Annual Report from Space Physics Data Facility (SPDF)

https://spdf.gsfc.nasa.gov

Lan Jian, Robert Candey, Tamara Kovalick

NASA Goddard Space Flight Center, MD, USA

International Heliophysics Data Environment Alliance (IHDEA) Meeting

ESAC, Spain 17 October 2024

Heliophysics Missions

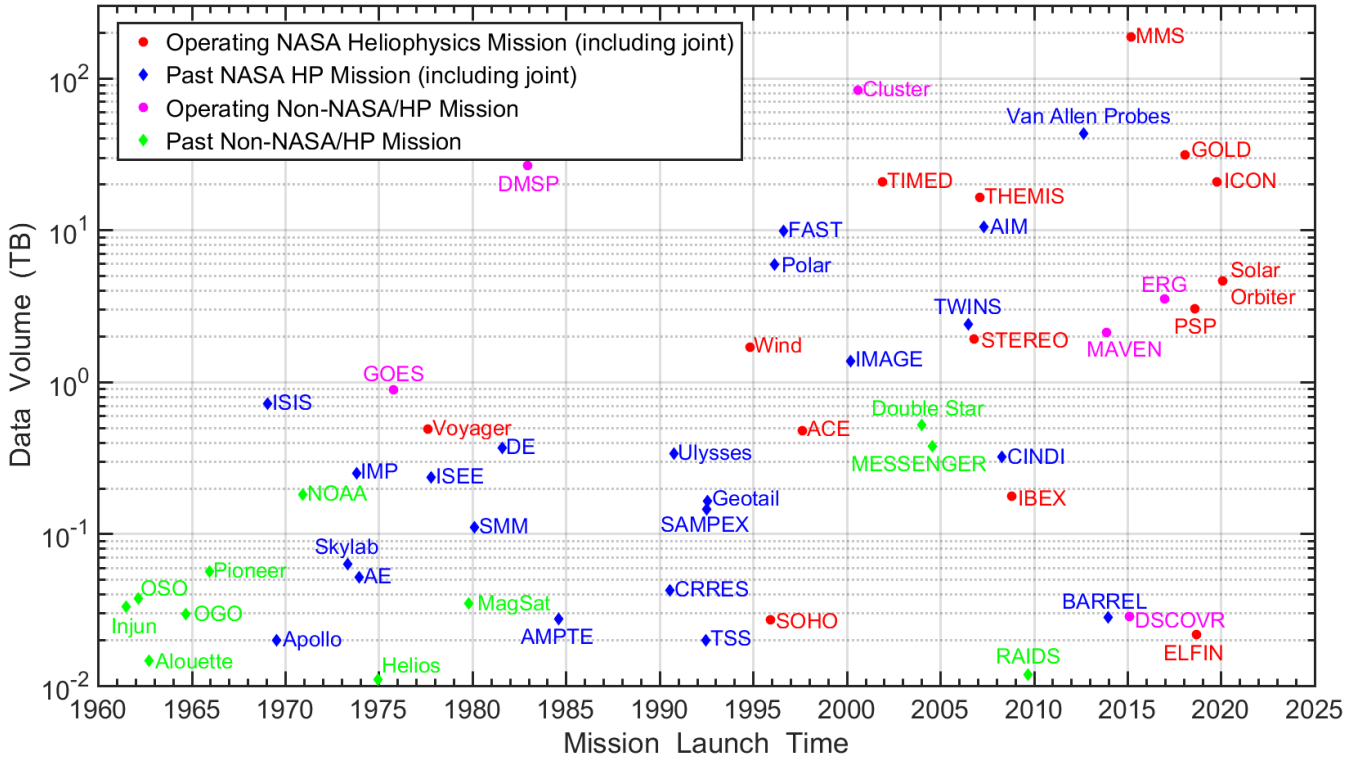
Introduction

- SPDF team include 10 curation scientists (mostly part time, covering a wide range of heliophysics domains) and 13 technical and development staff
- Established in 1990s, SPDF is now the **active and final archive of non-solar or space physics data** (mostly in situ data) from NASA Heliophysics missions, including joint missions with other US or foreign agencies, to enable correlative and collaborative research across discipline and mission boundaries
- SPDF also archives other data **relevant to NASA Heliophysics Science Objectives** (often per the request of missions/projects)
  - Related data from planetary missions (e.g., MAVEN, New Horizons)
  - Heliophysics data from satellites of NOAA, DoD, or other agencies (e.g., GOES, DMSP)
  - Ground-based magnetometers, aurora cameras, radars, etc., which are funded by NSF or other agencies (the majority are archived at relevant facilities other than SPDF)
- Besides local archival, data are backed up in Iron Mountain and NCCS on premises

Mission Data Volume Archived at SPDF

SPDF data holdings span 100+ missions over 65 years

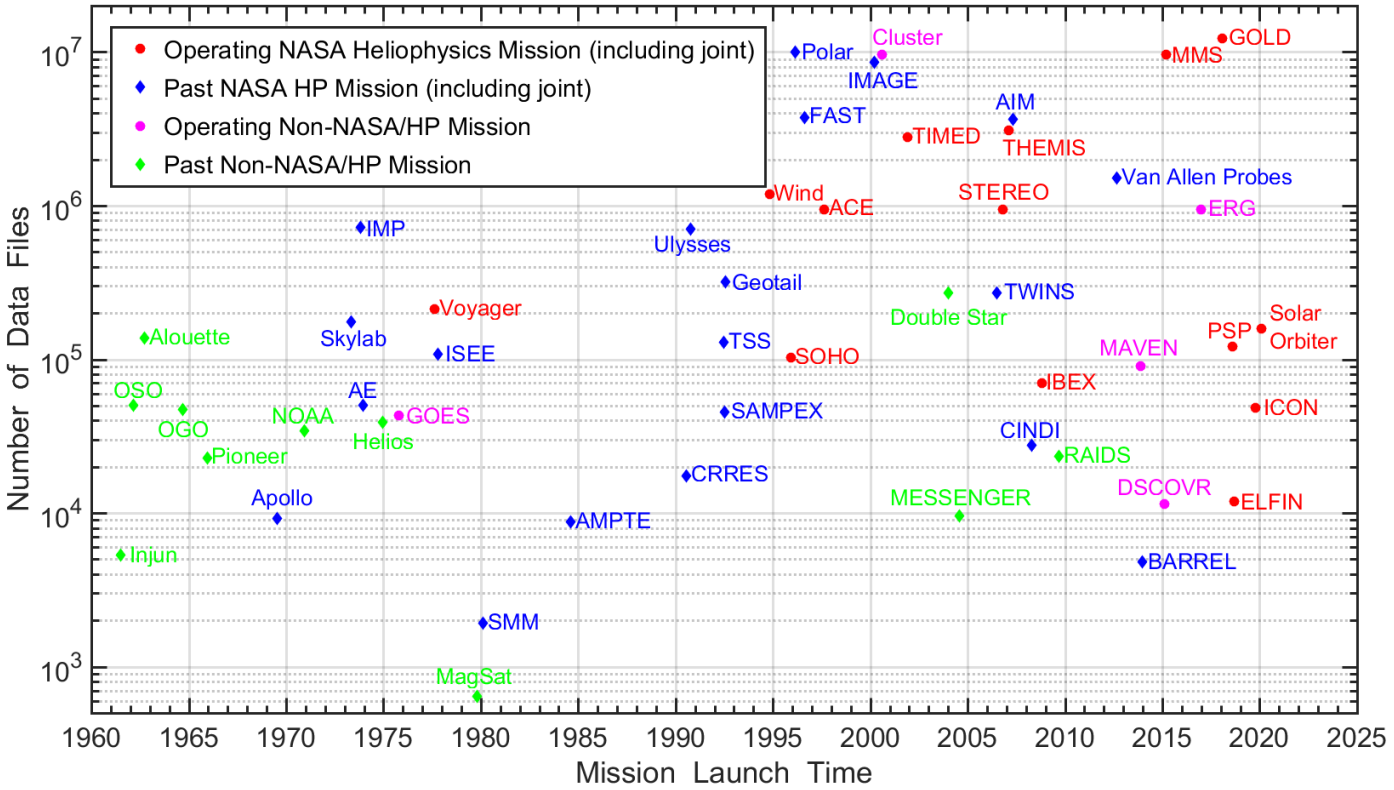
Covering **from the Sun to the local interstellar medium** including magnetosphere – ionosphere – thermosphere – mesosphere (M-ITM) of Earth and other applicable planets



File Numbers of Mission Data Archived at SPDF

~3000 CDF datasets, ~5000 non-CDF datasets, ~100M files, ~600 TB data

Guidelines for data delivery: https://spdf.gsfc.nasa.gov/archive_newdata_reqt.html



SPDF's Science-Enabling Services

- **Coordinated Data Analysis Web (CDAWeb)**
  - Data access through web browser, API, IDL, Python
  - Interface for browsing, correlating, and displaying data (audio and movie for special cases) from 60+ missions or mission groups and multiple instruments
  - *Inventory plot and usage statistics* for mission data
- **Satellite Situation Center (SSC)**
  - Data access through web browser, API
  - ~160 missions, orbit/ground track displays and queries
  - Coordinate transformation tools
- **OMNI Web (including COHO Web)**
  - Solar wind plasma, magnetic field, and energetic particle data at the nose of Earth's bow shock and other locations of the heliosphere
  - Interface for plotting, filtering, and statistical analysis
- Critical Infrastructures for the **Heliophysics Data Environment**
  - Common Data Format (CDF): self-describing science file format (cdf.gsfc.nasa.gov)
  - International Solar-Terrestrial Physics (ISTP) Metadata Standards for CDF (NetCDF) data including global and variable attributes
  - Heliophysics Data Portal: discipline-wide data inventory and access service

SPDF Data Archiving in the Past Year

- ✓ Automated ingestion pipeline for > 75 missions out of over 160 missions for a total of ~4,000 datasets and ~600 TB (ingestion and usage logs in https://cdaweb.gsfc.nasa.gov/publiclogs/)
- ✓ Recent average monthly data ingestion rate: ~0.7 million files, ~14 TB data
- ✓ Assisting Phase F of AIM and ICON missions to finalize their data archiving
- ✓ Added new datasets from GOLD, PSP, Solar Orbiter, MMS, IBEX, DMSP, and other missions
- ✓ Added new missions: **BioSentinel, Endurance, AWE** (in progress)
- ✓ Continuing the population of OMNI, COHO, SSC databases

Update of Cluster & Solar Orbiter Data

- **Cluster**: 771 CDF datasets and some canned graphics files, 11M files, 86 TB
  - In the past 3 years, Pertti Makela combed through 1960 datasets from Cluster Science Archive (CSA) with the help of Harri Laakso, and selected 771 to be archived at SPDF based on their science use, excluding datasets without metadata or of limited interest like engineering
  - New ingestion scripts for CDF datasets were set in addition to the previous ones that SPDF has been getting for years from CSA and other sources
  - Many datasets still in CDAWeb test site
    - Majority: variable **VALIDMIN/MAX** values are not defined to be applied across the dataset, but rather for a given file → Pertti is running software across each dataset to determine the correct values and adding them to CDAWeb master CDFs
    - Other issues with the FILLVAL value and other important attributes
    - Once these items are corrected and the display types are set up, the datasets will be moved to public CDAWeb
- **Solar Orbiter**: SPDF ingested from GSFC's mirroring site of Solar Orbiter Archive, 0.17M data files, 5 TB
  - 101 datasets in public CDAWeb, master CDFs were changed earlier this year for some datasets due to changes of variables
  - 13 datasets in test CDAWeb because of issues such as frequent changes of dimension size for variables

CDF Status and Recent Development

- CDF 3.9.1 C-library to be released soon with improved buffer overflow warnings and compression functions, updated zlib, and better support for very long attribute and variable names
- Continued CDF support and general development, with added features
- Improving documentation, beginner's guides, adding to Wikipedia CDF entry
- Establishing Steering Committee for ISTP Metadata Guidelines, standardizing ISTP Metadata Guidelines with version control and governance at https://github.com/IHDE-Alliance/ISTP_metadata

Robert Candey's talk tomorrow

- Updated CDFML and its corresponding JSON representation with cdf.xsd to use more specific datatypes (e.g., xs:dataTime, xs:integer, xs:float, etc.) instead of just xs:string
- Investigated multi-thread support in CDF C library but the performance was worse due to maintaining the locks during read/write operations, so not released
- External groups have written CDF libraries in C++, JavaScript

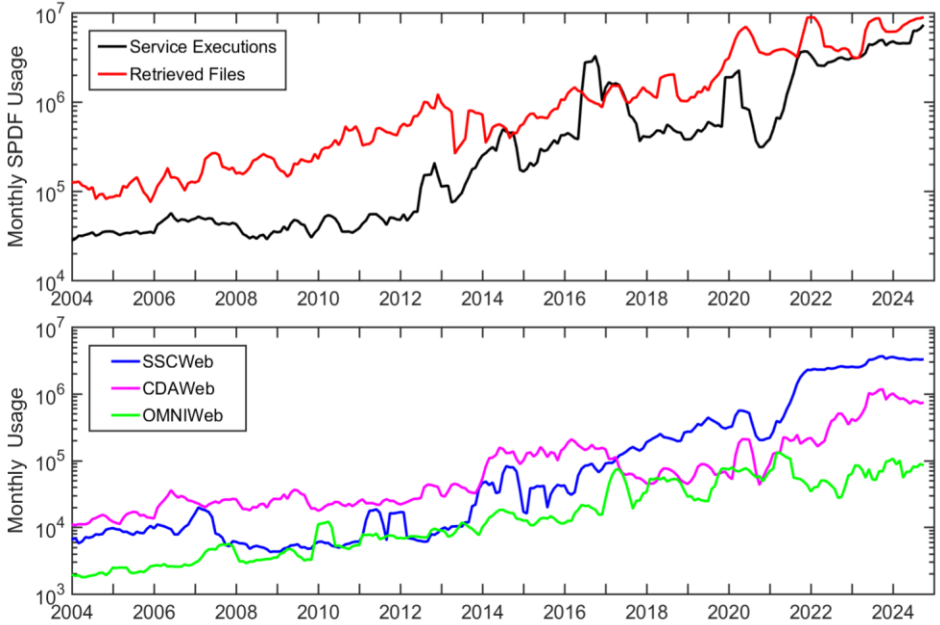
CDF Plans

- Releasing **new CDF website** with quick-start guides and more responsive design
- Releasing new ISTP metadata editor (SKTeditor) in JavaScript, adding Space Physics Archive Research and Extract (SPASE) metadata creation
- Supporting CDFs in cloud object storage
- Defining CDF MIME type and international standard
- Apache 2 license in place of current custom license
- Adding support for CDF to command line NetCDF tools, such as NCO, NCAR, ANTS, NCtools
- CDF gap checker to write filename, variable name, begin and end time, number of records, and any gaps greater than a certain amount (G-good, M-missing, F-fill, R-outside range, B-backward time)
- Adding CDF support to Octave, Gnu Data Language (GDL), Excel, Ruby, C++, WebWinds, LinkWinds, Opendap, SWIG.org

Eric Grimes's talk tomorrow

SPDF Other Activities in the Past Year

- Released **new SPDF website** following US Government web design system (USWDS) in late 2023
- Released Plot Walk and JavaScript-based 4D Orbit Viewer
- Working on new SSC and CDF websites
- Copying science level data from SPDF/CDAWeb to HelioCloud, supported HDRL outreach
- Improved CDAWeb plot and display: adjustable X and Y scales, autoscale time axis for intermittent (burst mode) data

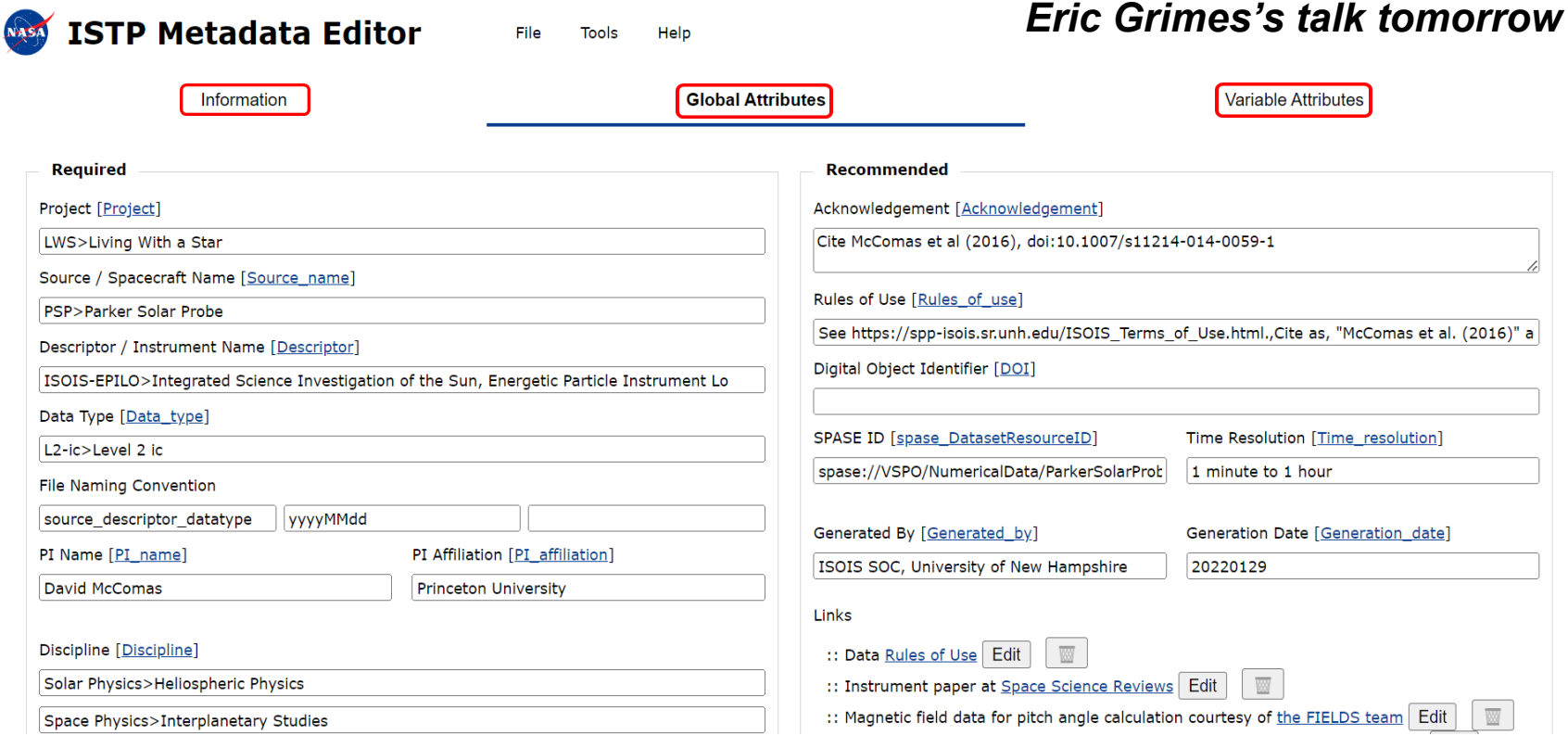


About 50% of *JGR-Space Physics* and *Space Weather* papers in 2023 acknowledged SPDF services and data https://spdf.gsfc.nasa.gov/Acknowledgements.html

Development of ISTP Metadata Editor

New JavaScript based web-browser tool to help users
create/update CDF datasets with ISTP and SPASE metadata

Eric Grimes's talk tomorrow

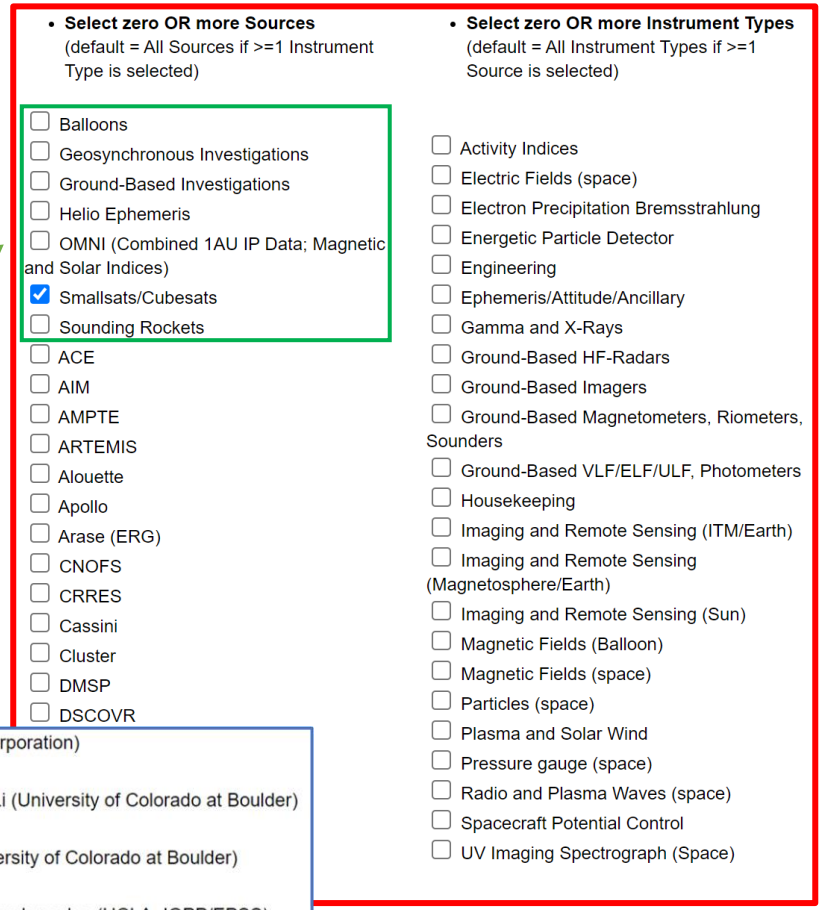


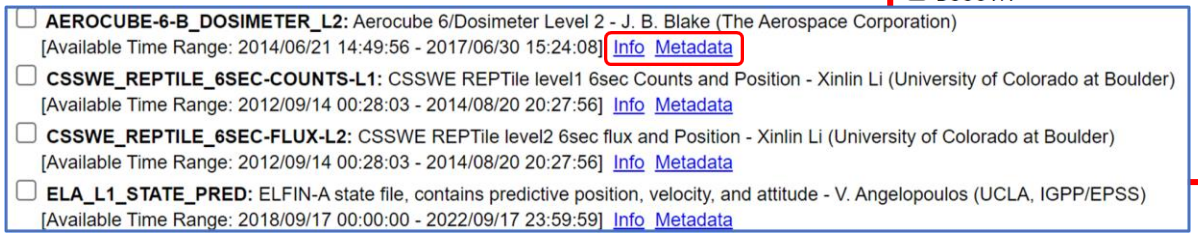
60+ Missions or Mission Groups →

Coordinated Data Analysis Web (CDAWeb)

https://cdaweb.gsfc.nasa.gov/

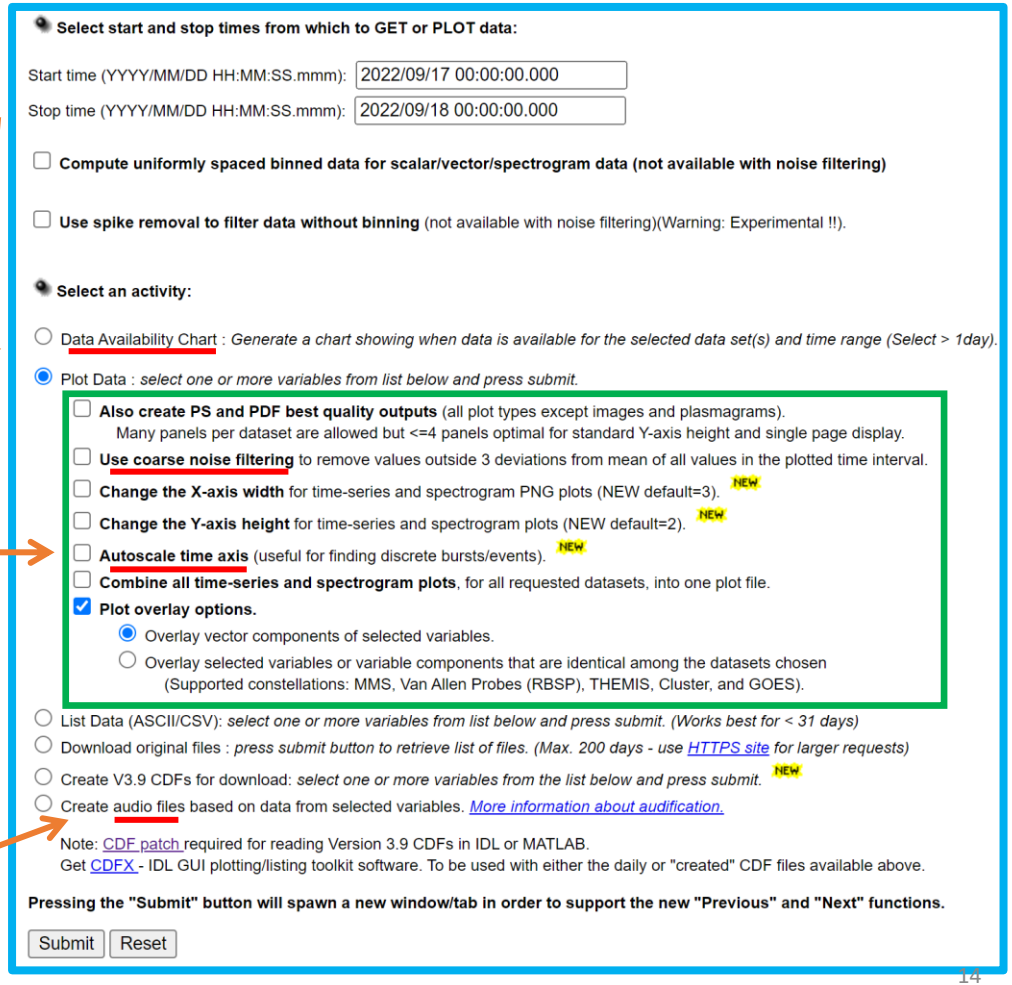
- **Special data source groups**: balloons, ground-based investigations, cubesats, sounding rockets, etc.
- Enable Systems Science: cross-mission, multi-instrument science
- Present dataset view rather than individual files
- 70% of 2744 datasets in CDAWeb have SPASE records; 57% have DOIs





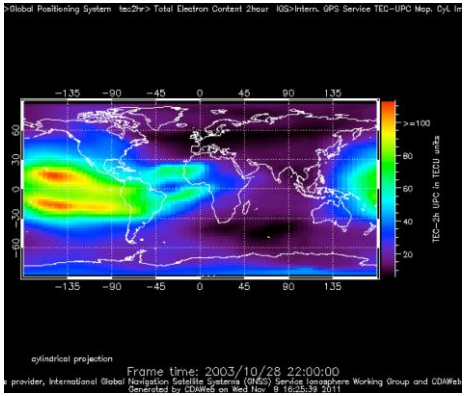
CDAWeb Data Explorer

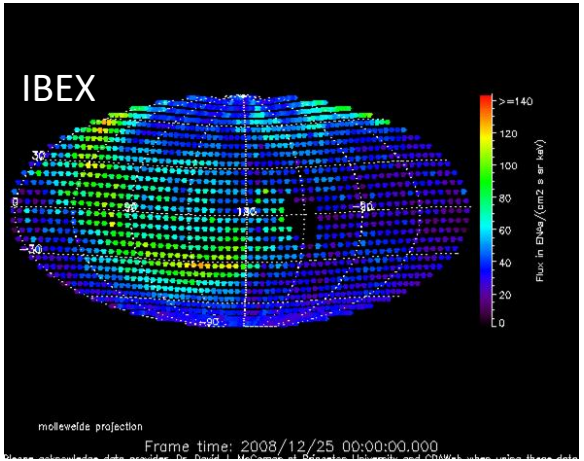
- Time interval is automatically set by the last available day of the selected dataset(s)
- Remove spikes or filter coarse noise
- **Plot data availability**
- Adjust X and Y lengths for plotting
- **Auto scale time axis for finding discrete bursts or events**
- Overlay vector components of selected variables, or selected variables that are identical among multiple datasets
- Output a subset or a superset of datasets in CDF, ASCII/CSV, JSON
- Create **audio** and **movie** files for selected variables



Example Parameter Displays in CDAWeb

GPS International GNSS Service
Total Electron Content

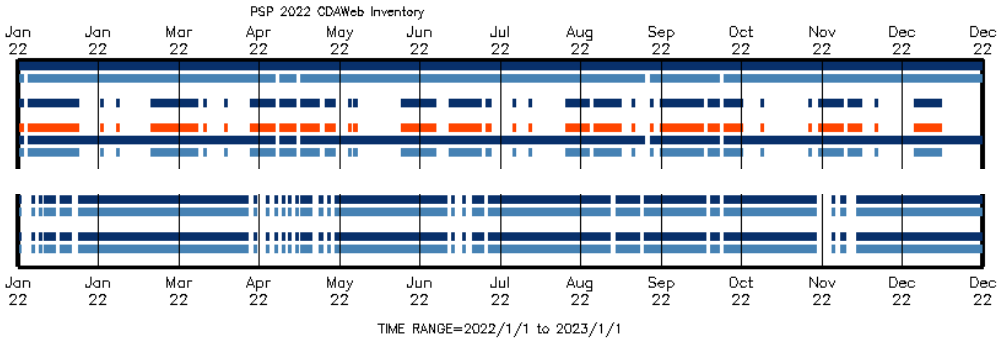




More at https://cdaweb.gsfc.nasa.gov/about.html

Inventory Plot for Mission Datasets

PSP_COHO1HR_MERGED_MAG_PLASMA
PSP_FLD_L2_AEB
PSP_FLD_L2_DFB_AC_BPF_DV12HG
PSP_FLD_L2_DFB_AC_BPF_DV34HG
PSP_FLD_L2_DFB_AC_BPF_SCMULFHG
PSP_FLD_L2_DFB_AC_BPF_SCMUMFHG
PSP_FLD_L2_DFB_AC_SPEC_DV12HG
PSP_FLD_L2_DFB_AC_SPEC_DV34HG
PSP_FLD_L2_DFB_AC_SPEC_SCMDLFHG
...
PSP_SWP_SPI_SF00_L3_MOM
PSP_SWP_SPI_SF00_L3_MOM_INST
PSP_SWP_SPI_SF01_L2_8DX32EX8A
PSP_SWP_SPI_SF0A_L3_MOM
PSP_SWP_SPI_SF0A_L3_MOM_INST



https://cdaweb.gsfc.nasa.gov/sc_inventory_plots/

Each supported dataset also provides links to IDL and Python code examples for downloading and working with the data files independently (outside of the CDAWeb system)

CDAS Web Service Client Code Examples

The following web service client code examples demonstrates how to access data from the AEROCUBE-6-B DOSIMETER L2 dataset from particular programming environments.

Jupyter Notebook on Binder

The following link launches a Python Jupyter Notebook that demonstrates using the cdasws library to access AEROCUBE-6-B DOSIMETER L2 data in a Jupyter Notebook. It is merely an example and does not show all the capabilities of the library. You should edit the code to suit your needs.

launch binder

cdasws Python Library

The following code demonstrates using the cdasws library to access AEROCUBE-6-B
should edit the code to suit your needs.

```
#  Install these prerequisites once before executing the example code:
#  Option 1.
#    Install CDF from https://cdf.gsfc.nasa.gov/
#    pip install -U spacepy
#    pip install -U cdasws
#  Option 2.
#    pip install -U xarray
#    pip install -U cdflib
#    pip install -U cdasws

from cdasws import CdasWs
cdas = CdasWs()

#  Edit the following vars, time variables, and printing to suit your
environment
# (spacepy or cdflib) and needs.
vars =
['alt','lat','lon','XYZ_GEO','dos1l','dos1m','dos1rate','dos2l','dos2m'
```

cdasws IDL Library

The following code demonstrates using the cdasws library to access AEROCUBE-6-B DOSIMETER L2 data in IDL. It is merely an example and does not show all the capabilities of the library. You should edit the code to suit your needs.

```
compile_opt idl2
savFilename = filepath('spdfcdas.sav', /tmp)
oUrl = obj_new('IDLnetUrl')
;  For IDL installations with old root certificates
oUrl->setProperty, SSL_VERIFY_PEER=0
savFilename = oUrl->get(filename=savFilename,
url='https://cdaweb.gsfc.nasa.gov/WebServices/REST/spdfcdas.sav')
restore, savFilename

;  Edit the following vars and time variables to suit your needs.
vars =
['alt','lat','lon','XYZ_GEO','dos1l','dos1m','dos1rate','dos2l','dos2m','dos2rat
e','dos3l','dos3m','dos3rate','flag','Sample_Rate','Lm_IGRF','Bmag_IGRF','MLT_IG
RF','InvLat_IGRF','Lm_OPQ','Bmag_OPQ','MLT_OPQ','InvLat_OPQ','Loss_Cone_Type','B
xyz_GEO','Beq','I','K','K_Z','Lstar','Lstar_Z','Hmin','Hmin_Z','Loss_Cone_Near',
'Loss_Cone_Far','B100N','LAT100N','LON100N','B100S','LAT100S','LON100S','Alpha',
'Alpha_X','Alpha_Y','Alpha_Eq','Beta','Beta_X','Beta_Y','Phi_B','OmegaXYZ_GEO','
B_spin','Spin_Sun','Dist_In_Track','Lag_In_Track','Dist_Cross_Track_Horiz','Dist
_Cross_Track_Vert','Dist_Total','alt_10Hz','lat_10Hz','lon_10Hz','dos1l_10Hz','d
os1m_10Hz','dos1rate_10Hz','dos2l_10Hz','dos2m_10Hz','dos2rate_10Hz','dos3l_10Hz
','dos3m_10Hz','dos3rate_10Hz','flag_10Hz','Subcom_10Hz','Lm_OPQ_10Hz','Bmag_OPQ
_10Hz','MLT_OPQ_10Hz','InvLat_OPQ_10Hz','Loss_Cone_Type_10Hz','K_Z_10Hz','Lstar_
Z_10Hz','Hmin_Z_10Hz','Alpha_10Hz','Beta_10Hz','Dist_In_Track_10Hz','Lag_In_Trac
```

Copy code to clipboard | Download code

More information about using this library is available from the following:

- IDL library description cdasws
- Jupyter IDL notebook example
- Application Programming Interface description API

Alternative data access methods https://cdaweb.gsfc.nasa.gov/alternative_access_methods.html

Update of Webpages

Linking SPDF Services with Missions

https://spdf.gsfc.nasa.gov/data_orbits.html (Partial Screenshot Below)

160+ missions

Click an SPDF service name to check mark (✔) the spacecraft whose data are available.
Click a spacecraft name to check mark (✔) the SPDF services with its data.
See Info for New Users for more information about these data services.

GO =Go to Service's Home Page ⓘ=Show Source Info *= Orbit/Trajectory Data Only

DATA SERVICES:

✔ CDAWeb GO

High-resolution, current space physics data with graphics and listings from many missions.

OMNIWeb Plus GO

Hourly-averaged solar wind magnetic field and plasma, etc.

✔ GIFWalk GO

Browse pregenerated data and orbit plots

✔ SPDF FTPS Site GO

Read FTP to FTPS information.

✔ SPDF HTTPS Site GO

ORBIT/ TRAJECTORY SERVICES:

✔ SSCWeb Services GO

Display and download trajectory

SOURCE SPACECRAFT:

ACE ⓘ
Active* ⓘ
Aerocube ⓘ
Aeros ⓘ
AIM ⓘ
Akebono* ⓘ
Alouette1 ⓘ
Alouette2 ⓘ
AMPTE ⓘ
APEX-MAIN* ⓘ
Apollo ⓘ
Aqua ⓘ
Ariel-4 ⓘ
Arase (ERG) ⓘ
ARCAD ⓘ
ARTEMIS ⓘ
ASTRID II* ⓘ
AE ⓘ
Aura ⓘ
Aureol2 ⓘ
BARREL ⓘ
BepiColombo ⓘ
BioSentinel ⓘ

Explorer ⓘ
FAST ⓘ
FIREBIRD* ⓘ
Formosat ⓘ
Freja* ⓘ
Galileo* ⓘ
Gateway* ⓘ
GCOM W1 ⓘ
Genesis ⓘ
Geotail ⓘ
Giotto* ⓘ
GOCE ⓘ
GOES18 ⓘ
GOES19 ⓘ
GOLD ⓘ
GPS ⓘ
GMS 3 ⓘ
GRACE* ⓘ
Granat ⓘ
Hawkeye ⓘ
Helios ⓘ
Hinode ⓘ
Hinotori ⓘ
IBEX ⓘ

Landsat ⓘ
LANL ⓘ
LRO ⓘ
LUNA ⓘ
Magsat ⓘ
MAP ⓘ
Mariner 10 ⓘ
Mars ⓘ
MAVEN ⓘ
MESSENGER ⓘ
MGS ⓘ
Microlab 1 ⓘ
Mir* ⓘ
MMS ⓘ
MRO ⓘ
MSL ⓘ
MSX* ⓘ
Munin ⓘ
Neptune ⓘ
New Horizons ⓘ
NOAA* ⓘ
Oersted ⓘ
OGO ⓘ
Ohzora ⓘ

San Marco ⓘ
Saturn ⓘ
SCATHA* ⓘ
SDO ⓘ
SET-1/DSX ⓘ
SMILE ⓘ
SNOE ⓘ
SOHO ⓘ
Solar Orbiter ⓘ
SORCE ⓘ
Spartan-A ⓘ
Spitzer ⓘ
SPORT ⓘ
Sputnik 1 ⓘ
STEREO ⓘ
Suisei ⓘ
Swarm ⓘ
Tatiana ⓘ
THEMIS ⓘ
TIMED ⓘ
TRACE ⓘ
TSS-1R ⓘ
TWINS ⓘ
UARS* ⓘ

OTHER DATA SOURCES:

Planet & Comet Positions

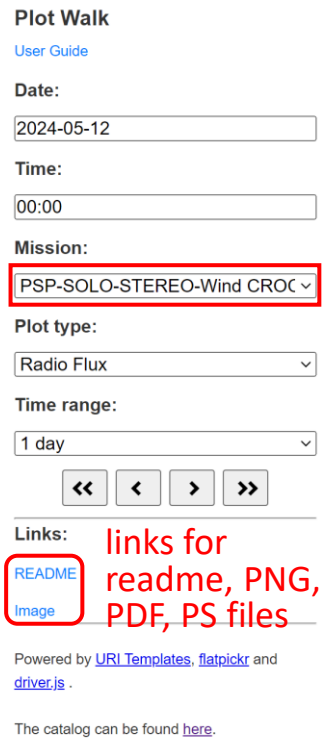
Ground-based

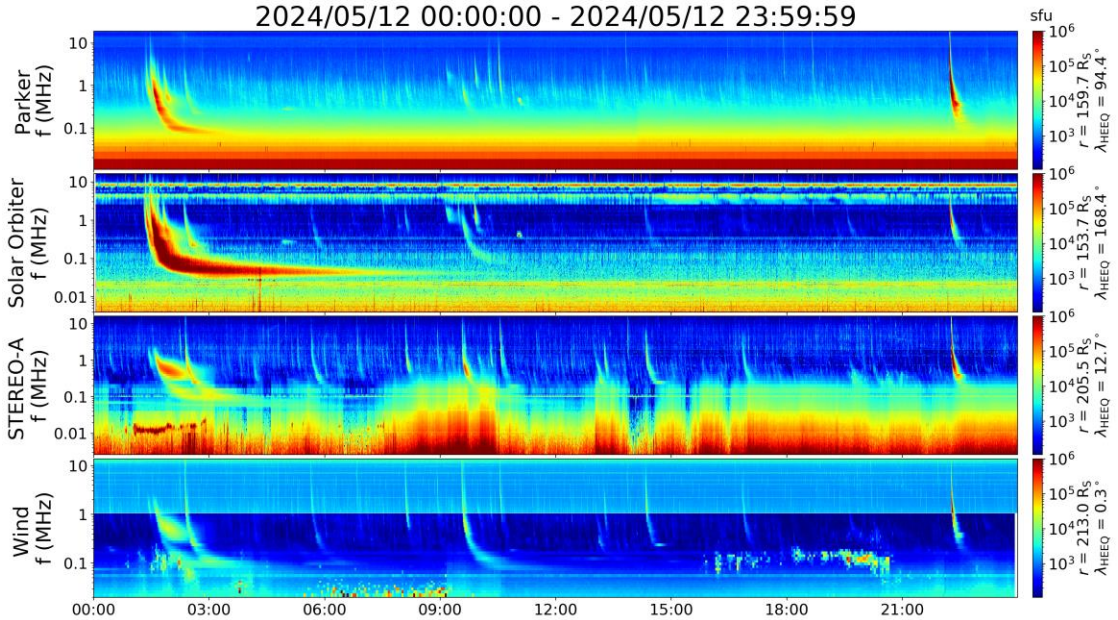
Activity Indices

Plot Walk for Pre-Generated Plots

https://spdf.gsfc.nasa.gov/plot_walk/

Summary or quick-look plots from 20+ missions (12.5 million plots)

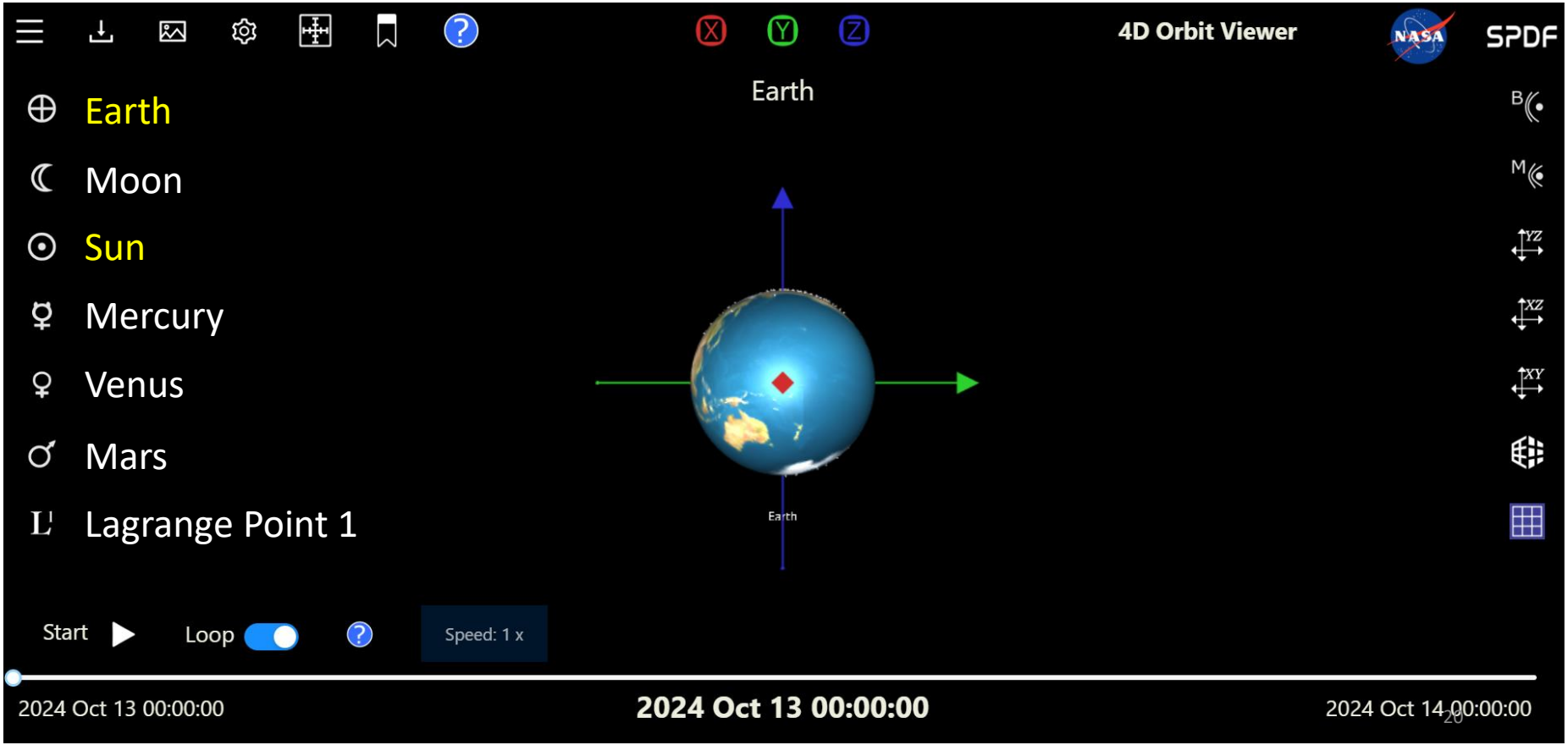




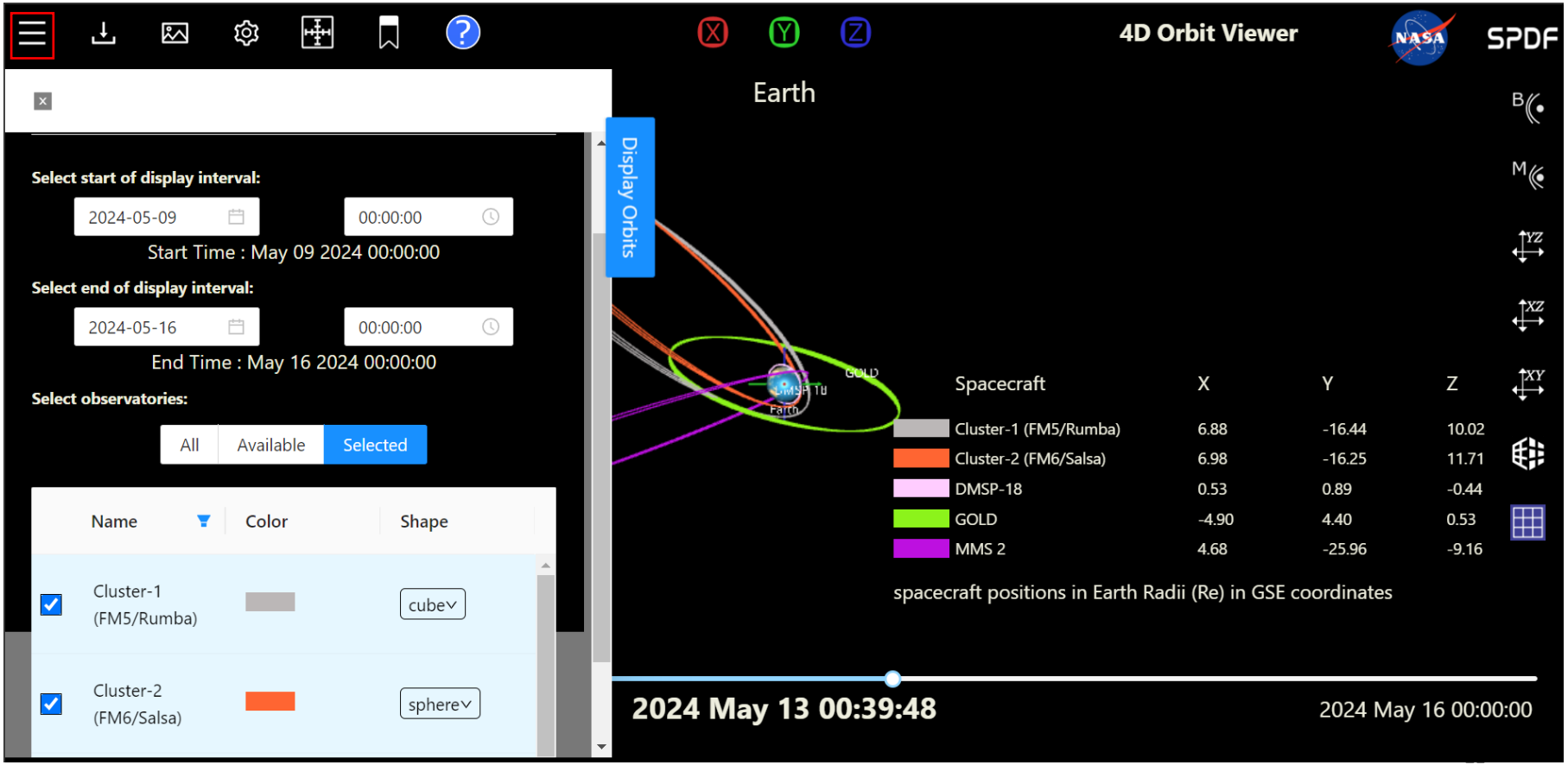
Coordinated Radiodiagnostics of CMEs and Solar flares (CROCS) plots using radio data from PSP, Solar Orbiter, STEREO A, and Wind spacecraft

4-D Orbit Viewer (160+ Spacecraft)

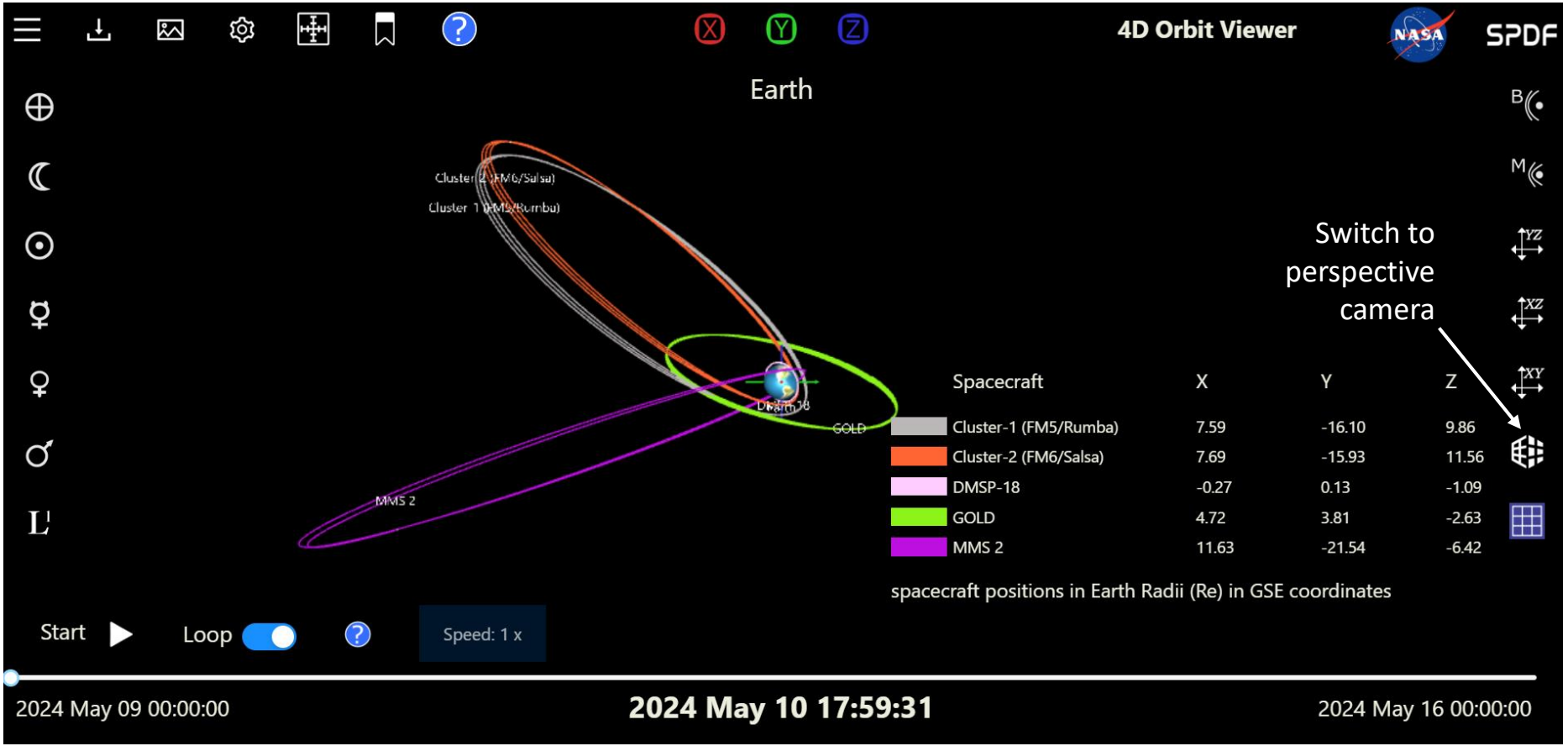
https://sscweb.gsfc.nasa.gov/4dorbit/



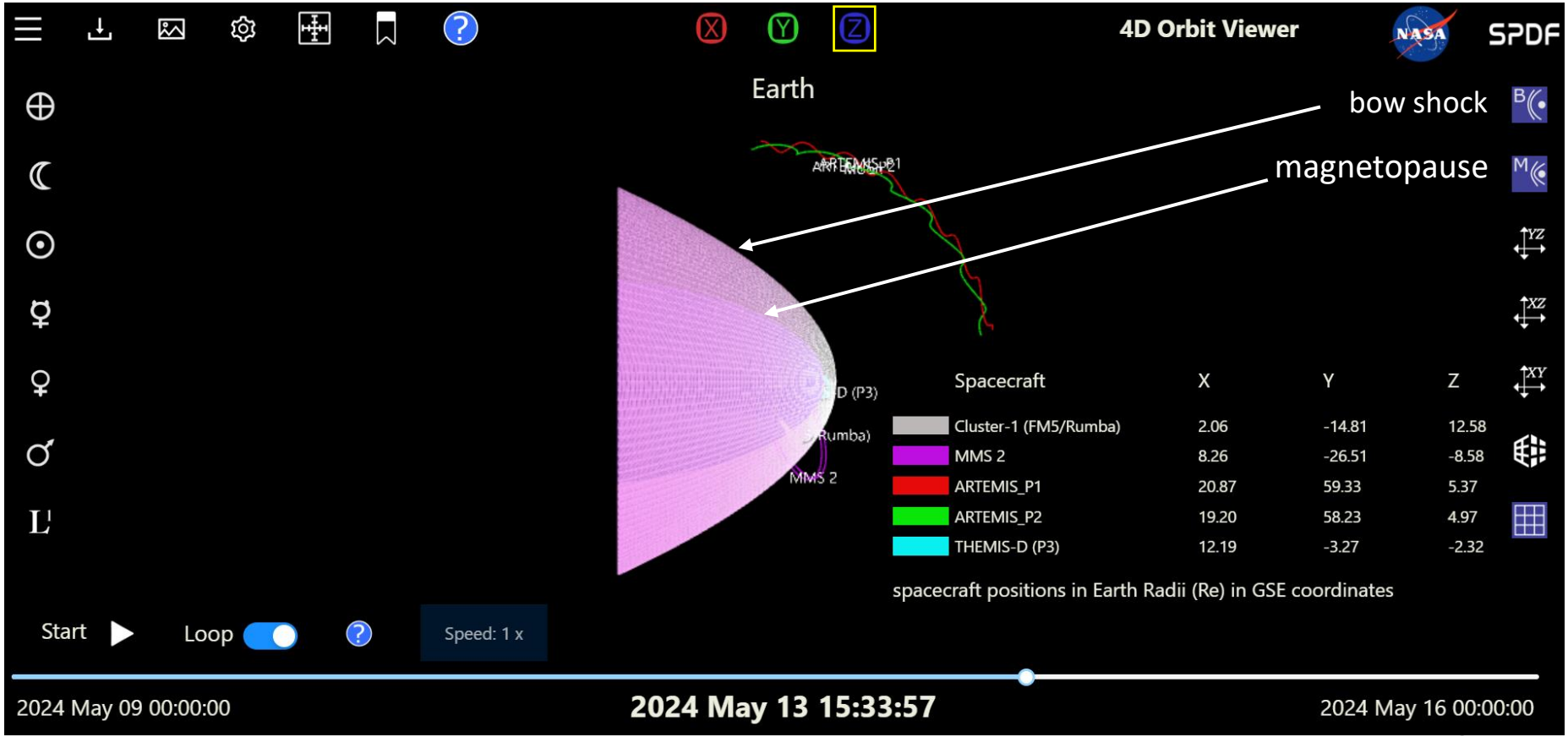
4-D Orbit Viewer: Time and Spacecraft Selection



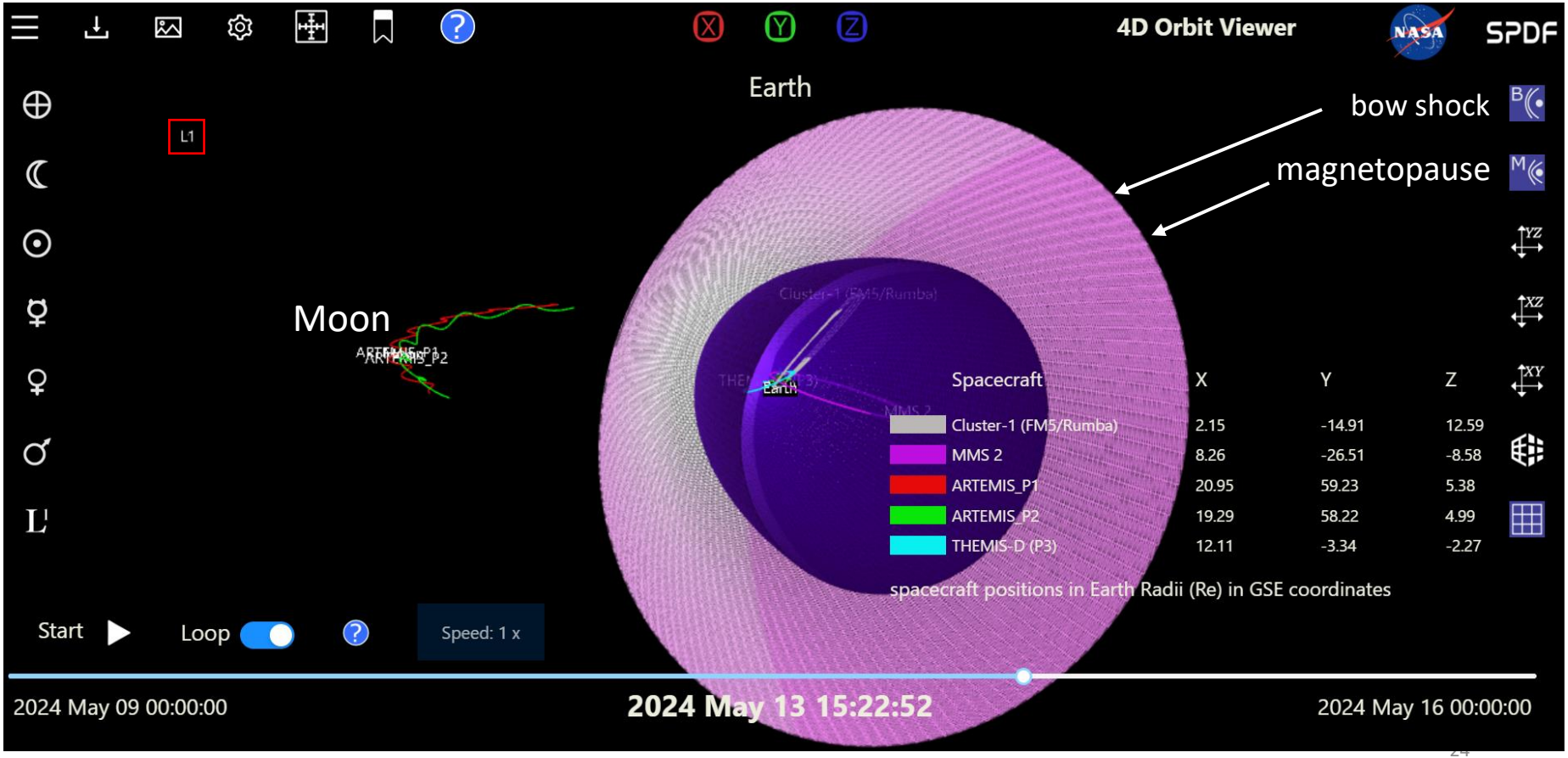
4-D Orbit Viewer: Different Perspectives



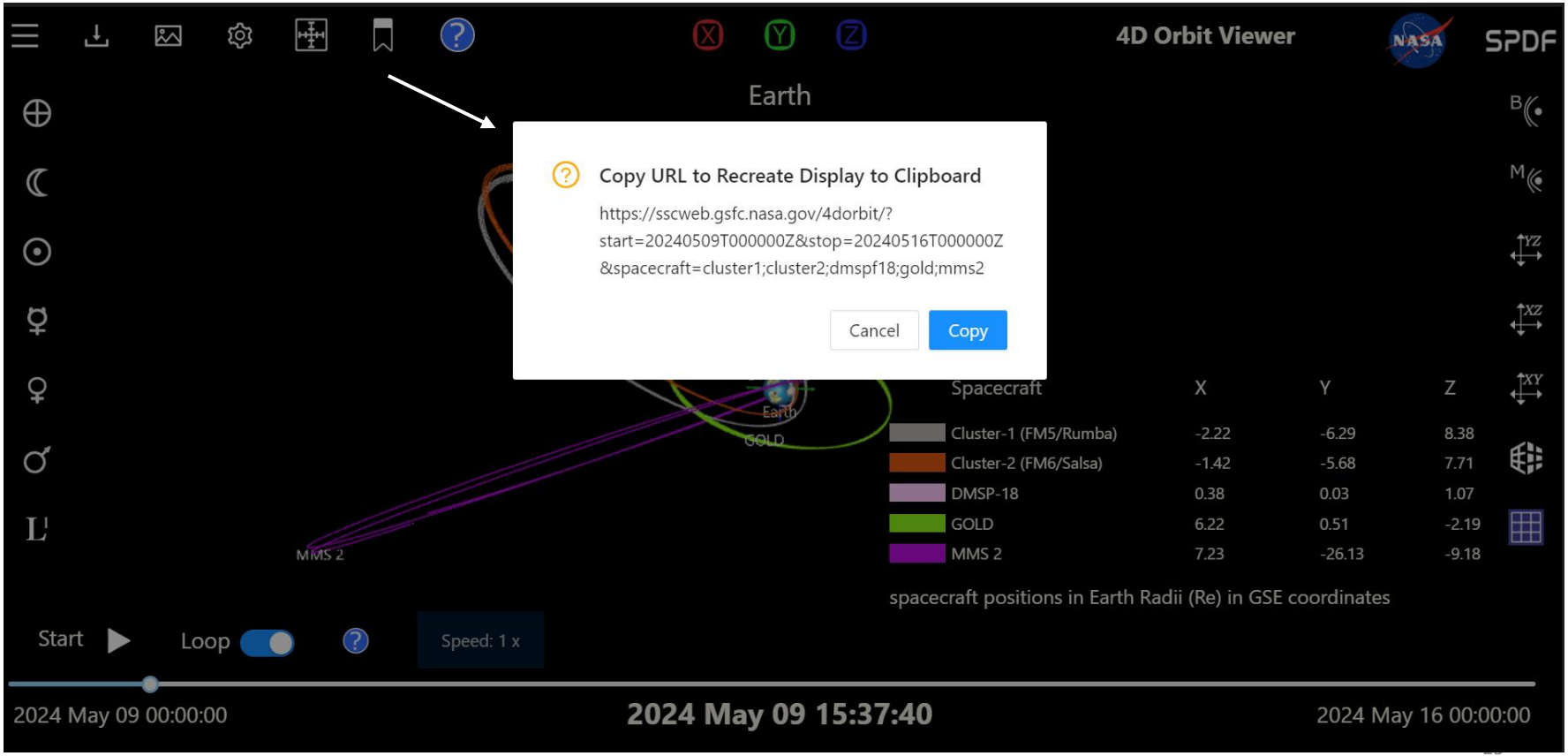
4-D Orbit Viewer: Bow Shock and Magnetopause



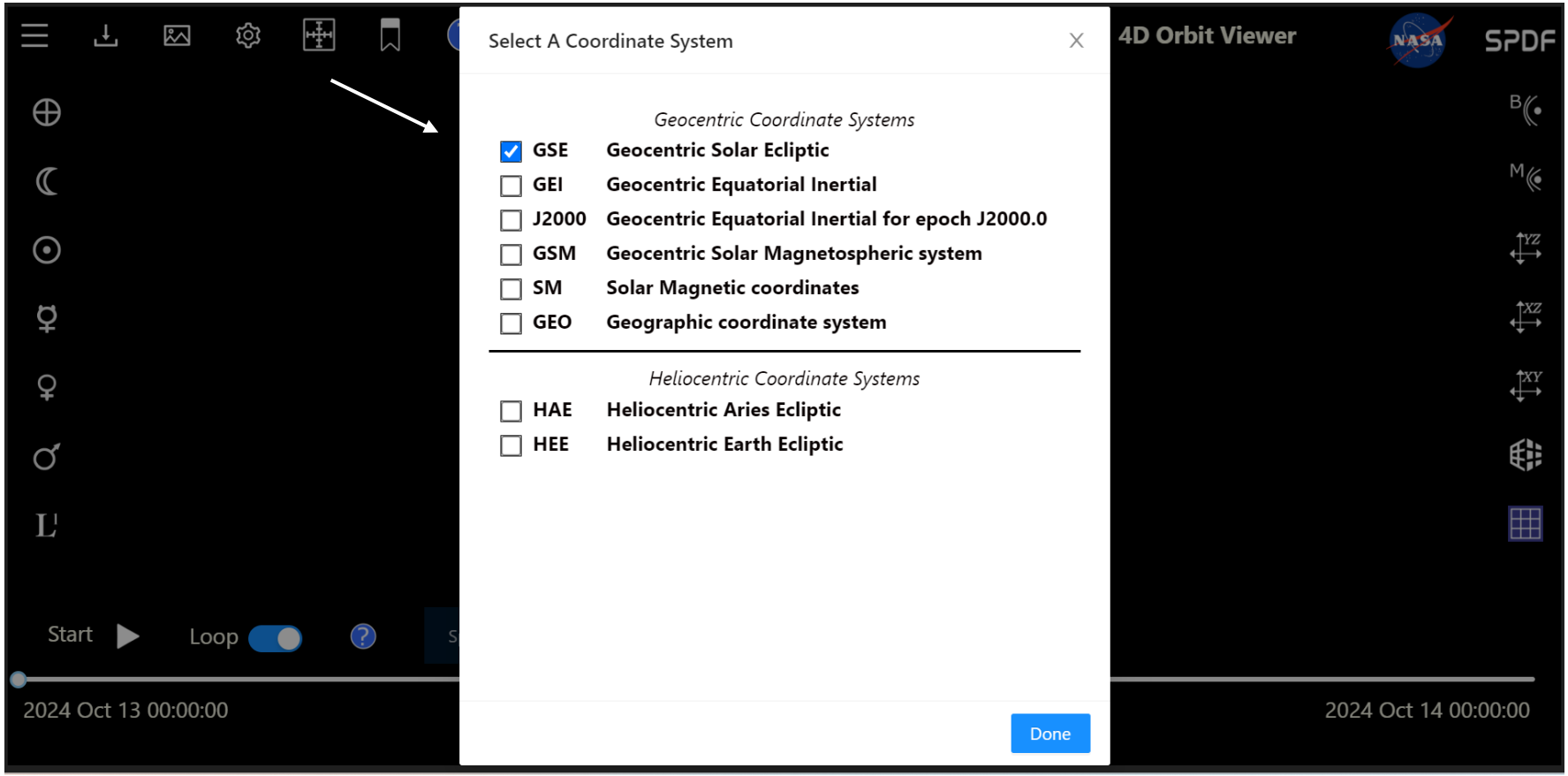
4-D Orbit Viewer: Rotation of Coordinates



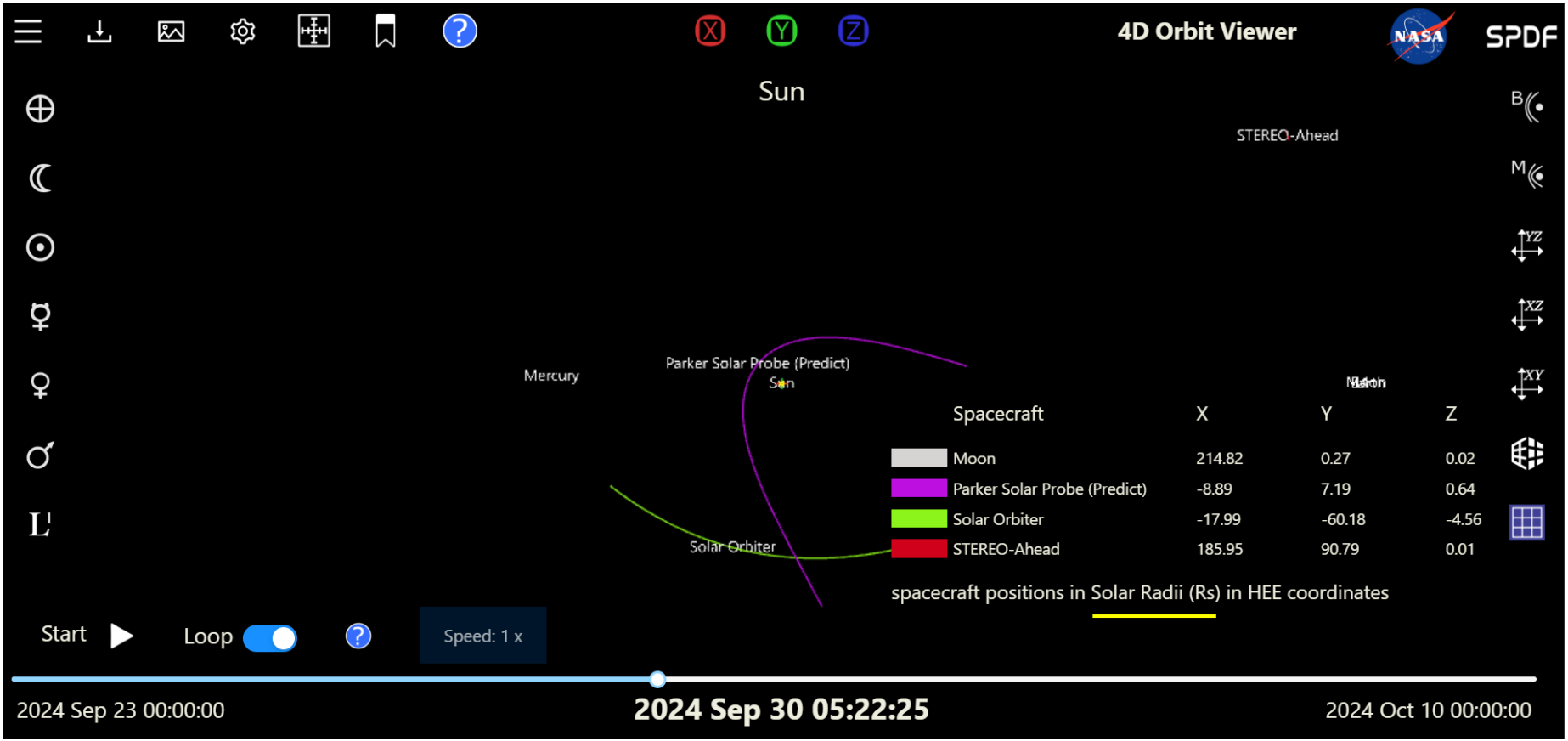
4-D Orbit Viewer: Bookmark the URL

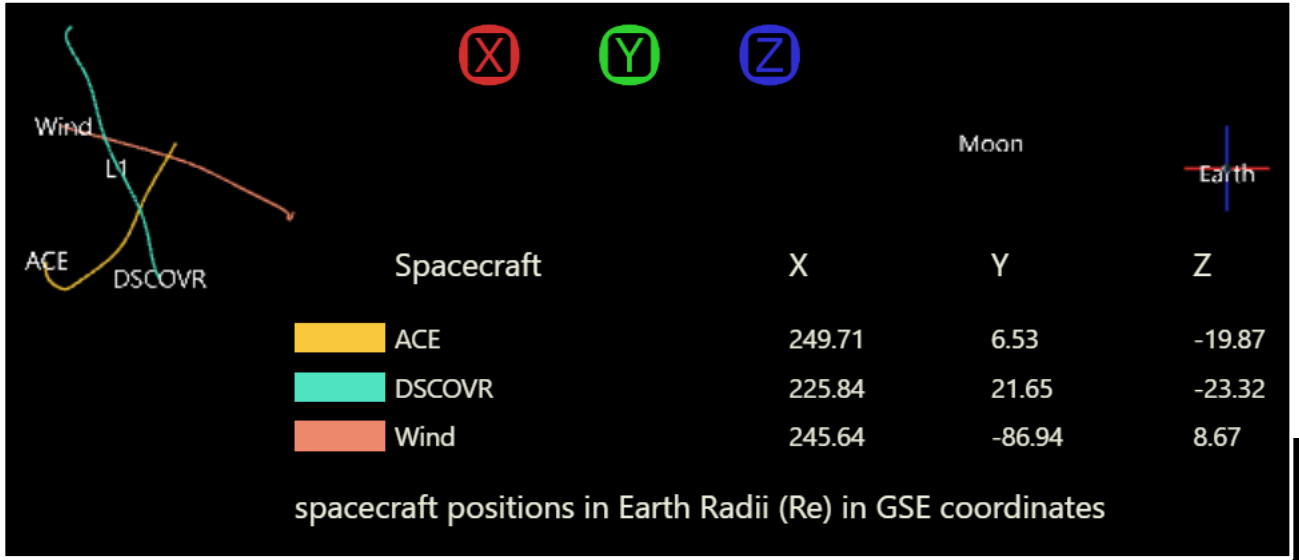


4-D Orbit Viewer: Different Coordinates



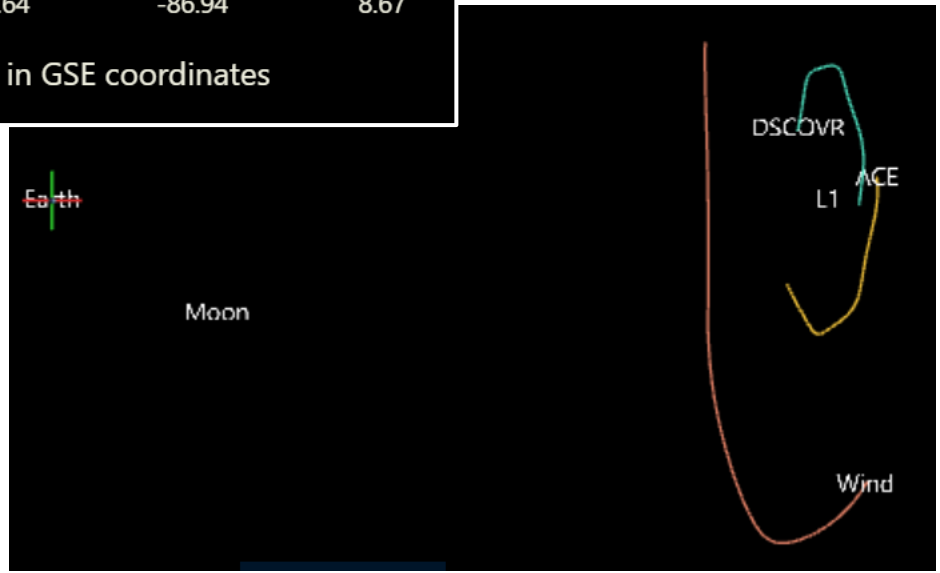
4-D Orbit Viewer: Heliocentric Coordinates





Centered at L1
Along Y axis

Centered at L1
Along Z axis



SPDF Plans

- Data archiving and service for the **upcoming new missions** (EZIE, IMAP, TRACERS, Carruthers)
- Higher-resolution orbit data for SSC
- Improving SPDF documentation and the archival of mission documents
- Standardizing ISTP Metadata Guidelines with version control, etc.
- Releasing new SKTeditor in JavaScript, including defining SPASE metadata at the same time as defining the internal metadata and structure of the CDF/NetCDF dataset
- Adding more planetary coordinates and simulated physical boundaries (e.g., magnetopause) to the new 4D Orbit Viewer
- Releasing new SSC and CDF websites following USWDS, with quick start guides and tutorials
- Designing new OMNI Web and CDAWeb sites
- Making web services for event lists of burst mode data or science events, using CDAWeb and SSCWeb to find previous/next burst or event and better serve them
- HTML5/JavaScript-based browser interface for CDAWeb/SSCWeb, building on the 4D orbit viewer, expanded to add interactive data plotting and sonification tied to the orbit display, perhaps with data glyphs along the orbits as well (using JSON output from CDAWeb and SSC web services)

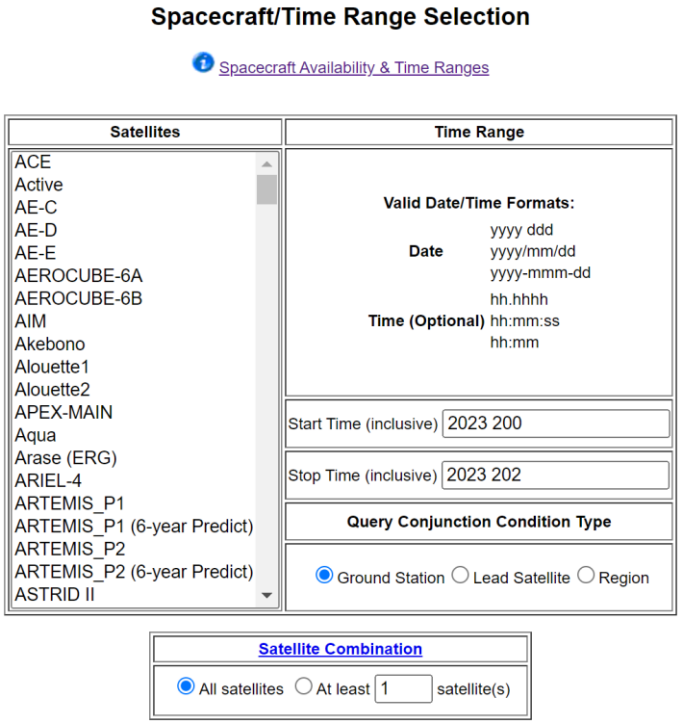
Backup

Alternative Data Access Methods

https://cdaweb.gsfc.nasa.gov/alternative_access_methods.html

- CDAWeb Browser interface for plots,listings and CDFs (both original and creation of sub/super-set CDFs)
- HTTPS access to original CDFs
  - Master CDFS (binary)
  - SkeletonTables (ascii)
- (FTPS required) FTPS access to original data CDFs
- REST Web Services interface and example use.
- SOAP Web Services interface and example use.
- Python CDAS web services package and ai.cdas package.
- Scripts to keep up-to-date on data files for particular datasets:
  - Example Perl code
  - Periodically download the CDAWeb HTTPS listing file and compare with a previous file (then download the differences). See Bash shell script below.
- IDL software - reads CDFs and selected Netcdfs, plots, lists and creates subset/superset CDFs (the software that underlies CDAWeb).
  - CDAWlib readme and example code.
  - Direct link to CDAWlib software.
- Extremely easy to use IDL software that provides direct access to CDAWeb held data files.
- CREADER : Load data from CDAWeb directly into IDL 7.1+ with specified variables, and optional spike editing and binning and interpolation that produce uniform time series at a chosen time resolution.
- SPEDAS : Space Physics Environment Data Analysis Software - IDL-based plotting, analysis, and data downloading tools.
- pySPEDAS : Python implementation of many SPEDAS plotting, analysis, and downloading tools.
- Autoplot.org **(non-NASA)** interface to public CDAWeb database
- HAPI streaming protocol using HAPI IDL, Python, and other clients from CDAWeb HAPI.

Spacecraft Situation Center (SSC) Conjunction Query



Example Using THEMIS Mission

https://sscweb.gsfc.nasa.gov/examples/THEMIS_queries/

themis_conjunction2 Magnetic conjunction of THEMIS-5 (lead satellite) with at least 3 other THEMIS

themis_fast2 Magnetic conjunction of at least 2 THEMIS satellites with FAST (lead satellite)

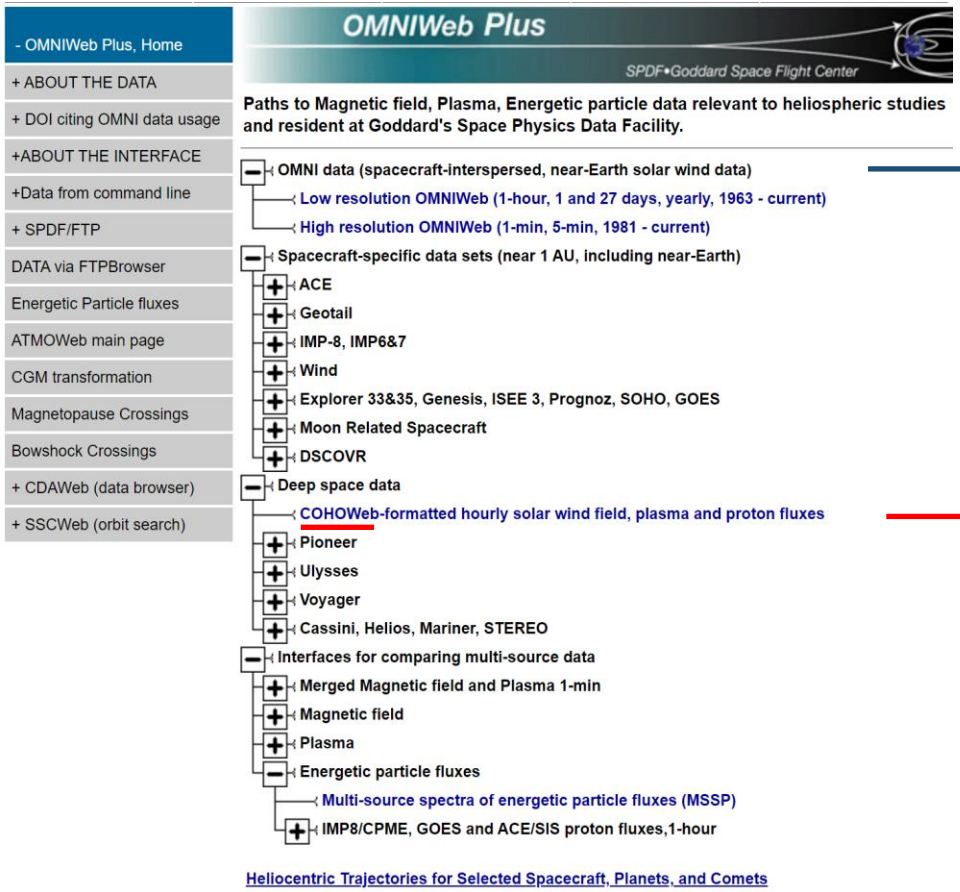
themis_goes11a Magnetic conjunction of at least two satellites (THEMIS 1-5, GOES 13) with GOES 11

themis_goes13a Magnetic conjunction of at least two satellites (THEMIS 1-5, GOES 11) with GOES 13

themis_ground_stations1 Magnetic conjunction of at least 2 THEMIS satellites with one of 4 THEMIS ground stations during 2008 doy=1-5

OMNI Web

https://omniweb.gsfc.nasa.gov/



- OMNI Data: Database of solar wind magnetic field and plasma parameters mapped to the nose of the Earth's bow shock
- Based on a large volume of quality-controlled satellite measurements (since Nov. 1963)
- **COHOWeb**: Solar wind field, plasma, and energetic particle fluxes in other locations of the heliosphere, especially useful for planetary studies and heliospheric model validation
- Interface for plotting, filtering, and downloading the data