

Space Physics Data Facility (SPDF)

Jonathon Smith, Lan Jian, Robert Candey,
Tamara Kovalick,
and the SPDF Team

NASA Goddard Space Flight Center, Greenbelt, MD

CEDAR Workshop June 23, 2022
Austin, TX

Introduction to SPDF

spdf.gsfc.nasa.gov



- ❑ SPDF is the **active and final archive of non-solar data** from NASA heliophysics missions and collaborative missions with other US and foreign agencies
 - Facilitate scientific analysis of multi-instrument and multi-mission data sets
 - Enhance the science return of many missions, providing context with other missions
 - Facilitate **open science** and long-term archiving
 - Make data available via many access methods (HTTP, FTP, REST, HAPI)
- ❑ We also archive other data **relevant to NASA heliophysics science objectives**
 - Related data from *planetary missions*, such as MAVEN, New Horizons
 - Heliophysics data from some NOAA and DoD satellites, such as GOES, DSCOVR, LANL
 - Non-US missions such as Arase and Formosat upon request
 - Ground-based magnetometers, aurora cameras, radars, etc., which are funded by NSF or other agencies
- ❑ We work closely with missions (some since their early development) on data issues and planning, particularly in implementing **ISTP* data standards**

*INTERNATIONAL SOLAR TERRESTRIAL PHYSICS

Over 200 Missions/Projects Supported by SPDF

* Only orbit data available

ACE	Cluster	GOES	Kepler	Parker Solar Probe	Spitzer
Active*	Cosmos 900	GOLD	LANL	Phobos	Sputnik 1
Aeros	C-NOFS	GPS	LRO	Pioneer	STEREO
AIM	CRRES	GMS 3	LUNA	Pioneer 10	Suisei
Akebono*	CSSWE	GRACE*	Magsat	Pioneer 11	Swarm
Alouette1	Dawn*	Granat	MAP	Pioneer Venus	Tatiana
Alouette2	DEMETER*	Hawkeye	Mariner 10	Polar	THEMIS
AMPTE	DMSP	Helios	Mars	Prognoz	TIMED
APEX-MAIN*	Double Star*	Hinode	MAVEN	Reimei	TRACE
Apollo	DSCOVR	Hinode	MESSENGER	Rosetta*	TWINS
Aqua	DE	IBEX	Microlab 1	RHESSI	UARS*
Ariel-4	Equator-S	ICON	Mir*	SAMPEX	Ulysses
Arase (ERG)	Explorer	IMAGE	MMS	Sakigake*	Van Allen Probes
ARCAD	FAST	IMP 7	MRO	San Marco	Vega
ARTEMIS	FIREBIRD*	IMP 8	MSL	SCATHA*	Venera
ASTRID II*	Formosat	IMP_early	MSX*	SDO	Viking
AE	Freja*	Interball	Munin	SET-1/DSX	Voyager
Aura	Galileo*	ISEE	New Horizons	SMILE	Voyager 1
Aureol2	GCOM W1	ISEE 3-ICE	NOAA*	SNOE	Voyager 2
BARREL	Genesis	ISIS	Oersted	SOHO	Wind
BepiColombo	Geotail	ISS	OGO	Solar Orbiter	XMM-Newton
CALIPSO	Giotto*	Jason 2	Ohzora	SORCE	Yohkoh*
Cassini*	GOCE	Juno	PARASOL	Spartan-A	Zond
Cassiope					

Total: ~3000 data sets, ~40 million data files, ~400 TB data

Recent average **monthly** data ingestion rate: ~0.6 million data files, ~14 TB data

Science-Enabling Services of SPDF

1. Coordinated Data Analysis Web (CDAWeb)

- Primary SPDF data service for mission data
- Present data set view rather than individual data files
- List, plot, subset, and download data in CDF or ASCII format

2. Satellite Situation Center (SSCWeb)

- List and plot the orbits of multiple s/c in a variety of coordinate systems
- Query for satellite-satellite and satellite-ground station conjunction.
- Include most heliospheric satellites and many ground stations.
- **4D Orbit Viewer:** Interactive 4D animation of orbits

3. OMNI Web and COHO Web

- Magnetic field, solar wind plasma, and energetic particle data in various locations of the heliosphere, especially the OMNI data mapped to Earth's bow shock
- Interface for plotting, filtering, and downloading the data

1. Coordinated Data Analysis Web (CDAWeb)

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

~80 Missions/Sources

- Enable multi-mission, multi-instrument science
- Present data set view rather than individual data files
- List, plot, and correlate data
- Download full or a subset of data in CDF or ASCII format
- Sometimes netCDF

• Select zero OR more Sources
(default = All Sources if ≥ 1 Instrument Type is selected)

- ACE
- AMPTE
- ARTEMIS
- Alouette
- Apollo
- Arase (ERG)
- BepiColumbo
- CNOFS
- CRRES
- Cassini
- Cluster
- DMSP
- DSCOVR
- Dawn
- Dynamics Explorer
- Equator-S
- FAST
- Formosat
- GOES
- GOLD
- GPS
- Galileo
- Genesis
- Geotail
- Giotto
- Hawkeye
- Helios
- IBEX
- ICON
- IMAGE

• Select zero OR more Instrument Types
(default = All Instrument Types if ≥ 1 Source is selected)

- Activity Indices
- Electric Fields (space)
- Electron Precipitation Bremsstrahlung
- Energetic Particle Detector
- Engineering
- Ephemeris/Attitude/Ancillary
- Gamma and X-Rays
- Housekeeping
- Imaging and Remote Sensing (ITM/Earth)
- Imaging and Remote Sensing (Magnetosphere/Earth)
- Imaging and Remote Sensing (Sun)
- Magnetic Fields (Balloon)
- Magnetic Fields (space)
- Particles (space)
- Plasma and Solar Wind
- Pressure gauge (space)
- Radio and Plasma Waves (space)
- Radio and Plasma Waves (space), Electric Antennas
- Spacecraft Potential Control
- UV Imaging Spectrograph (Space)
- Ground-Based HF-Radars
- Ground-Based Imagers
- Ground-Based Magnetometers, Riometers, Sounders
- Ground-Based VLF/ELF/ULF, Photometers

CDAS Web Service Client Codes for Python and IDL

CDAWeb Data Selector

- To go forward to plot, list and retrieve your selected data, press the "submit" button directly below or at the bottom of this page.
- For any special notes on usage of a given data set, please click on that data set name below.
- As needed to select the datasets of actual interest to you:

- manually check/uncheck one or more data sets from the list below OR
- [Click here to CLEAR All checkboxes](#), OR
- [Click here to SELECT All checkboxes](#)

Submit

- AC_OR_SSC**: ACE GSE Positions @ 12 min resolution - SSC/SSCWeb (NASA's GSFC)
[Available Time Range: 1997/08/25 17:48:00 - 2022/08/01 23:48:00] [Info](#) [Metadata](#)
- AC_OR_DEF**: ACE Daily GSE and J2000 GCI Position Data - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/26 00:00:00 - 2022/06/08 00:00:00] [Info](#) [Metadata](#)
- AC_AT_DEF**: ACE Hourly RTN, GSE and J2000 GCI Attitude direction cosines - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/26 00:00:00 - 2022/06/08 01:00:00] [Info](#) [Metadata](#)
- AC_H2_CRIS**: ACE/CRIS Cosmic Ray Isotope Spectrometer 1-Hour Level 2 Data - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/27 00:00:00 - 2022/05/18 23:00:00] [Info](#) [Metadata](#)

Link to **SPASE* Record**, main description

CDAS Web Service Client Code Examples

The following web service client code examples demonstrates how to access data from the [AC_OR_SSC](#) dataset from particular programming environments.

cdasws Python Library

The following code demonstrates using the cdasws library to access [AC_OR_SSC](#) data in Python. It is merely an example and does not show all the capabilities of the library. You 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 = ['GSE_LAT', 'GSE_LON', 'RADIUS', 'XYZ_GSE', 'XYZ_GSEO']
time = ['2022-08-01T21:48:00.000Z', '2022-08-01T23:48:00.000Z']
status, data = cdas.get_data('AC_OR_SSC', vars, time[0], time[1])

# If spacepy was installed
print(data['GSE_LAT'])
print(data['GSE_LAT'].attrs)
#print(data)
# ...
# see https://spacepy.github.io/datamodel.html

# If xarray and cdflib was installed
#print(data.data_vars['GSE_LAT'].values)
#print(data.data_vars['GSE_LAT'].attrs)

#print(data)
# ...
```

Copy code to clipboard

Download code

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

- PyPI description [cdasws](#)
- Jupyter Python [notebook example](#)
- Application Programming Interface description [API](#)

**Space Physics Archive Search and Extract*

Metadata → Skeleton Table

Global & Variable Attributes

CDAWeb Data Selector

- To go forward to plot, list and retrieve your selected data, press the "submit" button directly below or at the bottom of this page.
- For any special notes on usage of a given data set, please click on that data set name below.
- As needed to select the datasets of actual interest to you:

- manually check/uncheck one or more data sets from the list below OR

[- Click here to CLEAR All checkboxes, OR](#)

[- Click here to SELECT All checkboxes](#)

- AC_OR_SSC**: ACE GSE Positions @ 12 min resolution - SSC/SSCWeb (NASA's GSEFC)
[Available Time Range: 1997/08/25 17:48:00 - 2022/08/01 23:48:00] [Info](#) [Metadata](#)
- AC_OR_DEF**: ACE Daily GSE and J2000 GCI Position Data - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/26 00:00:00 - 2022/06/08 00:00:00] [Info](#) [Metadata](#)
- AC_AT_DEF**: ACE Hourly RTN, GSE and J2000 GCI Attitude direction cosines - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/26 00:00:00 - 2022/06/08 01:00:00] [Info](#) [Metadata](#)
- AC_H2_CRIS**: ACE/CRIS Cosmic Ray Isotope Spectrometer 1-Hour Level 2 Data - E. C. Stone (California Institute of Technology)
[Available Time Range: 1997/08/27 00:00:00 - 2022/05/18 23:00:00] [Info](#) [Metadata](#)

```
! Skeleton table for the "ac_or_ssc_00000000_v01.cdf" CDF.
! Generated: Thursday, 19-May-2022 11:44:16
! CDF created/modified by CDF V3.8.0
! Skeleton table created by CDF V3.8.1_0

#header

                                CDF NAME: ac_or_ssc_00000000_v01.cdf
                                DATA ENCODING: NETWORK
                                MAJORITY: ROW
                                FORMAT: SINGLE

! Variables  G.Attributes  V.Attributes  Records  Dims  Sizes
! -----
!          4/4          28          27          0/z    0
! CDF_COMPRESSION: None
! (Valid compression: None, GZIP.1-9, RLE.0, HUFF.0, AHUFF.0)
! CDF_CHECKSUM: None
! (Valid checksum: None, MD5)
! CDF_LEAPSECONDLASTUPDATED: 20150701

#GLOBALAttributes

! Attribute      Entry      Data      Value
! Name           Number     Type      -----
! -----
"TITLE"          1:         CDF_CHAR  { "SSC ORBIT CDF" } .
"Project"        1:         CDF_CHAR  { "SSC" } .
"Discipline"     1:         CDF_CHAR  { "Space " -
                    "Physics>Interplanetary " -
                    "Studies" } .
"Source_name"    1:         CDF_CHAR  { "ACE" } .
"Data_version"  1:         CDF_CHAR  { "1" } .
"ADID_ref"       1:         CDF_CHAR  { "NSSD0110" } .
"Logical_file_id" 1:         CDF_CHAR  { "ac_or_ssc_00000000_v01" } .
"Data_type"      1:         CDF_CHAR  { "OR>Orbit" } .
"Descriptor"     1:         CDF_CHAR  { "SSC>Satellite Situation " -
                    "Center Ephemeris" } .
"TEXT"          1:         CDF_CHAR  { "GROUP 1  Satellite " -
                    "Resolution  Factor" }
                    2:         CDF_CHAR  { "          ace          " -
                    "720          1" } .
```

CDAWeb Data Explorer

Automatically set as the last available day of the selected datasets

Options:

bin averaging, spike removal, noise filtering, overlay plotting, audification, animation

Select start and stop times from which to GET or PLOT data:

Start time (YYYY/MM/DD HH:MM:SS.mmm):

Stop time (YYYY/MM/DD HH:MM:SS.mmm):

Compute uniformly spaced binned data for scalar/vector/spectrogram data (not available with noise filtering) ^{NEW}

Binning interval:

Method to handle missing values: Use Fill Value Interpolate

Spike removal method:

[More information about binning is available here.](#)

Use spike removal to filter data without binning (not available with noise filtering)(Warning: Experimental !!).

Select an activity:

Plot Data : select one or more variables from list below and press submit.

Also create PS and PDF best quality outputs (all plot types except images and plasmagrams).
Many panels per dataset are allowed but <=4 panels optimal for standard Y-axis height and single page display.

Use coarse noise filtering to remove values outside 3 deviations from mean of all values in the plotted time interval.

Increase the Y-axis height for time-series and spectrogram plots. ^{NEW}
multiply by:

Combine all time-series and spectrogram plots, for all requested datasets, into one plot file.

Plot overlay options. ^{NEW}

Overlay vector components of selected variables.

Overlay selected variables or variable components that are identical among the datasets chosen
(Supported constellations: MMS, Van Allen Probes (RBSP), THEMIS, Cluster, and GOES).

List Data (ASCII/CSV): select one or more variables from list below and press submit. (Works best for < 31 days)

Download original files : press submit button to retrieve list of files. (Max. 200 days - use [HTTPS site](#) for larger requests)

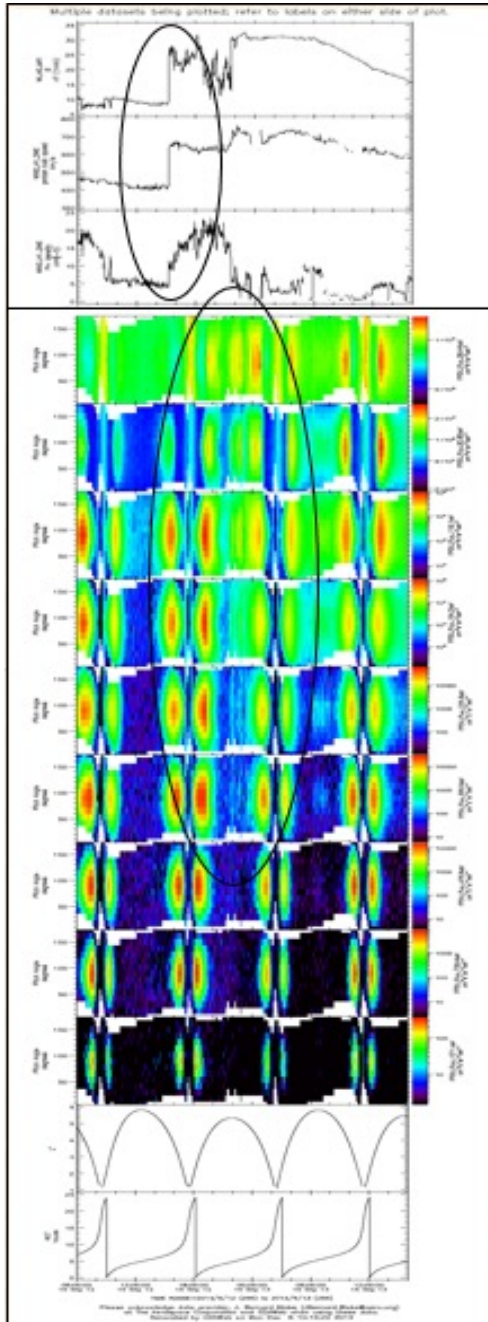
Create V3.8 CDFs for download or Autoplot demonstration: select one or more variables from the list below and press submit.

Create audio files based on data from selected variables. ^{NEW}

[More information about audification is available here.](#)

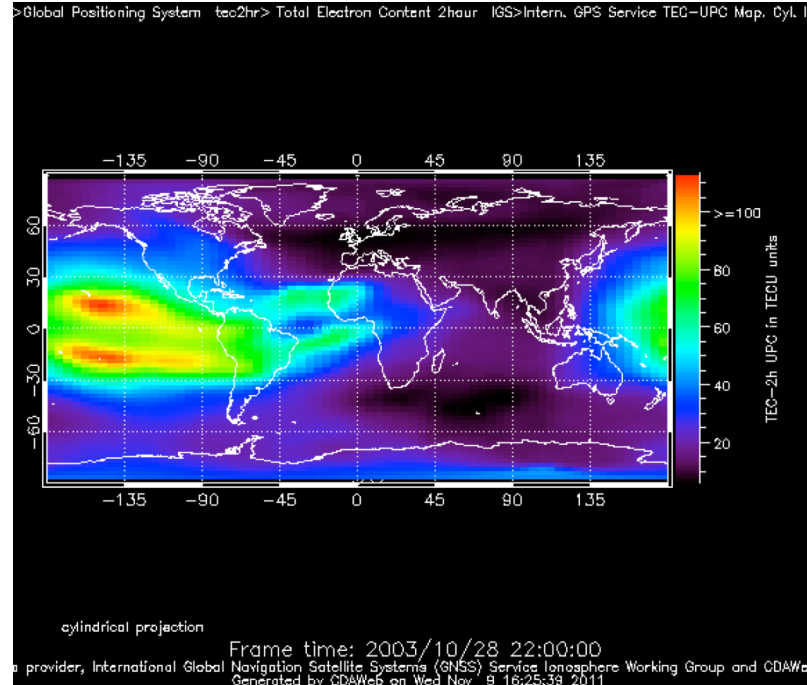
Note: [CDF patch](#) required for reading Version 3.8 CDFs in IDL or MATLAB.
Get [CDFX](#) - IDL GUI plotting/listing toolkit software. To be used with either the daily or "created" CDF files available above.

^{NEW} Pressing the "Submit" button will spawn a new window/tab in order to support the new "Previous" and "Next" functions.

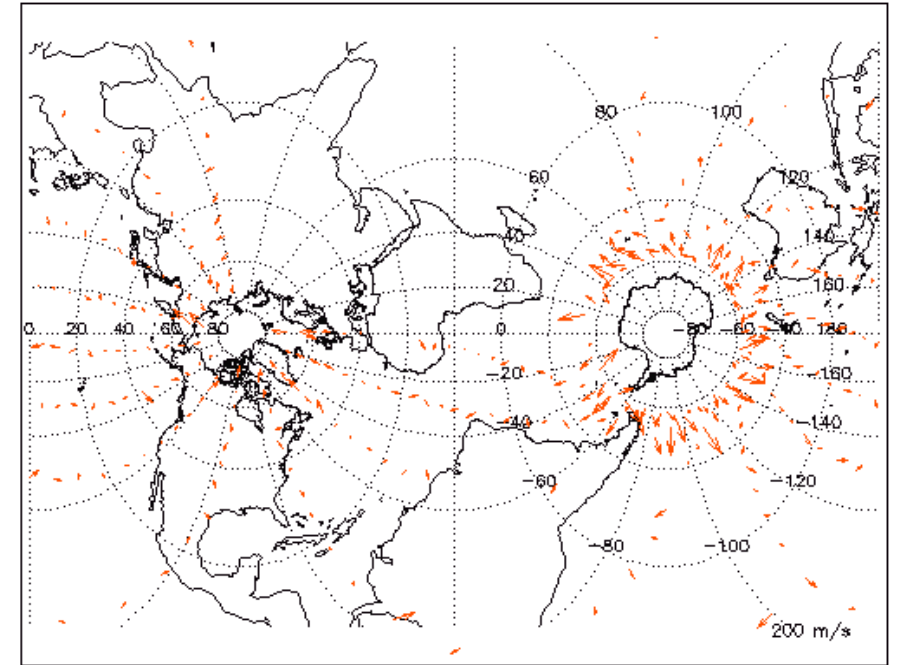


Example Plots from CDAWeb

Enable the service for **multi-dimensional variables** by setting up virtual variables



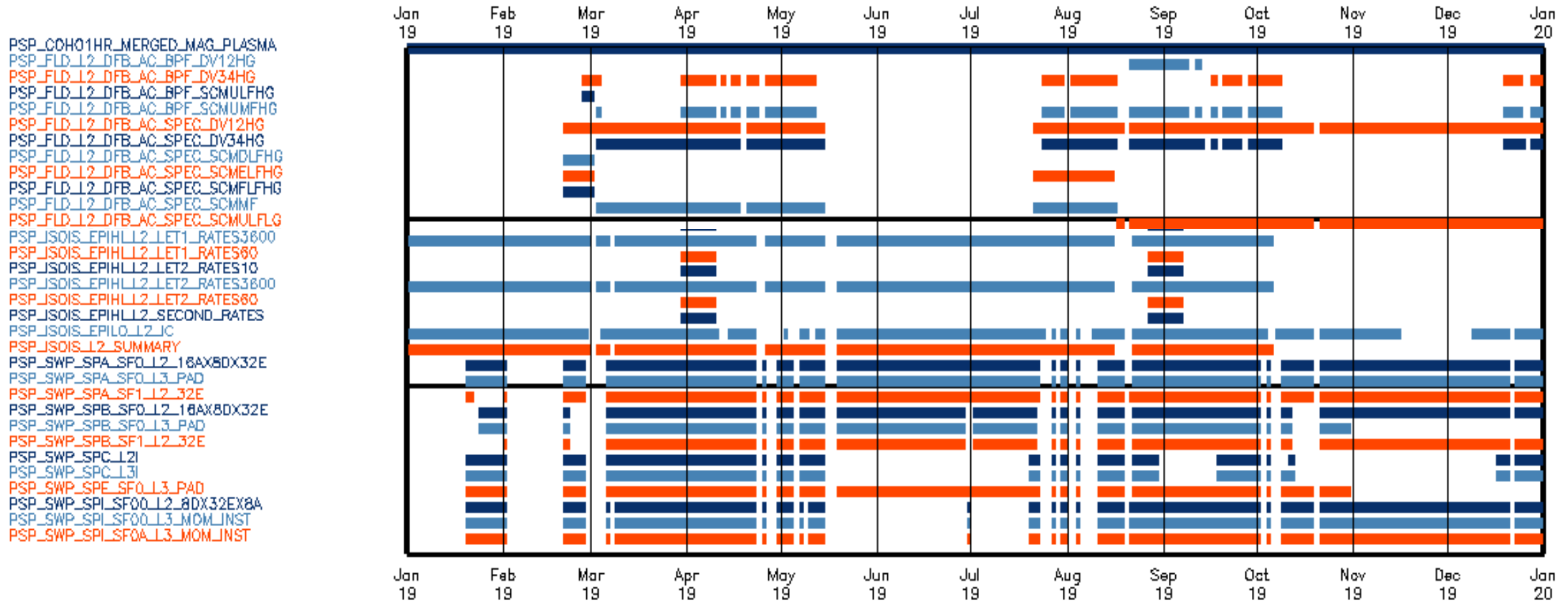
GPS International GNSS Service
Total Electron Content



TIMED/TIDI Wind Vectors Movie
Transverse Mercator Projection

Additional Resources at CDAWeb

Part of the Inventory Plot for Parker Solar Probe (PSP) Data in 2019



At CDAWeb, the inventory plots are available for every mission and they are updated daily.

There are also usage statistics for all the data sets.

Why metadata conventions

- Leverage standardized self-describing data formats, metadata for datasets and parameters, time conventions, and dataset and file naming 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 provider's experience 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 distinguish variables

Some standards and conventions

- **SPASE** <http://www.spase-group.org> dataset descriptions for easy searching
- **ISTP/IACG/SPDF Guidelines** for global and variable attributes
https://spdf.gsfc.nasa.gov/sp_use_of_cdf.html
 - SKTEditor metadata creation tool <https://spdf.gsfc.nasa.gov/skteditor>
- **Dataset naming and file naming** recommendations
http://www.tds.org/Recommended_file_and_data_collection_naming_practices
and filenaming templates http://tds.org/uri_templates
- **CDF** <https://cdf.sci.gsfc.nasa.gov> scientific data format (including its new Python library <https://github.com/MAVENSDC/cdfplib>)
 - Time variable types
https://cdf.sci.gsfc.nasa.gov/html/leapseconds_requirements.htm
- **netCDF** <https://www.unidata.ucar.edu/software/netcdf/>
- **FITS** <https://fits.gsfc.nasa.gov/>
- **UDunits** www.unidata.ucar.edu/software/udunits/

Contact me with any questions: jonathon.m.smith@nasa.gov

ISTP/SPDF Guidelines Structure and Metadata Concepts

- **ISTP/IACG Guidelines (mid1990s) and subsequent extensions by SPDF define a limiting set of implementation standards for CDFs**
 - 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)
- **Terminology: “Skeleton” CDF is a CDF with structure and metadata defined but no data, so it can be used as a template from which to build a data file**
- **CDAWeb additional concepts: “Master” CDFs and “Virtual” Variables**
 - “Master” CDF is the use of a “skeleton” CDF to insert supplemental or updated metadata for CDFs as a dataset

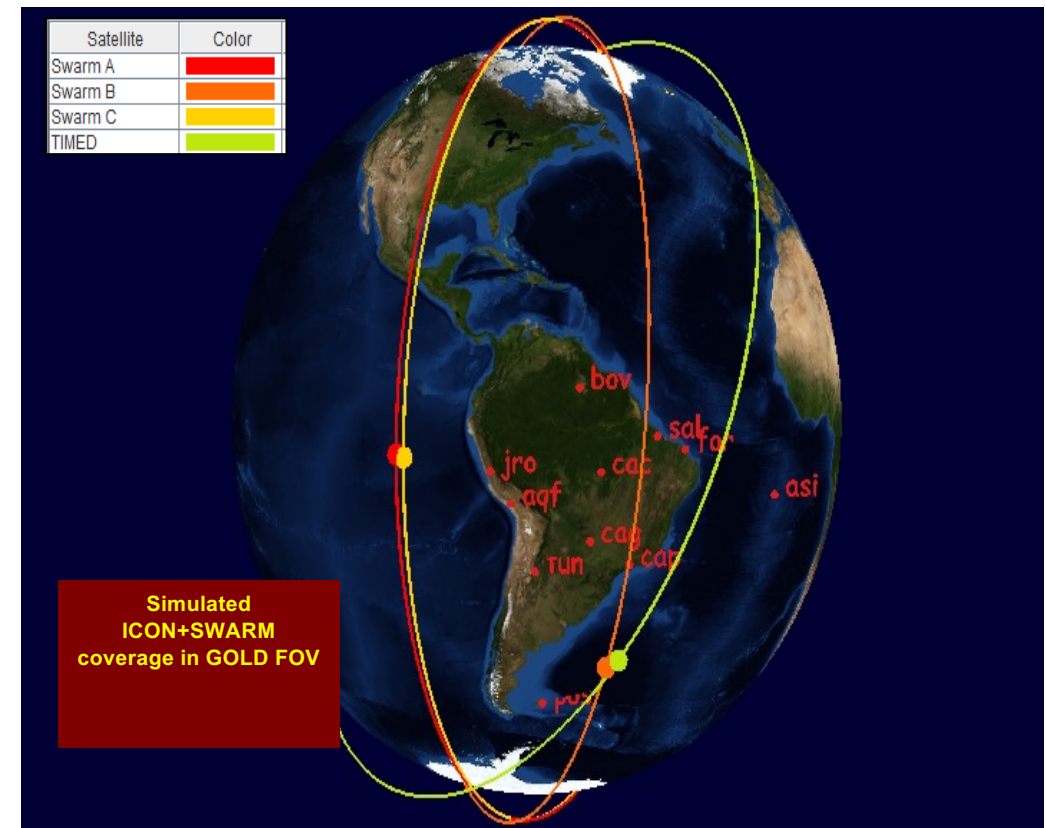
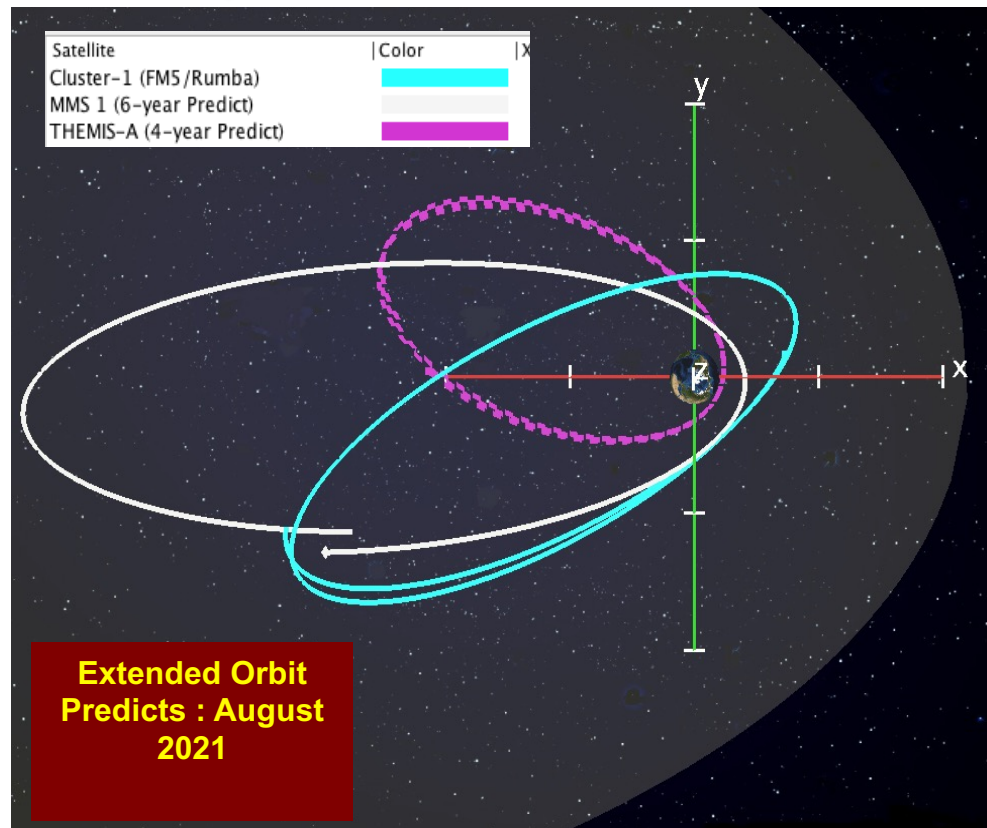
ISTP/SPDF Metadata Elements

- **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
- **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 8byte float; usually UTC but not leap seconds
 - EPOCH16 picoseconds from 0AD in two 8byte float; usually UTC but not leap seconds
- **ISTP/SPDF Guidelines online at**
https://spdf.gsfc.nasa.gov/sp_use_of_cdf.html

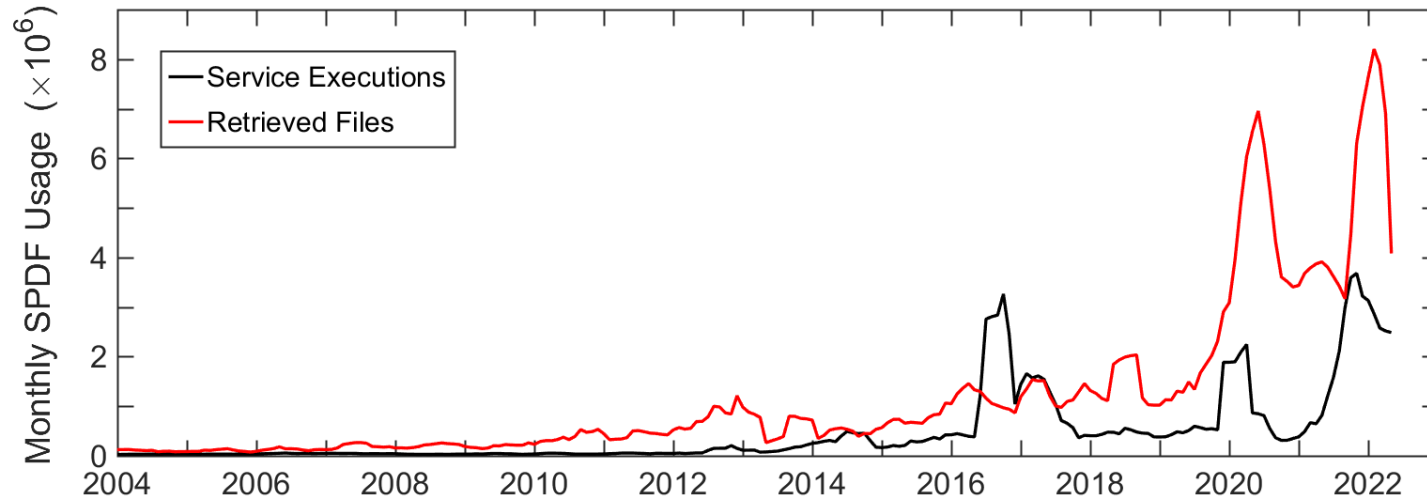
2. Satellite Situation Center (SSCWeb)

Upgrade

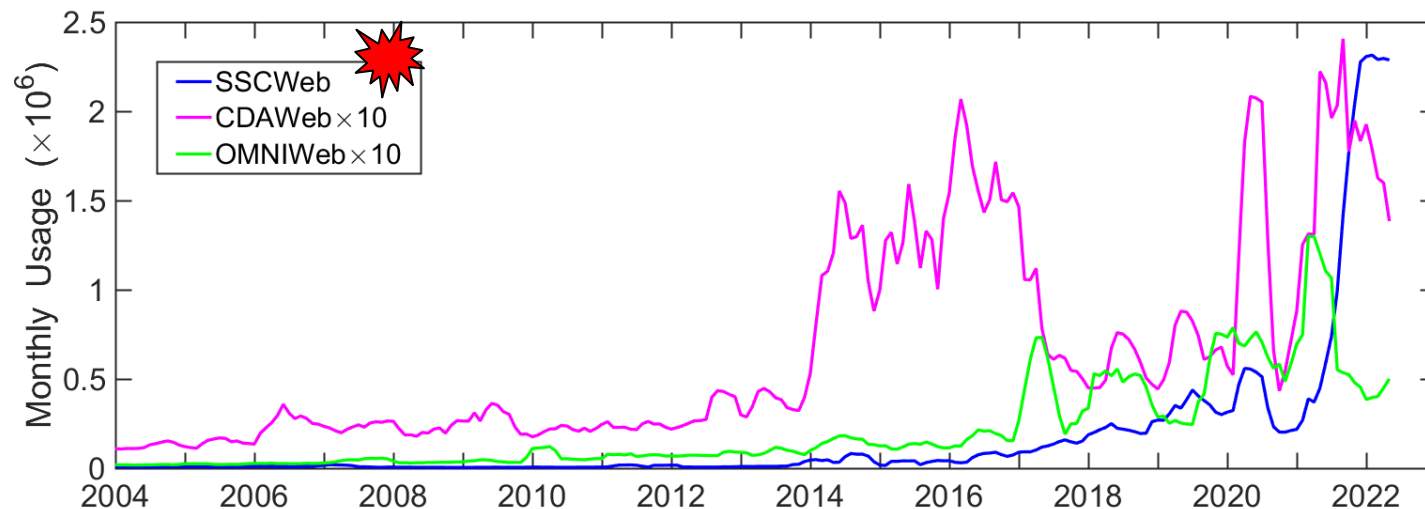
- Include most heliospheric satellites and many ground stations
- List and plot orbits of multiple s/c in a variety of coordinate systems
- **4D Orbit Viewer:** Interactive 4D animation of orbits
- Query for satellite-satellite and satellite-ground station conjunctions



Extensive Use of SPDF Data & Services



- Significant increase of SSCWeb usage since 2021 is probably related to CCMC and other users for model-observation comparison



- In 2021, ~40% of papers in AGU's *JGR Space Physics* and *Space Weather* journals acknowledged SPDF services and/or data