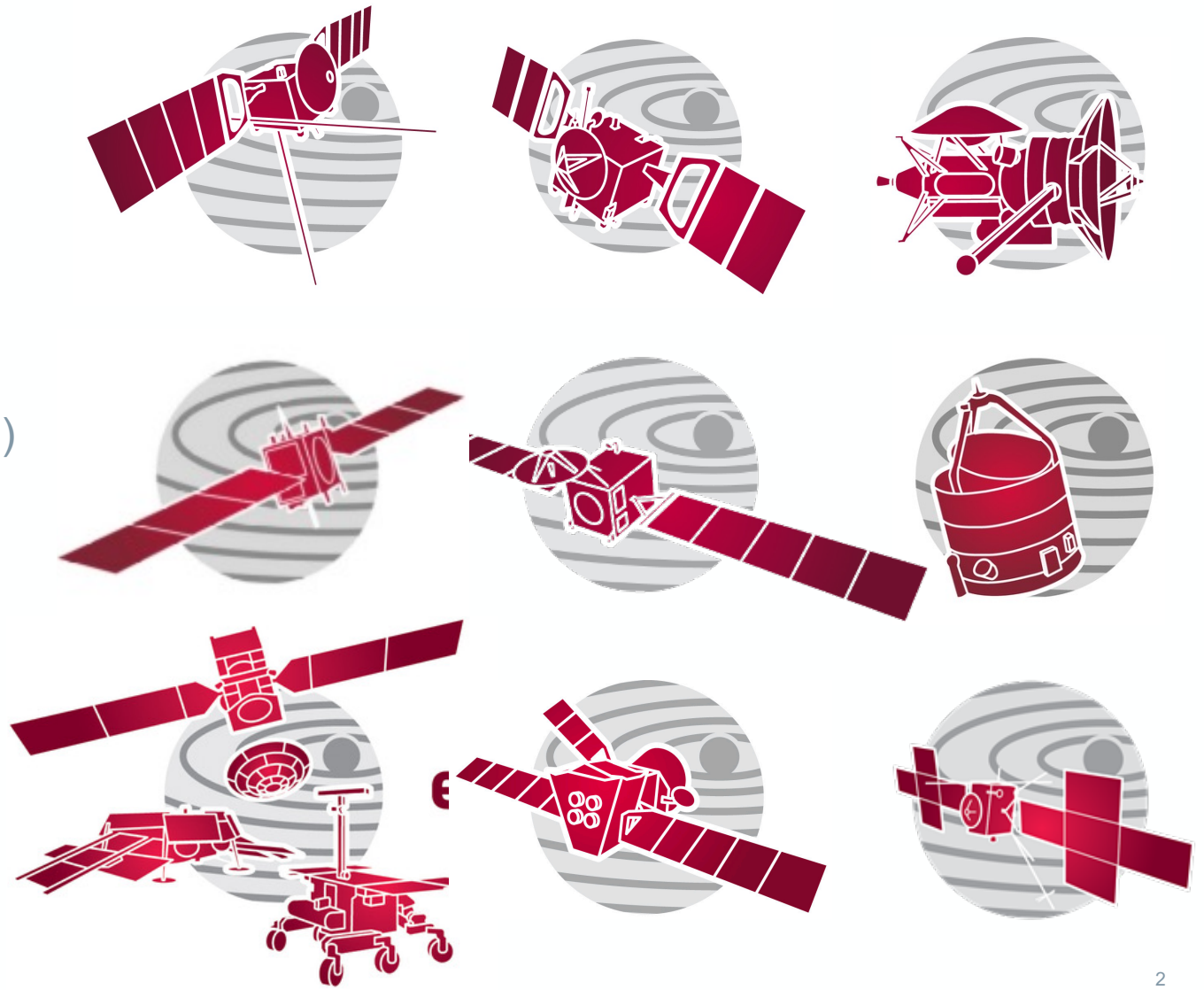
ESA's Planetary Science Archive: From data preservation to data exploitation

PSA

What is the PSA?

- PSA is a very multi-mission archive
  - saves cost on archive development
  - enables multi-mission search
  - single point of access
- With the upcoming missions (e.g. JUICE)
  - 10 space missions
  - >100 instruments (more with sub-units etc.)
  - huge diversity
    - of missions (orbiters, landers, flybys)
    - of instruments
    - or data formats and types
- This comes at a cost
  - hard to describe data uniformly
  - a lot is left as an exercise to the reader



The PSA Team

Planetary instrument teams

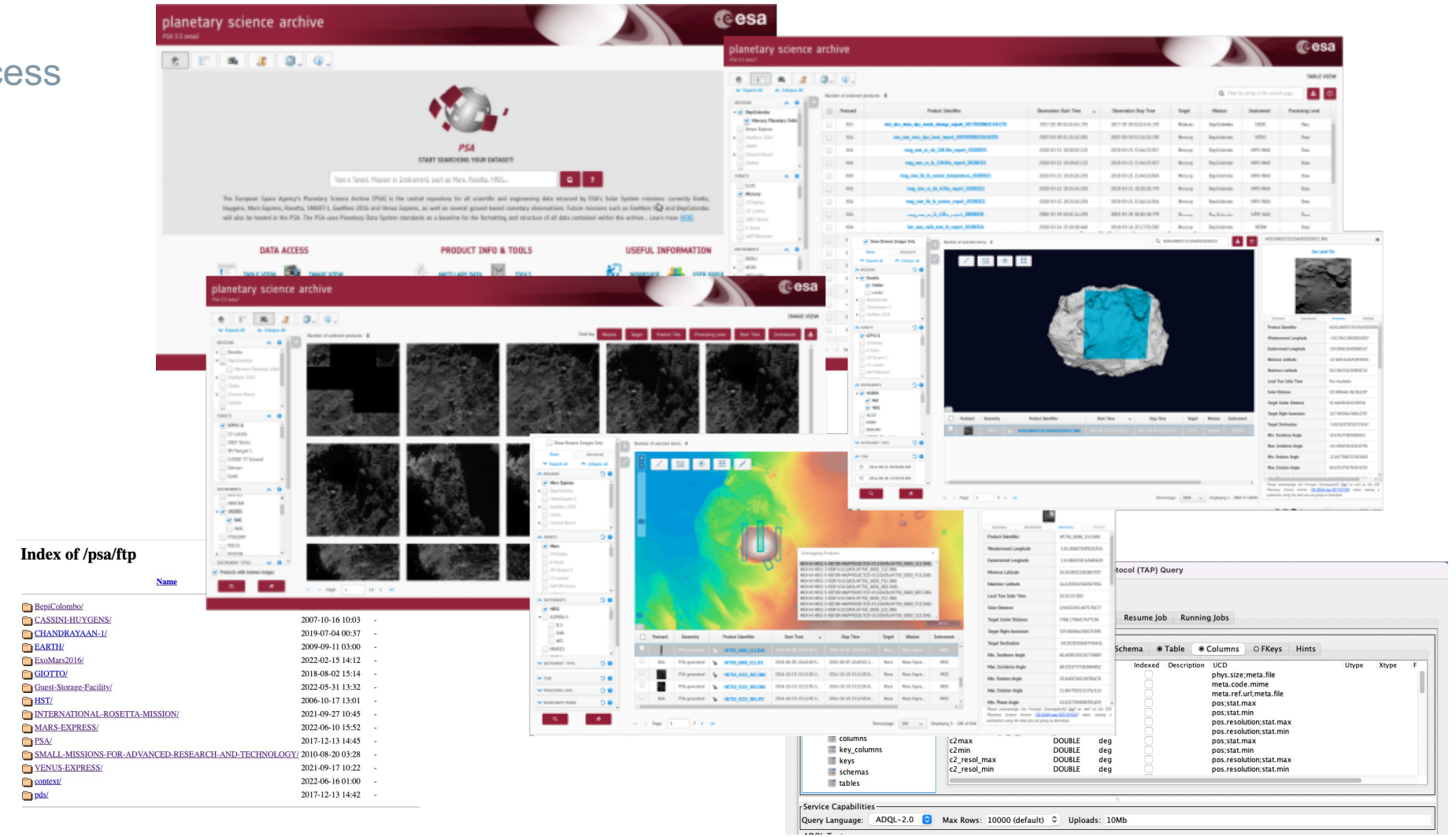
Archive scientists

PSA developers

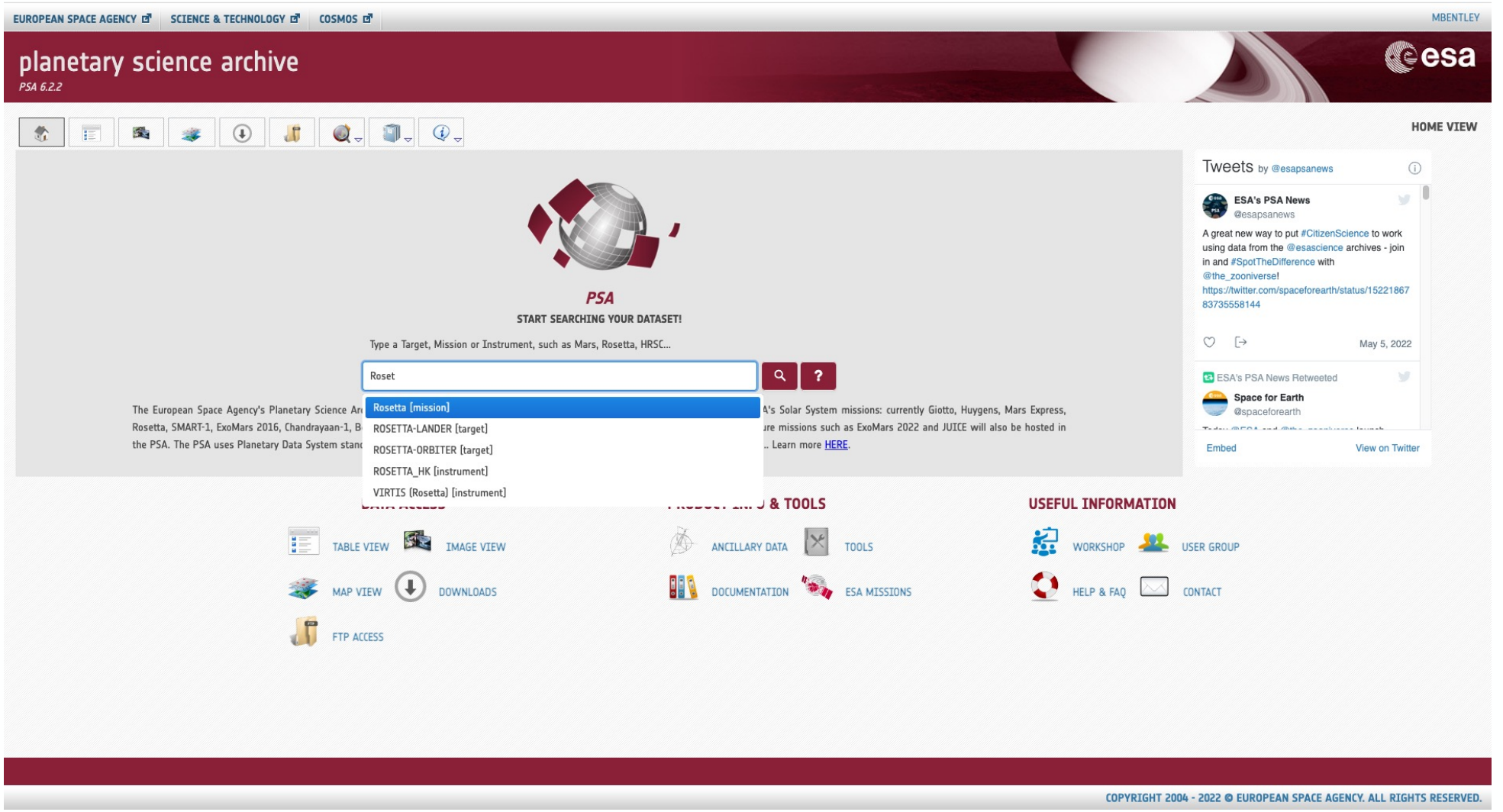
Scientific community

Current services

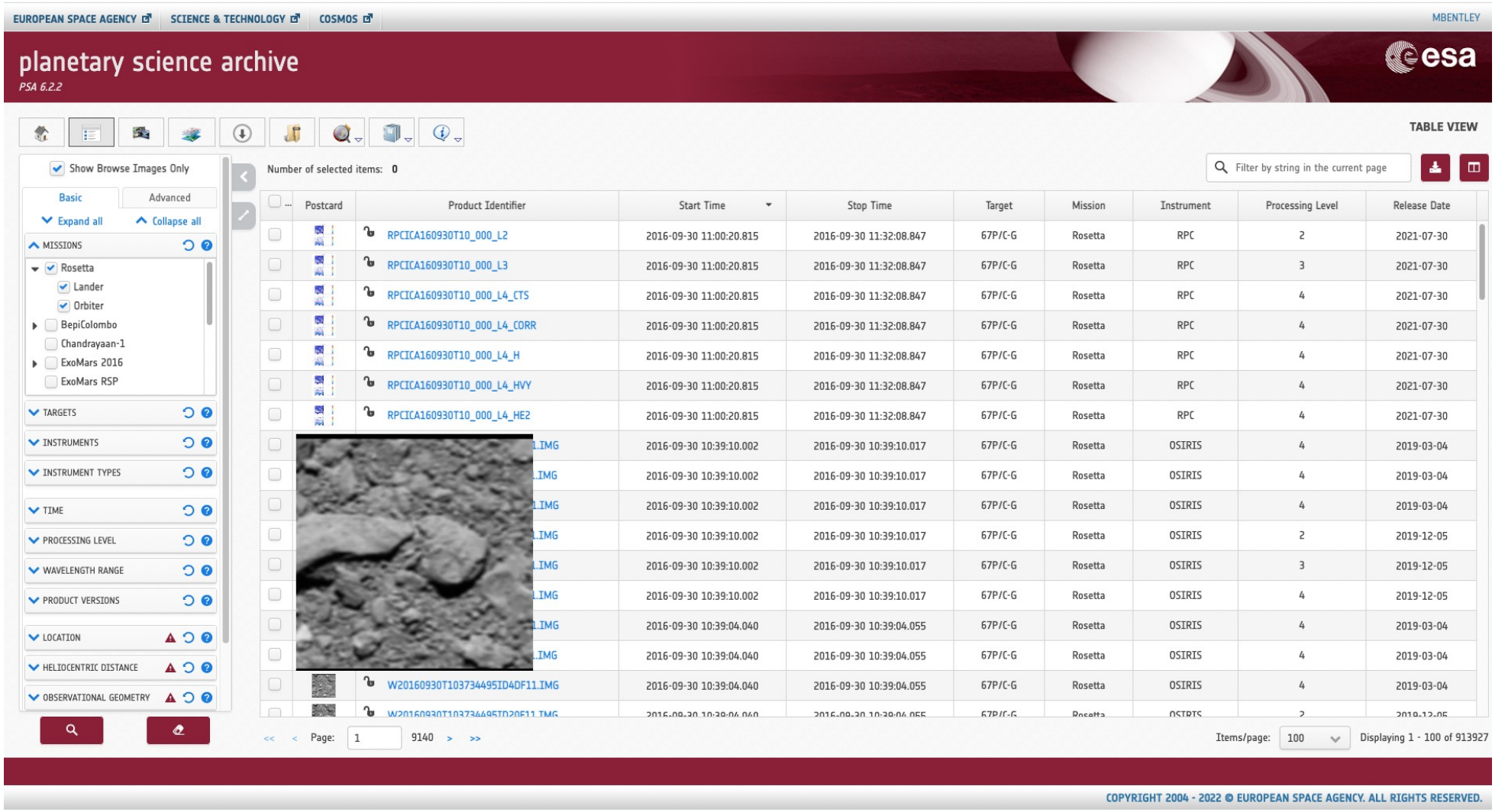
- FTP
  - and web (https) access
- Web UI
  - public + proprietary
  - table view
  - image gallery
  - 2D map view
  - comet view
- Machine access
  - PDAP
  - EPN-TAP



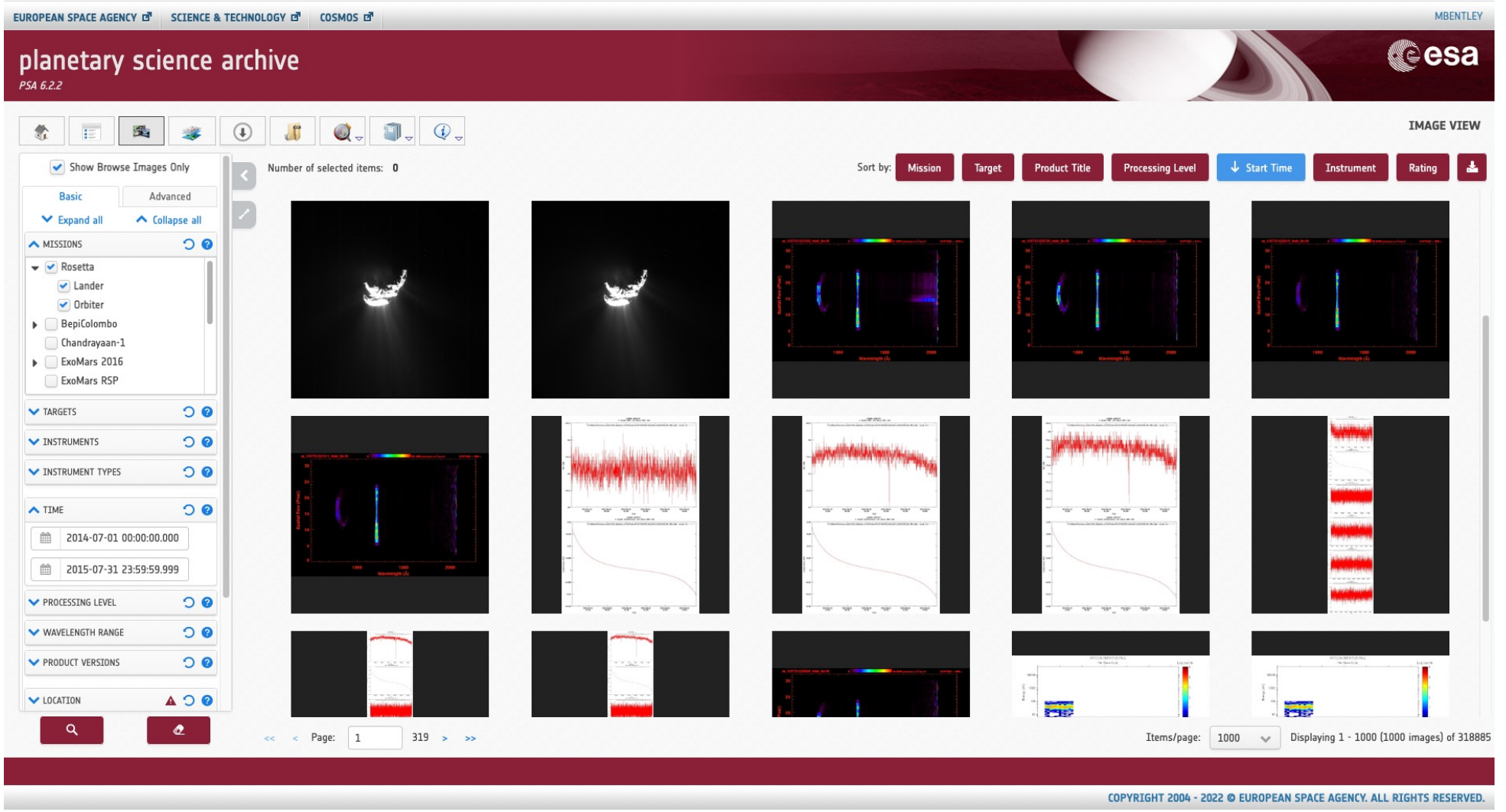
The home page



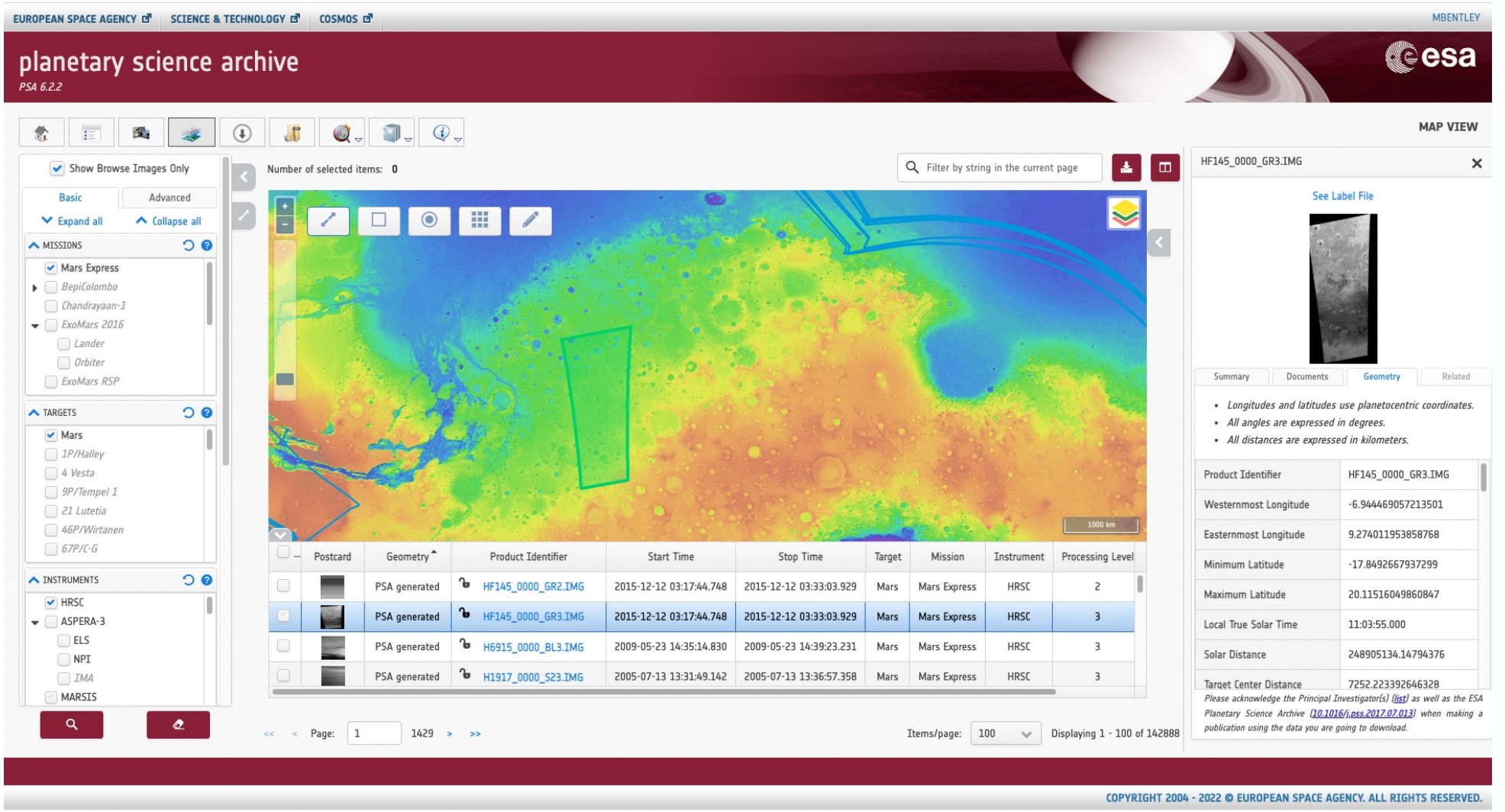
The table view



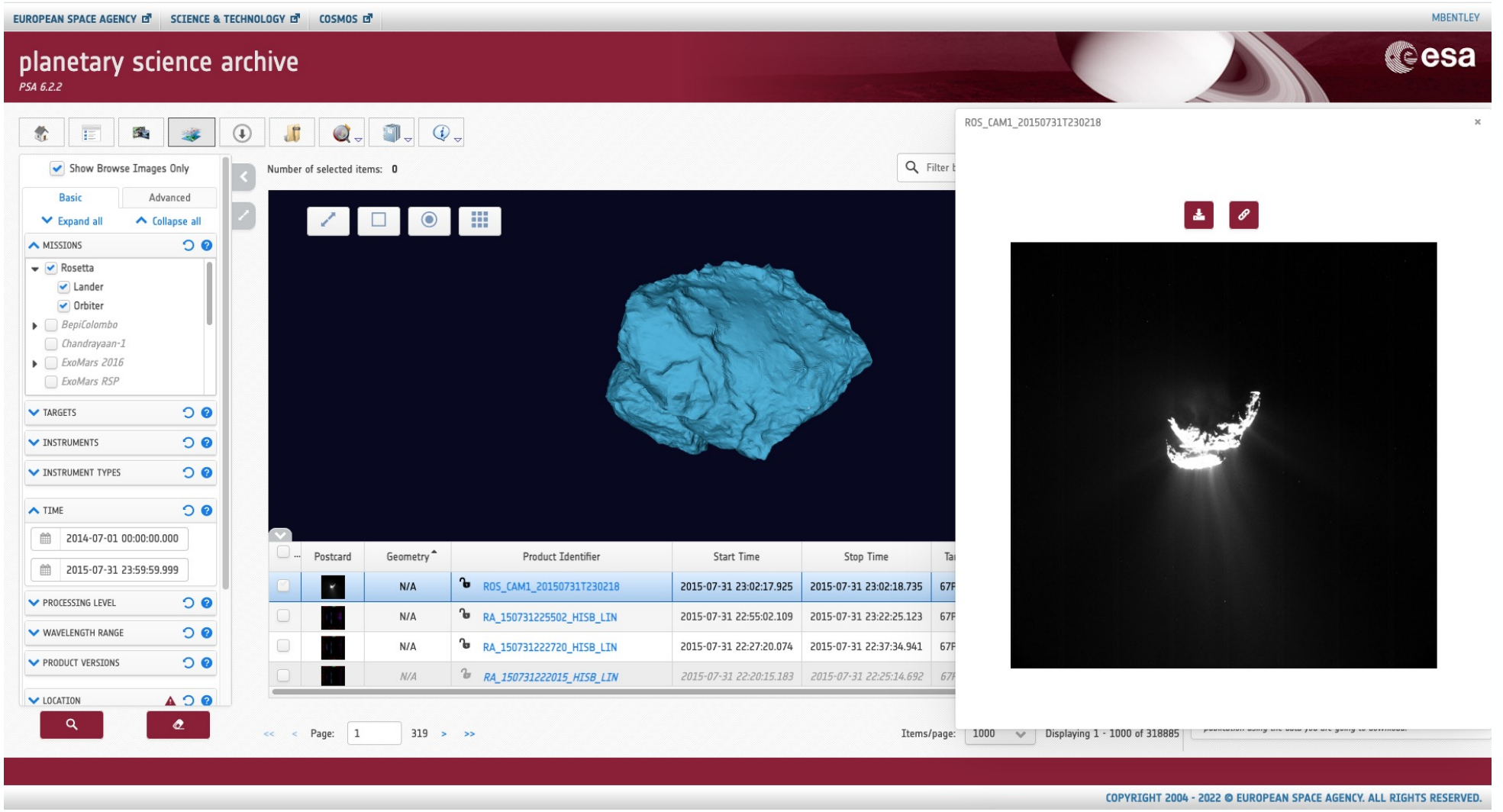
The gallery view



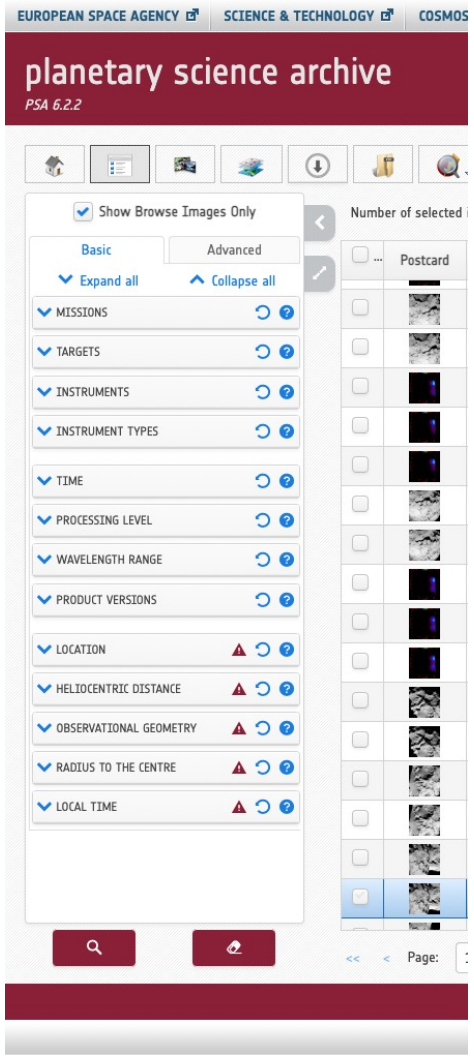
2D map view

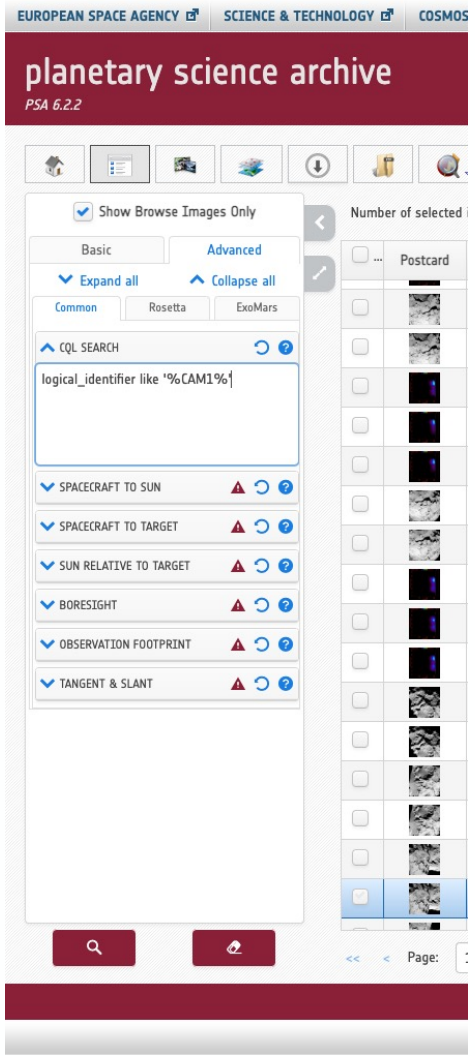

EUROPEAN SPACE AGENCY
SCIENCE & TECHNOLOGY
COSMOS
MBENTLEY
planetary science archive
PSA 6.2.2
MAP VIEW
Show Browse Images Only
Basic
Advanced
Expand all
Collapse all
MISSIONS
Mars Express
BepiColombo
Chandrayaan-1
ExoMars 2016
Lander
Orbiter
ExoMars RSP
TARGETS
Mars
1P/Halley
4 Vesta
9P/Tempel 1
21 Lutetia
46P/Wirtanen
67P/C-G
INSTRUMENTS
HRSC
ASPERA-3
ELS
NPI
IMA
MARSIS
Number of selected items: 0
Filter by string in the current page
HF145_0000_GR3.IMG
See Label File
Summary
Documents
Geometry
Related
Longitudes and latitudes use planetocentric coordinates.
All angles are expressed in degrees.
All distances are expressed in kilometers.
Product Identifier HF145_0000_GR3.IMG
Westernmost Longitude -6.944469057213501
Easternmost Longitude 9.274011953858768
Minimum Latitude -17.8492667937299
Maximum Latitude 20.11516049860847
Local True Solar Time 11:03:55.000
Solar Distance 248905134.14794376
Target Center Distance 7252.223392646328
Please acknowledge the Principal Investigator(s) (list) as well as the ESA Planetary Science Archive (10.1016/j.pss.2017.07.013) when making a publication using the data you are going to download.
Postcard
Geometry
Product Identifier
Start Time
Stop Time
Target
Mission
Instrument
Processing Level
PSA generated HF145_0000_GR2.IMG 2015-12-12 03:17:44.748 2015-12-12 03:33:03.929 Mars Mars Express HRSC 2
PSA generated HF145_0000_GR3.IMG 2015-12-12 03:17:44.748 2015-12-12 03:33:03.929 Mars Mars Express HRSC 3
PSA generated H6915_0000_BL3.IMG 2009-05-23 14:35:14.830 2009-05-23 14:39:23.231 Mars Mars Express HRSC 3
PSA generated H1917_0000_S23.IMG 2005-07-13 13:31:49.142 2005-07-13 13:36:57.358 Mars Mars Express HRSC 3
1000 km
<< < Page: 1 1429 > >>
Items/page: 100
Displaying 1 - 100 of 142888
COPYRIGHT 2004 - 2022 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.

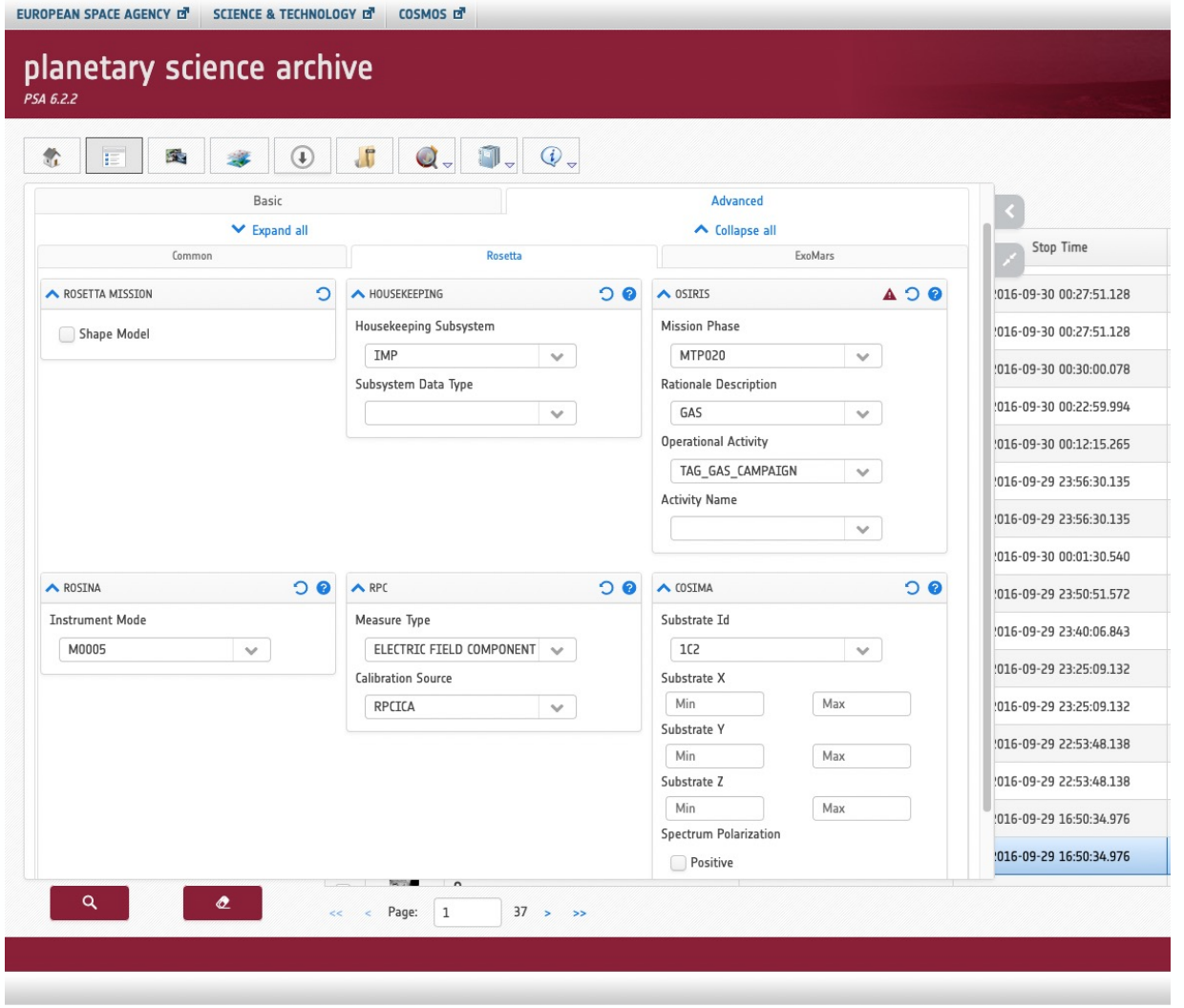
67P View


EUROPEAN SPACE AGENCY
SCIENCE & TECHNOLOGY
COSMOS
MBENTLEY
planetary science archive
PSA 6.2.2
Show Browse Images Only
Basic
Advanced
Expand all
Collapse all
MISSIONS
Rosetta
Lander
Orbiter
BepiColombo
Chandrayaan-1
ExoMars 2016
ExoMars RSP
TARGETS
INSTRUMENTS
INSTRUMENT TYPES
TIME
2014-07-01 00:00:00.000
2015-07-31 23:59:59.999
PROCESSING LEVEL
WAVELENGTH RANGE
PRODUCT VERSIONS
LOCATION
Number of selected items: 0
Postcard
Geometry
Product Identifier
Start Time
Stop Time
N/A
ROS_CAM1_20150731T230218
2015-07-31 23:02:17.925
2015-07-31 23:02:18.735
N/A
RA_150731225502_HISB_LIN
2015-07-31 22:55:02.109
2015-07-31 23:22:25.123
N/A
RA_150731222720_HISB_LIN
2015-07-31 22:27:20.074
2015-07-31 22:37:34.941
N/A
RA_150731222015_HISB_LIN
2015-07-31 22:20:15.183
2015-07-31 22:25:14.692
Page: 1 319
Items/page: 1000
Displaying 1 - 1000 of 318885
ROS_CAM1_20150731T230218
COPYRIGHT 2004 - 2022 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.

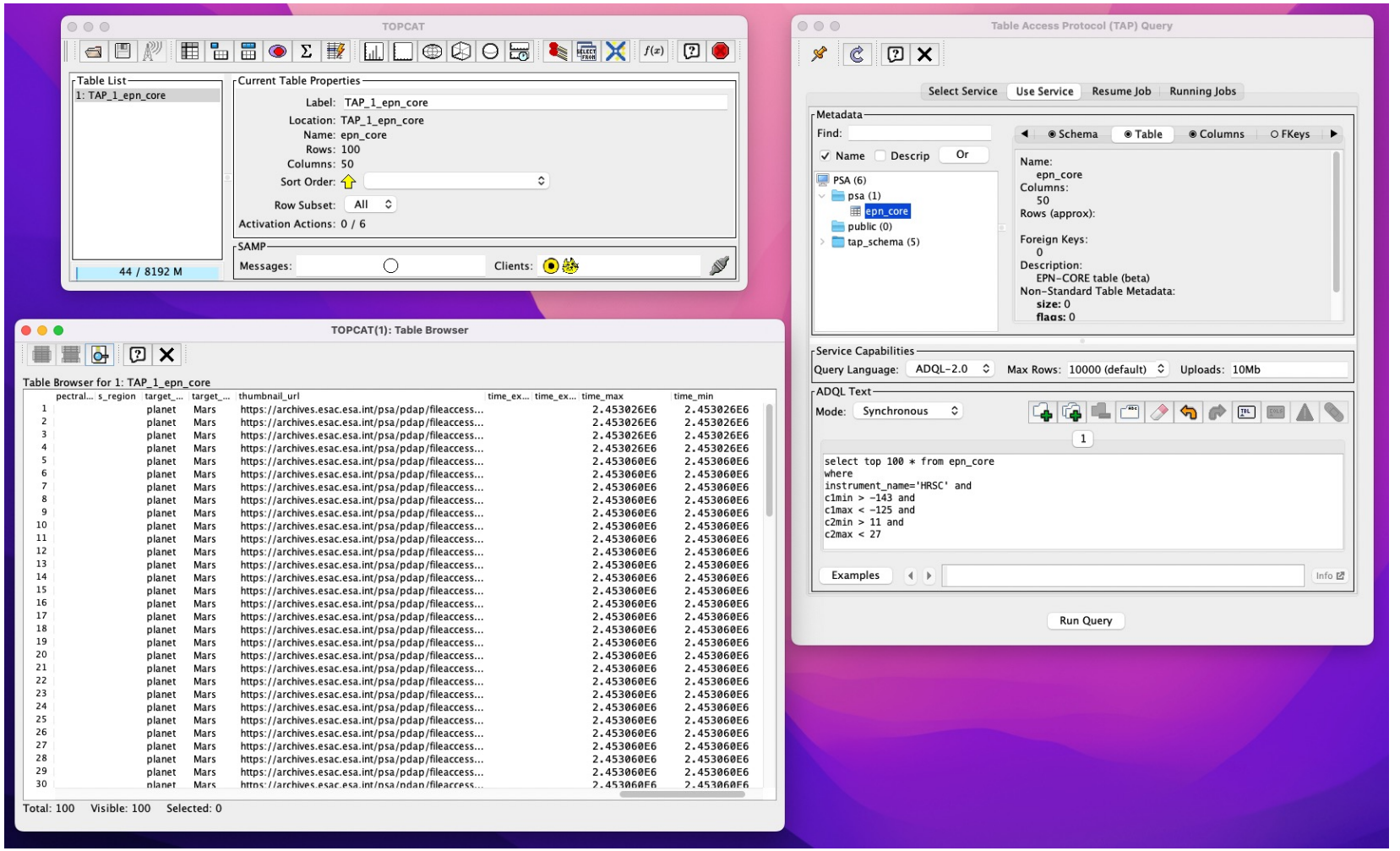
Common features: search and filter







EPN-TAP


TOPCAT
Table List
1: TAP_1_epn_core
44 / 8192 M
Current Table Properties
Label: TAP_1_epn_core
Location: TAP_1_epn_core
Name: epn_core
Rows: 100
Columns: 50
Sort Order:
Row Subset: All
Activation Actions: 0 / 6
SAMP
Messages:
Clients:
TOPCAT(1): Table Browser
Table Browser for 1: TAP_1_epn_core
pectral... s_region target_... target_... thumbnail_url time_ex... time_ex... time_max time_min
1 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453026E6 2.453026E6
2 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453026E6 2.453026E6
3 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453026E6 2.453026E6
4 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453026E6 2.453026E6
5 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
6 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
7 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
8 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
9 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
10 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
11 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
12 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
13 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
14 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
15 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
16 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
17 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
18 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
19 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
20 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
21 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
22 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
23 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
24 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
25 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
26 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
27 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
28 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
29 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
30 planet Mars https://archives.esac.esa.int/psa/pdap/fileaccess... 2.453060E6 2.453060E6
Total: 100 Visible: 100 Selected: 0
Table Access Protocol (TAP) Query
Select Service Use Service Resume Job Running Jobs
Metadata
Find:
Name Descrip Or
Schema Table Columns FKeys
PSA (6)
psa (1)
epn_core
public (0)
tap_schema (5)
Name:
epn_core
Columns:
50
Rows (approx):
Foreign Keys:
0
Description:
EPN-CORE table (beta)
Non-Standard Table Metadata:
size: 0
flags: 0
Service Capabilities
Query Language: ADQL-2.0
Max Rows: 10000 (default)
Uploads: 10Mb
ADQL Text
Mode: Synchronous
1
select top 100 * from epn_core
where
instrument_name='HRSC' and
c1min > -143 and
c1max < -125 and
c2min > 11 and
c2max < 27
Examples
Info
Run Query

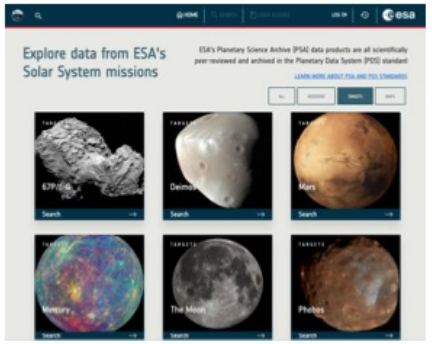
User feedback

- User feedback is solicited in various ways
  - at conferences and workshops
  - through the PSA User Group
  - through dedicated activities (e.g. UX workshop, surveys)
  - through the PSA helpdesk
- Key messages:
  - user experience needs to be improved
    - consistency in user interface, better docs, fewer crashes
  - generally users can find data, but then what?
    - they get a zip with a bunch of random files in a complex directory structure
    - how should they proceed?
  - downloading large quantities of data is painful
  - limited meta-data search (only specific meta-data ingested)

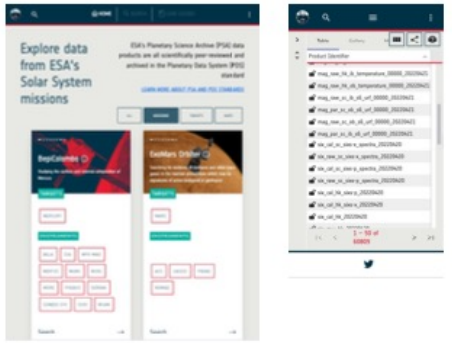
Where we stand

- Preservation
  - ensured through
    - PDS standards
    - hard work of instrument teams and archive scientists
    - commitment to long-term storage
  - PDS3 → PDS4 conversion will be important to bring legacy data into modern world
- Discovery
  - basic search works well across missions/instruments/targets
  - APIs work, but are limited in scope and not efficient for data science
  - GEOGEN (geometry calculation tool) is helping geospatial search
  - DOIs being issued at dataset (PDS3) and collection (PDS4) level
- Exploitation
  - multi-mission archive means we don't "dig into" the data (visualise, calculate etc.)
  - most of the work is left as an exercise for the reader

Moving towards data exploitation



The new UI design accounts for user feedback and is responsive



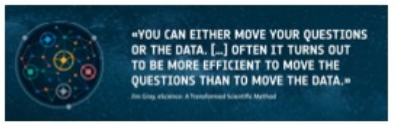
It will adapt to the screen size on computers, tablets and phones

User Interface

Two TAP services will be available: (PSA TAP and EPN-TAP) and a TAP-compliant download service

Integration of the PDS registry and API will allow search by arbitrary meta-data across all PDS4 products

APIs



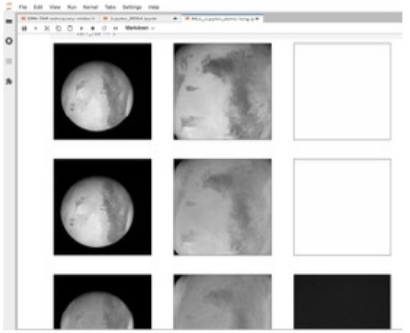
DataLabs allows desktop and web applications to be run next to the archive

A pre-configured Jupyter environment will be provided to work with PSA data

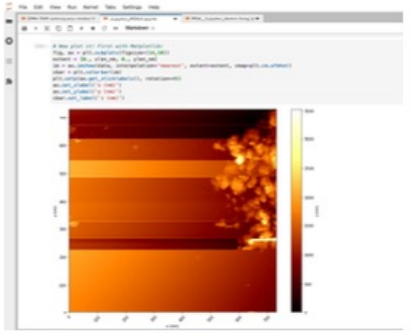
TAP will be updated to allow users to locate and load "local" archive data

Eventually products can also be sent from the user interface to DataLabs

DataLabs



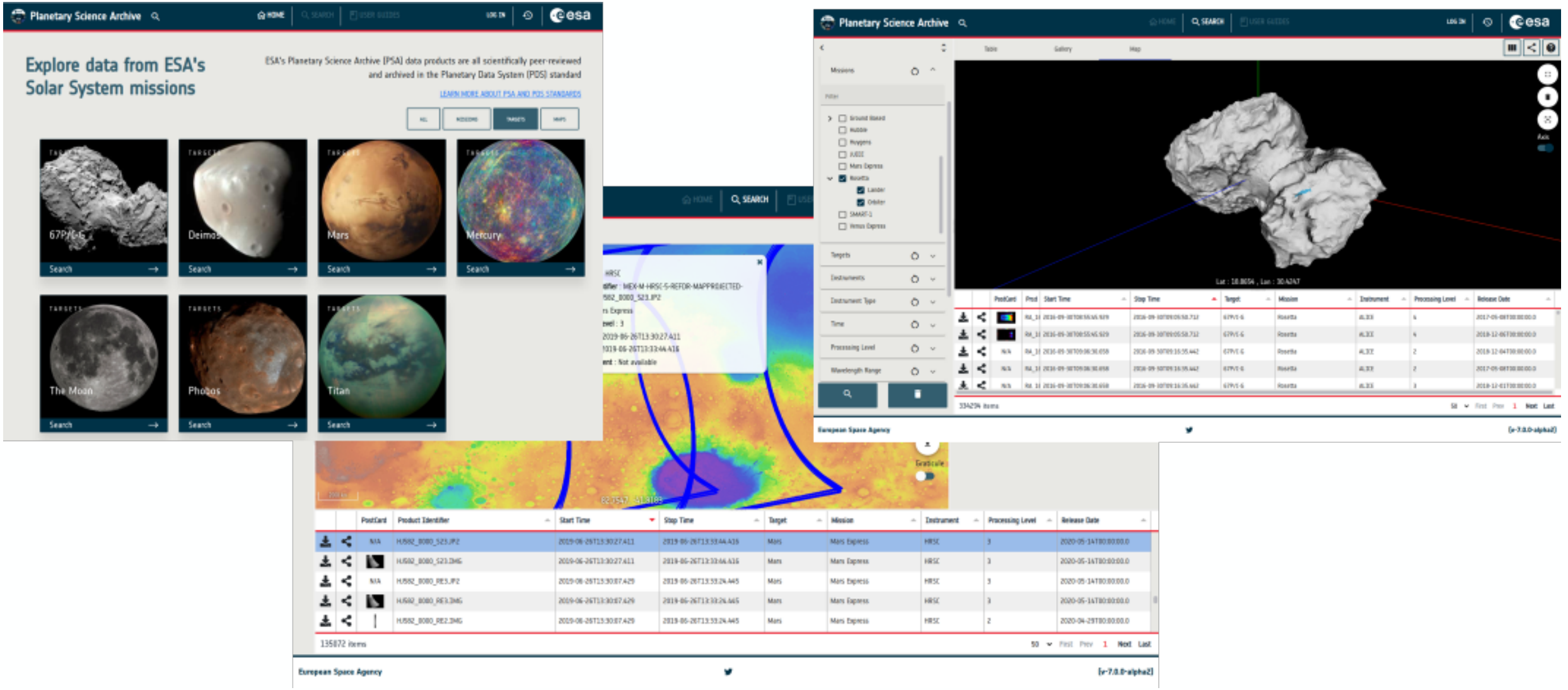
Tutorials will be developed with the instrument teams to demonstration how to find/slice/dice/visualise data products



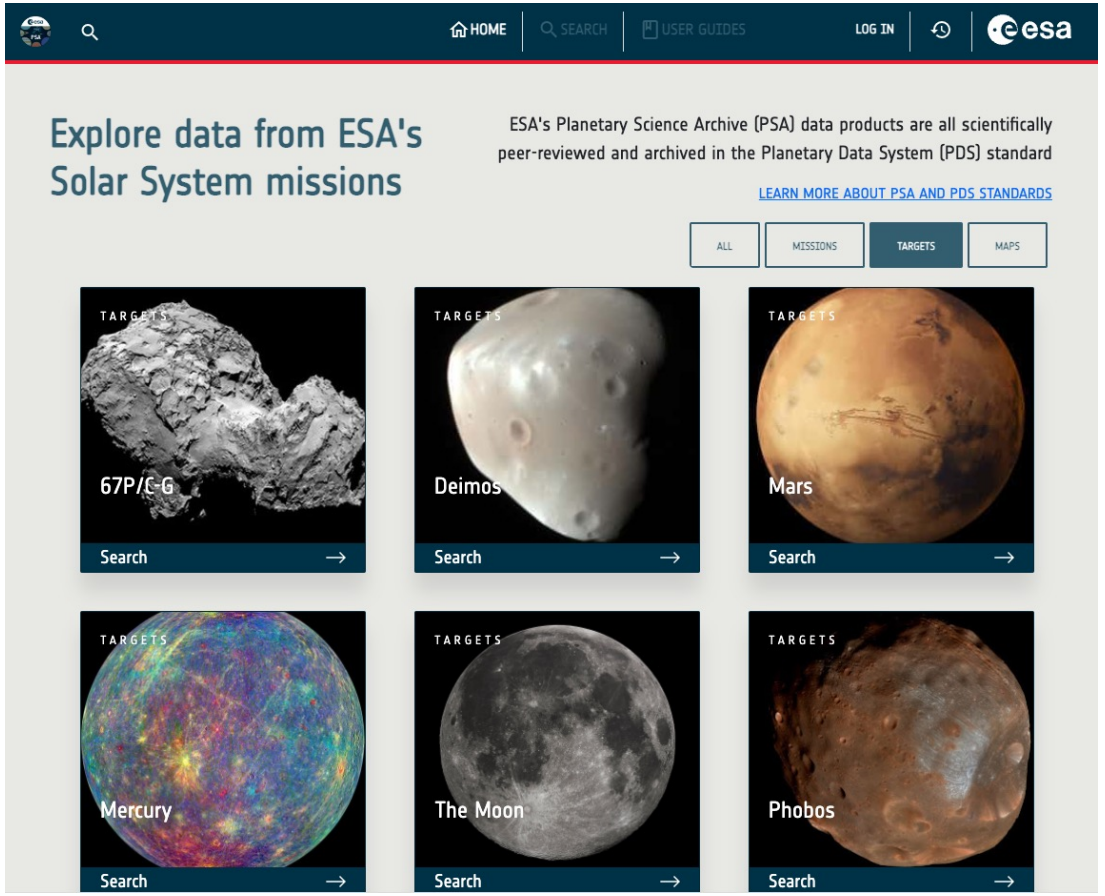
Each tutorial will be based on open-source tools and code and be runnable on DataLabs to dive into the data

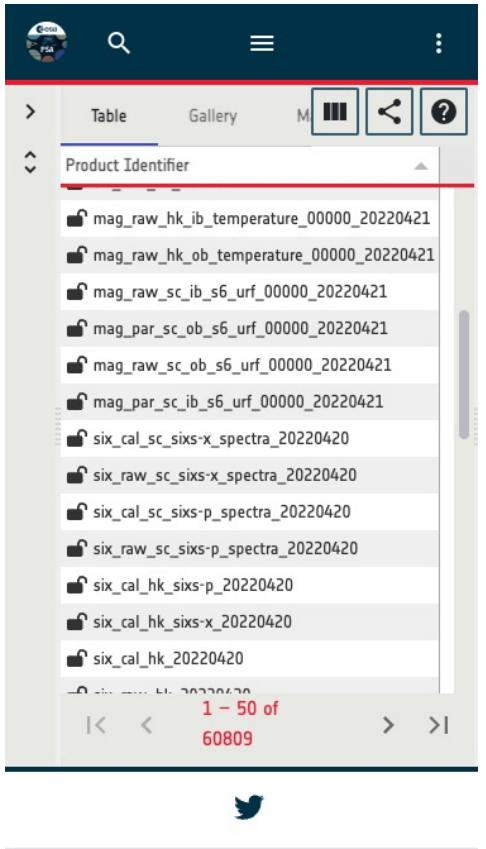
Tutorials

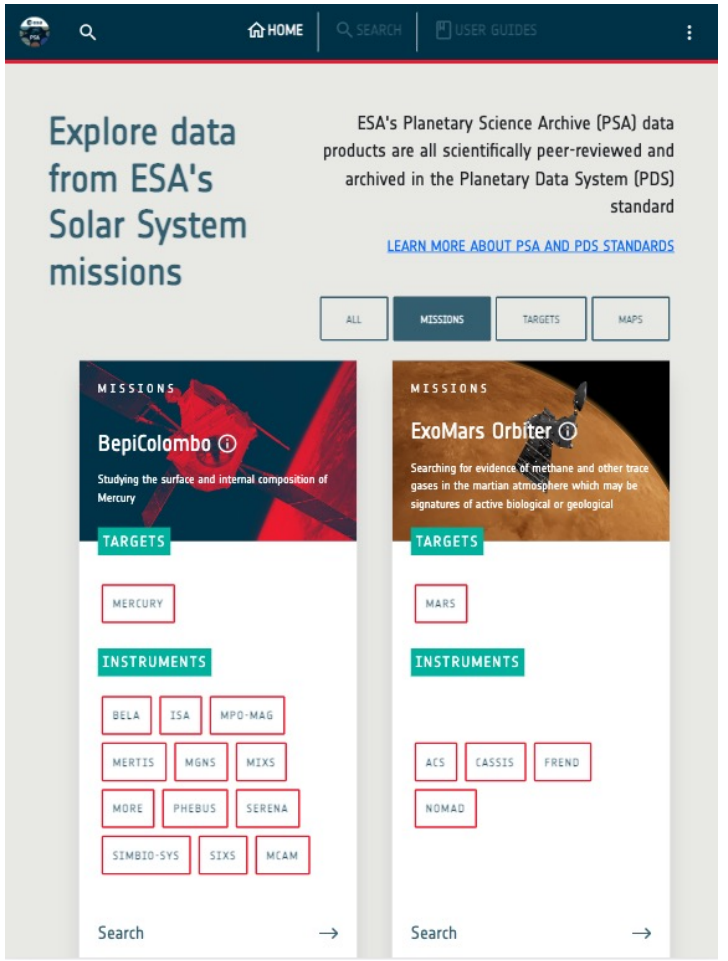
User interface refresh



Responsive design







New features coming soon…

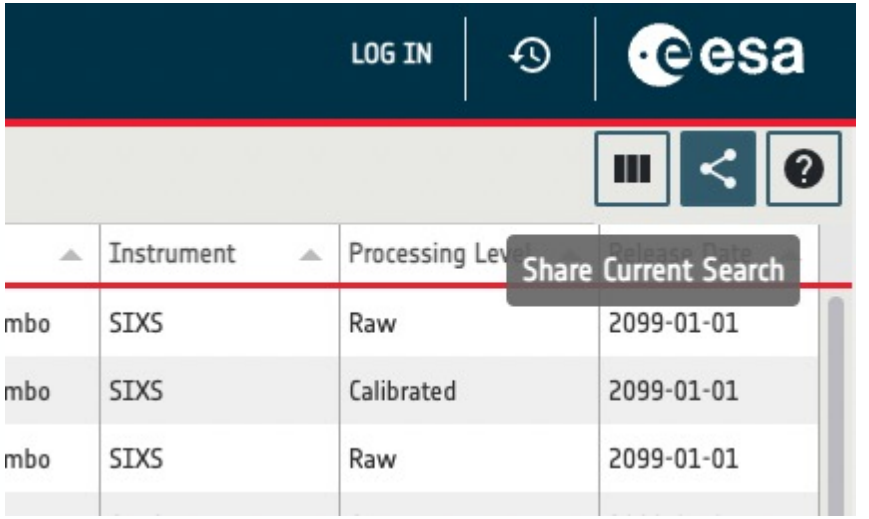
- Sharable links to searches and products

- Improved download manager
  - retrieve linked products in PDS3 and PDS4 labels
    - e.g. all calibrated products listed as forming the input for a derived product
    - very powerful in PDS4 with linking by LIDVID

- Improved access to private data (for instrument teams)
  - SFTP for each PDS4 bundle

- Improved access to auxiliary data
  - via TAP and download manager

- Asynchronous download (allows for larger numbers of files)

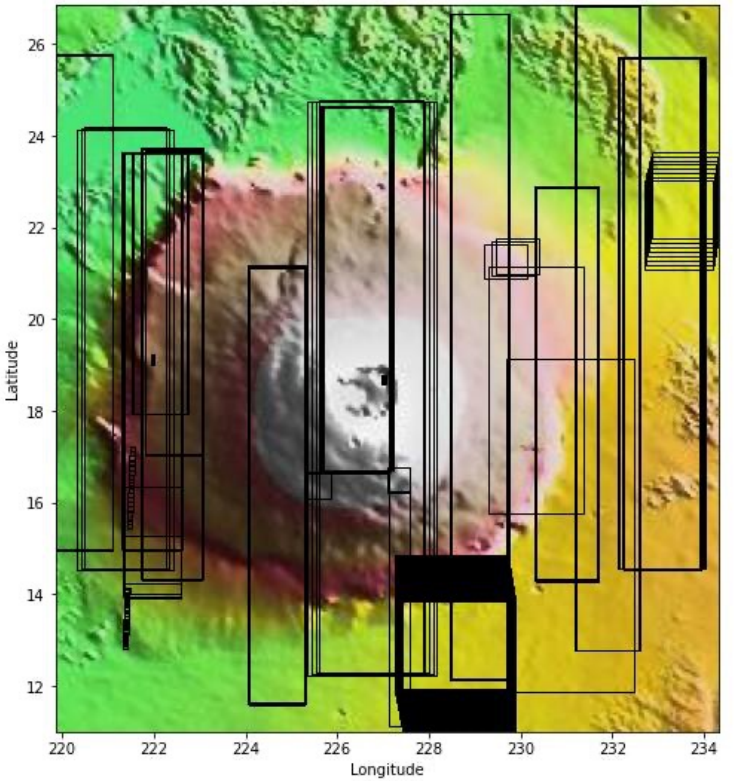


Improved APIs

- Adding additional data to EPN-TAP
  - copy of GEOGEN data into TAP
    - bounding box done
    - footprints coming soon
      - via s_region field
  - adding custom fields to support
    - direct https access to data/label (as well as download)
    - path to file in DataLabs environment (see later!)
  - eventually better mapping of EPN-TAP and PDS fields

- Make EPN-TAP updates realtime
  - currently refreshed daily

```
SELECT granule_uid, c1min, c1max,
c2min, c2max from psa.epn_core where
(c1min >= -143 and c1max <= -125 AND
c2min >= 11 AND c2max <= 27) and
instrument_name = 'HRSC'")
```



Improved APIs

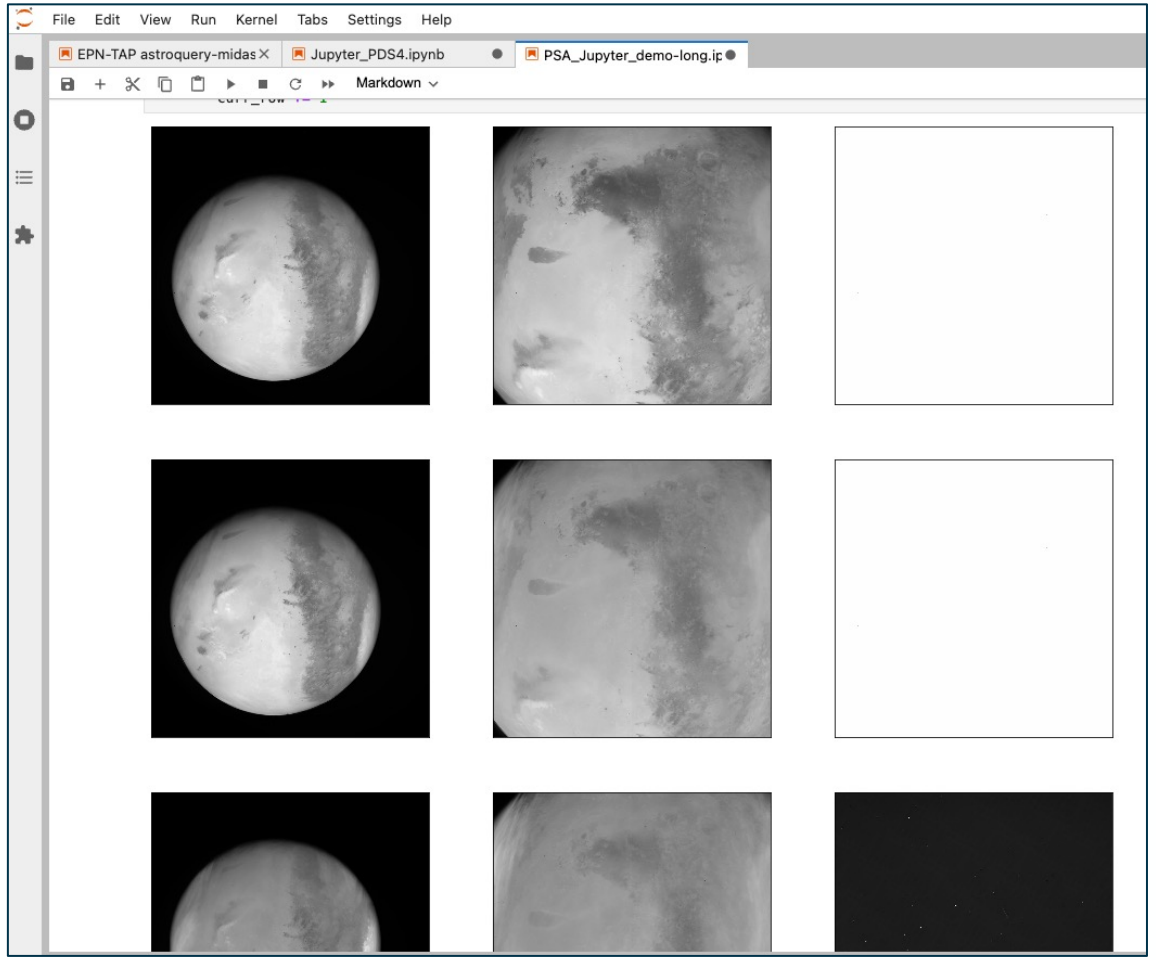
- Deprecation of PDAP
  - used for meta-data mirroring in the past, no significant traffic of late
  - not updated to better handle PDS4
  - limited query syntax

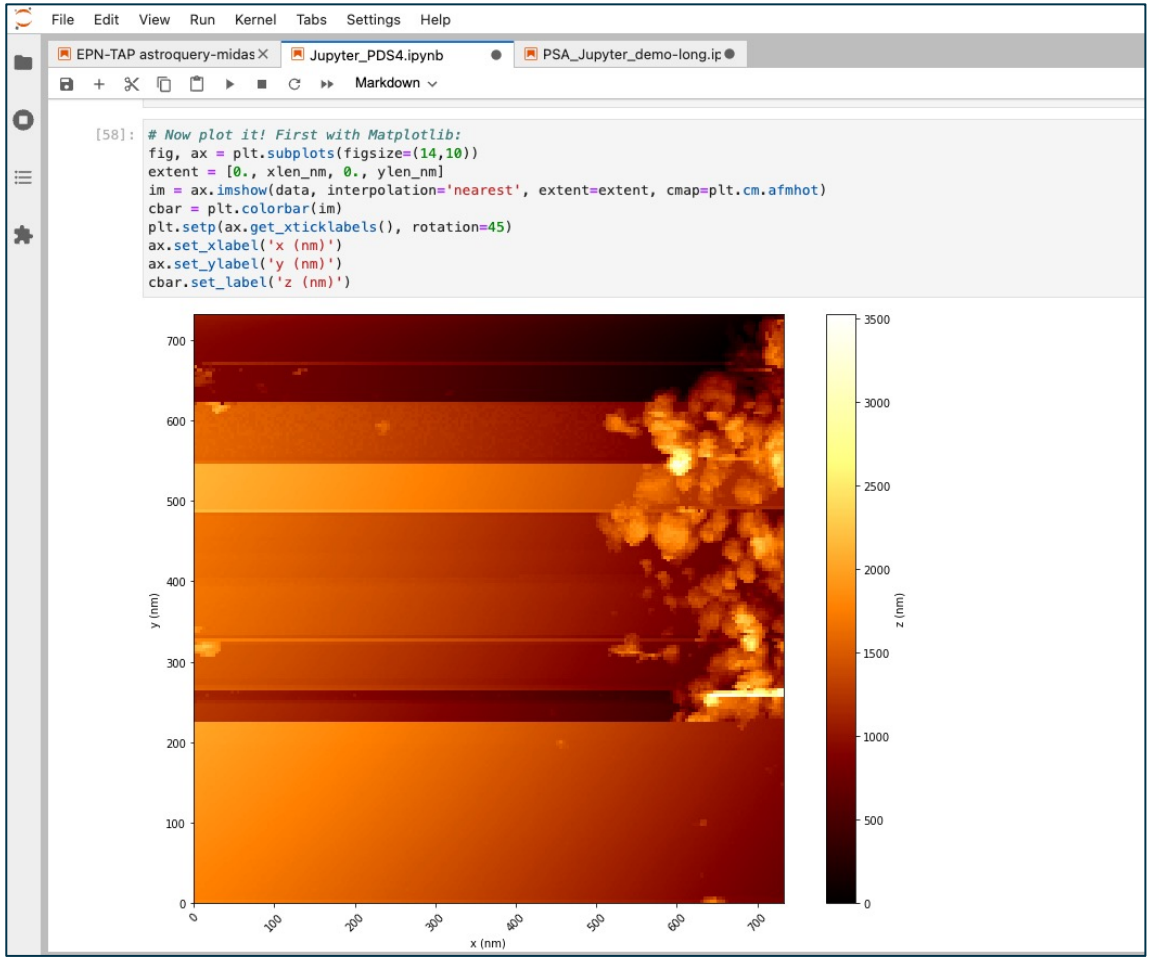
- Adoption of PDS registry and API
  - under discussion with PDS engineering node
  - benefits include
    - integrates with other partners for single entry-point search
    - searches arbitrary meta-data using OpenSearch
  - eventually can be integrated with UI to offer powerful search

Integration with ESA DataLabs

- See talk S03/M13
    - if you missed it, watch the recording, or ask in the Teams channel!
- Integration with DataLabs will happen in several steps:
    - mount public data into a DataLabs volume
    - create a PSA DataLab with standard tools
        - pds4_tools, pdr, astropy, astroquery, etc.
    - add the product path in EPN-TAP custom field
        - so that users can query and load data directly
    - add UI integration (“fine grained” integration)
- Initial use cases for DataLabs will be based around data tutorials
- Community input and feedback is needed for new use cases
    - please get in touch with ideas, wishes etc.

Data tutorials





Summary

- The Planetary Science Archive
  - is a multi-mission and instrument archive allowing search across the Solar System
  - preserves and exposes (meta-) data to the scientific community
  - but it offers limited scope for data exploitation
- Improvements coming soon (and soon-ish) include:
  - new, modern, user interface incorporating user feedback
  - updated APIs better-prepared for data science workflows and searching *all* meta-data
  - integration with ESA DataLabs for on-premises analysis and big data
  - data tutorials offering “click to run” examples of finding, slicing, dicing, plotting data
- Your feedback is needed, so please get in touch with your comments and use cases
  - in particular for the new developments, where we still have time to take them into account
  - mark.bentley@esa.int
- PSA: https://psa.esa.int