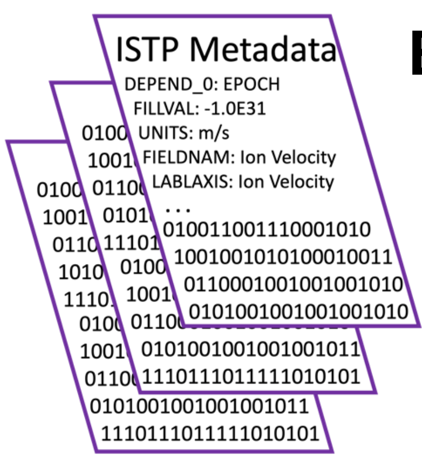
AGU Fall 2024
IN13B-2165

NASA Space Physics Data Facility (SPDF) Data Archive Services as infrastructure for Open Data Access, Correlative and Collaborative Research, and Common Standards

https://spdf.gsfc.nasa.gov



R Candey, S Boardsen, C Butterfield, A Cruz, S Fooks, L Garcia, E Grimes, B Harris, T Helvey-Kasulke, L Jian, R Johnson, A Koval, T Kovalick, H Leckner, M Liu, S Lyatsky, P Makela, K Marshall, N Papitashvili, J Smith, J Sun, R Yurow

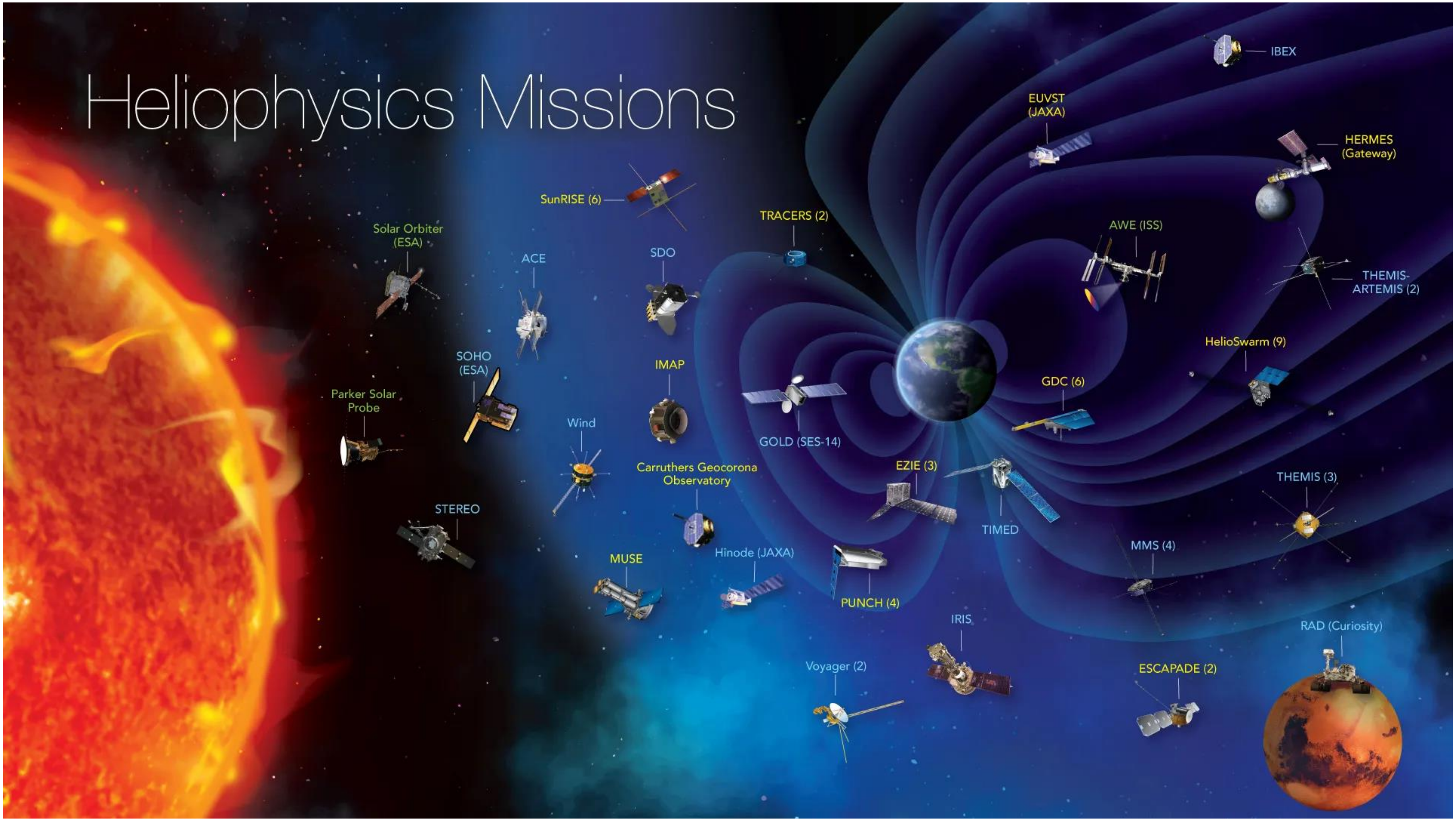
and the rest of SPDF Team

NASA Goddard Space Flight Center, Greenbelt, MD, USA

2024 Dec. 9

Presentations at https://spdf.gsfc.nasa.gov/pub/documents/SPDF/presentations/

Heliophysics System Observatory



HDRL Organizational Chart

HP Data and Model Consortium / HDMC

Brian Thomas (PS),
Tressa Helvey-Kasulke (PM)

Overall management of the HDRL.

Registries, Metadata and DOIs for all digital resources; SPASE Data Model.
Heliophysics Data Portal (HDP)
Python and other software integration (PyHC).
Analysis and visualization services ((Py)SPEDAS, Autoplot).
Data upgrades and related services.
HelioCloud initiative with data and software from all groups.

All activities within the various HDRL components are interrelated.

Solar Data Analysis Center / SDAC

Jack Ireland (PS)

Final Active Archive of solar data for Solar Dynamics Observatory (SDO) and other NASA missions.
Virtual Solar Observatory (VSO) data access.
Helioviewer. SolarSoft. SunPy.
High Performance Computing for NASA HP.

Space Physics Data Facility / SPDF

Robert Candey (PS), Lan Jian (DPS)

Final Active Archive of non-solar data for NASA HP (and other related) missions.
Coordinated Data Analysis Web (CDAWeb) data browsing & access; Web Service access.
OMNIWeb data production and service.
SSCWeb and 4-D spacecraft orbit viewer.
Common Data Format (CDF). ISTP metadata standards.

Key Collaborators

Center for HelioAnalytics

AI/ML and Adv Analytics Development, Mission-Enabling Tech, Community Resources, User Testing.

Community Coordinated Modeling Center / CCMC

Data-model comparisons; Registry of models and output; “Kamodo” enabled visualization.

The NASA Heliophysics Digital Resources Library (HDRL.gsfc.nasa.gov) coordinates the efforts of the NASA Heliophysics archives (including SPDF) and other data-related groups to increase discoverability and usability of data and model results, software and services.

Where the System Observatory Comes Together

- The HDRL enables the scientific analysis goals of the Heliophysics System Observatory:
  - Metadata and other data-related standard
  - Provisioning and curation of scientific big data from many sources
  - Support for data analysis in multiple computational environments
  - Design and implementation of a collaborative open science infrastructure
  - Improving access to the data of the Heliophysics System Observatory (HSO) and NASA-funded research projects
  - Improving cross-mission and observation-model comparison, machine learning and other large-scale and collaborative analysis
  - Increasing discoverability and usability of data and model results, software and services, with more complete metadata and provenance and quality control

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 and 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 and netCDF data including global and variable attributes
  - **Heliophysics Data Portal**: discipline-wide data inventory and access service (heliophysicsdata.gsfc.nasa.gov/)

SPDF Provides Multiple Services and Access Methods

- Direct file downloads via FTPS and HTTPS https://spdf.gsfc.nasa.gov/pub/data/
- Orbit and ground track displays/queries via SSCWeb and 4D Orbit Viewer
- CDAWeb services:
  - Data files, plots and listings
  - Web service interfaces (REST, SOAP, IDL, Matlab, Java, Python) https://cdaweb.gsfc.nasa.gov/WebServices/
  - HAPI (Heliophysics API) https://cdaweb.gsfc.nasa.gov/hapi
  - Autoplot autoplot.org/help#CDAWeb
  - Other methods such as IDL https://cdaweb.gsfc.nasa.gov/alternative_access_methods.html
- Calibrated data copied to HelioCloud.org in AWS for cloud-based analysis
- The **SPASE** (Space Physics Archive Search and Extract http://www.spase-group.org/) team use the master CDFs to generate SPASE IDs and descriptions for all datasets, to add entries to the **Heliophysics Data Portal** https://heliophysicsdata.gsfc.nasa.gov and mint DOIs for each dataset

Linking SPDF Services with Missions

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

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.

160+ missions

GO =Go to Service's Home Page — i =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

jupyter CdasWsGenericExample (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code | Download | GitHub | Binder | **Memory:** 239.3 MB / 4 GB

Jupyter notebook calling CDAWeb web services

Get Dataset Information

The following code demontrates how to get information about a dataset.

In [3]:

```
#print(dataset)
datasets = cdas.get_datasets(idPattern=dataset)
ds_info = datasets[0]
#print(ds_info)
print(ds_info['Id'], ':', ds_info['Label'])
ds_time_interval = ds_info['TimeInterval']
print('Time range:', ds_time_interval['Start'], 'to',
      ds_time_interval['End'])
print('Princple Investigator:', ds_info['PiName'], ':',
      ds_info['PiAffiliation'])
print('Notes:', ds_info['Notes'])
if 'Doi' in ds_info and ds_info['Doi']:
    print('DOI:', ds_info['Doi'])
    print('DOI landing page:',
          cdas.get_doi_landing_page_url(ds_info['Doi']))
```

```
AEROCUBE-6-B_DOSIMETER_L2 : Aerocube 6/Dosimeter Level 2 - J. B. Blake (The Aerospace Corporation)
Time range: 2014-06-21T14:49:56.000Z to 2017-06-30T15:24:08.000Z
Princple Investigator: J. B. Blake : The Aerospace Corporation
Notes: https://cdaweb.gsfc.nasa.gov/misc/NotesA.html#AEROCUBE-6-B_DOSIMETER_L2
DOI: 10.48322/49dd-na02
DOI landing page: https://doi.org/10.48322/49dd-na02
```

Get An Example Time Interval

The following code gets a small example time interval.

In [4]:

```
example_interval = cdas.get_example_time_interval(dataset)
print('Example time interval:', example_interval)
```

```
Example time interval: 2017-06-30T13:24:08+00:00 2017-06-30T15:24:08+00:00
```

Get Dataset Variable Names

The following code demonstrates how to get a dataset's variable names.

In [5]:

```
var_names = cdas.get_variable_names(dataset)
print('Variable names:', var_names)
```

```
Variable names: ['alt', 'lat', 'lon', 'XYZ_GEO', 'dos1l', 'dos1m', 'dos1rate', 'dos2l', 'dos2m', 'dos2rate', 'dos3
l', 'dos3m', 'dos3rate', 'flag', 'Sample_Rate', 'Lm_IGRF', 'Bmag_IGRF', 'MLT_IGRF', 'InvLat_IGRF', 'Lm_OPQ', 'Bmag_
```

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 A
should edit the code to suit your needs.

```
#  Install these prerequisites once before executing the exa
#  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
environment
# (spacepy or cdflib) and needs.
vars =
['alt','lat','lon','XYZ_GEO','dos1l','dos1m','dos1rate','dos
```

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

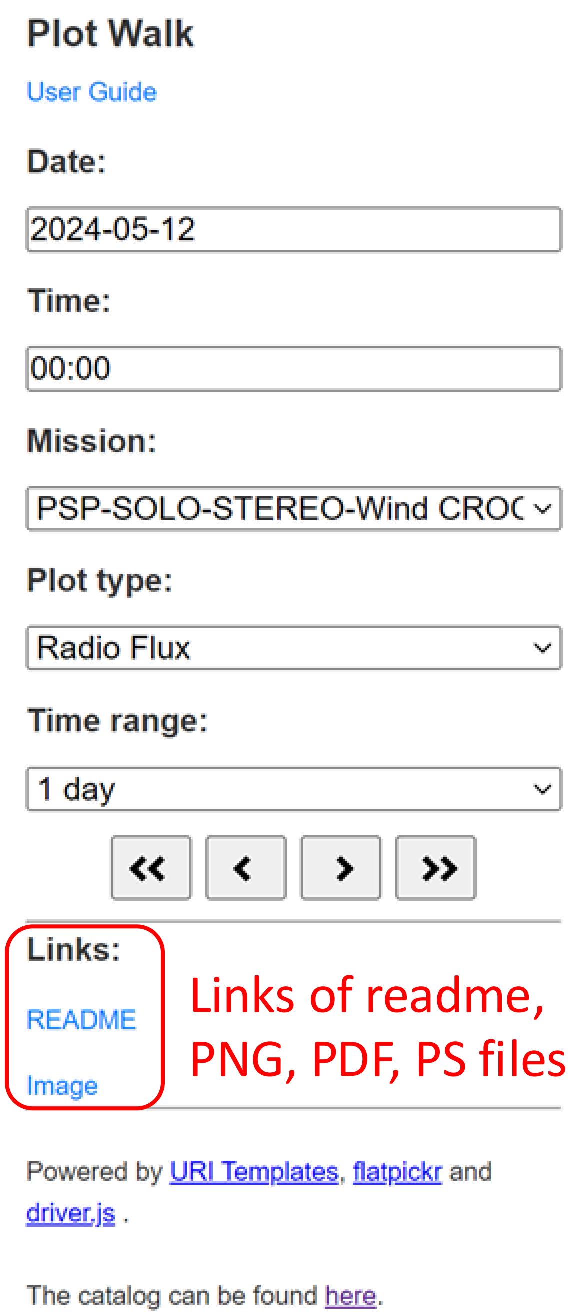
Alternative data access methods

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

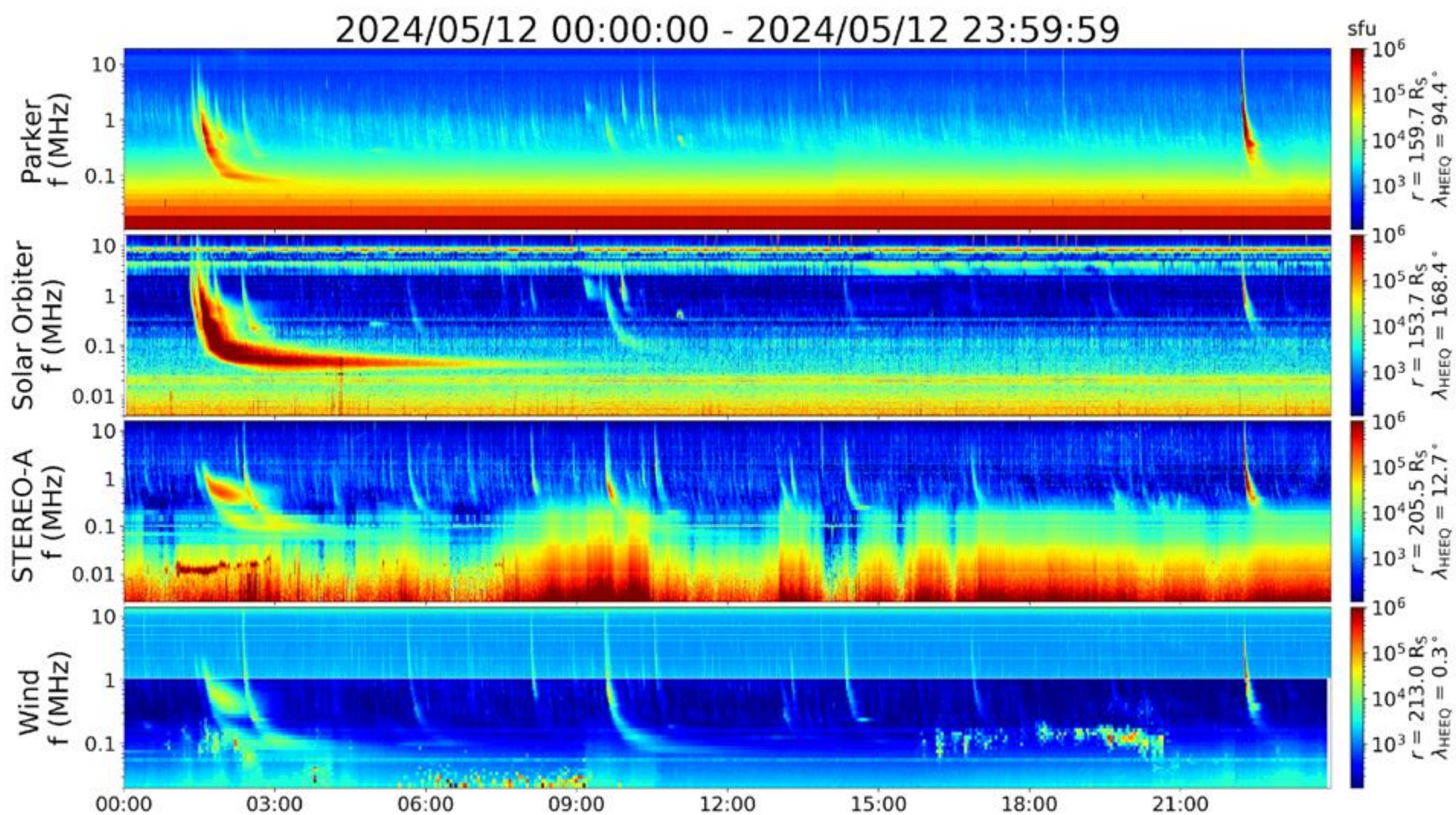
Plot Walk for Pre-Generated Plots

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

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



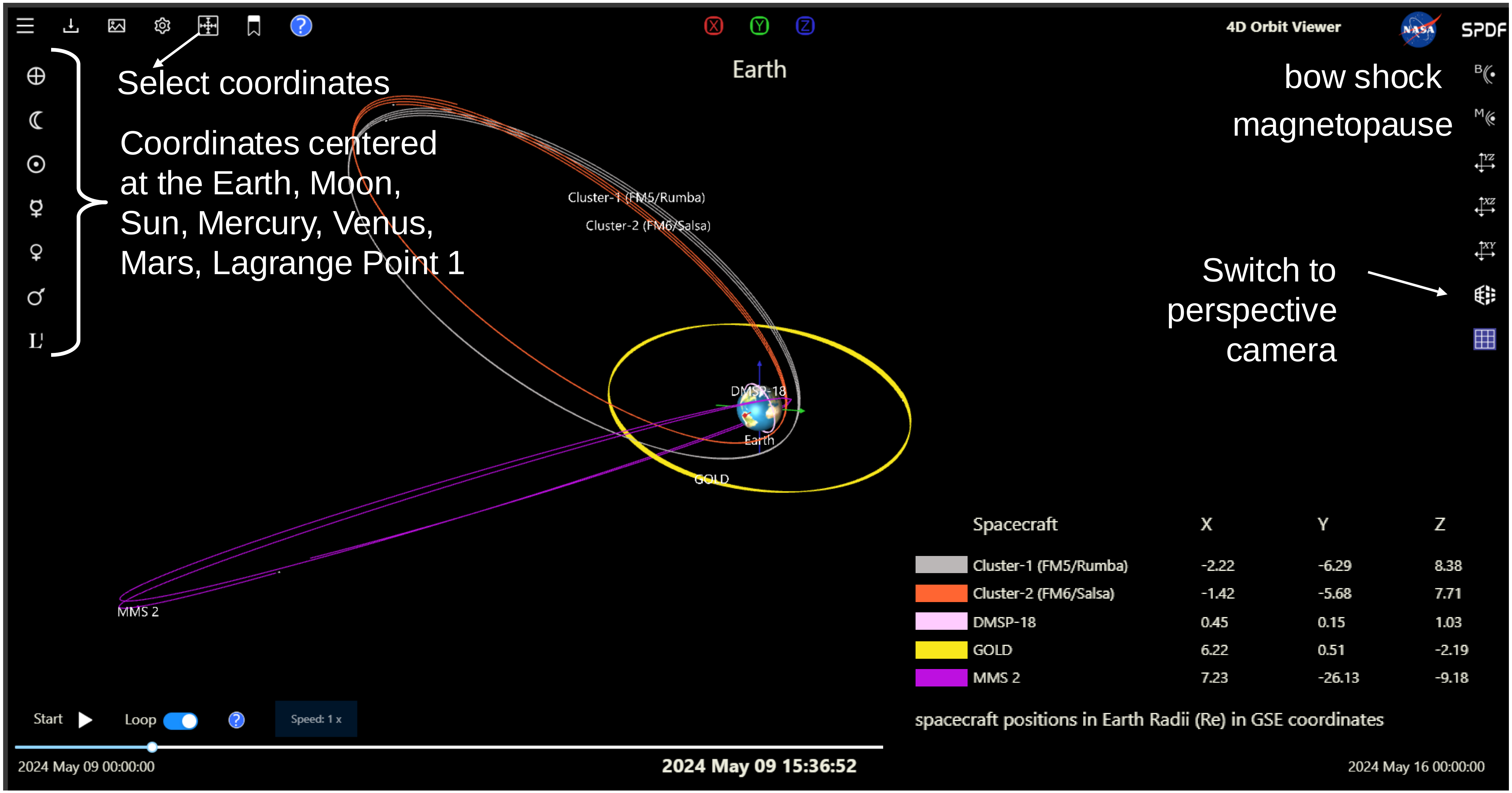
Links of readme, PNG, PDF, PS files



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

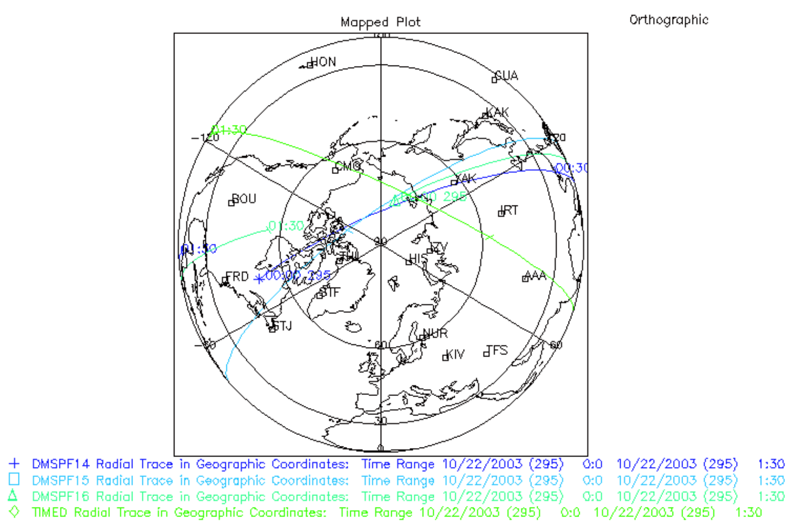
4-D Orbit Viewer (160+ Spacecraft)

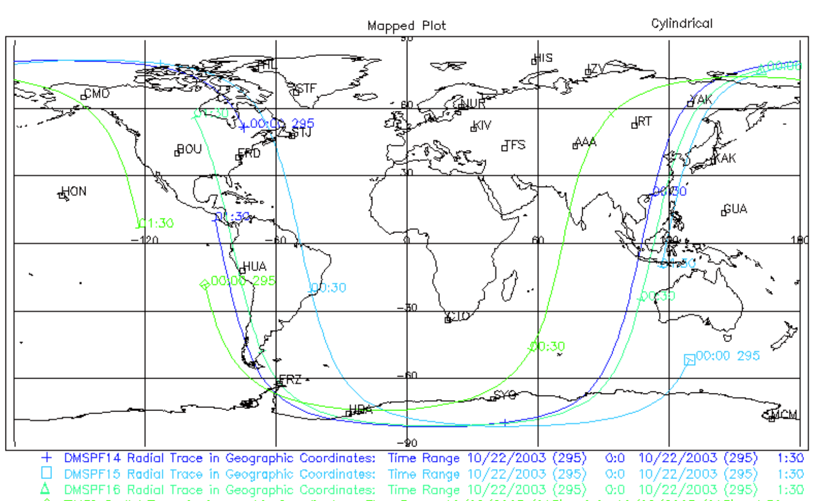
https://sscweb.gsfc.nasa.gov/4dorbit/



SSCWeb
Multi-satellite or satellite to ground station magnetic conjunctions

Orbit plots for TIMED and DMSP 14, 15, 16 in a Polar or Cylindrical Projection. Ground station 3-letter codes included.





Listing of times when the magnetic footpoints of TIMED, DMSP 13, 14, 15 or 16 crossed the Arecibo ground station.

Time yyyy	ddd	hh.hhhhh	Sat.	GEO Lat	GEO Long	Radius (km)	Trace GEO Lat	Trace GEO Long	ArcLen (km)	Ground Stations:
2003	292	0.01667	dmspf16	-32.78	304.10	7230	19.60	294.66	8199	Arecibo
2003	292	0.03333		-29.31	303.11	7230	17.01	294.61	6898	Arecibo
2003	292	0.23333	dmspf16	12.51	293.05	7229	19.05	291.77	1097	Arecibo
2003	292	0.31667	dmspf14	-32.87	297.28	7221	17.44	290.81	7650	Arecibo
2003	292	1.23333	dmspf15	-28.26	306.59	7219	17.71	296.86	6875	Arecibo
2003	292	1.41667	dmspf15	10.19	297.54	7219	17.57	295.85	1173	Arecibo
2003	292	9.41667	timed	-35.35	296.89	7004	17.40	290.50	8028	Arecibo
2003	292	9.63333	dmspf13	-28.67	306.75	7233	18.11	296.85	7045	Arecibo
2003	292	11.13333	dmspf13	12.61	291.12	7228	19.03	289.98	1084	Arecibo
2003	292	11.36667	dmspf16	-31.39	302.55	7229	18.10	293.97	7538	Arecibo
2003	292	11.66667	dmspf14	-33.90	295.55	7227	17.79	289.73	7937	Arecibo
2003	292	12.53333	dmspf15	-29.56	307.38	7221	18.82	297.08	7377	Arecibo
2003	292	14.03333	dmspf15	11.53	291.65	7221	18.18	290.46	1101	Arecibo
2003	292	22.28333	dmspf13	11.54	300.06	7217	18.76	298.24	1160	Arecibo
2003	292	23.81667	dmspf16	-29.87	306.48	7231	18.70	296.49	7409	Arecibo

Listing of times of magnetic conjunction between TIMED and Doublestar 1, or DMSP 15 or 16.

Time yyyy	ddd	hh.hhhhh	Satellite	GEO Lat	GEO Long	Radius (km)	Trace GEO Lat	Trace GEO Long	ArcLen (km)	Lead Sat. Dist.	Lead Sat. Name
2004	15	10.58333	polar	13.96	76.64	38152	69.34	73.69	41635	34567	timed
2004	15	10.58333	timed	68.49	76.85	6992	69.13	77.44	539		
2004	16	0.53333	doublestar1	-28.18	165.25	38715	-58.65	151.85	34985	32959	timed
2004	16	0.53333	timed	-58.42	151.78	7003	-58.85	151.19	544		
2004	16	6.08333	dmspf15	74.37	195.14	7205	75.16	196.12	754	239	timed
2004	16	6.08333	timed	73.96	192.34	6992	74.57	192.98	539		
2004	16	22.25000	dmspf16	74.43	291.52	7215	74.79	290.13	759	239	timed
2004	16	22.25000	timed	73.81	290.53	6992	74.07	289.54	535		

Standards and Conventions

- **SPASE** *http://www.spase-group.org* dataset descriptions for easy searching
- **Heliophysics Data Portal** *https://heliophysicsdata.sci.gsfc.nasa.gov*
- **ISTP/IACG/SPDF Guidelines** for global and variable attributes *https://spdf.gsfc.nasa.gov/sp_use_of_cdf.html*
  - ISTP metadata Editor *https://spdf.gsfc.nasa.gov/skteditor*
  - Defining additional standard attributes: Cluster, THEMIS, RBSP (PRBEM), MMS, etc.
- **Dataset naming and file naming** recommendations *https://spdf.gsfc.nasa.gov/guidelines/filenaming_recommendations.html*

  and file naming templates:

  https://github.com/hapi-server/uri-templates/wiki/Specification **$Y/data_$Y_j_idx.cdf**
- **CDF** *https://cdf.gsfc.nasa.gov* scientific data format (including pure Python library *https://github.com/MAVENSDC/cdflib*)
  - Time variable types *https://cdf.gsfc.nasa.gov/html/leapseconds_requirements.html*
- **netCDF** *https://www.unidata.ucar.edu/software/netcdf/*
- **FITS** *https://fits.gsfc.nasa.gov/*
- UDunits www.unidata.ucar.edu/software/udunits/
- Tools enabled by standards: CDAWeb and CDAWlib IDL/Python library, Autoplot http://autoplot.org, SPEDAS *http://spedas.org* IDL/Python library

Why Metadata Conventions

- Leverage standardized self-describing data formats, metadata for datasets and parameters, time conventions, and dataset and filenaming conventions to enable effective data analysis and browsing using generic easy-to-use software and web services
- Restricting metadata representations limits the number of equivalent possibilities with which software must deal, and thus fosters **interoperability**
- Conventions standardize ways to name things, represent relationships, and locate data in space and time
- Enables developing applications with powerful extraction, regridding, analysis, visualization, and processing capabilities
- Abstracts general data models to represent data semantics
- Embody data provider’s knowledge and capture the meaning in data and make data semantics accessible to humans as well as programs
- Provide higher-level abstractions such as coordinate systems, standard names for physical quantities for comparing different data and distinguishing variables

ISTP Guidelines Structure and Metadata Concepts

- ISTP/IACG Guidelines (mid 1990s) and subsequent extensions by SPDF define implementation standards for CDFs and NetCDFs
  - Include general file naming conventions
  - Data is time-ordered and time-identified; times vary by record
  - Set of required and suggested metadata (details on next slide)
  - Variable attributes can point to other variables by name and carry arguments
    - Attributes thus carry information about relationships among variables
    - Variables can carry metadata (e.g., labels for dimensional variables)
  - Global attributes provide overall context of the dataset
  - Missions add their own metadata requirements

- **CDAWeb additional concepts: "Master" CDFs and "Virtual" Variables**
  - "Master" CDF is the use of a "skeleton" CDF (structure and metadata but no data) to insert supplemental or updated metadata for CDFs as a dataset
  - "Virtual" variables are computed variables, using specialized CDF attributes to link defined variables and routines within CDAWeb/CDAWlib

History of ISTP Metadata Guidelines

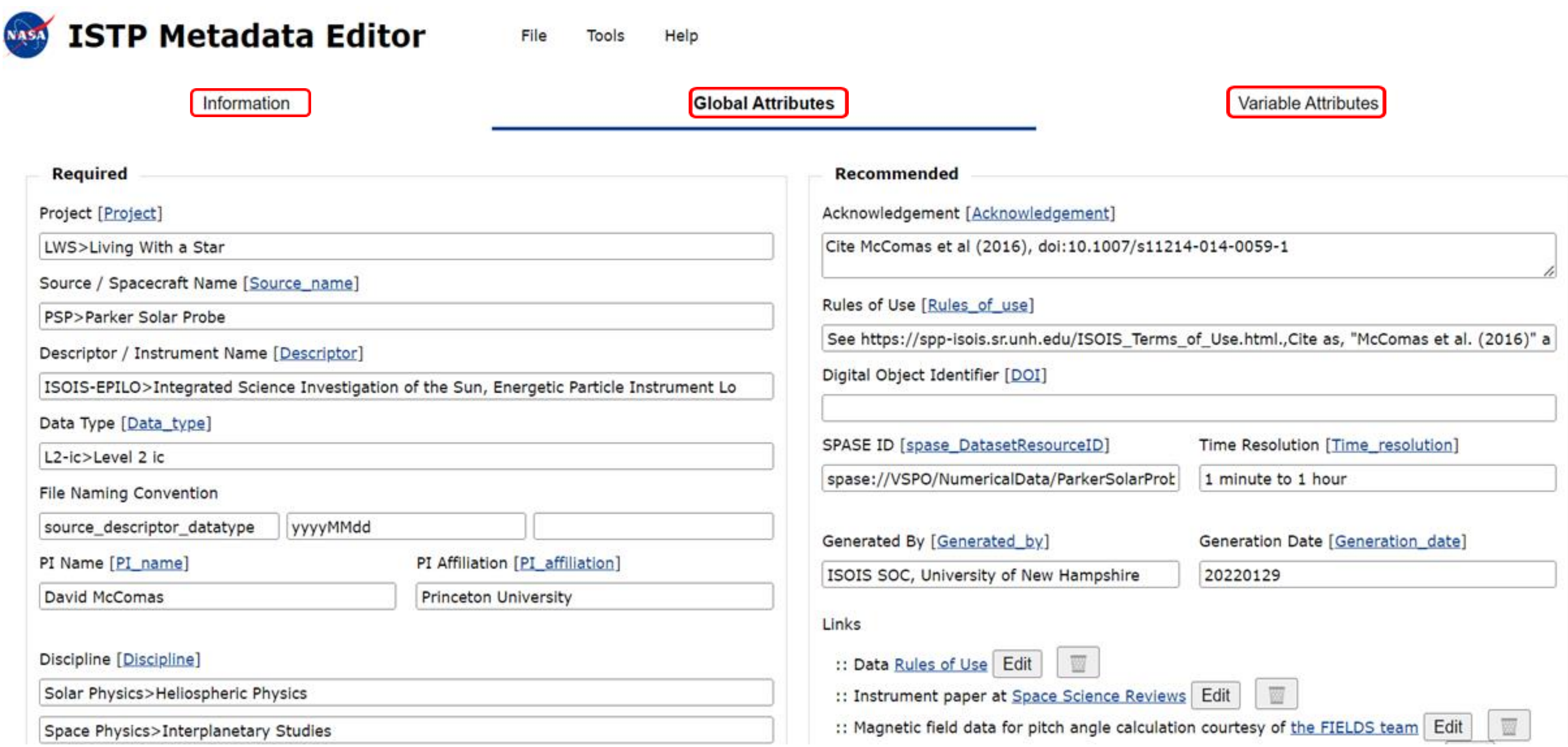
- International Solar-Terrestrial Physics (ISTP) for coordinated, simultaneous investigations of the Sun-Earth space environment in 1990s
  - NASA
  - European Space Agency (ESA)
  - Institute of Space and Astronautical Science (ISAS) of Japan
  - Academy of Sciences (Russia)/Rosaviakosmos
- Science community developed the ISTP Metadata Guidelines and other conventions for describing and naming the datasets
- Later adopted by the Interagency Consultative Group (IACG)
- Missions add more attributes: Cluster, THEMIS, RBSP (PRBEM), MMS, etc.
- Widely-used data display and analysis tools depend on the guidelines
- Used to populate SPASE metadata

ISTP Metadata Tools and Usage

- Current ISTP/IACG/SPDF Guidelines
  https://spdf.gsfc.nasa.gov/sp_use_of_cdf.html
- Move updates to Github in markdown
  https://github.com/IHDE-Alliance/ISTP_metadata/tree/main/v1.0.0
- ISTP Metadata Editor https://spdf.gsfc.nasa.gov/skteditor
  originally in Java and now in Javascript by Eric Grimes
- **PDS added CDF-A as a standard format**, which is CDF with ISTP Guidelines and two SPASE attributes, but no compression or sparse variables
- Not just in CDFs, but used in netCDFs such as for GOES, ICON, GOLD
  and could be used in other self-describing formats
- Added SPASE and DOI global attributes to CDAWeb datasets via Master CDFs and shown in CDAWeb interface

Development of ISTP Metadata Editor

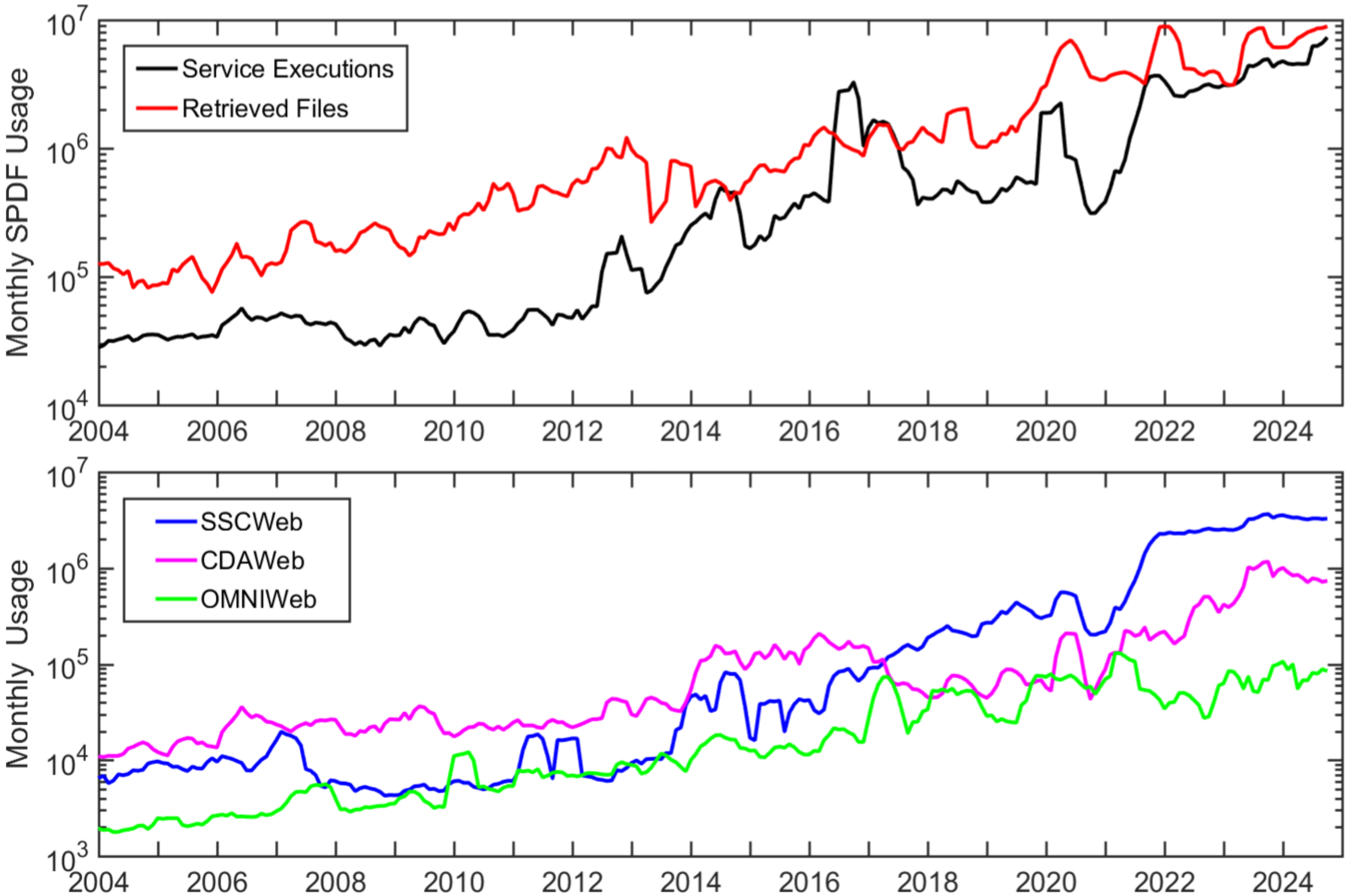
- New JavaScript based web-browser tool to help users create/update CDF datasets with ISTP and SPASE (Space Physics Archive Research and Extract) metadata
- Includes validator function, that can also validate against mission-specific requirements



ISTP Metadata Development

- Formalize Steering Committee
- Define governance and update processes
- Review version on Github, add changes identified in the past few years, and general dataset creation recommendations and lessons-learned
- Create Github issues for various additions, such as author list for DOIs, DOI, Variable_display_order, Variable_display_indent_level, Associated_parent_variable, Dataset_group, Mission_parent) and from Cluster/Solar Orbiter: Representation, Tensor_order, Coordinate_systems, Rotation_matrices, Unit_quarternion
- Consider requirements specific to model results
- Add explanatory material from the CF Conventions (https://cfconventions.org/) that also apply in heliophysics
- Add content from **mission-specific documents** that reference the ISTP guidelines
- Add crosswalk with SPASE metadata
- Better document Guidelines on Github with mission-specific metadata as well, but want to keep flexible for interactions with missions and enabling framework for CDAWeb services

Summary



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

- SPDF archives and serves observational space physics data to promote correlative and collaborative research across discipline and mission boundaries
- SPDF provides three main science-enabling services: CDAWeb, SSCweb, and OMNIweb
- CDAWeb database (quality-controlled, ~3000 datasets) is also widely used outside of SPDF (e.g., Autoplot, HelioCloud, AMDA)
- Steering committee is forming to improve ISTP metadata to add more attributes and crosswalk with SPASE metadata
- The mission never ends with SPDF

Backup

Abstract

- NASA's Space Physics Data Facility (SPDF https://spdf.gsfc.nasa.gov) is the non-solar archive for data from NASA missions and other data relevant to NASA heliophysics science objectives, from the mesosphere into the furthest reach of deep-space exploration, including from Parker Solar Probe, Solar Orbiter, ICON, MMS, Van Allen Probes, THEMIS/ARTEMIS, GOLD, ACE, Cluster, IBEX, Voyager, Geotail, Wind and many others. SPDF serves the global science community with multi-project, cross-disciplinary access to space physics data to enable correlative and collaborative research across discipline and mission boundaries with more than 100 operating and past missions/projects. SPDF provides three main science-enabling services: Coordinated Data Analysis Web (CDAWeb), Satellite Situation Center Web (SSCWeb), and OMNI Web. SPDF also supports standards such as the Common Data Format (CDF) self-describing science format and the International Solar-Terrestrial Physics (ISTP) metadata standards. Highlights include recent updates to SPDF webpages, new browser-based 4-D orbit viewer, improved pre-generated plot browsing system, copying the archives into the HelioCloud.org for cloud-based collaborative data analysis, and describing SPDF datasets with Space Physics Archive Search and Extract (SPASE) records as well as Digital Object Identifiers (DOIs). SPDF supports the Heliophysics Data Portal (HDP) discipline-wide data inventory based on the SPASE-group.org metadata. These updates facilitate the production, access, analysis, and citation of space physics data and other digit resources, thus ultimately support Open Science. The NASA Heliophysics Digital Resources Library (HDRL.gsfc.nasa.gov) coordinates the efforts of the NASA Heliophysics archives (including SPDF) and other data-related groups to increase discoverability and usability of data and model results, software and services.

Dataset creation: Understand the Data to be Loaded

- What are the key data quantities
  - What is their definition/meaning?
  - How are they going to be named?
    - N.B. MMS parameter naming convention: scId_instrumentID_paramName
      - »
- Understand (at the dataset level)
  - Dimensionality and dependencies
  - Variance with time and dimension
    - ISTP conventions allow >1 time variable in a file
    - Carry slowly-varying data as variables rather than in attributes
- General rule is to capture relationships in the structure
  - Otherwise capture relationships in variable attributes
  - Want relationships to be logically-structured and machine-readable
    - Available for more general-purpose codes to exploit
- Let CDF/NetCDF deal with mechanics of efficient data storage
  - Once more: lay out data by what's science logical and useful
    - E.g. methods to handle slowly-varying data include setting "sparse=sRecords.PREV" in CDFs

ISTP Metadata Elements

- **Variable attributes required for automated processing:**
  - Catdesc for longer variable description
  - Depend_0 points to time variables
  - Depend_1, 2, 3 point to variables that describe other dimensions
  - Fieldnam short variable name for plots
  - Fillval values indicating missing or bad data
  - Lablaxis/Labl_ptr for axis and column titles
  - Units/Unit_ptr
  - Validmin/max for valid data range
- **ISTP metadata independent of CDF and easily used in other self-describing science formats like CEFs, netCDFs and HDFs, and probably FITS and ASDF**
- **CDF Time variable types**
  - **CDF_TIME_TT2000** nanoseconds from J2000 in Terrestrial Time in 8 byte integer handles leap seconds and is well-defined; UTC conversion requires up-to-date leap second table (last value stored in CDF header as a check)
  - EPOCH milliseconds from 0AD in 8-byte float; usually UTC but not leap seconds
  - EPOCH16 picoseconds from 0AD in two 8-byte float; usually UTC but not leap seconds

Tool to Create/Edit a CDF/NetCDF File Compliant to ISTP Standard

- SKTeditor is a Java, web-start application, soon to be in JavaScript
  - Guide designers to good choices consistent with ISTP guidelines
  - Create new CDF/NetCDF or check/correct then modify existing skeleton file
- Guided by the interface flow, add or edit
  - Scalar and higher-dimensional variables, multiple time variables
    - Times as cdf_epoch or preferably cdf_time_tt2000
  - Variable attributes (descriptions, labels, units, display_type)
  - Global attributes and file naming
  - Virtual variables (functions in CDAWlib, compute values on-the-fly)
      - »
- Checking and validation functions
  - Against ISTP standards
  - For PRBEM, MMS or other specified project compliance reporting
- New JavaScript SKTeditor plans to add capability to add SPASE metadata at the same time when creating a dataset
  - Incorporate Lee Bargatze's ADAPT business logic to reduce effort

netCDF Issues

- No predefined time variable types
  - Time not always the unlimited dimension
  - CDAWeb adds CDF_TIME_TT2000 virtual variables for netCDF datasets, computed from various time schemes (base time, time units)
- CDAWeb adds missing Fillval, Validmin/max, Var_type, depend_0, and other attributes
- netCDF to CDF converter adds attributes to store version, dimensions, sizes, compression, chunking, and string (not character) information
- Compression requires careful block size determination
- CDF to netCDF converter converts time variables to binary or encoded string forms
- Supports only netCDF4 model with no groups or user-defined variable types