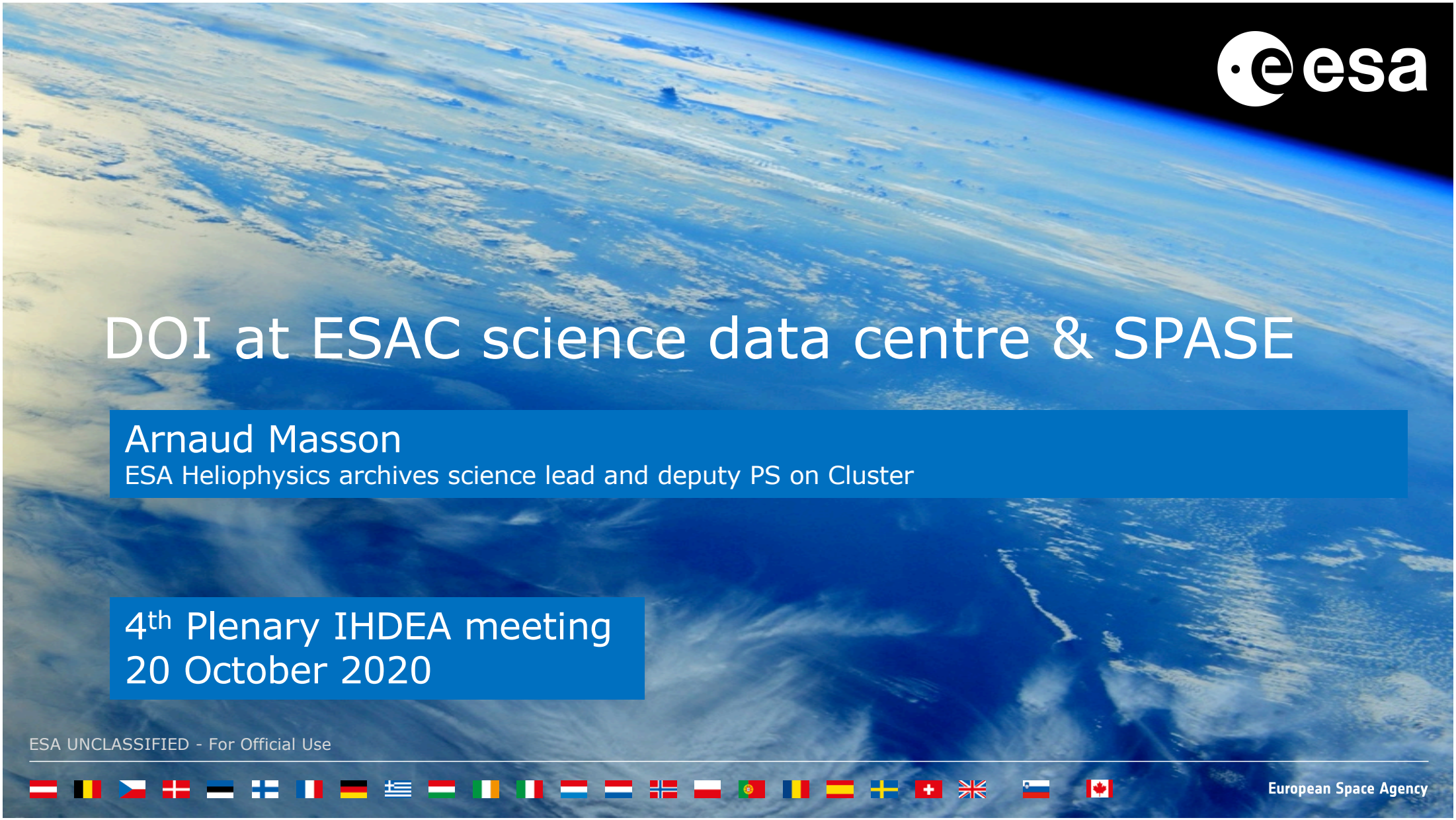


DOI at ESAC science data centre & SPASE

Arnaud Masson
ESA Heliophysics archives science lead and deputy PS on Cluster

4th Plenary IHDEA meeting
20 October 2020

Outline

What do we want?

- Direct link from data in papers to our archives
- Improved findability of our datasets
- Better tracking of their usage
- Improved reproducibility of published results

Outline

Introduction: DOI?

1. Why assigning DOIs to experiments on ESA heliophysics s/c?
2. How to get a DOI indexed by Google dataset search?
3. Direct link to the data and findability on Google dataset search
4. Who cares about heliophysics?
5. Next steps and open questions

DOI?

DOI is a persistent URL like this one https://doi.org/10.5270/esa-yrl3swm

Leading to a DOI (HTML) landing page

managed by ESA at the ESAC Science Data Centre (ESDC) in Madrid

A dataset provided by the European Space Agency

soho

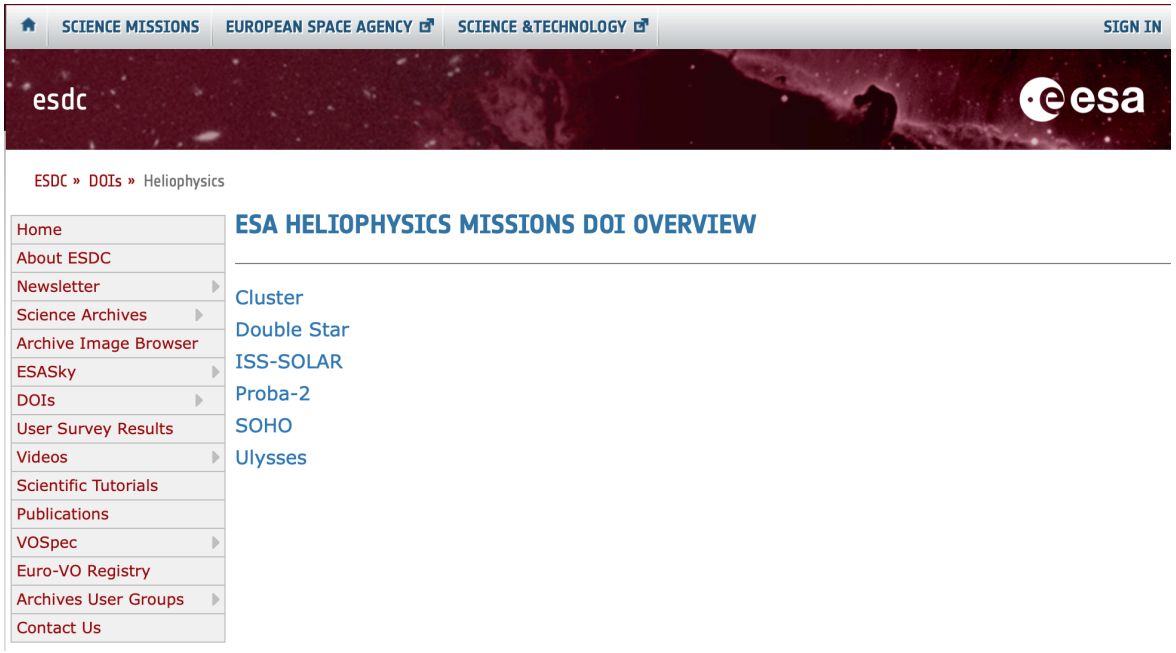
esa

Name	COSTEP, Comprehensive Suprathermal and Energetic Particle Analyzer
Mission	SOHO
URL	https://www.cosmos.esa.int/web/soho/mission-long-files
DOI	10.5270/esa-yrl3swm
Abstract	The COSTEP experiment on SOHO forms part of the CEPAC complex of instruments that perform studies of the suprathermal and energetic particle populations of solar, interplanetary, and galactic origin. Specifically, the LION and EPHIN instruments are designed to use particle emissions from the Sun for several species (electrons, protons, and helium nuclei) in the energy range 44 keV/particle to > 53 MeV/n as tools to study critical problems in solar physics as well as fundamental problems in space plasma and astrophysics. EPHIN is able to obtain energy spectra and achieve isotope separation for light nuclei.
Description	COSTEP key scientific data products are mission long files of calibrated proton and helium fluxes at either 1 minute, 5 minutes, 30 minutes, 1 hour or 1 day cadence, estimated by the EPHIN detector. The protons and Helium energy bands cover an energy range from 4.3 to 53 MeV/n. Daily files from EPHIN and LION are also available. While EPHIN is functioning nominally, LION is impaired since shortly after launch, with increased noise.
Publication	Müller-Mellin, R., et al., COSTEP - Comprehensive Suprathermal and Energetic Particle Analyser, Sol. Phys., 162, 483–504 (1995); https://doi.org/10.1007/BF00733437
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 (PI's), three by PI's from the US. Large engineering teams and more than 200 co-investigators from many institutions supported the PI's 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	Heber, B., Principal Investigator, University of Kiel, Germany, heber@physik.uni-kiel.de The SOHO/COSTEP/EPHIN project is supported under various grants by the German Federal Ministry of Economics and Technology.
Publisher	

1. Why assigning DOIs to ESA Heliophysics missions experiments?

- Link publications to the data
- Improve traceability of data usage
- Acknowledging the work of PI teams
- Make all heliophysics experiments findable on Google Dataset Search
- Possibly increase the data usage

cosmos.esa.int/web/esdc/doi/heliophysics



These missions overall carry
47 experiments => 47 DOI

A dataset provided by the European Space Agency

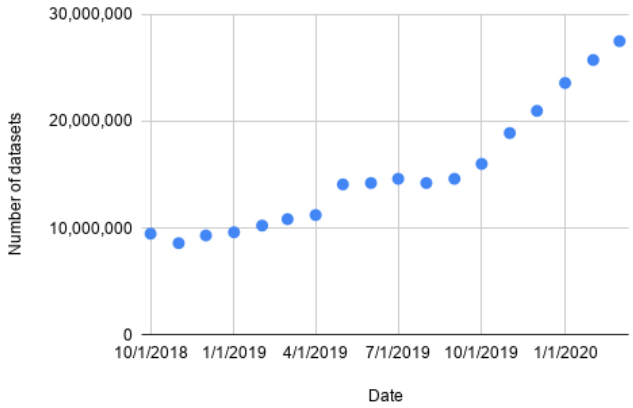
Name	MAG, MAGnetometer
Mission	Solar Orbiter
URL	http://soar.esac.esa.int/
DOI	10.5270/esa-
Abstract	The Solar Orbiter magnetometer is a conventional dual fluxgate design. Two sensors are accommodated on the spacecraft boom: MAG-IBS and MAG-OBS. A dual sensor configuration provides redundancy and, since they are at different distances from the spacecraft body (approx. 1 m for IBS and 3 m for OBS), also allows gradiometer magnetometer characterisation of spacecraft signals in flight. The instrument noise floor has been successfully tested at 10 pT Hz−1/2 at 1 Hz.
Description	Calibrated magnetic field data in RTN coordinates and in the spacecraft reference frame. Field vector components are given units of nanoteslas and in RTN coordinates, where R is the Sun-spacecraft axis, T is the cross product of the solar rotation a and R, and N is the cross product of R and T. During cruise phase, MAG is operating at 1 vector/s cadence. After the cruise phase, MAG is expected to operate continuously at 16 vectors/s cadence (normal mode) except during 1 hour per day at 128 Hz or during 2 hours at 64 Hz (Burst mode). Alternatively, burst modes will be triggered in coordination with other in-situ instruments' burst modes.
Publication	Horbury, A., et al., The Solar Orbiter magnetometer, Astron. Astrophys., 2020; DOI: doi.org/10.1051/0004-6361/201937257
Temporal Coverage	2020-05-01 - present
Mission Description	Solar Orbiter is a mission of international collaboration between ESA and NASA. It explores the Sun and the heliosphere fron close up and out of the ecliptic plane. Launched on 10 February 2020, it aims to address the overarching science question: how does the Sun create and control the Heliosphere – and why does solar activity change with time? To answer it, the Solar Orbiter spacecraft is cruising to a unique orbit around the Sun, eventually reaching a minimum perihelion of 0.28 AU, and performing measurements out of the ecliptic plane: reaching 18° heliographic latitude during its nominal mission phase, and above 30° during its extended mission phase. It carries six remote sensing instruments to observe the Sun and the solar corona, and four in-situ instruments to measure the solar wind, its thermal and energetic particles, and electromagnetic fields Müller, D., O.C.St. Cyr, I. Zouganelis, et al., Astron. Astrophys., 2020; DOI: doi.org/10.1051/0004-6361/202038467 Müller, D., Marsden, R.G., St. Cyr, O.C. et al., Solar Orbiter, Sol. Phys., 285, 25–70 (2013); doi.org/10.1007/s11207-012-008
Creator Contact	Prof. T. Horbury, Principal Investigator, Imperial College, United Kingdom, t.horbury@imperial.ac.uk
Publisher And Registrant	European Space Agency

2. What is GDS? How to get indexed?

Google Dataset Search (GDS) is a new search engine from Google, launched in January 2020.
https://datasetsearch.research.google.com

To get indexed, GDS requires to include a **JSON script** with at least **two compulsory properties**: name and description

```
<script type=application/ld+json>
{
@context: http://schema.org/
@type: Dataset
name: GOLF
description: Global Oscillations at Low Frequencies
}
</script>
```



Google Dataset Search

2. How to get indexed?

Adding more properties in the JSON script enables us to

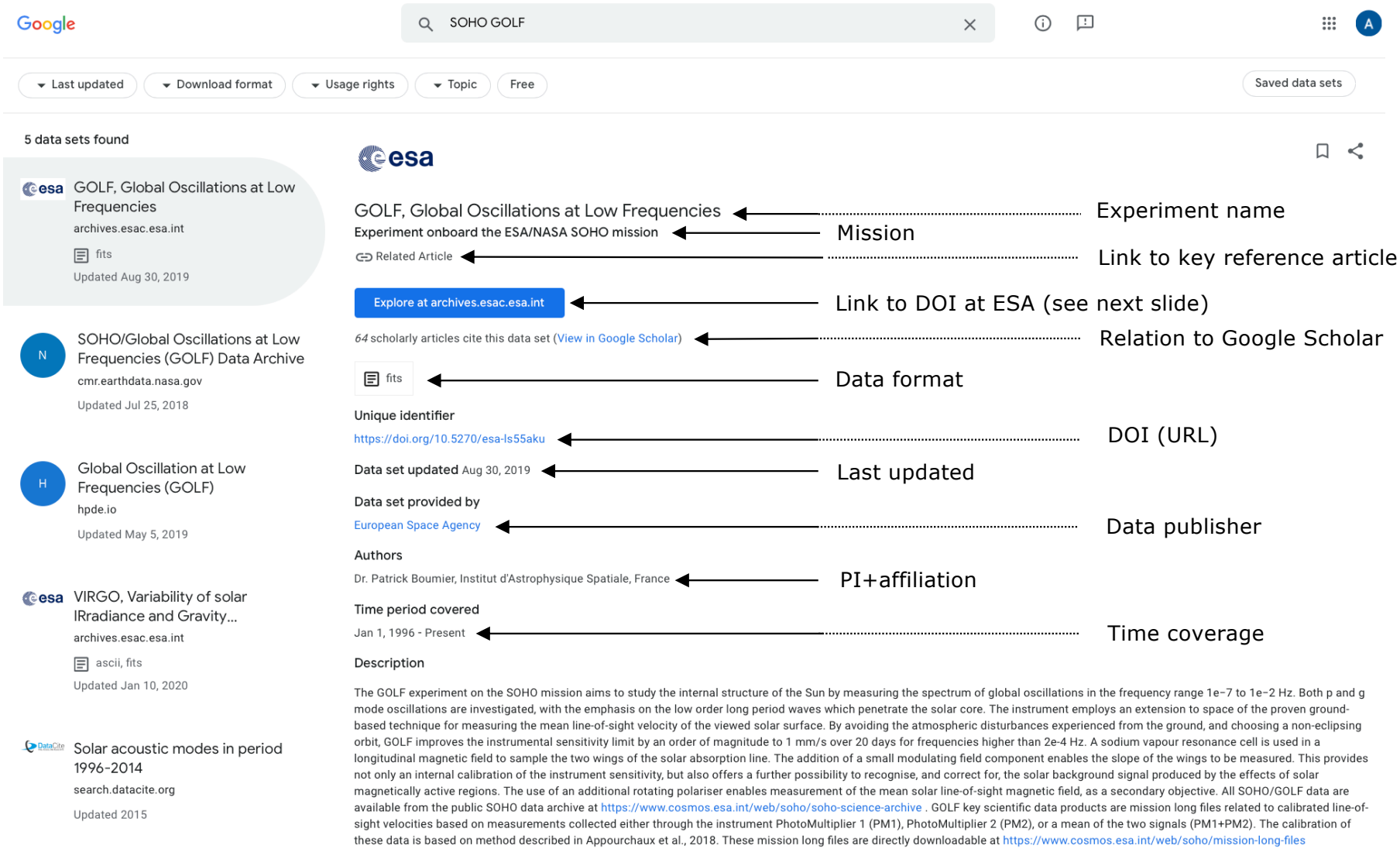
- Point to the DOI landing page maintained by ESDC
- Make Google dataset search display a detailed description of the dataset
- Acknowledge the PI
- Acknowledge the data publisher (i.e. ESA)

2. How to get indexed?

Schema.org property used in Heliophysics DOI landing pages JSON script	Google Dataset Search compliance	Remark
@context	Compliant	
@type	Compliant	
name	Compliant	
alternate name	Compliant	This has been used to specify the mission, as sometimes the same type of experiment with the same name has been used on multiple missions like FGM, PEACE or ASPOC. It could also be used to specify the names of the different instruments of an experiment.
citation	Compliant	Key scientific paper listed, those agreed with PS/PI. At the moment, only one reference is linked (May 2020)
identifier	Compliant	this should be the DOI address
creator	Compliant	The creator or author of a dataset. To uniquely identify individuals, Google advises to use ORCID ID as the value of the `sameAs` property of the `Person` type. To uniquely identify institutions and organizations, use ROR ID. Implemented when available.
publisher	Compliant	Contains web address and new ESA logo
distribution	Compliant	Direct link to data download. Added when relevant, like for mission long files for SOHO.
temporalCoverage	Compliant	Correct format implemented
dateModified	Compliant	Added when dataset was updated recently, or last update is known
description	Compliant	Abstract content added before dataset description content, double quotes and HTML tags for hyperlinks removed. Link to the ESDC archive added.
keywords	Compliant	Keywords agreed with PS/PI
audience	Compliant	Manually edited per mission to fit scientific communities

Table 1. Schema.org properties used in the Heliophysics DOI landing pages

Please search ‘SOHO GOLF’ @ https://datasetsearch.research.google.com

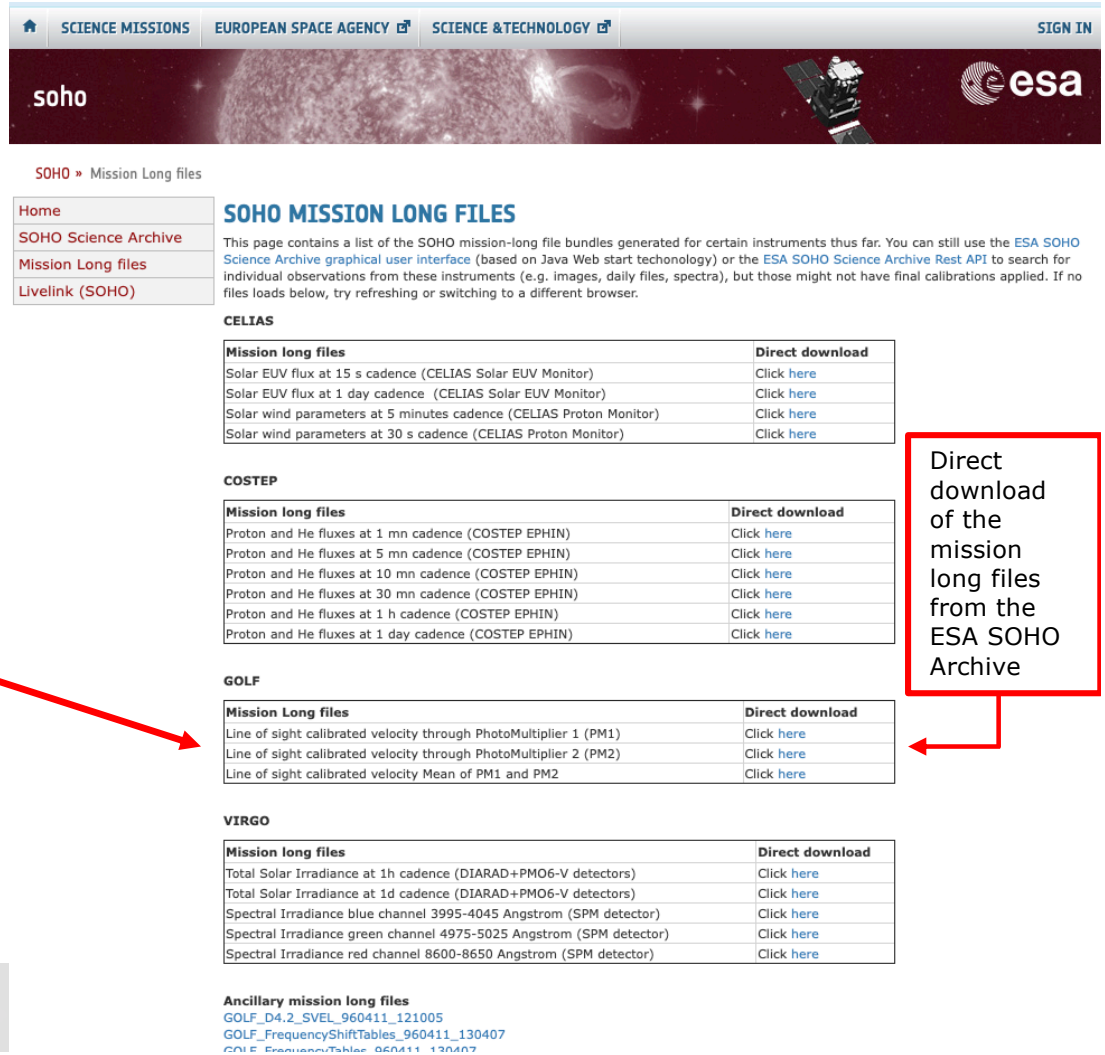


3. Direct link to the data

A dataset provided by the European Space Agency

soho

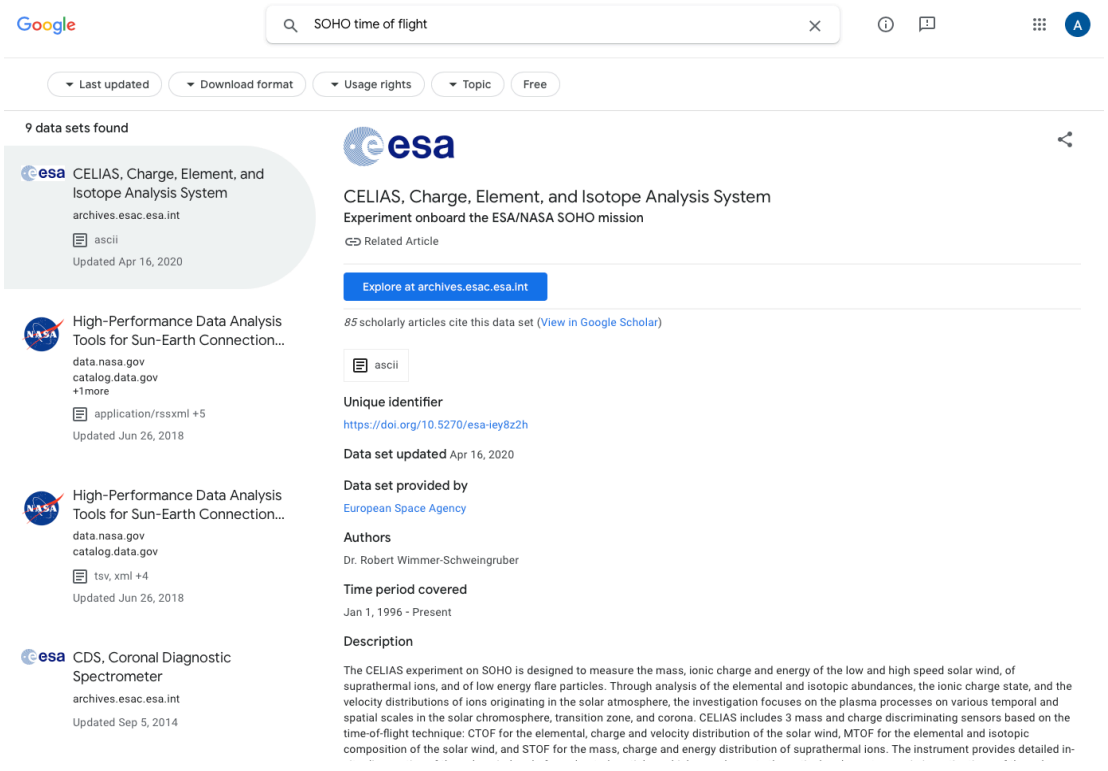
Name	GOLF, Global Oscillations at Low Frequencies
Mission	SOHO
URL	https://www.cosmos.esa.int/web/soho/mission-long-files
DOI	10.5270/esa-ls55aku
Abstract	The GOLF experiment on the SOHO mission aims to study the internal structure of the Sun by measuring the spectrum of global oscillations in the frequency range 10^{-7} to 10^{-2} Hz. Both p and g mode oscillations are investigated, with the emphasis on the low order long period waves which penetrate the solar core. The instrument employs an extension to space of the proven ground-based technique for measuring the mean line-of-sight velocity of the viewed solar surface. By avoiding the atmospheric disturbances experienced from the ground, and choosing a non-eclipsing orbit, GOLF improves the instrumental sensitivity limit by an order of magnitude to 1 mm s^{-1} over 20 days for frequencies higher than 2.10^{-4} Hz. A sodium vapour resonance cell is used in a longitudinal magnetic field to sample the two wings of the solar absorption line. The addition of a small modulating field component enables the slope of the wings to be measured. This provides not only an internal calibration of the instrument sensitivity, but also offers a further possibility to recognise, and correct for, the solar background signal produced by the effects of solar magnetically active regions. The use of an additional rotating polariser enables measurement of the mean solar line-of-sight magnetic field, as a secondary objective.
Description	GOLF key scientific data products are mission long files related to calibrated line of sight velocities based on measurements collected either through the instrument PhotoMultiplier 1 (PM1), PhotoMultiplier 2 (PM2), or a mean of the two signals (PM1+PM2). The calibration of these data is based on method described in Appourchaux et al., 2018.
Publication	Gabriel, A.H., et al., Global Oscillations at Low Frequency from the SOHO mission (GOLF), Sol. Phys., 162, 61–99 (1995); https://doi.org/10.1007/BF00733427 Appourchaux, T., et al., Searching for g modes. I. A new calibration of the GOLF instrument, A&A, 617,

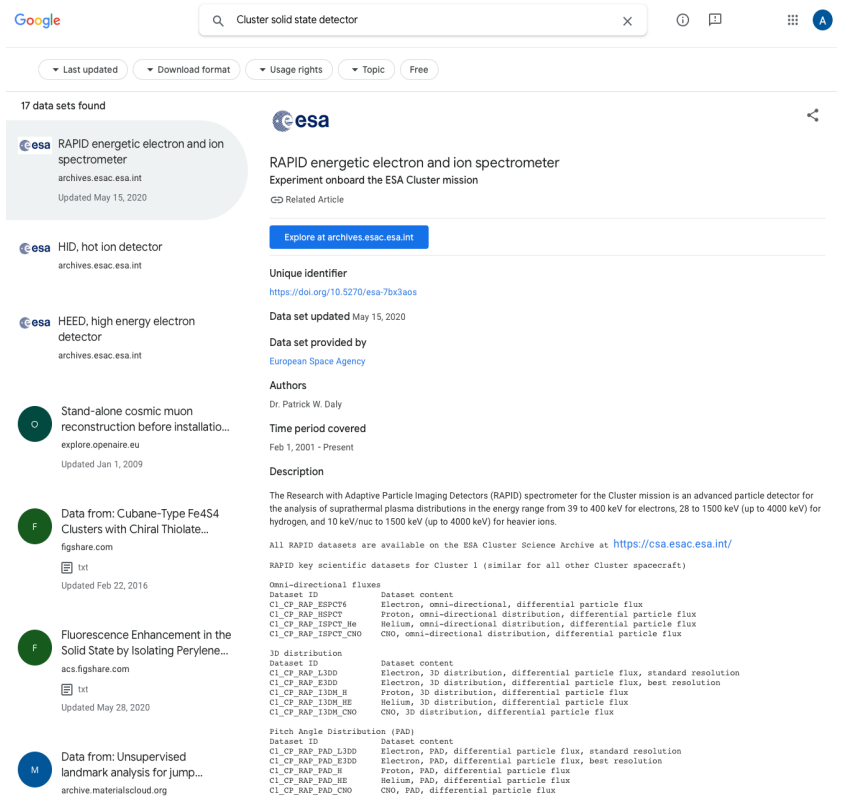


Direct download of the mission long files from the ESA SOHO Archive

https://datasetsearch.research.google.com

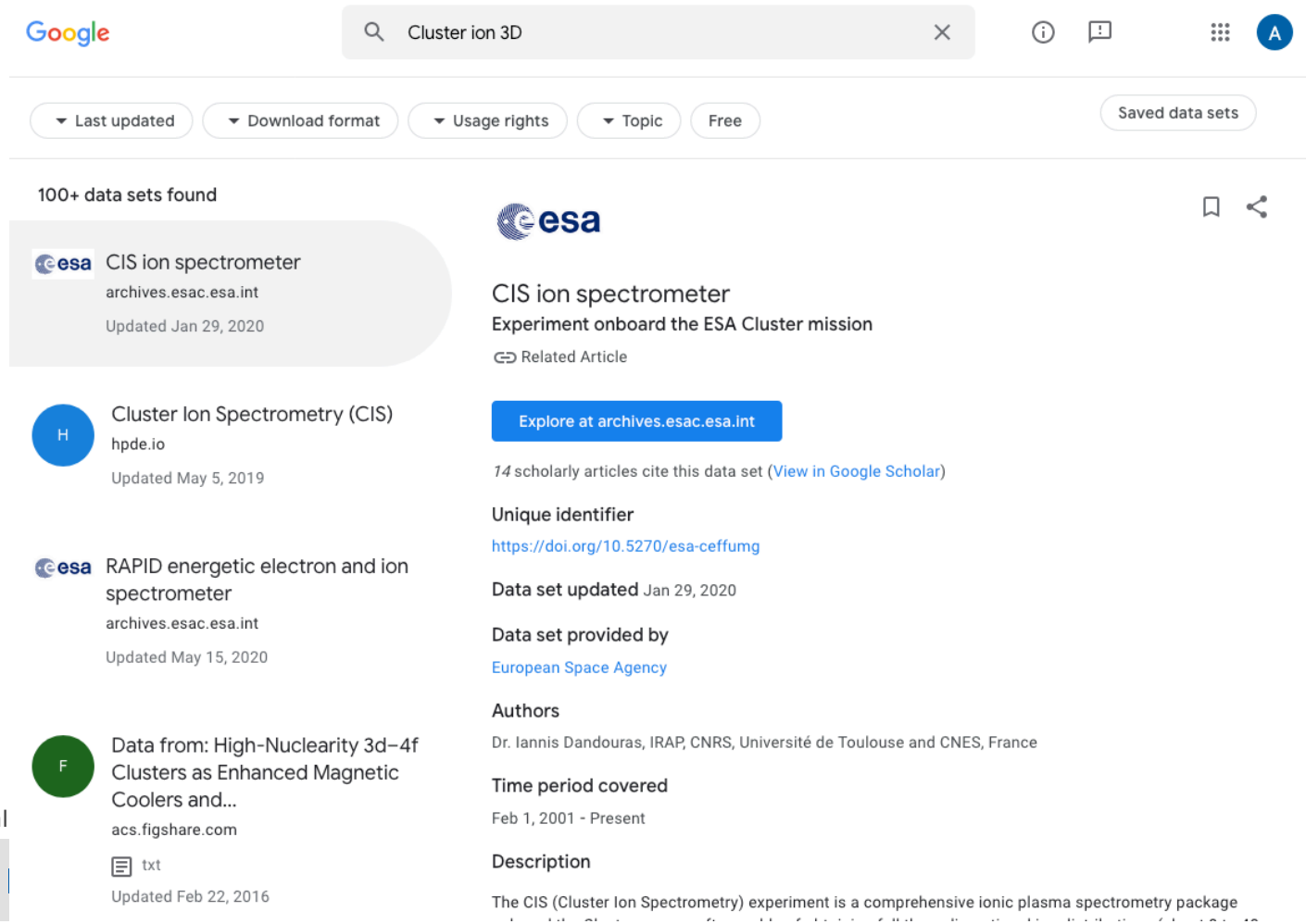
Finding datasets by measurement technique





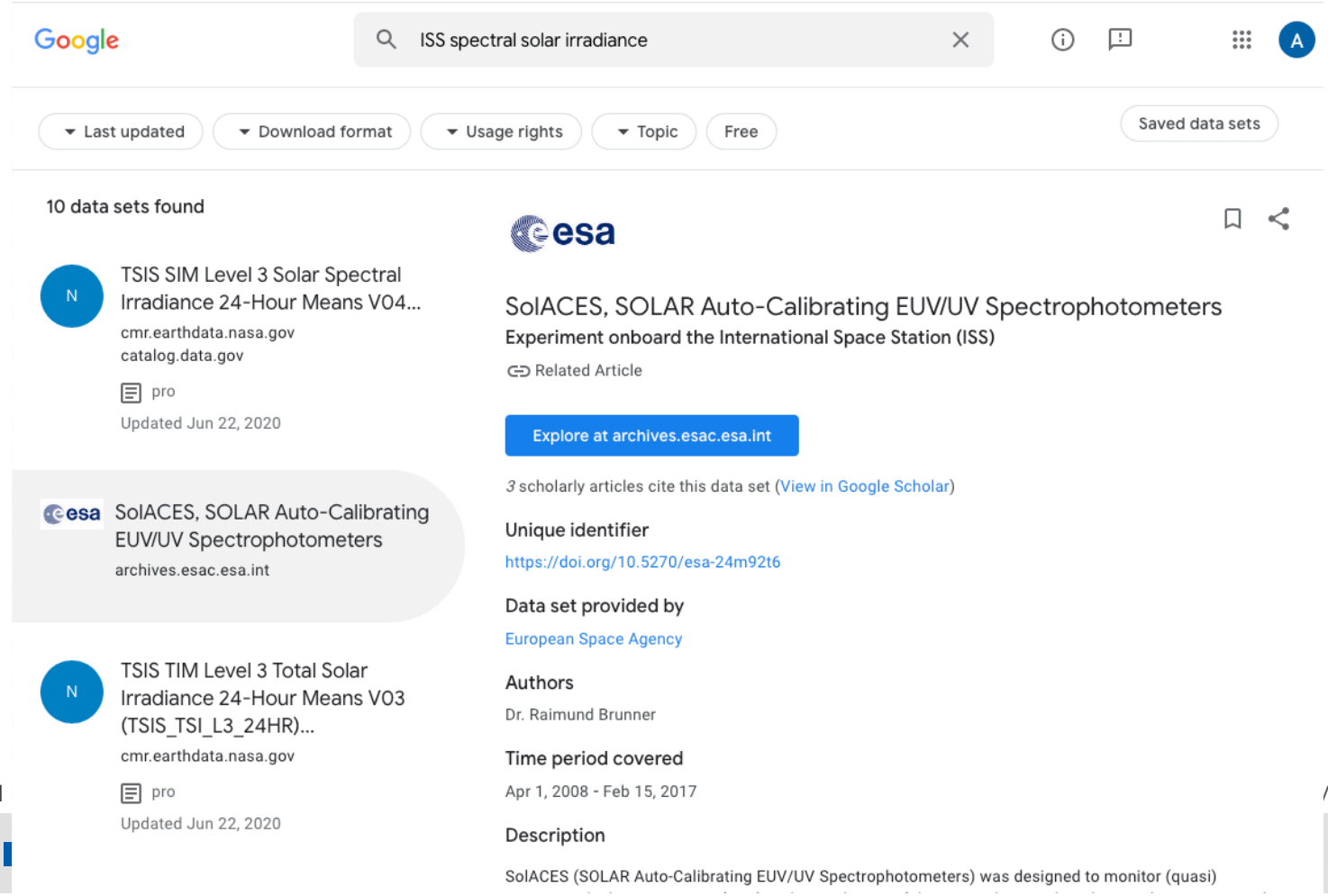
https://datasetsearch.research.google.com

Finding datasets by type of measurements

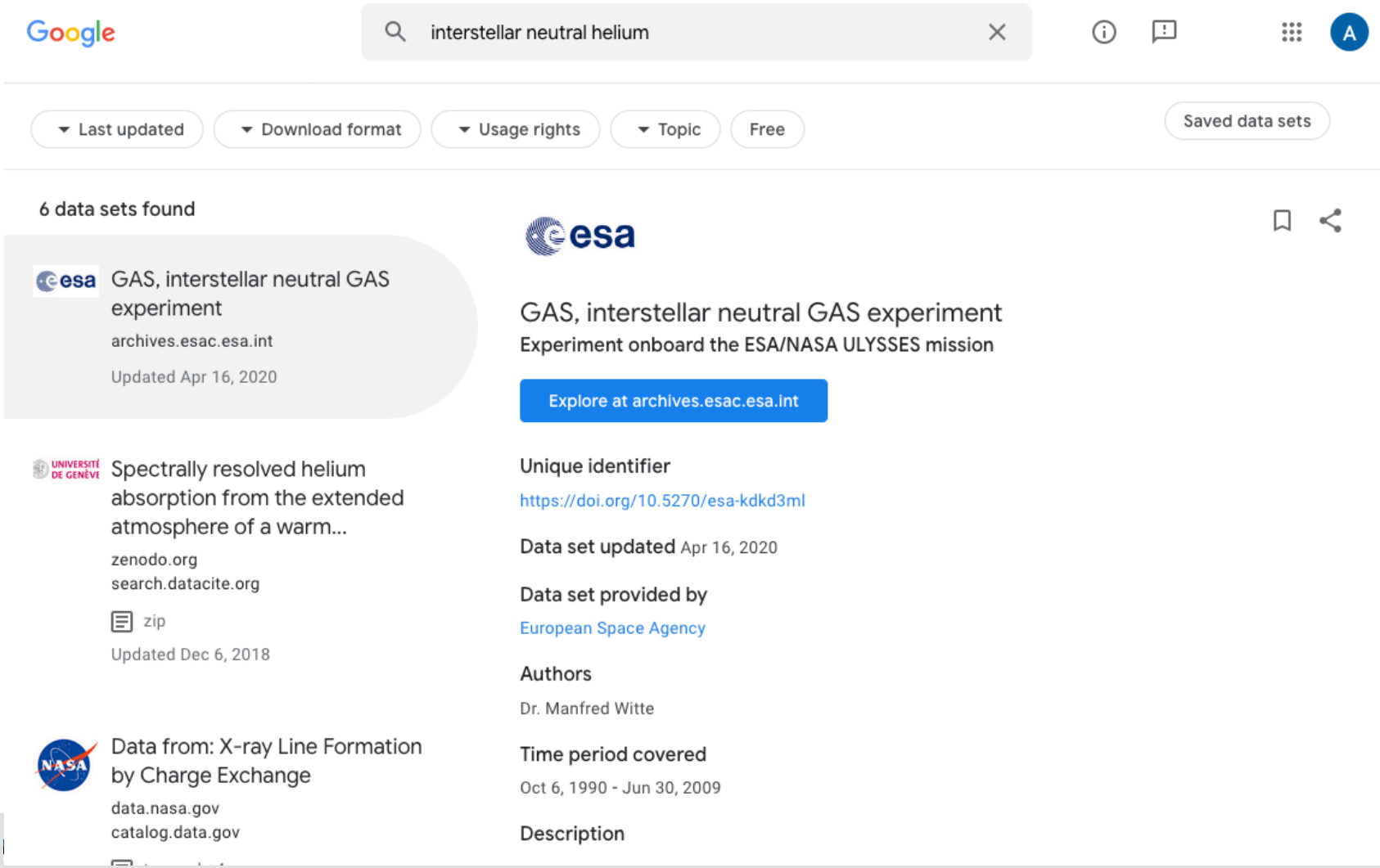


https://datasetsearch.research.google.com

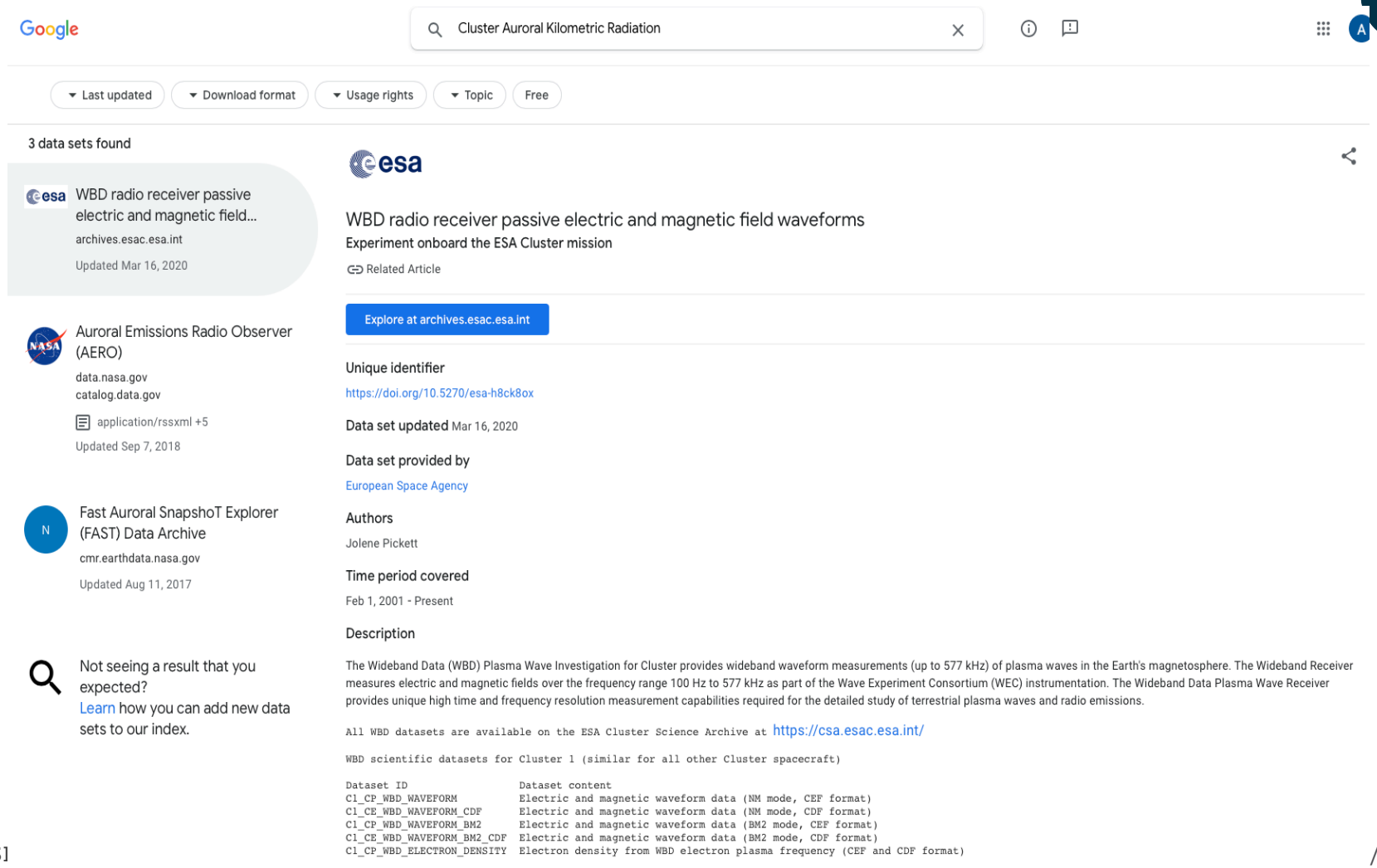
by scientific phenomena



https://datasetsearch.research.google.com

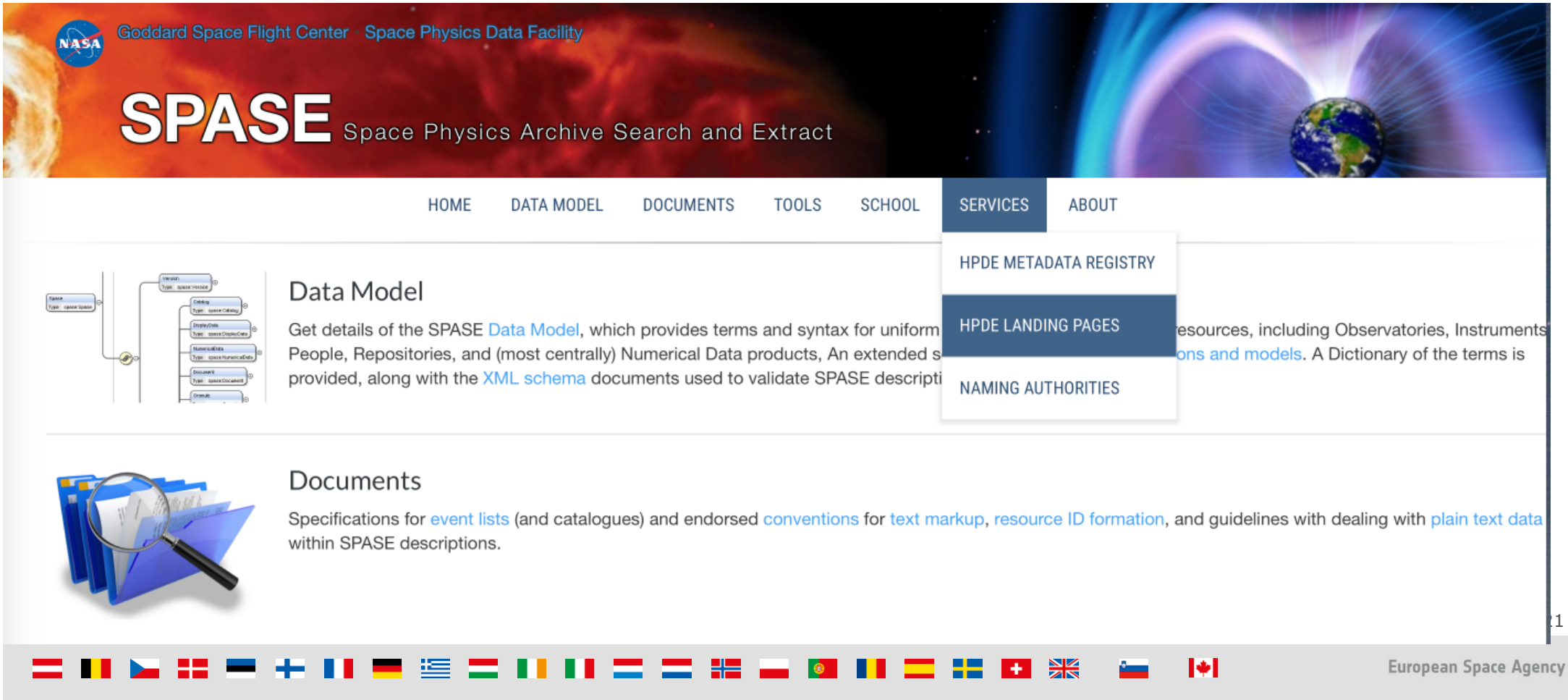


https://datasetsearch.research.google.com



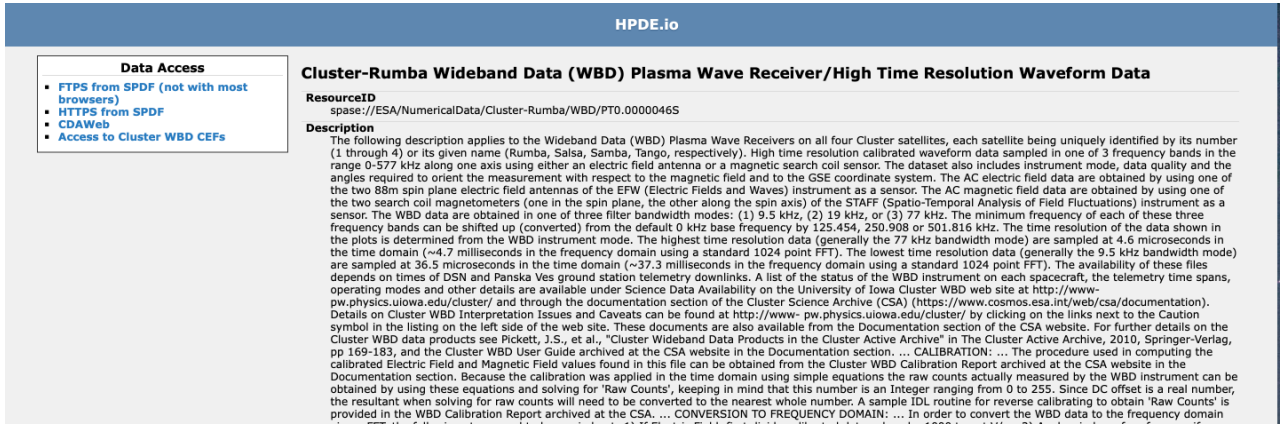
SPASE HPDE DOI landing pages

https://spase-group.org/



SPASE HPDE DOI landing page for WBD

https://hpde.io/ESA/NumericalData/Cluster-Rumba/WBD/PT0.0000046S.html



HPDE.io

Data Access

- FTPS from SPDF (not with most browsers)
- HTTPS from SPDF
- CDAWeb
- Access to Cluster WBD CEFs

Cluster-Rumba Wideband Data (WBD) Plasma Wave Receiver/High Time Resolution Waveform Data

ResourceID
spase://ESA/NumericalData/Cluster-Rumba/WBD/PT0.0000046S

Description
The following description applies to the Wideband Data (WBD) Plasma Wave Receivers on all four Cluster satellites, each satellite being uniquely identified by its number (1 through 4) or its given name (Rumba, Salsa, Samba, Tango, respectively). High time resolution calibrated waveform data sampled in one of 3 frequency bands in the range 0-577 kHz along one axis using either an electric field antenna or a magnetic search coil sensor. The dataset also includes instrument mode, data quality and the angles required to orient the measurement with respect to the magnetic field and to the GSE coordinate system. The AC electric field data are obtained by using one of the two 88m spin plane electric field antennas of the EFW (Electric Fields and Waves) instrument as a sensor. The AC magnetic field data are obtained by using one of the two search coil magnetometers (one in the spin plane, the other along the spin axis) of the STAFF (Spatio-Temporal Analysis of Field Fluctuations) instrument as a sensor. The WBD data are obtained in one of three filter bandwidth modes: (1) 9.5 kHz, (2) 19 kHz, or (3) 77 kHz. The minimum frequency of each of these three frequency bands can be shifted up (converted) from the default 0 kHz base frequency by 125.454, 250.908 or 501.816 kHz. The time resolution of the data shown in the plots is determined from the WBD instrument mode. The highest time resolution data (generally the 77 kHz bandwidth mode) are sampled at 4.6 microseconds in the time domain (~4.7 milliseconds in the frequency domain using a standard 1024 point FFT). The lowest time resolution data (generally the 9.5 kHz bandwidth mode) are sampled at 36.5 microseconds in the time domain (~37.3 milliseconds in the frequency domain using a standard 1024 point FFT). The availability of these files depends on times of DSN and Panska Ves ground station telemetry downlinks. A list of the status of the WBD instrument on each spacecraft, the telemetry time spans, operating modes and other details are available under Science Data Availability on the University of Iowa Cluster WBD web site at http://www-pw.physics.uiowa.edu/cluster/ and through the documentation section of the Cluster Science Archive (CSA) (https://www.cosmos.esa.int/web/csa/documentation). Details on Cluster WBD Interpretation Issues and Caveats can be found at http://www- pw.physics.uiowa.edu/cluster/ by clicking on the links next to the Caution symbol in the listing on the left side of the web site. These documents are also available from the Documentation section of the CSA website. For further details on the Cluster WBD data products see Pickett, J.S., et al., "Cluster Wideband Data Products in the Cluster Active Archive" in The Cluster Active Archive, 2010, Springer-Verlag, pp 169-183, and the Cluster WBD User Guide archived at the CSA website in the Documentation section. ... CALIBRATION: ... The procedure used in computing the calibrated Electric Field and Magnetic Field values found in this file can be obtained from the Cluster WBD Calibration Report archived at the CSA website in the Documentation section. Because the calibration was applied in the time domain using simple equations the raw counts actually measured by the WBD instrument can be obtained by using these equations and solving for 'Raw Counts', keeping in mind that this number is an Integer ranging from 0 to 255. Since DC offset is a real number, the resultant when solving for raw counts will need to be converted to the nearest whole number. A sample IDL routine for reverse calibrating to obtain 'Raw Counts' is provided in the WBD Calibration Report archived at the CSA. ... CONVERSION TO FREQUENCY DOMAIN: ... In order to convert the WBD data to the frequency domain

ObservedRegion
Earth.NearSurface.AuroralRegion

ObservedRegion
Earth.NearSurface.PolarCap

ObservedRegion
Earth.NearSurface.Ionosphere

ObservedRegion
Heliosphere.Inner

Caveats
See the Cluster WBD WBD data Interpretation Issues and Caveats documents at the web site ht
left hand side of the page. The documents can also be retrieved from the Documentation sectior

Keywords
Time Series
Waveform
AKR
Auroral hiss
Auroral Kilometric Radiation
Chorus
Continuum radiation
Equatorial Noise
Kilometric Continuum radiation
Plasmaspheric Hiss
Terrestrial Kilometric Radiation
TKR
Very Long Baseline Interferometry
VLBI
Langmuir Waves
Type III Solar Bursts
Electostatic Solitary Waves
ESW
Whistlers
Ion Acoustic Waves
Bernstein Waves
Upper Hybrid Waves
Lower Hybrid Waves

4. Who cares about heliophysics?

Planetary DOIs status

Guest Storage Facility is working and already contains 2 datasets with DOI directly linked to refereed papers (high value added products)

Thousands of landing pages already created, one per dataset, DOI assignement on-going

Astronomy DOIs status

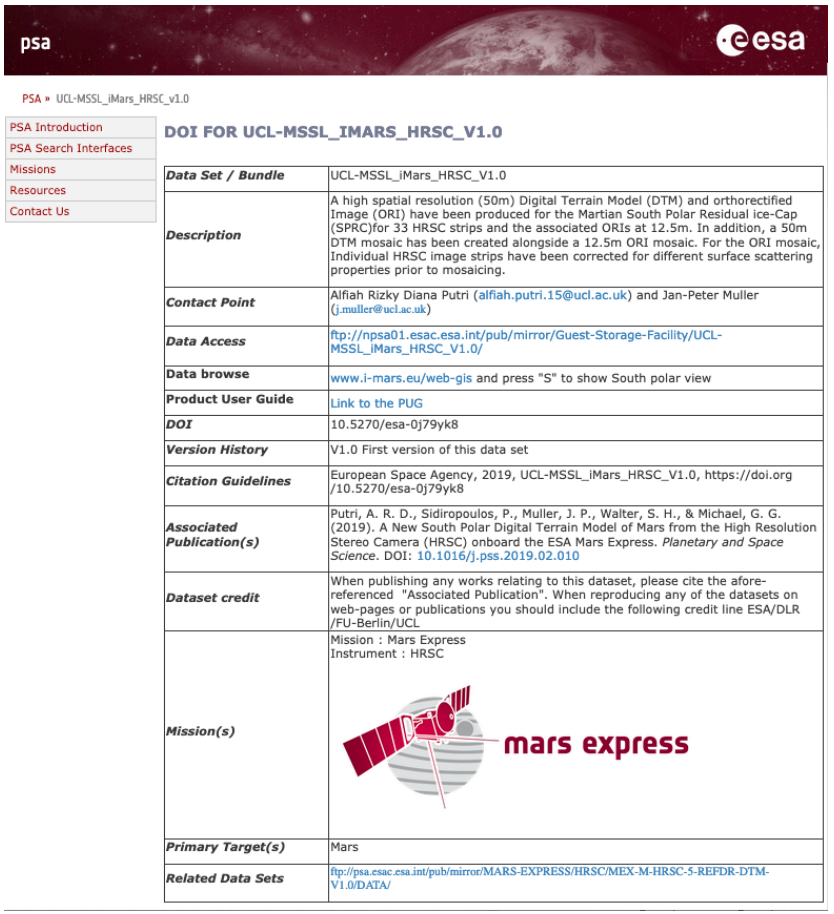
One per catalogue for survey missions (e.g. Planck, GAIA)

One per observing proposal for observatory type missions (e.g. XMM, Herschel)

A significant number of landing pages have already been generated

JSON automatic generation (keyword extraction):

AI techniques may be envisaged (MAUI, RAKE, spaCy, Alchemy…) instead of TF-IDF



psa

PSA » UCL-MSSL_iMars_HRSC_v1.0

- PSA Introduction
- PSA Search Interfaces
- Missions
- Resources
- Contact Us

DOI FOR UCL-MSSL_IMARS_HRSC_V1.0

Data Set / Bundle	UCL-MSSL_iMars_HRSC_V1.0
Description	A high spatial resolution (50m) Digital Terrain Model (DTM) and orthorectified Image (ORI) have been produced for the Martian South Polar Residual ice-Cap (SPRC)for 33 HRSC strips and the associated ORIs at 12.5m. In addition, a 50m DTM mosaic has been created alongside a 12.5m ORI mosaic. For the ORI mosaic, Individual HRSC image strips have been corrected for different surface scattering properties prior to mosaicing.
Contact Point	Alfiah Rizky Diana Putri (alfiah.putri.15@ucl.ac.uk) and Jan-Peter Muller (j.muller@ucl.ac.uk)
Data Access	ftp://npsa01.esac.esa.int/pub/mirror/Guest-Storage-Facility/UCL-MSSL_iMars_HRSC_V1.0/
Data browse	www.i-mars.eu/web-gis and press "S" to show South polar view
Product User Guide	Link to the PUG
DOI	10.5270/esa-0j79yk8
Version History	V1.0 First version of this data set
Citation Guidelines	European Space Agency, 2019, UCL-MSSL_iMars_HRSC_V1.0, https://doi.org/10.5270/esa-0j79yk8
Associated Publication(s)	Putri, A. R. D., Sidiropoulos, P., Muller, J. P., Walter, S. H., & Michael, G. G. (2019). A New South Polar Digital Terrain Model of Mars from the High Resolution Stereo Camera (HRSC) onboard the ESA Mars Express. *Planetary and Space Science*. DOI: 10.1016/j.pss.2019.02.010
Dataset credit	When publishing any works relating to this dataset, please cite the afore-referenced "Associated Publication". When reproducing any of the datasets on web-pages or publications you should include the following credit line ESA/DLR/FU-Berlin/UCL
Mission(s)	Mission : Mars Express Instrument : HRSC mars express
Primary Target(s)	Mars
Related Data Sets	ftp://psa.esac.esa.int/pub/mirror/MARS-EXPRESS/HRSC/MEX-M-HRSC-5-REFDR-DTM-V1.0/DATA/

5. Next steps and open questions

Closer to the data usage

Short time periods from a few datasets are often used in refereed papers (case studies)

Short compared to decades long files

1. On-demand DOI could link papers directly to the data bundle used (via TAP, Rest API)
2. Such a service is available e.g. for Hubble data at the Canadian Astronomy Data Centre
3. Saving the version of the data used in papers could improve reproducibility
4. Include data quality? Which ones should be used? Need for international standard?

See Ruggeri, K., et al., (2020), Standards for evidence in policy decision-making, Nature Research Social and Behavioural Sciences, 399005; go.nature.com/2zdTQIs

Conclusion

1. DOIs related to 47 experiments onboard ESA heliophysics spacecraft are registered
2. This increases ESA commitment to the FAIR data principles and acknowledge PI work
3. All of them are findable on Google dataset search
4. A related peer reviewed manuscript is under revision (COSPAR Adv. Space Res.)
5. Discussion with editors to eventually enforce their usage has started
6. Long term goal: provide on-demand DOI and save the version of data used in papers

Questions? Arnaud.Masson@esa.int