

# **ESA Data Discovery Portal**

Christophe.Arviset@esa.int

ESA, Directorate of Science

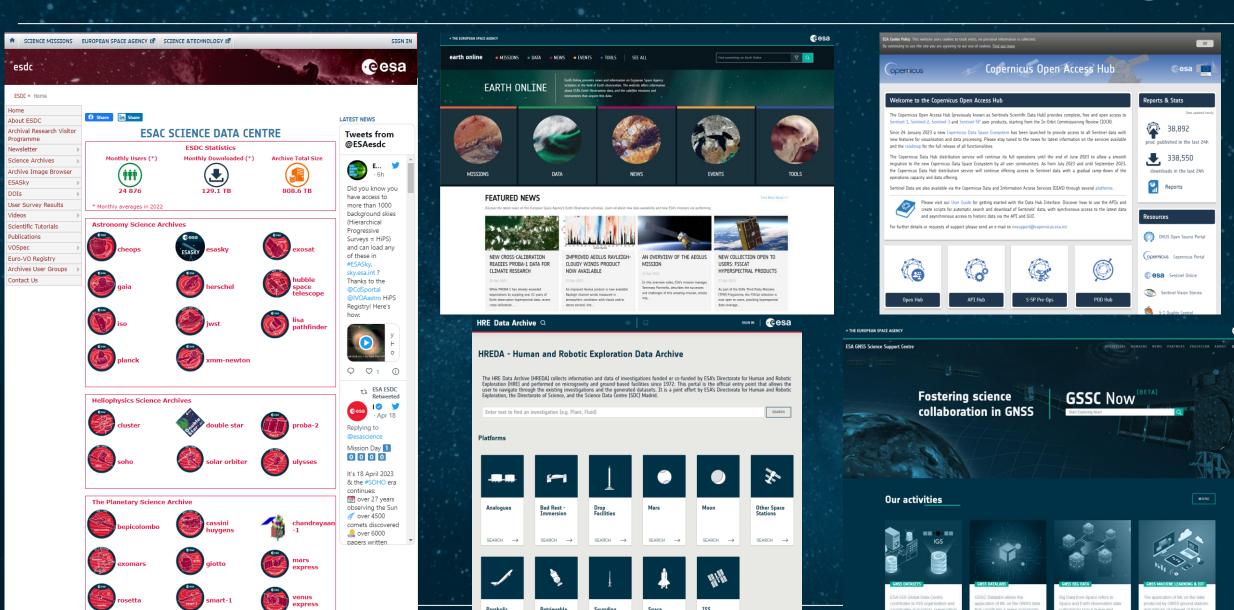
PV @ CERN, 03/05/2023

ESA UNCLASSIFIED – Releasable to the Public



## **Science Data Archives at ESA**





## Science Data Archives are more than JBOD\*



Complete data management system as per OAIS

### Implement FAIR principles

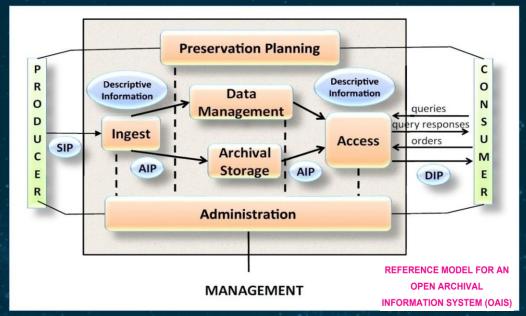
Interoperable with other external services

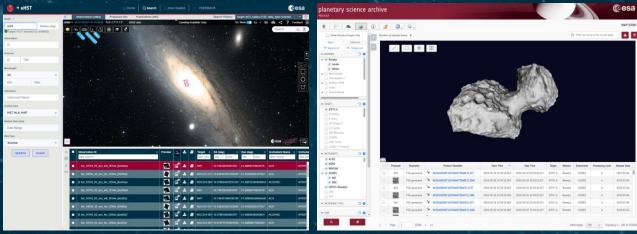
## Powerful and user friendly GUI

- Simple and complex searches
- Data Visualization

### Scriptable interfaces

- Query and download
- Link to data analysis tools





\*Just a bunch of disks

## Defining DOIs for every ESA dataset



DOI is a persistent and unique resource identifier pointing to <u>a landing page</u> that provides a description and a <u>download link</u> for the particular dataset

Define a common set of metadata for all ESA Datasets, regardless of the scientific discipline

DOIs defined through CrossRef and now through DataCite

Common Metadata				
Dataset name	Dataset Description			
Mission name and description	Dataset version			
Instrument	Publishing date			
Scientific domain	Publisher and associated information			
Temporal coverage	Credit and citation Guidelines			
Link to the actual data				

#### A dataset provided by the European Space Agency



Name	CELIAS, Charge, Element, and Isotope Analysis System
Mission	soно
URL	https://www.cosmos.esa.int/web/soho/mission-long-files
DOI	https://doi.org/10.5270/esa-iey8z2h
Abstract	The CELIAS experiment on SOHO is designed to measure the mass, ionic charge and energy of the low and high speed solar wind, of suprathermal ions, and of low energy flare particles. Through analysis of the elemental and isotopic abundances, the ionic charge state, and the velocity distributions of ions originating in the solar atmosphere, the investigation focuses on the plasma processes on various temporal and spatial scales in the solar chromosphere, transition zone, and corona. CELIAS includes 3 mass- and charge-discriminating sensors based on the time-of-flight technic. CTOF for the elemental, charge and velocity distribution of the solar wind, MTOF for the elemental and isotopic composition of the solar wind, and STOF for the mass, charge and energy distribution of suprathermal ions. The instrument provides detailed in situ diagnostics of the solar wind and of accelerated particles, which complements the optical and spectroscopic investigations of the solar atmosphere on SOHO. The Proton Monitor (PM) is a subsystem of the MTOF experiment allowing the measure of the solar proton velocities from 200 to 1200 km/s. CELIAS also contains a Solar Extreme ultraviolet Monitor, SEM, which continously measures the EUV flux in a wide band of 17-70 nm, and a narrow band around the 30.4 nm He II line. SEM is mounted on top of the STOF instrument and shares the same electronics.
Description	CELIAS key scientific data products are mission long files of calibrated  Solar EUV flux at 15 seconds and 1 day cadence, from the SEM detector Solar wind parameters at either 30 seconds, 5 minutes or Carrigton Rotation cadence, from the Proton Monitor detector Daily files of MTOF and STOF are available. CTOF is impaired since October 1996.
Publication	Hovestadt, D., et al., CELIAS - Charge, Element and Isotope Analysis System for SOHO, Sol. Phys., 162, 441–481 (1995); https://doi.org/10.1007/BF00733436
Temporal Coverage	1996 - current
Mission Description	SOHO, the Solar & Heliospheric Observatory, is a project of international collaboration between ESA and NASA to study the Sun from its deep core to the outer corona and the solar wind. SOHO was launched on December 2, 1995. The SOHO spacecraft was built in Europe by an industry team led by prime contractor Matra Marconi Space (now Airbus) under overall management by ESA. The twelve instruments on board SOHO were provided by European and American scientists. Nine of the international instrument consortia are led by European Principal Investigators (Pfs), three by Pfs from the US. Large engineering teams and more than 200 co-investigators from many institutions supported the Pfs in the development of the instruments and in the preparation of their operations and data analysis. NASA was responsible for the launch and is now responsible for mission operations. Large radio dishes around the world which form NASA's Deep Space Network are used for data downlink and commanding. Mission control is based at Goddard Space Flight Center in Maryland.  Domingo, V., Fleck, B. & Poland, A.I., The SOHO mission: An overview, Sol. Phys., 162, 1–37, 1995; https://doi.org/10.1007/BF00733425
Creator Contact	Wimmer-Schweingruber, R., Principal Investigator, University of Kiel, Germany, Email : wimmer Berger, L., Project Scientist, University of Kiel, Germany, Email : berger All Email : @pthysik.uni-Kiel.de

## Granularity level for DOI assignment – space science



Minting DOIs depends on the way data is organized for each mission type

#### Astronomy survey missions (e.g. Gaia, Planck, ...)

- Catalogue of astronomical objects, cosmology maps
- A few DOIs required

#### Astronomy observatory mission (e.g. HST, XMM - Newton, Herschel)

- Proposal for a particular scientific topic resulting in many observations
- A few thousands of DOIs

#### Planetary missions (e.g. Rosetta, Mars Express, BepiColombo)

- Dataset for an instrument for a particular mission phase
- A few thousands of DOIs

#### Heliophysics missions (e.g. SOHO, Solar Orbiter, Cluster)

- 1 DOI for each experiment throughout all mission lifetime
- A few tens of DOIs

Science Mission 🔻	Domain	~	DOIs	¥
HST	Astronomy		10901	
XMM - Newton	Astronomy		4270	
ISO	Astronomy		3877	
Rosetta	Planetary		3594	
Mars Express	Planetary		3512	
CHEOPS	Astronomy		2971	
Herschel	Astronomy		741	
Venus Express	Planetary		727	
SMART-1	Planetary		22	
SOHO	Heliophysics		13	
Cluster	Heliophysics		12	
Ulysses	Heliophysics		12	
Giotto	Planetary		10	
ExoMars TGO	Planetary		10	
Planck	Astronomy		9	
Cassini-Huygens	Planetary		8	
Lisa PathFinder	Astronomy		8	
Double Star	Heliophysics		8	
Solar Orbiter	Heliophysics		7	
Chandrayaan-1	Planetary		6	
Proba-2	Heliophysics		5	
Gaia	Astronomy		4	
ISS-SolACES	Heliophysics		1	

Around 30.000 DOIs

## ESA Data Discovery Portal – data.esa.int

esa

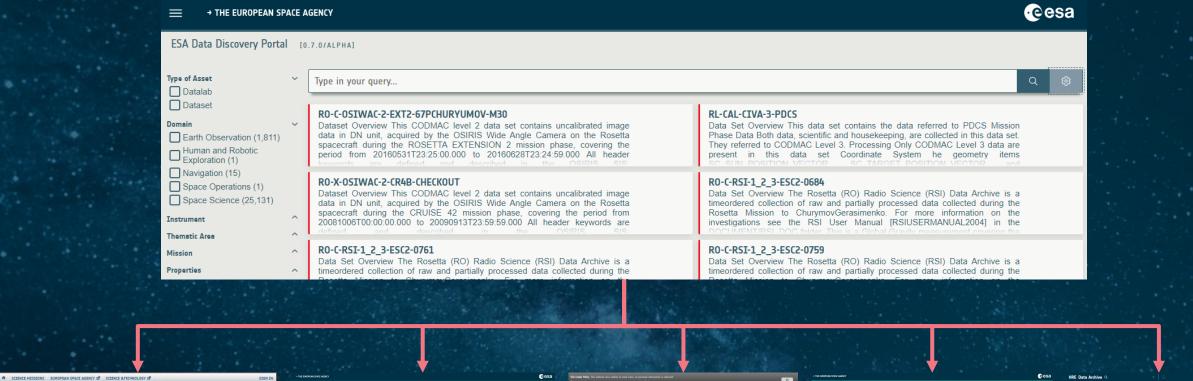
Tweets from @ESAesdc

more than 1000

Thanks to the

**ESAC SCIENCE DATA CENTRE** 







## **ESA Data Discovery Portal – main concepts**



## Common Portal for all ESA datasets from all disciplines

- ESA branding, homogenous description of ESA datasets through DOI (data.esa.int/asset/<reference>)
- Provides link to the corresponding mission data archive for dataset download
- (In progress) Provides access to tools / software available to analyse the datasets (datalabs)

## Simple Search Interface

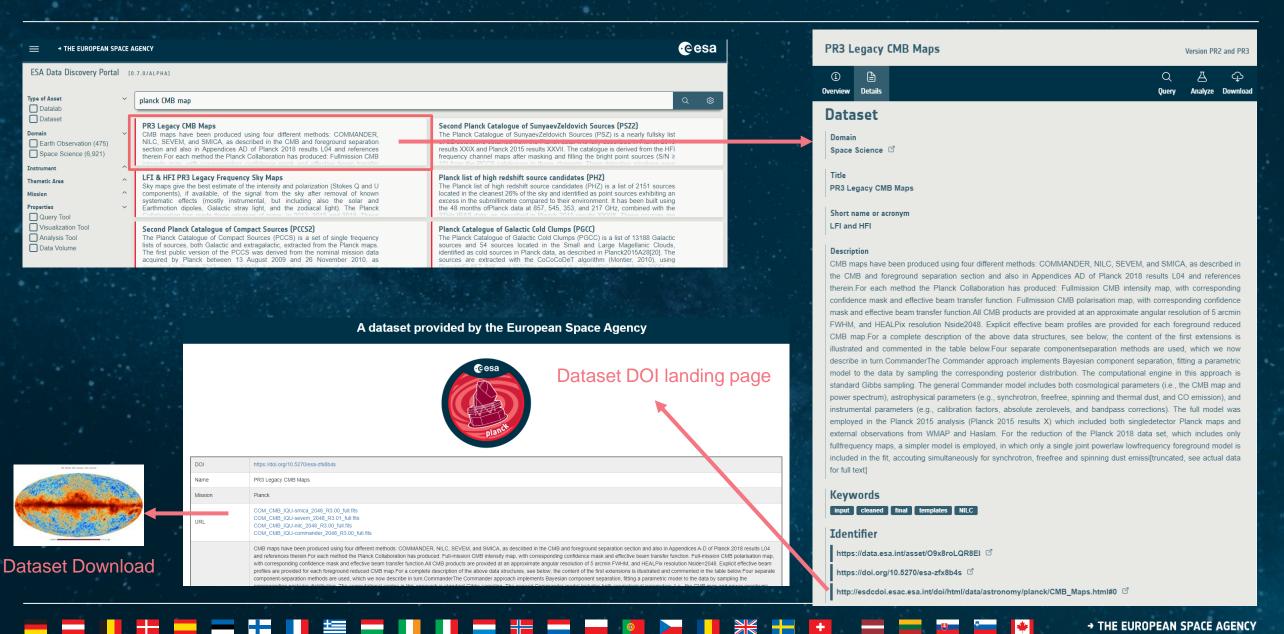
- Drill down your search through faceted search criteria selection
- Domain, Thematic area, Mission, Instrument

#### Not an actual data archive

- Does not actually archive datasets (but only common metadata)
- Does <u>not</u> supersede the existing mission data archives

## From ESA Data Discovery Portal to Dataset download





## Google Dataset Search - https://datasetsearch.research.google.com/



#### Released in September 2018, targeting

- Open data and open science
- Researchers, data geeks/scientists/journalists, students, ...
- 25M datasets referenced

All datasets referenced through schema.org vocabulary

Not clear if really to be used by scientific community which is probably better served through our mission data archives

But ESA datasets need to be properly referenced in Google to ensure access to correct and quality data

When minting DOI for ESA Dataset, add creation of schema.org for it (as great overlap with DOI metadata)

=> ESA Datasets then get referenced into Google Dataset Search



#### **User Support Centre**



#### What is Dataset Search?

Dataset Search is a search engine for data sets

Using a simple keyword search, users can discover data sets hosted in thousands of repositories across the web.



#### Our mission

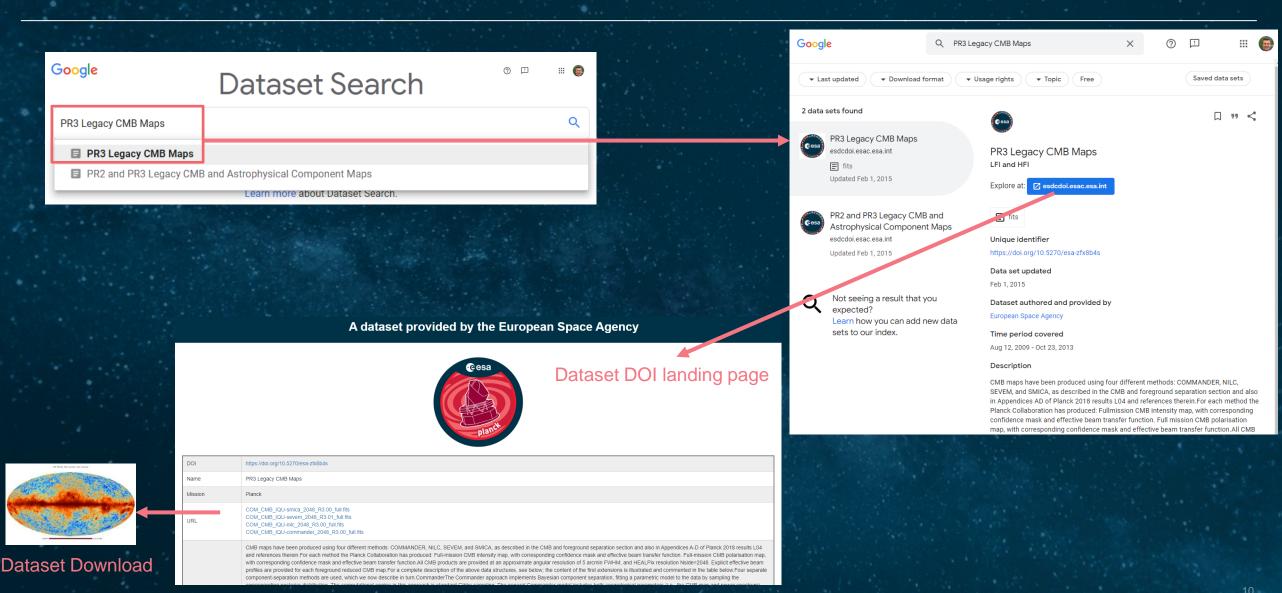
In addition to making datasets universally accessible and useful, Dataset Search's mission is to:

- Foster a data sharing ecosystem that will encourage data publishers to follow best practices for data storage and publication.
- Give scientists a way to show the impact of their work through citation of datasets that they have produced

As more dataset repositories use schema.org and similar standards to describe their datasets, the variety and coverage of datasets that users find in Dataset Search will continue to grow.

## From Google Dataset Search to ESA dataset download





## ESA Datalabs – datalabs.esa.int



Multi-Domain and Multi-Mission Digital Platform

#### **ESA Science Exploitation Platform**

Science Data available for researchers to work on it, made easy

#### Increase Science Operations Efficiency

- Reusable for fast implementation of Scientific Processing Pipelines
- Reusable for fast implementation of Scientific Analysis and Visualisation Tools

#### Collaboration and Open Science

- Share complex processing tools and data with your team
- Share your contributions with the community in ESA Datalabs
   Science AppStore



Do you know what is "ESA Datalabs"?

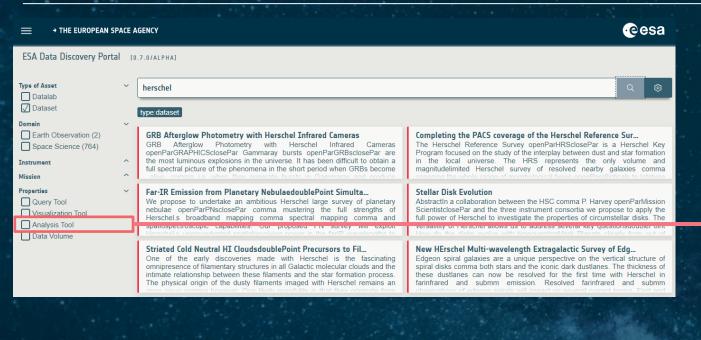


Yes, I'm familiar with ESA Datalabs. ESA Datalabs is a data analytics platform developed by the European Space Agency (ESA) that provides access to a wide range of satellite data and tools for data analysis and processing. The platform is designed to support researchers, scientists, and developers in their work, by providing them with access to data and tools that can help them analyze and make sense of the data they are working with. With its focus on enabling data-driven decision-making, ESA Datalabs aims to provide users with the tools they need to perform advanced data analysis and processing tasks, such as data visualization, data mining, and machine learning. The goal of ESA Datalabs is to help researchers, scientists, and developers unlock the full potential of the data they are working with and to support the development of new and innovative applications based on satellite data.



## From ESA Data Discovery Portal to ESA Datalabs

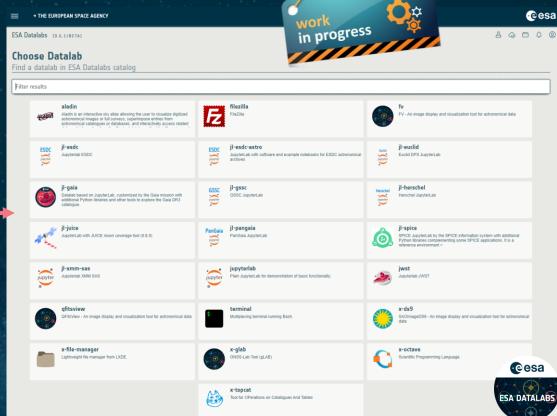




Search for Tools associated with data

Launch it directly into ESA Datalabs platform

- Data will automatically be loaded into your datalab
- Start analysing the data



## **Conclusions and next steps**



### ESA Data Discovery Portal (data.esa.int)

- Offers a unified entry point to all ESA datasets from all ESA directorates
- Is complementary to the existing ESA data archives where data resides

#### DOI minted for all ESA datasets

- Common metadata description for all
- Granularity depends on type of mission
- Almost complete for Space and Earth Science, in progress for other disciplines

ESA Data Discovery Portal expands to ESA Datalabs science exploitation platform

From the data to the visualization / analysis tools