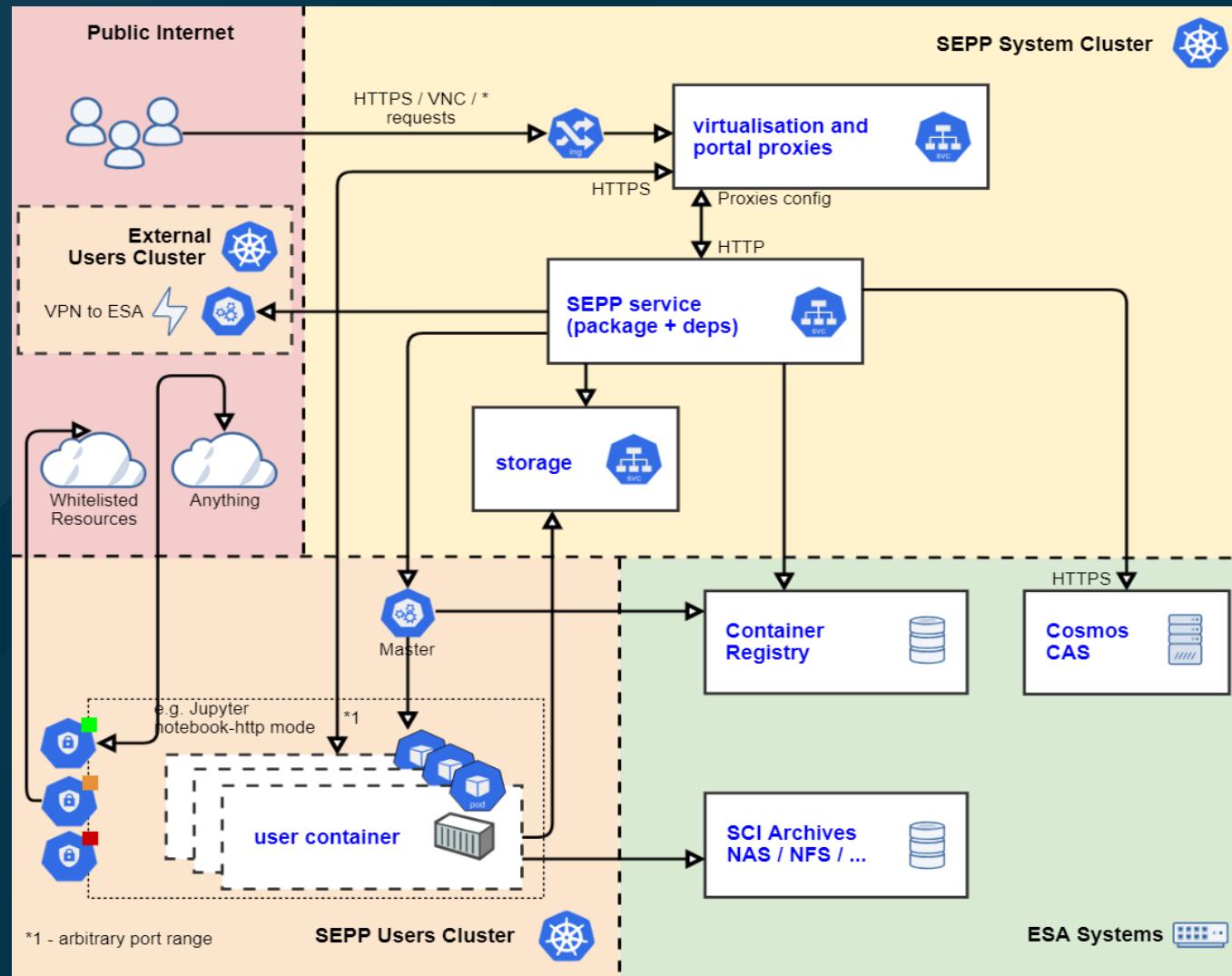
What's inside?

Reference Architecture
leverages on mainstream
technologies



P 9.18 Del Rio, Navarro et al.

Container Management with Kubernetes



GraphQL Interface



ESA Datalabs - Chromium

ESA Datalabs Not secure | datalabs.esa.int:3333/dlview/1

eesa | datalabs

History X GraphQL Prettify History

```
mutation {  
  createDatalab(sapId: "sepp/jupyterlab:v1") {  
    datalab {  
      id, base  
    }  
  }  
}
```

QUERY VARIABLES

Schema Mutations X

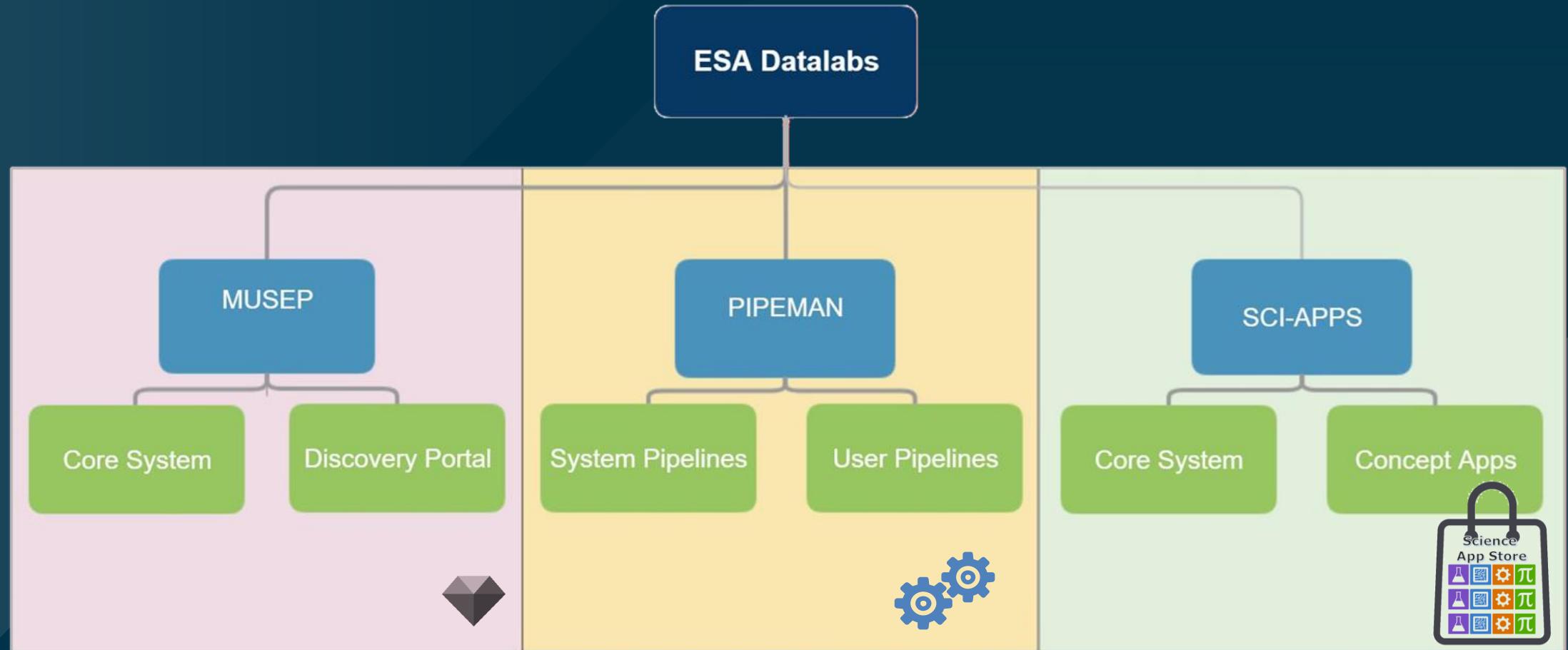
Search Mutations...

This is the root of all SEPP mutations.

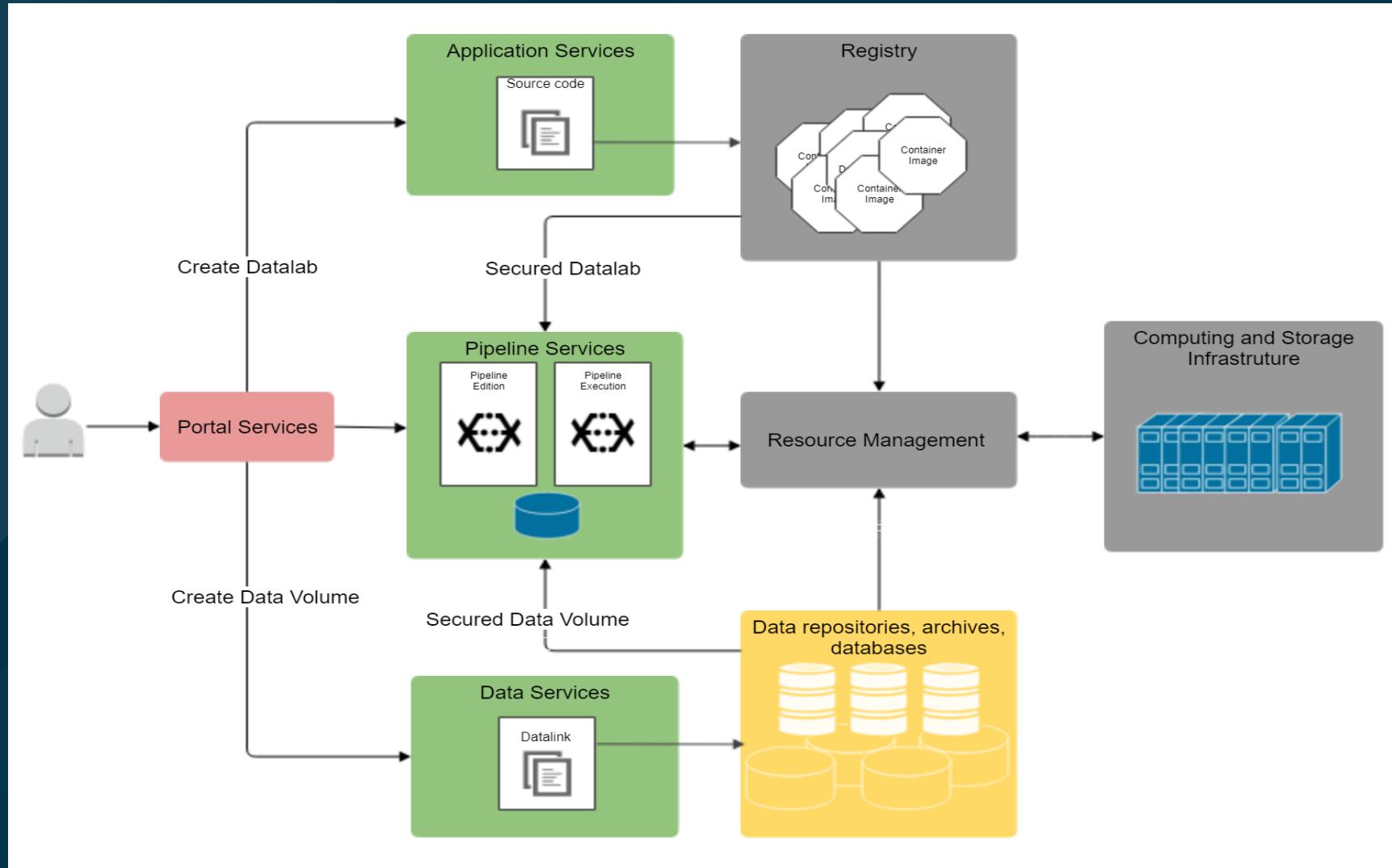
FIELDS

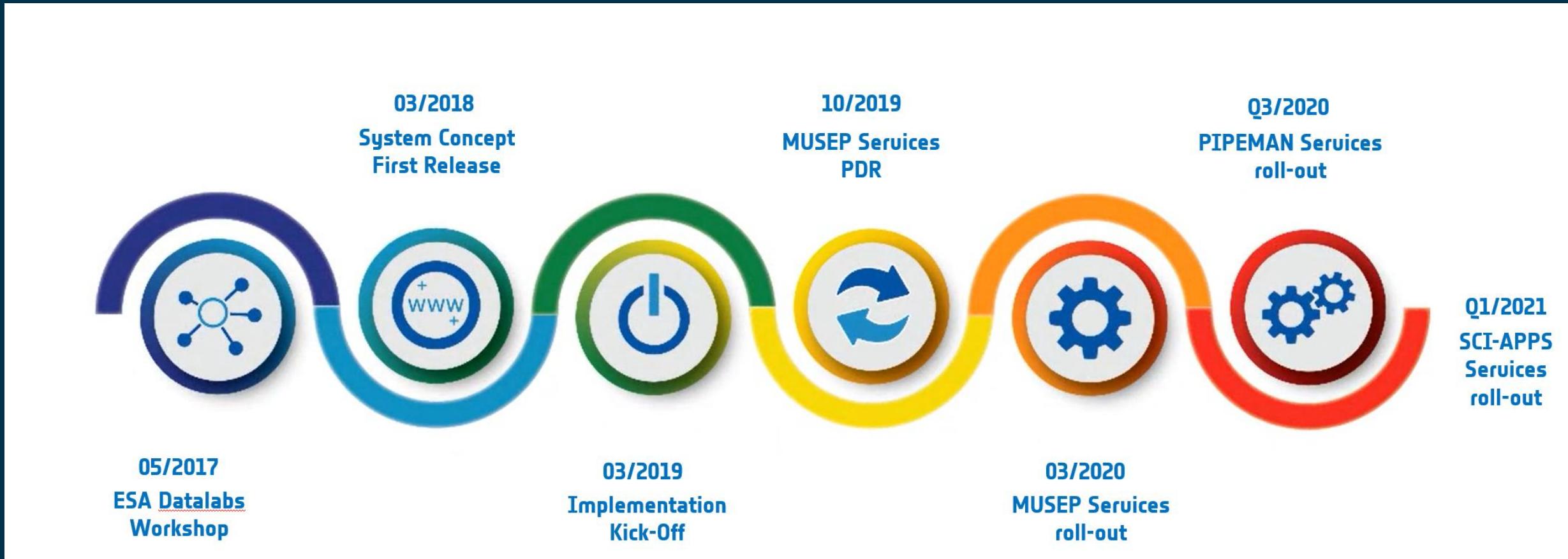
createDatalab(sapId: String): CreateDatalab
Create a new datalab for the current user

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



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

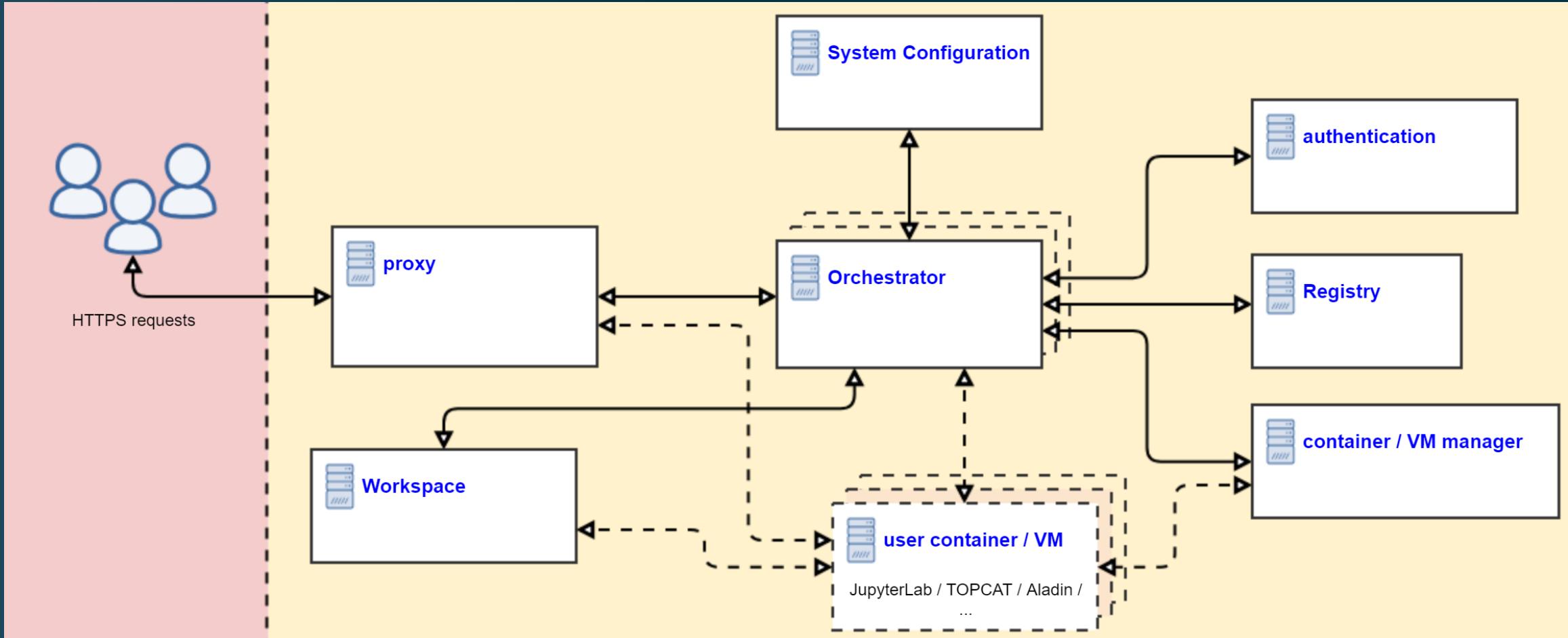
Thank You!



datalabs.esa.int

BackUp Slides

Container Management



≡ e esa | datalabs

Step 1 – Select a datalab from the catalog

Search X 🔍

Quick Launch New datalab
Launch New datalab: Advanced

Cancel Continue

Scrollable Grid or array of cards. Space for Icon, text, perhaps a link for details. A good example of this UI is Murano.

Item Title: 1
Item Subtitle: 1
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...

Item Title: 2
Item Subtitle: 2
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...

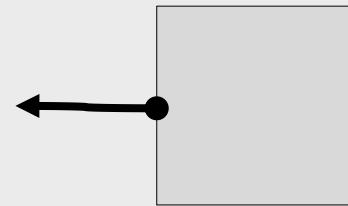
Item Title: 3
Item Subtitle: 3
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...

Search box for search inside catalogue

User can also select and click continue in this button

Step 1 – Select a type of Data Volume from the catalog

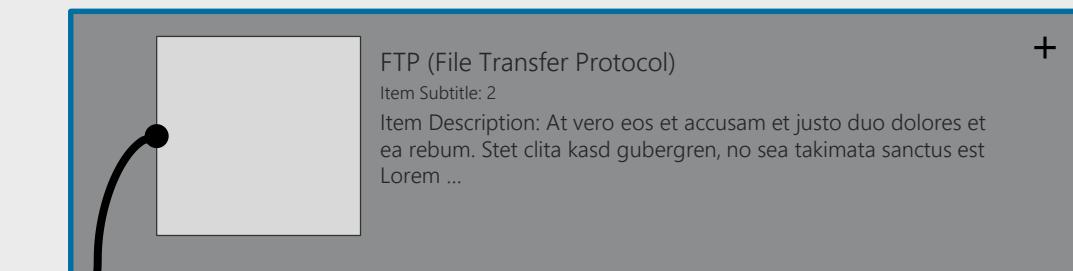
Predefined Collection
Should go to the search engine where the user can find a predefined collection and select one



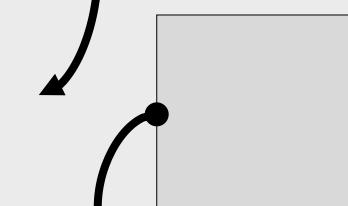
Predefined Collection
Item Subtitle: 1
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...

Consider also having predefined FTP servers and VO spaces, created beforehand

FTP and VO Space should lead to a configuration window where parameters can be set



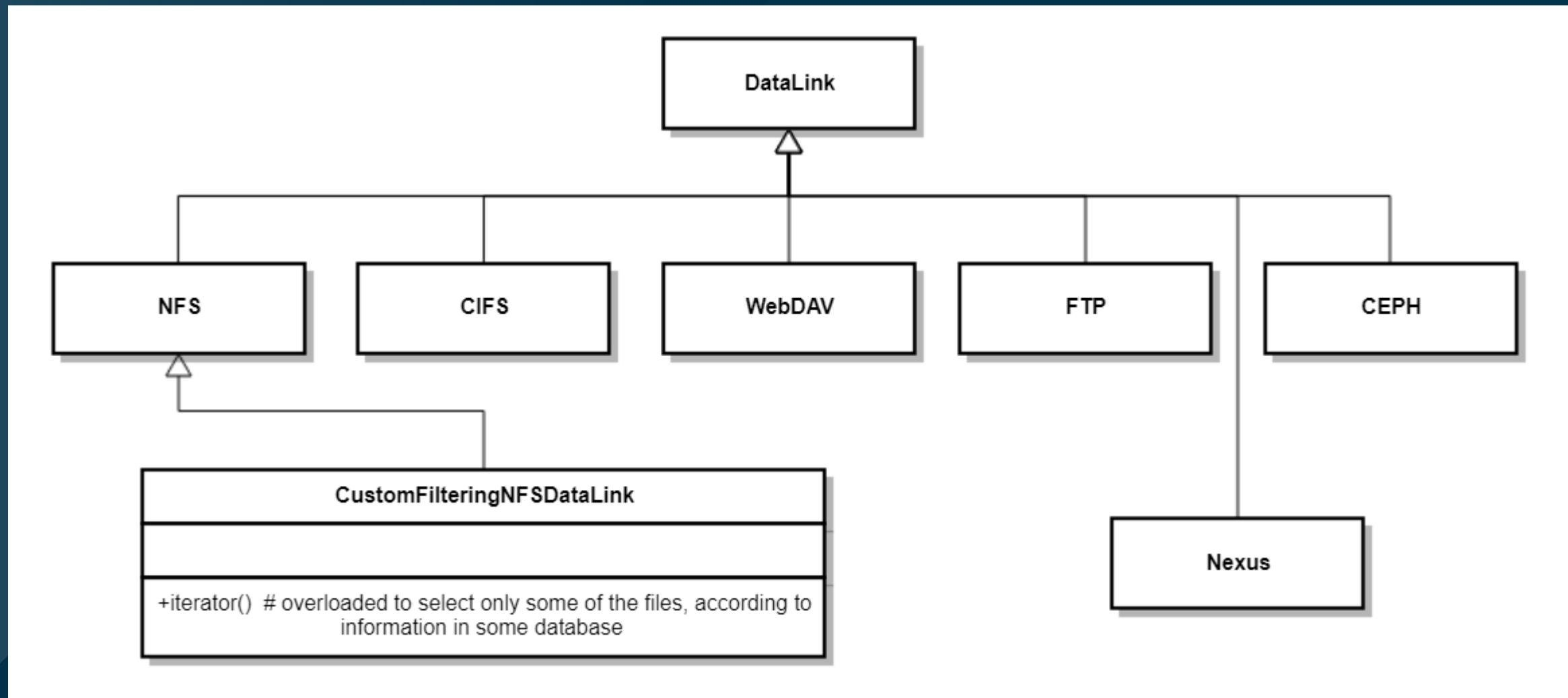
FTP (File Transfer Protocol)
Item Subtitle: 2
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...



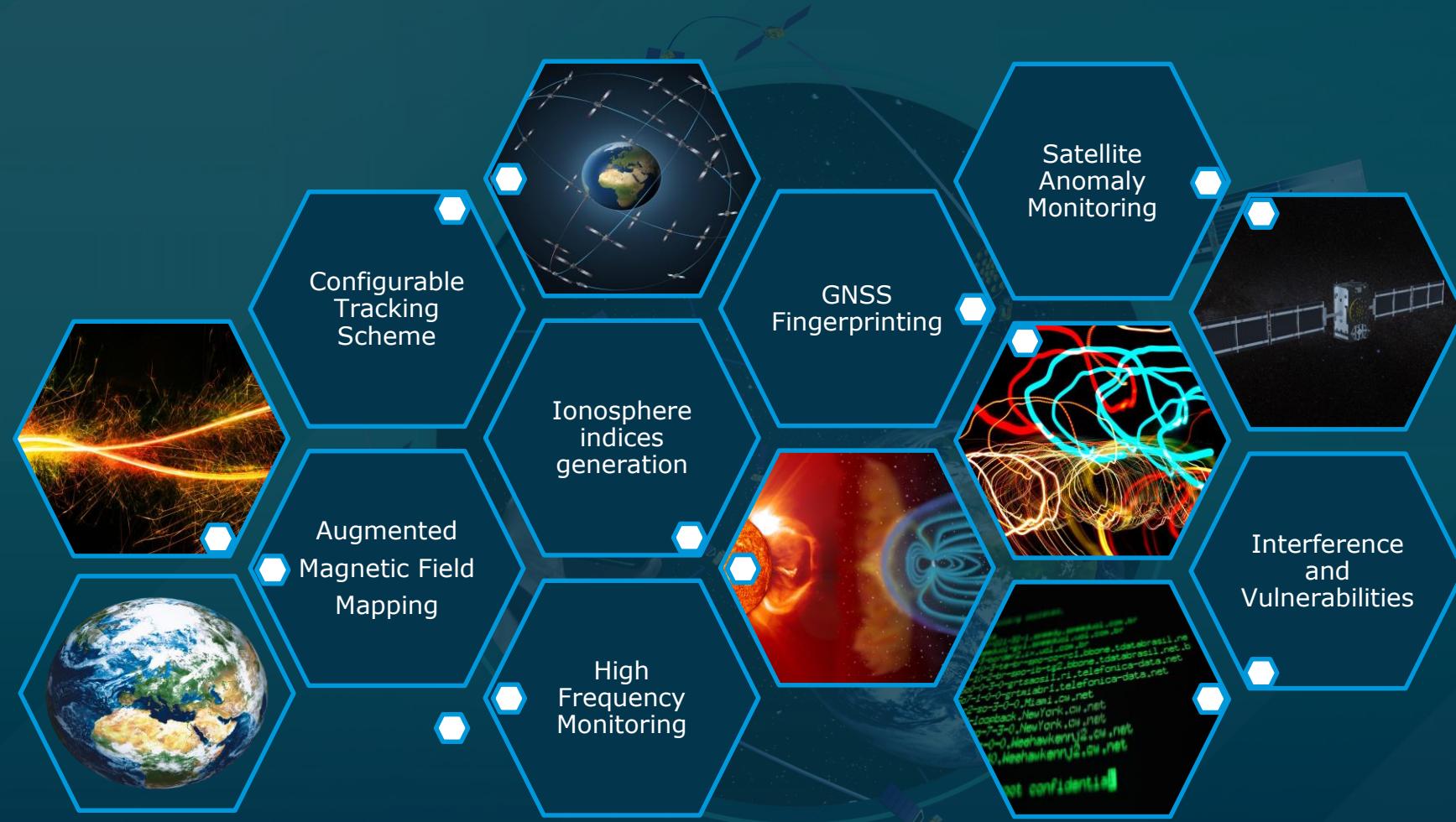
VO Space
Item Subtitle: 3
Item Description: At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ...

Cancel

Continue



GSSC Exploitation Platform Innovative Use Cases



GSSC Jupyter Notebook - Configuration



Visualisation Layer

- Matplotlib
- iPympl
- Beakerx
- iPyleaflet

GNSS Layer

- GSSC LoggerReader
- Georinex: load Rinex files
- Ephem: leap second calculation

Mathematical Layer

- Numpy: scientific computing
- Pandas: Data Analysis Library
- PyAstronomy: calculation of orbital parameters
- Pyproj: Reference Coordinate System Transformation

Core Layer

- NBconvert