Heliophysics Digital Resource Library (HDRL) Program Review Report

June 18th, 2026

Executive Summary

The Heliophysics Digital Resource Library (HDRL) serves as the unified federation of three core organizations: Solar Data Analysis Center (SDAC), Space Physics Data Facility (SPDF), and Data Integration Services for Heliophysics (DISH), dedicated to the acquisition, curation, preservation of Heliophysics data, and the development of services to enable the science goals of NASA's Heliophysics Division. By coordinating strategic efforts between these groups, HDRL ensures that NASA's vast observational heliodatasets remain accessible and well-documented for the research community. The program's strategy is driven by stakeholder needs and national Open Science mandates, such as SPD-41a and the Decadal Survey, focusing on four key pillars: high-quality services, sustainable software libraries, robust community engagement, and the development of essential data standards.

Operationally, HDRL leverages the specialized expertise of its member entities to provide comprehensive data lifecycle support. The SDAC acts as the primary repository for solar physics mission data, offering visualization and search tools like Helioviewer and the Virtual Solar Observatory (VSO), while the SPDF maintains permanent archives for space physics data, providing foundational infrastructure such as CDAWeb, SSCWeb, and OMNIWeb. Complementing these, DISH focuses on cross-cutting integration, driving cloud-based computing initiatives like HelioCloud, and supporting software development. This integrated approach enables researchers to navigate through diverse data environments, from archival solar records to real-time space weather observations.

Looking ahead, HDRL faces a period of rapid growth characterized by significant increases in data volumes and the necessity of integrating advanced AI/ML capabilities into discovery and analysis workflows. To sustain its impact, the program is prioritizing the automation of curation pipelines, the modernization of legacy infrastructures, and the expansion of Open Science practices. While challenges such as hardware costs, the complexity of integrating diverse mission-specific data characteristics, and the demands of big data analysis remain, HDRL's commitment to innovation and international collaboration ensures it remains the cornerstone of NASA's Heliophysics data environment.

1. Current Program Activities and Operations

The Heliophysics Digital Resource Library (HDRL) located at NASA Goddard Space Flight Center is a federation of three member organizations: the Solar Data Analysis Center (SDAC), the Space Physics Data Facility (SPDF), and the Data Integration Services for Heliophysics (DISH; formerly the Heliophysics Data and Model Consortium or HDMC). Figure 1.0. below shows the organization of HDRL and the main functions of each member entity.

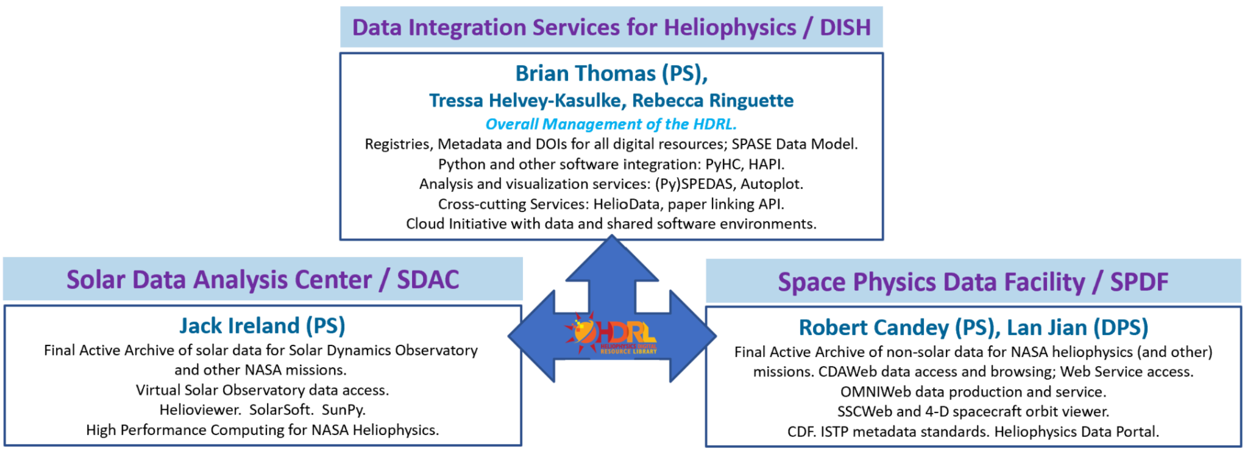


Figure 1.0. The organization chart of HDRL *listing the functions of each member entity.*

HDRL's core mission is to acquire, curate, register, preserve, and provide easy and uniform access to well-documented digital data in Heliophysics (Roberts et al 2023, HDP Data Policy 2025). Observational data is irreplaceable, and its preservation is the fundamental task of HDRL along with associated goals of providing the infrastructure which serves as a unified central source for these diverse data, creating and hosting the services which maximize the value of these data for research, and growing the community of practice for Heliophysics observational data (Heliophysics Archiving Strategy 2023). These mandates lead to a diverse portfolio of projects and activities which are coordinated through a shared vision and strategy between all HDRL members. The project scientist of HDRL ensures that regular meetings are held frequently between the project scientists and deputies of each of the HDRL member organizations and that strategic exercises are conducted to allow members to formulate a unified HDRL strategy. The project scientist for HDRL also serves as the primary point of contact for communicating HDRL-wide strategy and activities.

HDRL prioritizes its activities and processes through the formulation of a common strategy which is stakeholder driven. Stakeholders include NASA Heliophysics and space weather researchers, NASA Heliophysics Division (HPD), other US agencies engaged in Heliophysics, HDRL collaborators and international partners. HDRL engages these stakeholders through regular meetings, presentations and participation in relevant forums such as the Inter-agency Heliophysics Data Working Group (IHDWG), the International Heliophysics Data Environment Alliance (IHDEA) and its related working groups.

Beyond HDRL stakeholders, there are a number of strategic drivers which are relevant for HDRL strategy. Drivers include NASA SMD and HPD strategy and policy such as SPD-41a, US Governmental policies such as Open Data, Open Science and Gold Standard Science, and the Heliophysics Decadal Survey. Prioritization of the work in these strategic areas is done either at the individual member organization or at the HDRL-wide level depending on the scope and impact of the work. Prioritization is done by balancing stakeholder feedback and HDRL strategic drivers.

The overall HDRL strategy goal is to maximize the return on investment for HDRL users to discover, access, understand, and use NASA Heliophysics data. This is achieved in four strategic areas which include:

- **Standards**

This area involves the development of metadata and associated standards for data, services and software for Heliophysics research. Standards supported by HDRL include ISTP, SPASE, Heliophysics coordinate systems, units, events, Heliophysics research software environments, citation of Heliophysics research. In all cases our work in standards is done to improve FAIR and Open Science compliance for NASA data.

- **Services**

This area includes both human-led activities like curation and improvement of metadata and data as well as the development of software-based services such as Helioviewer, VSO, CDAWeb, OMNIWeb, SSCWeb, 4D Orbit Viewer, HelioData, HEK, HDP, and HelioCloud.

- **Software (libraries)**

This area involves the development of software libraries which support Heliophysics research and community data infrastructure. Examples in this area include Common Data Format (CDF), SolarSoft, CDFlib, CDAWlib, (py)SPEDAS, HAPI, and Autoplot.

- **Community (support & feedback)**

This area involves engagement with researchers and other stakeholders to both understand gaps in our capabilities and advertise new and improved HDRL capabilities. Examples include organized outreach such as the Heliophysics and PyHC Summer Schools, strategic workshops, and engagement with the community at meetings including GEM, SHINE, CEDAR, AGU, AMS, SPD, and the combined DASH/IHDEA meeting. This area also includes communities sponsored by HDRL which help to exchange information and best practices involving Heliophysics research. This last area includes funded efforts such as PyHC and HelioNauts, and it will soon include the support for NASA funded research data products as well.

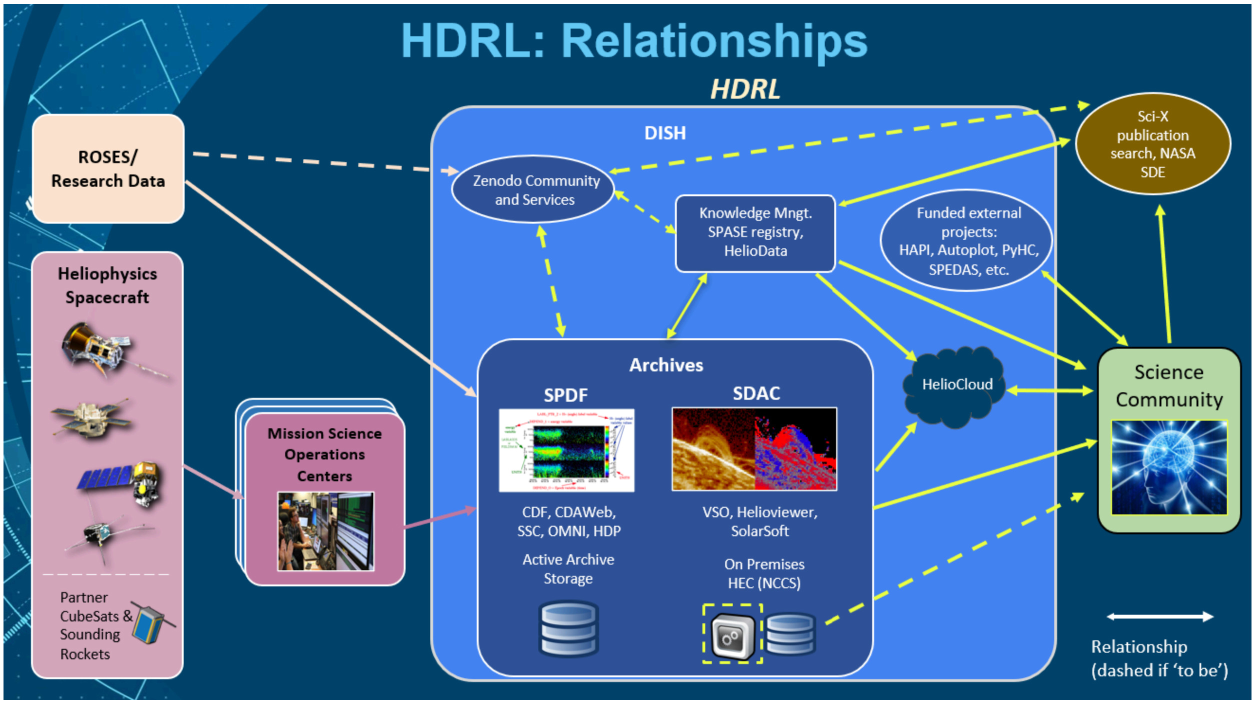


***Figure 1.0.1. Diagram of Relationships** between HDRL members (center), data providers (left hand side) and the science community (right hand side). Dashed lines and boxes indicate future capabilities and relationships.*

An overview schematic of the relationships between HDRL members, data providers, and the science community are shown in Figure 1.0.1 above. Details of core capabilities, current programs, projects and activities for each member organization are given in the sections below.

1.1. SDAC

The SDAC is NASA's designated repository of NASA solar physics mission data. The mission of the SDAC is to support the analysis of NASA's solar physics mission data. The SDAC has six core functions: data curation, data search and access, data visualization, provision of data analysis and processing software, mission science data archive hosting, and mission instrument science operations. Updates to SDAC capabilities are presented regularly to the community and feedback solicited at scientific meetings such as the AGU, SPD, TESS, and DASH/IHDEA.

1.1.1. Data Curation

The SDAC holds and curates data from previous and currently operating NASA solar physics missions. Missions typically copy science data products that they produce and make available via local repositories regularly and frequently to the SDAC during their Phase E, with the delivery of final science data products in Phase F. The exceptions to this are PUNCH, PADRE, and CODEX where the SDAC is the only source of publicly available data for these missions. The SDAC stores only data that is free from any restriction of use. Mission data, excepting SDO, is stored on a NetApp device (2.8PB usable). The SDAC also acts as a data distribution site for SDO data, holding a rolling 1.25PB cache of SDO data. Mission science data is currently being copied onto LTO-9 tapes in a SpectraLogic T950V cabinet with four tape drives. The tape library system is expandable, and will be used to create and hold offline copies of the SDO data. All SDAC and SPDF science data (excepting the rolling cache of SDO) have also been copied to SDAC procured storage at the IMCC (former NCCS, 20PB). This storage is co-located with the compute resources of the IMCC which include the DISCOVER supercomputer and a GPU cluster. Current SDAC plans call for the automated creation, storage, and provision of higher level AIA data products, such as PSF-removed and co-aligned images which do not exist in large quantities anywhere else. The co-location of large volumes of data and computational capabilities brings a new data analysis capability to Heliophysics research. SDAC, working with SDO/JSOC, has created a system that has been downloading and storing the daily production of SDO Level 0 data since November 2025. Finally, high-level PUNCH data is being uploaded to HelioCloud from the SDAC.

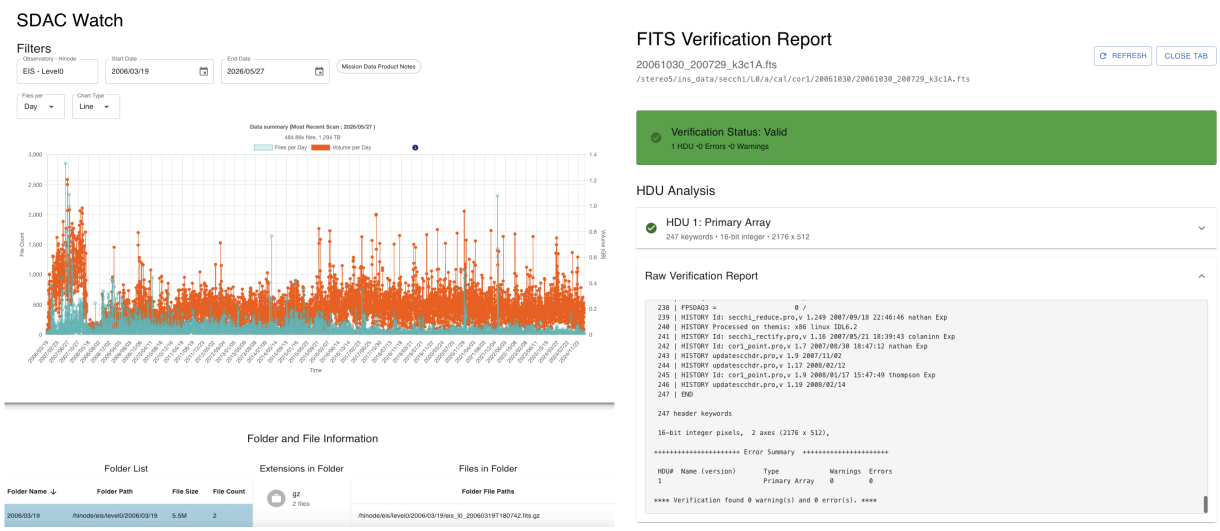


***Figure 1.1.1. Screenshots of the SDAC Watch** application showing number & volume of files per day (left) and an example FITS verification report.*

In the past two years, SDAC staff have developed the [SDAC Watch application]() to track every file at the SDAC obtained from mission archives, and to test every FITS file against the FITS standard (Figure 1.1.1). The number of FITS errors and warnings for all FITS files are stored and updated, and users can get a full report on FITS compliance for every FITS file held at the SDAC. The information collected by this application will be returned to missions so they understand exactly which files the SDAC has, and their quality. Further, this information will be used to provide data coverage details to VSO 2.0, and will be reconciled with Helioviewer Project image databases to ensure that Helioviewer clients have images of the image data at the SDAC. The SDAC Watch API is still under development and is not yet publicly available. SDAC Watch code will be made publicly available under a free and open source license.

The relationship between the SDAC and the missions has evolved. SDAC is now the primary public repository for data from the PUNCH, PADRE and CODEX missions. This represents a shift in the role of the SDAC, from secondary to primary data provider for these missions. This saves effort for the missions, as the SDAC now retains the role of making the data available, and can leverage existing SDAC projects such as the VSO and Helioviewer. This concept could be extended further by leveraging HDRL's collaboration with the IMCC.

1.1.2. Data Search and Access

The Virtual Solar Observatory (VSO) is supported by the SDAC and provides a single interface for data held at the SDAC and at other data archives around the world (Figure 1.1.2.1). It provides a unified interface to data holdings at multiple locations, permitting data search and download from those archives. This greatly reduces the burden on the user to find where the data are online, navigate user interfaces, and find and download the data they need. Development and maintenance of the VSO is a joint responsibility between the SDAC and the NSF's National Solar Observatory (NSO), although all funding for the effort comes from NASA. The VSO's impact is shown in Table 1.1.4.

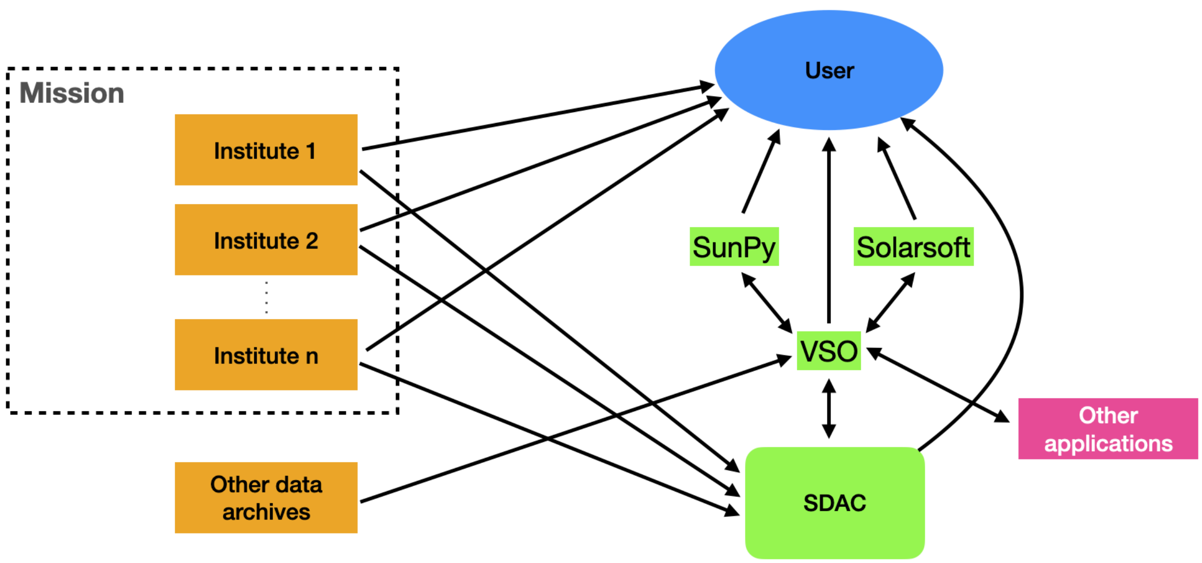


Figure 1.1.2.1. Data flow diagram *to the user from mission institutes, other data archives, SDAC, and VSO.*

The VSO has been in operation since 2002, and its API is publicly available. In the past two years VSO developers have been maintaining the VSO's existing Perl-based codebase while also re-developing and re-implementing the entire VSO. VSO 2.0 is based on twenty-plus years of experience operating the VSO and engaging with the solar community, modern software frameworks, and is being implemented in Python. Switching to industry-standard frameworks and Python as opposed to maintaining the Perl-based infrastructure positions VSO for continued success in the future. The VSO 2.0 API and the VSO website (see Figure 1.1.2.2) have been completely redesigned. We plan a gradual rollout of VSO 2.0 through 2026 and 2027 to enable continued service without interruption. VSO 2.0 will also take advantage of two DISH-led capabilities. First, users obtaining data through the VSO 2.0 will get dataset DOIs stored in SPASE records. This will make it easier for VSO 2.0 users to cite the data they download. Secondly, VSO 2.0 will use the Paper-Data-Linking API (Section 5.3.2, Table 1.3.5) to return science papers related to the data a user is searching for. This very powerfully supports the scientific community. The VSO 2.0 API follows the OpenAPI standard. The VSO 2.0 code and design will also be made publicly available under a free and open source license as much as is practicable, given that the VSO interacts with non-NASA archives with their own security requirements.

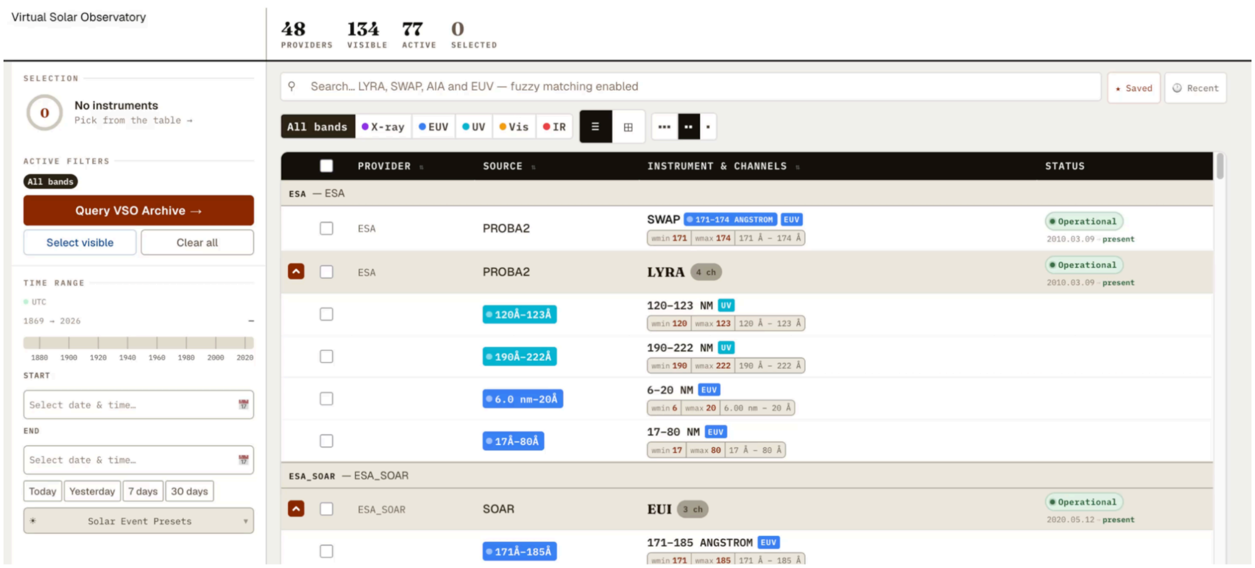


Figure 1.1.2.2. Screenshot of the new VSO 2.0-based web application*. This application is still in development and may change before initial release.*

1.1.3. Data Visualization

The Helioviewer Project, begun at the SDAC, is the major data visualization project supported by the SDAC. ESA also contributes to the development of the project. All Helioviewer Project code is free and open source. In 2025 Helioviewer Project users made over 400,000 movies and 4.8 million screenshots across all supported clients (Table 1.1.4). The two major clients supported by the Project are helioviewer.org and JHelioviewer. The web client helioviewer.org is supported and developed by the SDAC, and by analogy is “Google Maps for the Sun” (Figure 1.1.3). The Helioviewer Project hosts image data from multiple missions, covering a time range from September 1991 (Yohkoh) through to about 15 minutes behind real time (SDO). JHelioviewer, a downloadable standalone application, is supported by ESA, and is developed at ROB (Brussels, Belgium; Müller et al. 2017). JHelioviewer has powerful three-dimensional movie streaming capabilities and is aimed primarily at the professional scientific community. SDAC, ROB and IAS run the back-end server system required by JHelioviewer. JHelioviewer servers have also been installed at NOAA SWPC for internal use to support their operational needs.

All Helioviewer Project APIs are publicly available. A new [Helioviewer Events API] (Figure 1.1.3) unifies event descriptions from three different source catalogs and makes them searchable, providing simplified infrastructure for Helioviewer Project clients and a new, powerful cross-event list search capability for scientists. SDAC also directly supports the mobile version of helioviewer.org and student.helioviewer.org, which is specifically aimed at students in the classroom. IAS in Orsay, France, runs a publicly available mirror of helioviewer.org.

The Helioviewer Project is collaborating with the CCMC to provide additional information via helioviewer.org. Currently helioviewer.org displays data from CCMC's DONKI database and the Flare Scoreboard (Figure 1.1.3, left). This will shortly be extended to show data from the CCMC's WSA model predictions and PFSS field lines. Conversely, the CCMC is using the Helioviewer API to show images of solar data in their OpenSpace application.

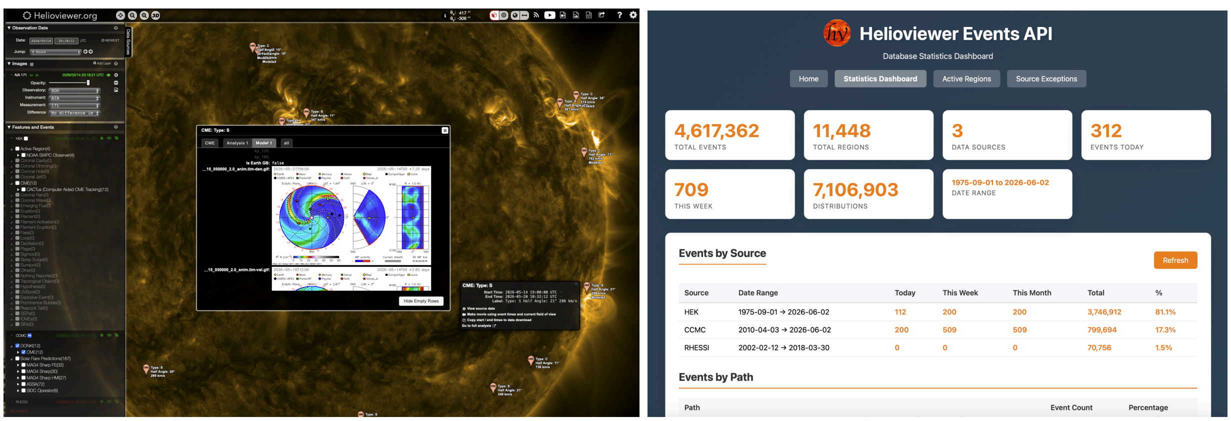


Figure 1.1.3. Screenshot of helioviewer.org interface *showing CCMC model runs (left), and screenshot of Helioviewer Events API dashboard showing data sources.*

1.1.4. Provision of Data Analysis and Processing Software

The SDAC supports the continued maintenance of the Solarsoft system (Freeland & Handy 1998). The SolarSoft system is a set of integrated software libraries, databases, and system utilities which provide a common programming and data analysis environment for solar physics. It is implemented primarily in the IDL programming language. It is publicly available and free to download. As well as being heavily used by the scientific community for over 30 years, almost all existing solar mission science data processing pipelines use the Solarsoft environment to process their science data. SDAC supports Solarsoft by funding LMSAL to provide maintenance for the system. In the past decade, SunPy (The SunPy Community et al., 2020) has become popular as a scientific data analysis environment, complementing the use of Solarsoft. Since Solarsoft is fundamental to many operating mission science data processing pipelines, and those pipelines are unlikely to be rewritten to other languages, SDAC is committed to providing maintenance for Solarsoft as long as a significant number of NASA missions use the Solarsoft system (Table 1.1.4).

Table 1.1.4 Impact of SDAC-supported software projects

Software	Description	Impact
VSO	Perl-based data discovery and download application	> 120k unique IPs, > 5M queries (2025), 125 data sources
Helioviewer	Data visualization clients, servers, image stores	> 400k movies, ~4.8M screenshots, 165M API requests (2025), 140 TB images.
Solarsoft	Data analysis and processing suite for solar physics, based largely in IDL	~ 20k IP addresses (estimated, 2025), almost all operating solar physics missions (2025)

1.1.5. Mission Science Data Archive Hosting

The SDAC provides science data archive hosting services for two operational NASA missions, SOHO and STEREO, found at the SOHO website and STEREO Science Center, for over 30 years. The SDAC also hosts Solar Orbiter servers that mirror the most recent versions of Solar Orbiter data held at ESAC. This function is distinct from Section 1.1.1 since the responsibility for the server procurement and operation, and data curation, lies entirely with the missions and not the SDAC. SDAC physically hosts the mission data archive servers, ensures that they conform to NASA security requirements, and helps diagnose issues in data transfer and curation. At the completion of each mission's Phase F, data curation responsibilities will be fully transferred to the SDAC.

1.1.6. Mission Instrument Science Operations

The SDAC provides instrument science operations support for the LASCO and EIT instruments onboard SOHO. SDAC physically hosts the instrument science operation computers, ensures that they conform to NASA security requirements, and helps diagnose issues in data transfer and curation. Responsibility for instrument planning, telemetry acquisition and processing, lies with SOHO.

1.2. SPDF

SPDF was established in 1992 and has been the active and permanent archive for non-solar Heliophysics data or space physics data for NASA since 2015. In addition to the ingestion and preservation of non-solar Heliophysics data, SPDF provides a wide range of data services and capabilities aimed at the space research community. As shown in Table 1.2., SPDF has built foundational infrastructure for Heliophysics data environments, including Common Data Format (CDF), ISTP Metadata Guidelines, and Heliophysics Data Portal (HDP), and provided many science-enabling services that offer fast access to current data across a wide range of missions and disciplines.

Table 1.2. SPDF software and services

Name	Description
SPDF Archive	Permanent archive of non-solar Heliophysics data and associated documentation, with over 2500 automated ingests daily
CDF	Scientific self-describing file format and libraries, including translations between HDF4, HDF5, FITS, netCDF, JSON, XML, and CDF
ISTP Metadata Guidelines	Standardized metadata for using and understanding self-describing science data files, and underlie generic analysis software like CDAWeb. ISTP Metadata Editor enables easy dataset layout and metadata population and verification
CDAWeb	Fast and simple data browsing interface and web services API
SSCWeb	Database and API for spacecraft orbits and complex queries for mission planning. The 4D Orbit Viewer enables easy 3D display of orbits through time (4D).
OMNIWeb	The widely used OMNI datasets that map solar wind plasma, magnetic field, and energetic particle data to the nose of Earth's bow shock along with many useful geomagnetic activity indices, and at other locations of the heliosphere.
CDAWlib	IDL software library powering CDAWeb and used in SPEDAS and other packages
Plot Walk	Fast browsing of mission-provided pre-generated summary plots and other images as well as SPDF-generated spacecraft orbit plots
Heliophysics Data Portal	Discipline-wide Heliophysics dataset search portal and API, based on SPASE records, sortable by resource types, measurement types, etc.

1.2.1. CDF

SPDF continues the development of CDF, a self-describing science file format. CDF was developed in 1982, providing a conceptual data abstraction for storing, manipulation, and accessing multi-dimensional datasets. The International Solar-Terrestrial Physics (ISTP) program (the flagship program for the space physics scientific community in the 1990s) adopted

CDF as its standard in the late 1980s, and CDF gradually became the standard for space physics data. The CDF library allows developers of CDF-based systems to easily create applications that permit users to slice data across multi-dimensional subspaces, access entire structures of data, perform subsampling of data, and access one data element independently regardless of its relationship to any other data element. CDF is supported by commercial and open source data analysis or visualization software such as IDL, MATLAB, and Python. SPDF maintains and upgrades CDF and its software libraries, and just released [CDF Version 3.9.2](). We also provide [tools for conversion]() among multiple science data file formats including CDF, FITS, netCDF, HDF, XML and ASCII.

1.2.2. ISTP Metadata

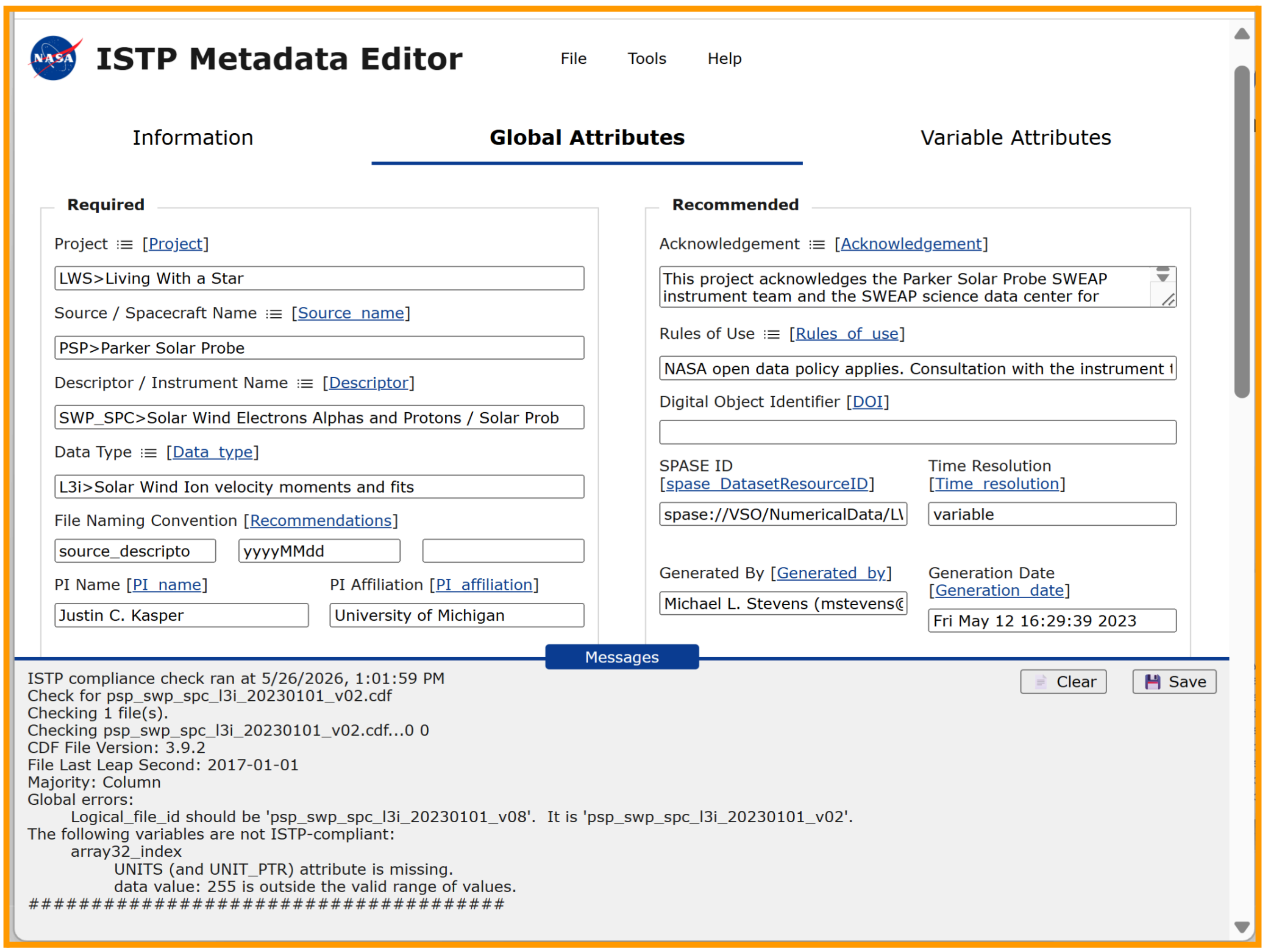


Figure 1.2.2.1. Screenshot of [ISTP Metadata Editor]() *showing (top) the global attributes and (bottom in gray box) the ISTP compliance check results which can be downloaded.*

SPDF developed and maintains the [ISTP Metadata Guidelines]() for the use in CDF and netCDF data, originally developed by the ISTP project as led by SPDF. A dataset following ISTP guidelines forms a logically complete and self-sufficient whole in terms of data and descriptions.

The underlying metadata standards allow the effective data sharing and cross-mission data analysis. See Section 4.3 for details.

SPDF has developed the ISTP Metadata Editor to help users design/edit CDF metadata or check its compliance with ISTP Metadata Guidelines. The original Java-based GUI application “SKTEditor” requires users to install Java software. A new JavaScript-based web application requiring no software installation was released in fall 2025. It has specific sections for information, (required, recommended, optional, and additional) global attributes, and variable attributes for editing, as shown in Figures 1.2.2.1 and 1.2.2.2, with itemized tutorials. It can open (including from online sources) and save as CDF, netCDF, Skeleton, JSON files. This editor can also export SPASE XML files based on the relationships between ISTP and SPASE metadata. There are tools to check for ISTP compliance, custom validation, and netCDF validation. Users can compare different versions of metadata for the same dataset and merge them using the tool as well. This new editor has become popular since the release and is used for 3500-5500 (500-700 unique) ISTP compliance checks per month recently.

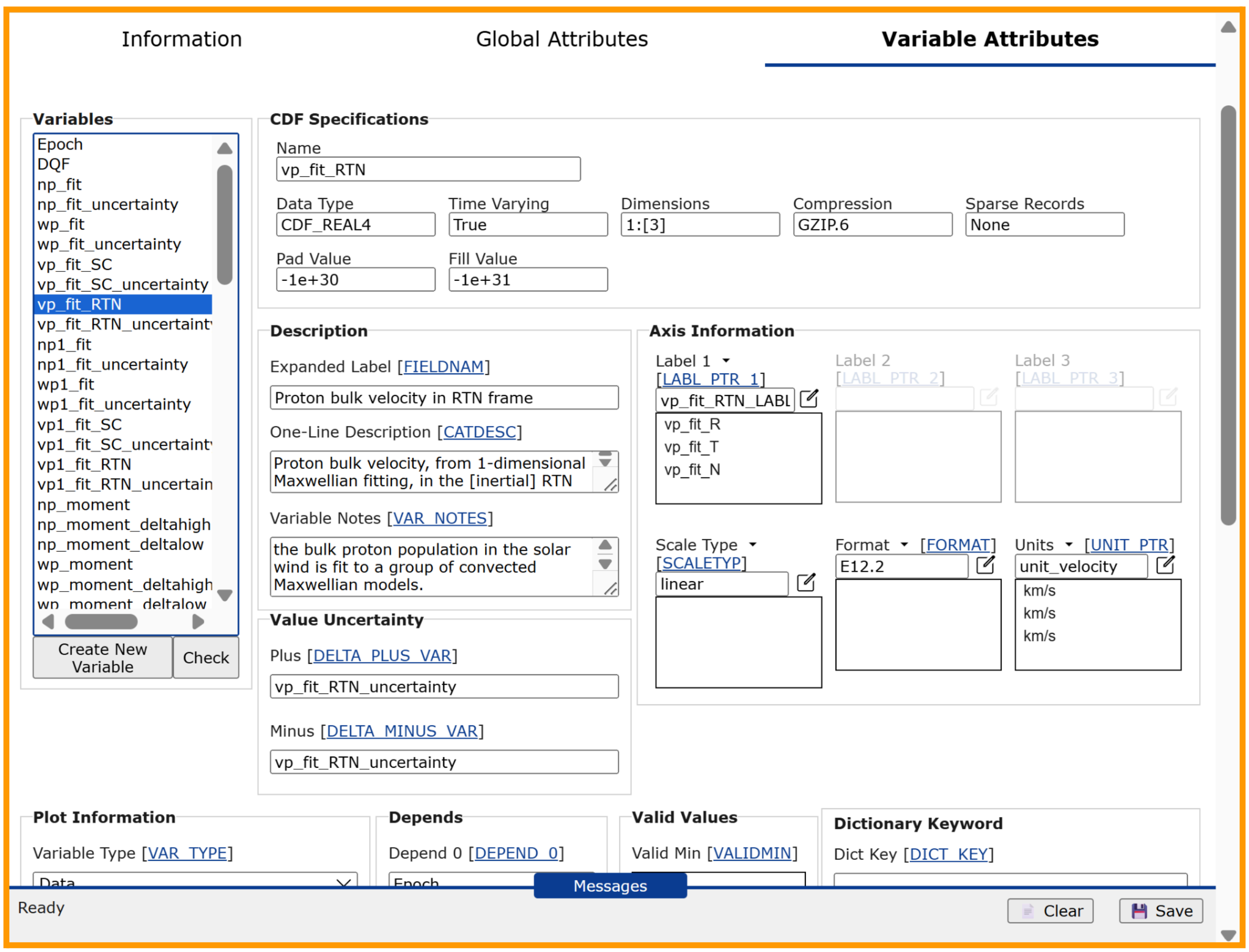


Figure 1.2.2.2. Screenshot of ISTP Metadata Editor *showing the details of variable attributes.*

1.2.3. Heliophysics Data Portal

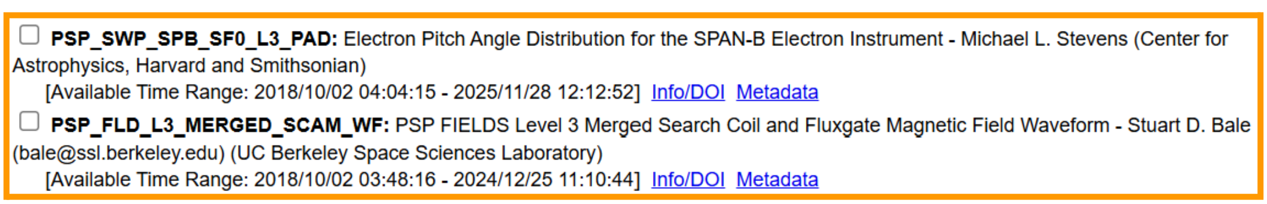
SPDF operates the Heliophysics Data Portal and associated API to provide discipline-wide data inventory and access service based on about 5400 SPASE records of catalogs, data collections, displays, and services. Users can search by resource type, measurement type, observatory group, instrument, observed region, spectral range, cadence, repository name, access rights, etc.

Besides the above three critical infrastructure, SPDF has been providing the following several main science-enabling services, and offering fast access to current data and plot holdings across a wide range of missions and disciplines, at high-resolution (in time, space and other scales like energy and frequency) and at the highest quality possible, via easy-to-use interfaces that balance simplicity with complexity.

1.2.4. Coordinated Data Analysis Web (CDAWeb)

CDAWeb provides data access through web browser, web services API, IDL, HAPI, Autoplot, and Python for more than 60 missions or mission groups. About 8000 datasets are archived at SPDF, and about 3000 of them are CDF or netCDF data meeting the ISTP metadata guidelines and able to be served at CDAWeb. CDAWeb is an interface for browsing, correlating, and displaying data, and users can request data listings, plots, audio and movies for selected data variables. There are 19 instrument types for users to down select data, including activity indices, dust and debris, electric fields, engineering, ephemeris, etc.

After selecting missions and instrument types, users reach the **CDAWeb Data Selector** where all datasets are listed with brief descriptions, PI, time range, links to Info/DOI and metadata for each dataset, as shown in Figure 1.2.4.1. Through the links, users would know how to cite the data, how to get the data using IDL/Python codes, and the content in the data without opening the CDF/netCDF data files using special tools.


PSP_SWP_SPB_SF0_L3_PAD: Electron Pitch Angle Distribution for the SPAN-B Electron Instrument - Michael L. Stevens (Center for Astrophysics, Harvard and Smithsonian)
[Available Time Range: 2018/10/02 04:04:15 - 2025/11/28 12:12:52] Info/DOI Metadata

PSP_FLD_L3_MERGED_SCAM_WF: PSP FIELDS Level 3 Merged Search Coil and Fluxgate Magnetic Field Waveform - Stuart D. Bale (bale@ssl.berkeley.edu) (UC Berkeley Space Sciences Laboratory)
[Available Time Range: 2018/10/02 03:48:16 - 2024/12/25 11:10:44] Info/DOI Metadata

Figure 1.2.4.1. The screenshot of CDAWeb Data Selector.

After selecting dataset(s), users use the **CDAWeb Data Explorer** to plot, list, or download data as shown in Figure 1.2.4. Users can average the data, remove spikes or filter coarse noise from the data. They can check the data availability, plot the data in varying axis scales, overplot vectors or the same variables of multiple satellites in the same panel. For intermittent data such as burst mode data, users can autoscale the time axis to the time interval when the data are available. CDAWeb supports several types of plots: (stacked) time series, spectrogram, and image. The output plot can be in PNG, PS, PDF formats. The output data can be in ASCII/CSV, CDF formats. Users can create new CDF files for the selected variables which could cross missions and instruments. For some datasets, users can create audio or movie files.

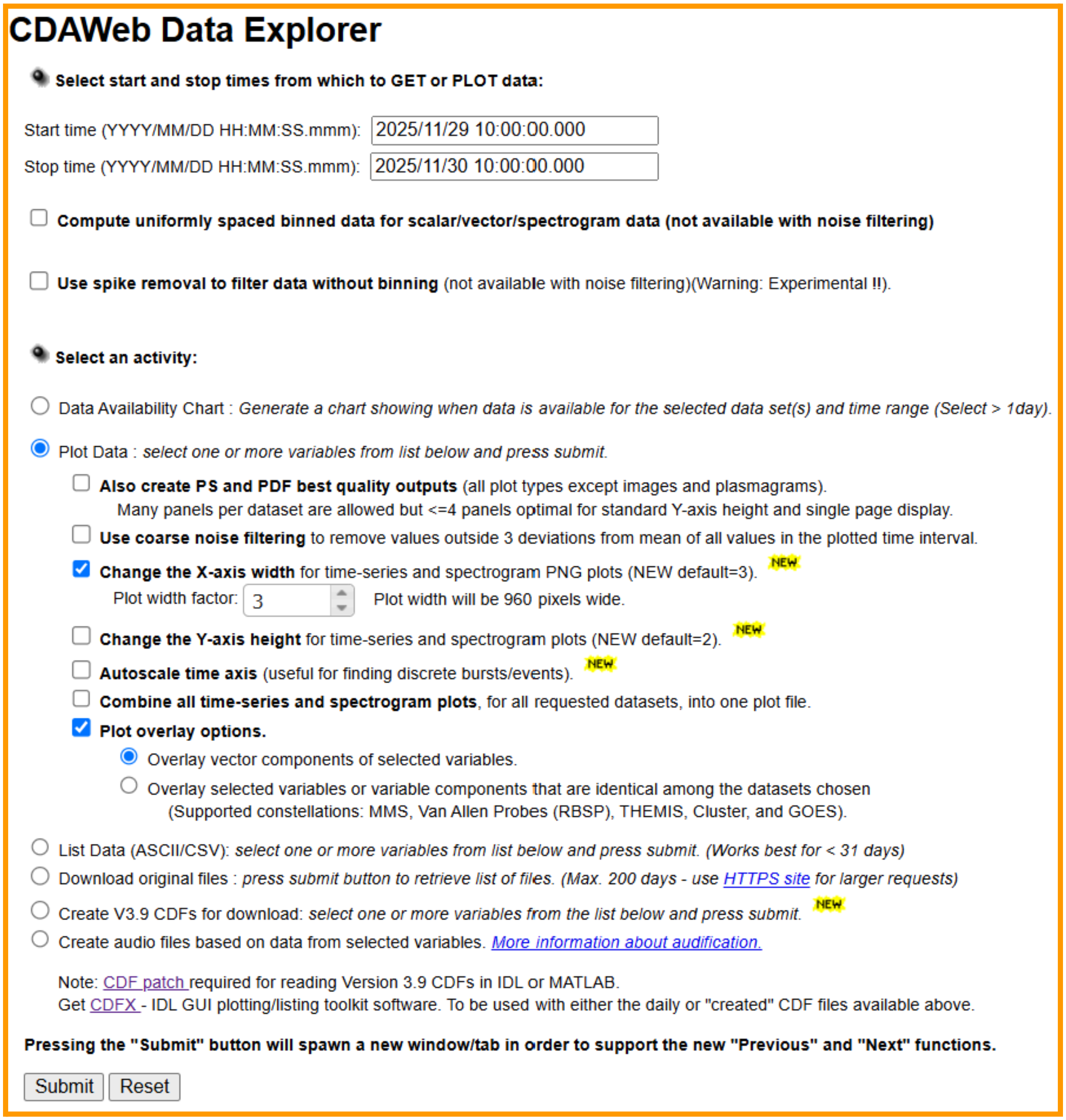


***Figure 1.2.4.2. Screenshot of CDAWeb Data Explorer** top panel, where many functions are listed. Users can download, plot, and create data using the interface.*

In CDAWeb Notes, all the datasets served at CDAWeb are listed with brief descriptions, data variables, modification history (if available from the metadata), a link to their DOIs (landing at SPASE records), and links to access codes in Python and IDL. CDAWeb also provides inventory plots (example in Figure 1.2.4.3) and usage statistics for mission data. A full redesign of CDAWeb is in progress.

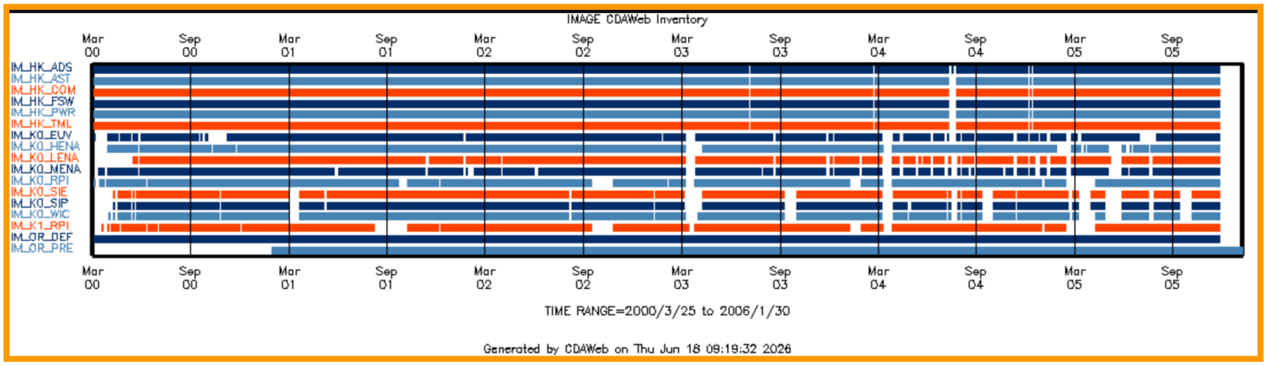


Figure 1.2.4.3. Screenshot of Inventory Plot *for IMAGE mission.*

1.2.5. Satellite Situation Center Web (SSCWeb)

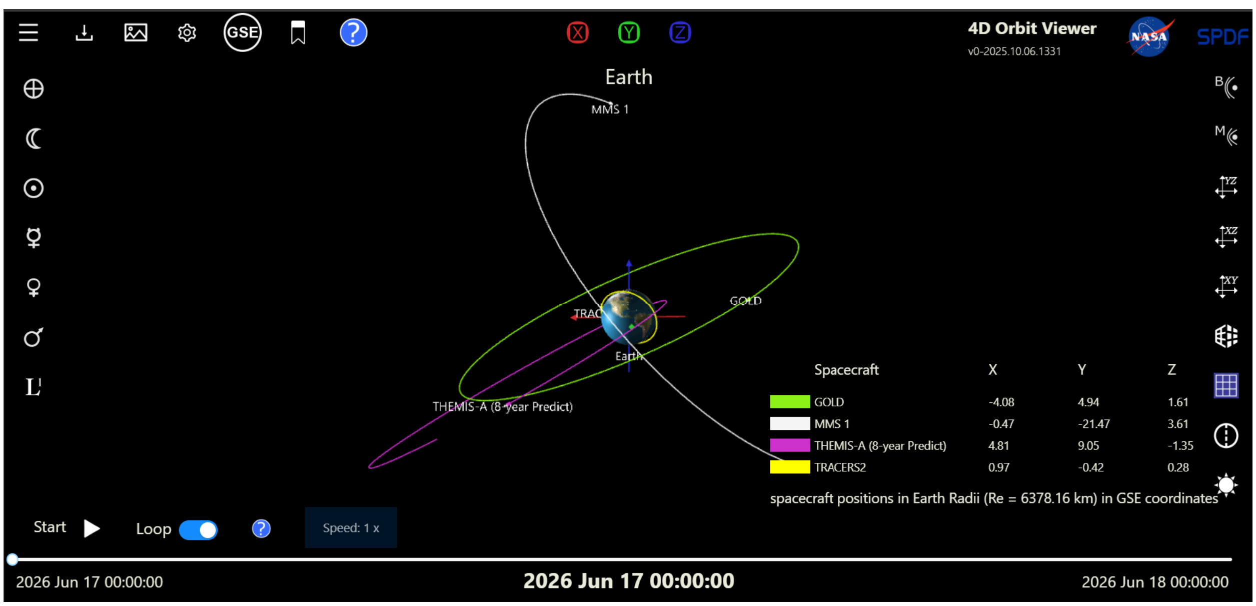


Figure 1.2.5. The screenshot of 4-D Orbit Viewer*. The left lists Earth, moon, Sun, Mercury, Venus, Mars, and Lagrange Point 1 (L1), which can be used as the center of the coordinates or display. The top menu includes the selection of spacecraft, time interval, setting, coordinates. Users can download orbit data, image or video from this viewer and save the setting by bookmarking the link. At the right side, users can add bow shock, magnetopause, and coordinate grids, switch to perspective camera by fixing at any selected spacecraft or celestial object, and add the terminator line on Earth.*

SPDF operates the SSCWeb, providing access to spacecraft orbit data through web browser and web services APIs. It hosts about 160 missions, about 200 spacecraft or celestial objects. Using SSCWeb, users can plot the spacecraft's orbit and make complex queries for conjunctions, geophysical regions, and connections to ground stations along magnetic field lines. SSCWeb also provides coordinate transform tools.

There are two ways to view 3D spacecraft orbit interactively with time as the 4th dimension: the original Java-based and standalone TIPSOD requiring downloading and installing the software package, new JavaScript browser-based 4D Orbit Viewer released in late 2023. In the browser-based tool, besides six geocentric coordinate systems and two heliocentric coordinate systems, the Selenocentric Solar Ecliptic coordinate has been recently added in the test version. We plan to add more planetary coordinates (Mars, Venus, etc.) in the near future.

1.2.6. OMNI Data and Web Services

SPDF has made the widely used OMNI datasets, mapping solar wind plasma, magnetic field, and energetic particle data to the nose of Earth's bow shock, and adding many useful geomagnetic activity indices, all spanning 1963-present (>= hourly resolution) or 1995-present (1 and 5 minute resolutions). OMNI data is selected to be one of the most impactful datasets by AGU in 2026. OMNIWeb is an interface for plotting, filtering and conducting simple statistical analysis of these data. OMNI and/or OMNIWeb were referenced in about 700 academic publications in 2025.

As a part of the OMNI web system, SPDF also provides COordinated Heliospheric Observations Web (COHOWeb) and maintains a collection of solar wind plasma, magnetic field, and energetic particle data at other locations of the heliosphere, using data from Helios 1/2, Ulysses, STEREO, PSP, Solar Orbiter, and other interplanetary missions. These long-term hourly or daily data are useful for studying the solar cycle variations and spatial variations of the heliospheric environment, and for space weather model input and validation. Furthermore, SPDF provides the hourly and daily heliocentric trajectories for about 50 selected spacecraft, planets, and comets based on JPL Horizons system.

1.2.7. Plot Walk

SPDF has upgraded the old GIFwalk to Plot Walk for easy browsing of pre-generated plots or images that missions have provided and spacecraft orbit plots that SPDF generates. About 17.5 million plots from about 30 missions and 350 plot types are served at Plot Walk. A simple JSON table provides the description of the plot datasets, along with the top URL for each collection, the filename patterns that provide the plot times, and pointers to associated images in different versions (PNG, GIF, JPG, PDF, Postscript, etc.).

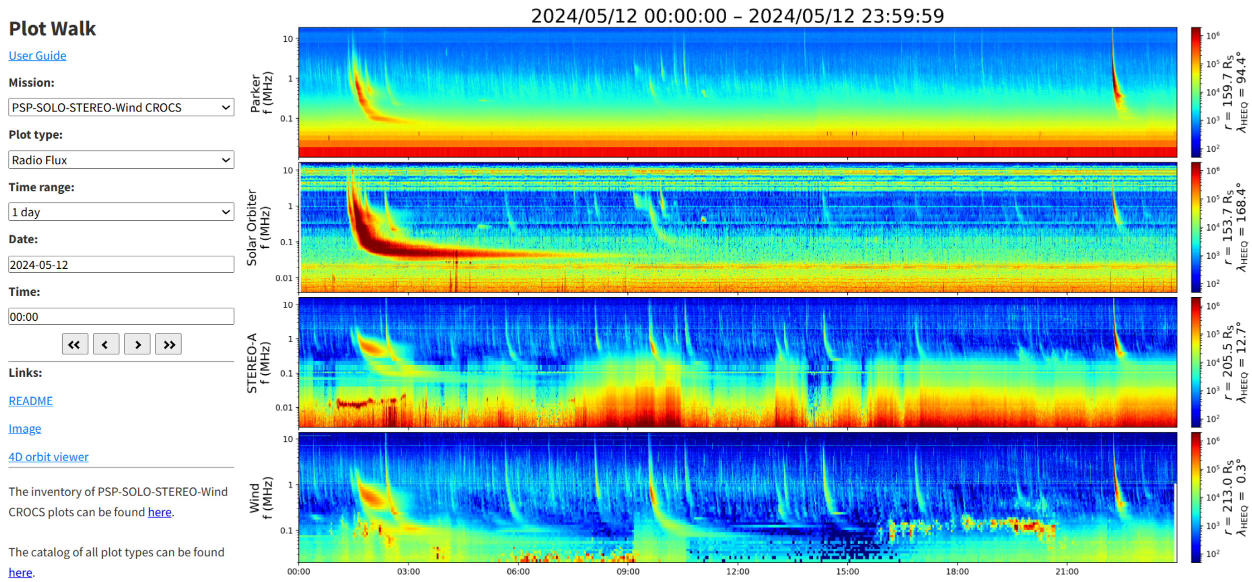


Figure 1.2.7. Screenshot of Plot Walk*. At the left side panel, users can select missions, time cadence, and relevant time (not necessarily the start time). The readme file, plot files of different formats are linked. The 1-day orbit video for the selected spacecraft and time in “4D orbit viewer” is linked, and the inventory plot of the associated mission or plot type is provided.*

To demonstrate the usefulness of the various SPDF services, assume a sample research scenario for studying solar events that produce energetic particles and effects at Earth. One can start by querying OMNIWeb for times when solar activity and energetic particle flux are high and querying the SSCWeb orbit location service for when Cluster, Geotail, Polar/IMAGE are in position to measure the cusp, magnetotail and the Earth’s aurora, respectively. Add a SSCWeb query for times when Polar and ground magnetometer stations are on the same field lines. For the selected times, use CDAWeb to browse data from these spacecraft, and add Wind and ACE field and plasma data to identify interplanetary shocks arriving at Earth. Use the SPASE-based Heliophysics Data Portal or HelioData website to find and retrieve SOHO/LASCO CME data at SDAC. Use the 4D Orbit Viewer to display the relative spacecraft positions and geophysical boundaries and to follow the magnetic footpoints of the satellites. Confirm auroral substorm activity by a quick browse of IMAGE FUV and TIMED GUVI data as movies showing the expanding and intensifying auroral oval. Finally, pull these data directly into one’s own analysis tool using our web services or simple HTTPS data file transfer to complete the analysis.

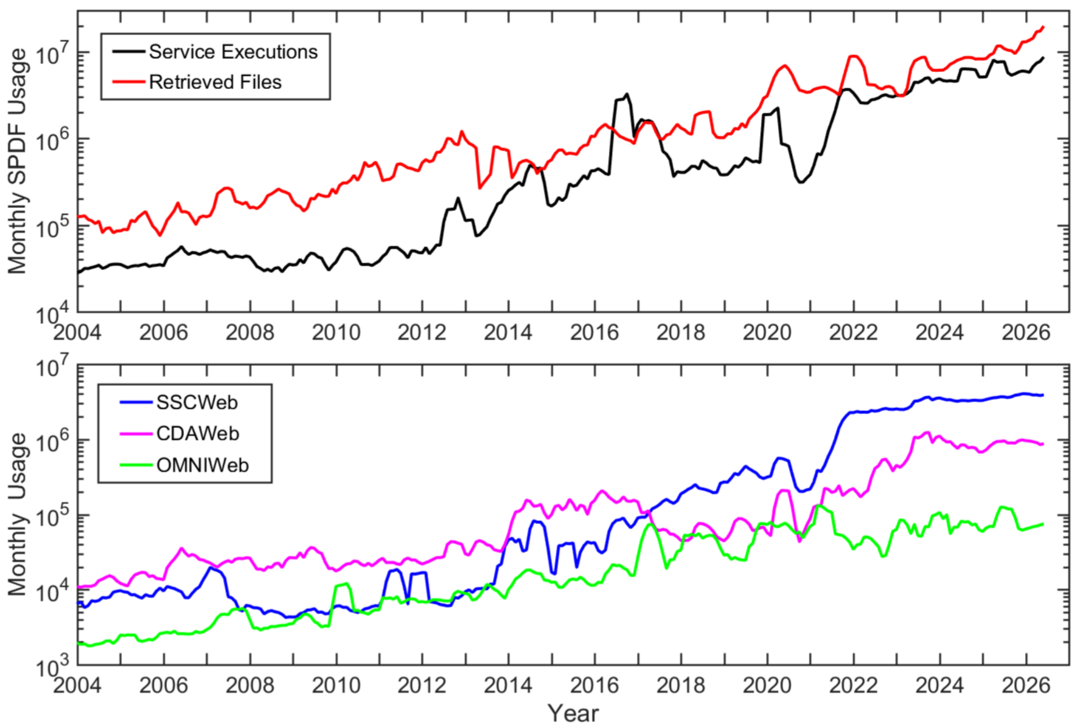


Figure 1.2. Statistics of monthly SPDF usage and services *from 2004 to present. (Top) The numbers of monthly SPDF service executions (black line) and retrieved CDF and netCDF files (red line). (Bottom) the monthly usage of SSCWeb (blue), CDAWeb (magenta), and OMNIWeb (green).*

The use of SPDF data and services has grown exponentially over years, as shown in Figure 1.2. above. Among the main services, SSCWeb is used 3-4 million times per month, and CDAWeb is used about one million times per month. The SPDF services were cited or acknowledged in about 40% of JGR Space Physics papers and 47% of Space Weather journal papers in 2025. Such yearly acknowledgement records are available online.

1.3. DISH

The mission of Data Integration Services for Heliophysics (DISH) is to fill gaps in HDRL to help better unify the solar- and space-physics data and capabilities supplied by the SDAC and SPDF and to provide capabilities which are needed by all HDRL members and should be coordinated at the HDRL level such as outreach and Open Science. DISH core capabilities may be broken down into seven areas which are detailed below. In each section we detail key projects and activities.

1.3.1. Standards Management

The standards management capability combines both the development and maintenance of Heliophysics standards which are needed for both solar- and space-physics and enable current and planned HDRL services.

A key standard being maintained is SPASE – the metadata de facto standard for describing Heliophysics data for discovery. SPASE maintenance involves the identification and incremental changes needed by NASA in order to fully and better curate the unified registered holdings of HDRL. An internal assessment of SPASE has motivated the creation of a new "version 3.0" SPASE standard to be better compliant with international standards and structures (e.g., DataCite, schema.org, DCAT) and NASA policy (Open Data, Open Science, Gold Standard Science, and SPD-41a) and to support needed capabilities which the current generation of SPASE format is unable to fully support, including solar physics metadata, metadata describing ground-based observations, terms used from FAIR vocabularies (e.g., the UAT), and improved interlinking between data, software, and publications.

DISH also supports the development of other needed standards in Heliophysics including events, curated and referenceable researcher software environments, Research Citation Objects[1] to support numerous citations to data and software in published literature, several dictionaries of Heliophysics terms including integration into the UAT (e.g., the addition of "space weather" and its subterms[2]), time space coordinates and associated transforms, designation of standard persistent identifiers for specified resource types (e.g., DataCite DOIs for datasets), service access layers and units. All of these standards support projects are done in conjunction with the international community (IHDEA and the Research Data Alliance) and HDRL partners. DISH supported Heliophysics standards projects are summarized in Table 1.3.

Resource discovery across concepts is a particularly powerful pathway to connect disparate datasets, software, and publications for increased usage by researchers. These connections are best supported by a foundation of taxonomies and dictionaries providing qualified connections between concepts and their definitions. Years of intensely collaborative work has recently attained the significant milestone of incorporating a rich organized taxonomy of nearly 200 region concepts with their connections into the Unified Astronomy Thesaurus (UAT), now providing coverage of the entire field of Heliophysics[3]. This work provides a solid foundation upon which HDRL will build concept-based connections between observational datasets, software, publications, and simulated datasets. Work continues to develop a similar structure for phenomena and parameter types and to develop definitions for these concepts in direct collaboration with the world-wide Heliophysics research community[4].

---

[1] Previously "Complex Citation Objects", https://doi.org/10.15497/RDA00130
[2] http://astrothesaurus.org/uat/2037
[3] https://github.com/astrothesaurus/UAT/issues/452
[4] The Heliophysics Semantics and Capabilities IHDEA working group, https://ihdea.net/groups.html

Table 1.3. DISH supported Heliophysics standards efforts

Standard(s)	Purpose / Need	Impact / Scope
SPASE	International metadata structure supporting discovery of data across Heliophysics (e.g., the HelioData website).	HDRL, IHDEA, SDE/OCSDO
Events	Unified events metadata for catalogs and services.	HDRL, IHDEA, SDE/OCSDO
Dictionaries and taxonomies (e.g., the UAT and Phenomena vocabulary)	Enabling topic-based discovery across services and partners.	Sci-X, HDRL, IHDEA, UAT, Center for HelioAnalytics, HelioKNOW Project, Observatoire de Paris
Software Environments	Reproducible research, publishable and citable. Open Science.	HDRL, IDHEA
Research Citation Objects	Streamlined structure to support a large number of citations to data and software in peer-reviewed science publications.	HDRL, RDA, COPDESS, DataCite
Service Access Layers	Unified search and data access (HAPI)	HDRL, IHDEA, SDE/OCSDO
Space-Time Coordinates	Unified definitions for space-time coordinates and associated transformations. Software and services.	HDRL, IDHEA
Standardized Persistent Identifiers	Designation of the standard persistent identifier to be used for a given resource type and the development of supporting workflows.	HDRL, IHDEA
Units	Unified definitions of units for helio data and other standards	HDRL, IDHEA, IVOA

1.3.2. Metadata Services and Operations

Metadata Operations at DISH provide operation of and improvements for Heliophysics discovery metadata curation, resource registration in the form of SPASE records, and related workflows and tools. The creation of reasonably rich and accurate SPASE records for HDRL-hosted resources forms the backbone of our discovery services, such as the HelioData website, and provides discovery interoperability with resources hosted by international Heliophysics repositories. The creation process critically depends on metadata collected by the SPDF and the SDAC, but must be supplemented with additional information necessary to support discovery and Open Science practices such as citation with DOIs. Current efforts heavily focus on modernizing, streamlining, and accelerating our aging workflows through collaboration with the IHDEA, NASA, and [Earth Science Informatics Partners (ESIP)] communities, development of AI tooling, and operationalization of new developments. This group also creates and maintains services and tools which aid in curation such as the SPASE metadata editor and pipelines which update the SPASE landing pages; see Section 5.3 for details on operational workflows for discovery metadata.

1.3.3. Cloud Computing

HEC (high end computing) is a shared initiative with the other HDRL members. DISH leads the cloud computing activities (on-premise HEC computing activities are described in Section 1.1.1.)

Cloud computing supports activities which leverage cloud technologies for Heliophysics research and HDRL service infrastructure. This area may be grouped into both operational and developmental activities which utilize the AWS-hosted NASA Sciencecloud.

Cloud computing operations involve supporting more than 180 Heliophysics researchers and over 30 scientific and HDRL infrastructure projects currently. Services which support researchers include the hosting and operation of an instance of the "HelioCloud" platform (see Figure 1.3.3.), the hosting and curation of a cache of over 1.2 PB of 'science ready' NASA Heliophysics mission data (see section below for more details), and supporting various researcher created projects hosted in the cloud.

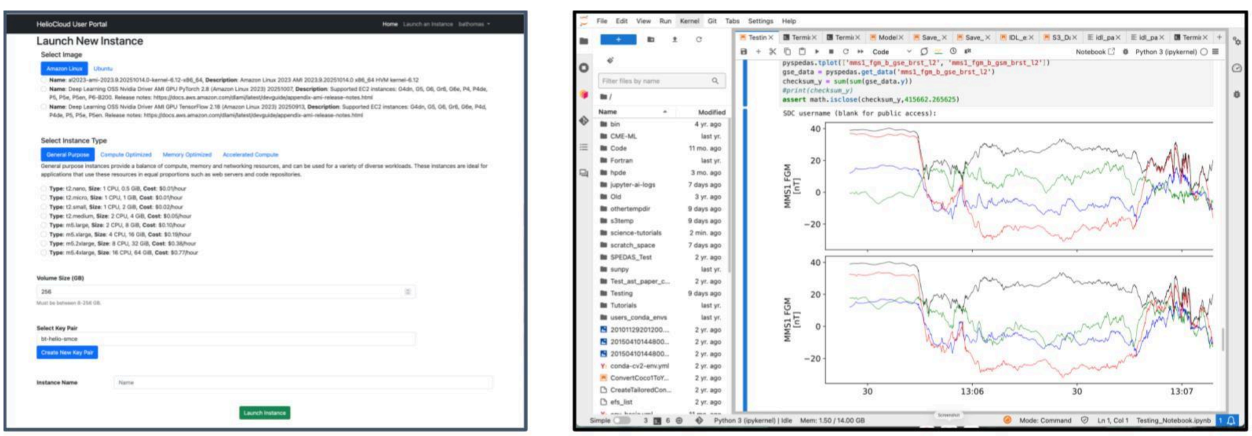


Figure 1.3.3. HelioCloud User Interface. *HelioCloud provides two browser-based user interfaces. The portal (lefthand side) allows researchers to create cloud-hosted virtual machines while the daskhub (righthand side) provides a notebooking and unix terminal environment which ships with curated Python and IDL Heliophysics software of use to researchers.*

The HelioCloud platform provides two browser-based user interfaces – the "Portal" and the "Daskhub". The Portal allows researchers to create their own remote machines which can be used to obtain significant computational resources for their science research as well as host services/resources which they may expose to their team and collaborators. The Daskhub provides a notebooking and UNIX terminal environment which facilitates team collaboration and provides substantial computational resources via 'bursting' virtual clusters. Daskhub also features 'ready to go' but customizable Python and IDL research environments which accelerate user's ability to collaborate and start work writing and using research code. Both the Portal and Daskhub capabilities may be used to readily access the substantial cloud-hosted NASA Heliophysics data and support aspects of both big data and collaborative research.

DISH also supports the development of the HelioCloud platform which is designed as an 'Open Architecture'. As such, it is open source and meant to be deployed at various Heliophysics institutions in their own AWS accounts. This simplifies the cost accounting and democratizes the access to these capabilities beyond what may be deployed at NASA. It further provides an accelerated ability for users at each institution to discover, access and utilize the data hosted by the others via the native 'cloudcatalog' software. Currently there are four deployed instances of HelioCloud (NASA/GSFC, JHUAPL, Parker Solar Probe Mission, and one planned at the CGS Drive Center). HelioCloud has also been used to support various outreach initiatives and workshops by HDRL and partners.

Finally DISH also utilizes cloud computing to host various HDRL infrastructure projects including the HelioData site and API and Helioviewer development along with supporting other HDRL activities such as SPASE resource registration and service development. Over 20 infrastructural projects are so hosted and supported currently.

1.3.4. Software Support

Table 1.3.4. DISH software project support

Software	Description	Impact
AutoPlot	Java-based science data visualization suite for Heliophysics	> 1000 users
HAPI	Heliophysics API, facilitates uniform access to Heliophysics datasets. Java and Python implementations.	> 10,000 datasets accessible
SPEDAS	IDL and Python analysis suite for Space physics	~1000 users

DISH supports software and communities which facilitate the analysis of NASA Heliophysics mission data. This support is provided as grants to organizations and members which are uniquely positioned to provide the needed capabilities. The current software project support roster is indicated in Table 1.3.4. above. This area also includes support for a "Python Heliophysics Community" or PyHC. PyHC provides a forum for community members who develop research software for Heliophysics in Python. Together these members have developed community standards and best practices for creating, maintaining, reusing and contributing to this type of software. There are over 85 member projects and over 100 active contributors in the PyHC community currently. Recently the PyHC community became an international working group in IHDEA which has further broadened the scope, engagement and value of this organization.

1.3.5. Software Service Development and Operations

DISH supports the development and maintenance of cross-cutting software services (such as APIs and websites) which help to better tie together and enable the current solar- and space-physics researchers and data and fill associated gaps. Table 1.3.5. provides a roster of the services currently supported. This area will soon include support for the creation and incorporation of Model Context Protocol (MCP)[5] service(s) to overlay existing DISH & HDRL capabilities. It is expected that this new capability will be available by the end of the current calendar year (CY26). One expected outcome of this new capability will be the creation of an 'AI mode' for discovery and tailored mission impact statistics pages on HelioData (see Figure 1.3.5. for a screenshot of the current prototype).

Table 1.3.5. DISH software services

Name	Description
HelioData	The helio.data.nasa.gov website provides a unified search and browse capability built on existing SDAC and SPDF services for solar- and space-physics datasets.
HelioData API	Provides a unified programmatic interface for unified solar- and space-physics search (https://api.Heliophysics.net/api). Supports an initiative with CCMC to allow researchers to discover relevant observational data for model runs.
HEK space physics prototype	Prototype to develop an ability to work with events in geo-mag space about the earth.
Helionauts	Supporting the hosting and maintenance of Discourse-based chat boards for Heliophysics researchers.
Paper Dataset Link API	Provides HelioData and other HDRL services access to linking metadata on which datasets are associated with research papers.
MCP dataset service (in development)	Will allow the creation of tailored mission impact statistics pages on HelioData (Figure 1.3.5) and Large Language Model (LLM)-like support for other HDRL services such as Helioviewer and VSO.

[5] Model Context Protocol : acts as a universal bridge for AI assistants and LLMs to connect to external tools, services, and databases at DISH such as the HelioData discovery API.

Another developing capability will be to interlink the indexed software found in the Heliophysics Software Search Interface (HSSI)[6] service with datasets indexed by Heliodata (more details appear in Sections 4.3 and 4.7 below).

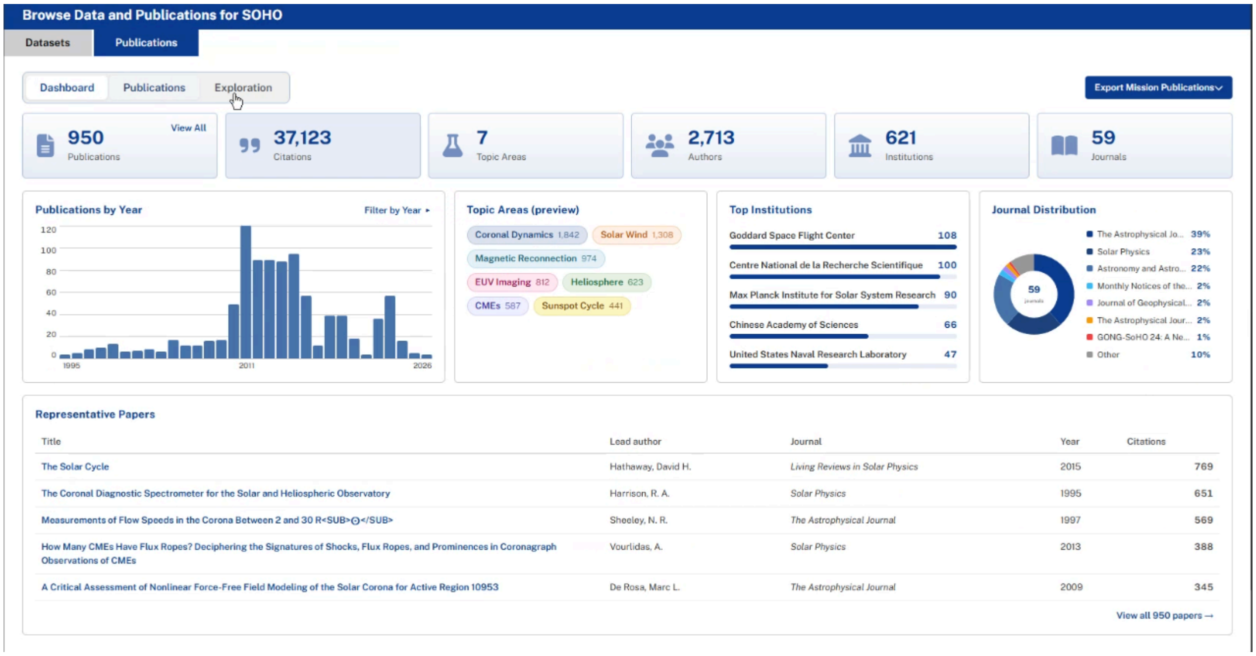


FIgure 1.3.5. *Screenshot of mission data pages on HelioData* *(in development). These pages allow for users to see publications and associated details for papers which have used NASA mission data. Note that this page represents the information contained in a sample subset of the SOHO mission papers (out of >7000 known papers).*

1.3.6. Open Science

DISH-supported efforts on Open Science seek to improve all HDRL member compliance with relevant policies in this area which include the previously mentioned Open Data, Open Science, Gold Standard Science and SPD-41a. In broad terms this area:

- Promotes and enables NASA-funded Heliophysics researchers and HDRL staff to adopt Open Science practices,
- Increases HDRL's metadata interoperability with international structures and standards,
- Leads the development of open science focused archiving policy for HDRL, and
- Prepares and strategizes infrastructure changes needed in support of Open Science practices.

[6] Heliophysics Software Search Interface, https://hssi.hsdcloud.org/

A sample of current projects in these areas with the contributing organizations are shown in Table 1.3.6. below.

Table 1.3.6. DISH Selected Open Science Projects

Project	Reason	Collaborators
Creating the semi-automated NASA Heliophysics Research Repository on Zenodo.	Provide support and guidance for NASA-funded Heliophysics researchers to more easily comply with SPD-41a, including related metrics and discovery.	Zenodo, AGU, AAS, scientists, RSEs
New semi-automated pipelines connecting SPASE to DataCite and Schema.org.	Improved metadata synchronization to increase discovery in internet search results and improve FAIR.	ESIP, ESA
New metadata mapping for datasets for [DataCite's API]().	Repair currently broken interoperability between DataCite's and Schema.org's metadata which negatively impacts discovery and attribution.	ESIP, DataCite
Assessing and improving the citation pipeline for HDRL.	Citations to data and software in peer-reviewed publications are not tracked, hampering impact measurement in Heliophysics.	Sci-X, COPDESS, CUA
Developing Research Citation Objects.	Contribute to RDA-led work improving support for large numbers of data and software citations in peer-reviewed publications, including Heliophysics.	Research Data Alliance
Streamlining science software metadata pipelines and citation availability.	Coordinating to streamline DOI creation for software to improve software discovery and citation accuracy, including for Heliophysics.	CodeMeta, Zenodo, Research Software Alliance, HSSI
Collaboratively publishing Open Science practices for missions.	NASA-funded Heliophysics missions have difficulty seeing opportunities to adopt Open Science practices.	Mission representatives
Codeveloping curated science platforms.	NASA-funded Heliophysics researchers lack resources to make their research transparent and computationally reproducible.	IHDEA
Increasing accessibility of DOIs in HDRL's discovery and data access services.	HDRL users have difficulty finding DOIs for the resources used, hampering proper citation of data and software in the literature.	VSO, ISTP, HAPI, SPDF, SDAC
Initiating the HDRL Strategy Meeting series.	HDRL Senior Management needs a way to collect input and drive collaboration.	SPDF, SDAC, HAPI, VSO

1.3.7. Outreach and Community Engagement

DISH supports outreach and engagement activities which both promote HDRL capabilities and Heliophysics research with NASA data and understand their use by the community. Activities in this area involve both the communication outwards of existing HDRL capabilities as well as gathering feedback from the community. Outreach also includes the support of targeted workshops which help the research community better understand and utilize HDRL-supported research capabilities.

Activities which involve the communication out and feedback on HDRL capabilities currently centers on participating in and supporting various meetings. Recent science meetings include AGU, AMS, TESS, DASH, IHDEA, GEM, SHINE and CEDAR. Meeting engagement may involve one or more of the following activities: hosting an information table/exhibit on one or more HDRL capabilities, 'office hours' for HDRL capabilities, posters, invited and submitted oral presentations and co-hosting agency town halls to engage the community. Non-science meeting engagements include meetings with technical stakeholders from other institutions at both national and international levels; DISH and SDAC support the DASH and IHDEA meetings in this regard. Finally, example targeted workshops DISH has supported include the NASA Heliophysics Summer School, the PyHC Summer School and the upcoming Surya foundation model workshop.

Upcoming work for outreach involves the development of "HDRL Working Groups" to collaboratively improve HDRL's services and datasets with the Heliophysics research community.

2. Engagement with NASA and External Partners

2.1. SDAC

SDAC actively collaborates with NASA and other national and international partners extensively to develop and implement community-wide capabilities (Table 2.1).

Table 2.1 SDAC partnerships and collaborations

Partner(s)	Nature of collaboration
All in-development and operating NASA solar physics missions and their associated institutions.	Development, implementation, and maintenance of data transfer and discoverability mechanisms. Solarsoft development. Helioviewer data visualization. Data search via VSO.
CCMC & MSFC	Provide space weather relevant products and model outputs via Helioviewer clients.
	Development of Surya foundation model derived products predicting the structure of the solar atmosphere (Sujit et al 2025).
NSF/NSO, Georgia State University	Design, development, implementation and maintenance of the VSO.
NOAA SWPC	Installation and operation of Helioviewer Project services.
Harvard & Smithsonian Center for Astrophysics, LMSAL, MLSO/HAO, ROB, SwRI, ESAC, NOAA, NRL	Provision of images for the Helioviewer Project.
ESA, IAS, ROB	Maintenance and development of the Helioviewer Project.
LMSAL	Development and maintenance of Solarsoft and the HEK. Development of a cross-Heliophysics feature/event standard.
Moon to Mars Office, Space Weather Program, Artemis Program, ESA	Maintenance of operational SOHO and STEREO science centers to provide low-latency data for space weather.
University of Oslo	Development of the SOLARNET standard.
IHDEA, IVOA	Standards development incorporating Heliophysics-specific improvements for metadata interoperability.
Inter-agency Heliophysics Data Working Group (IHDWG), with NSF, NOAA, USGS	US Interagency working group for standards and services in Heliophysics data environments.

2.2. SPDF

SPDF has been collaborating with NASA and other national or international partners extensively.

- SPDF has been working with NASA Heliophysics missions closely to ensure timely, accurate, and complete data archiving for their data and documentation, by designating curation scientists and meeting with mission teams regularly. For newer missions, the communications are established as early as Phase B when the missions are developing the mission data management plans. SPDF also provides user feedback to the data providers to improve the data quality and service.
- SPDF has archived data relevant to NASA Heliophysics Science Objectives from sources outside of NASA Heliophysics missions, for example, (1) data from planetary missions of NASA such as MAVEN and New Horizons, (2) from other US agencies such as GOES, SOLAR-1 data from NOAA, DMSP from DOD, data from ground-based magnetometers, aurora cameras, radars funded by NSF or other agencies. ESA, JAXA, and KASI have also asked for their data to be included, to take advantage of the many SPDF data services and access methods.
- CCMC has used SPDF's SSC database extensively for spacecraft locations, used OMNI and Wind near-real-time data, and many other science data from SPDF for their model input or validation.
- Some other data services, for instance, Autoplot and AMDA, use the CDAWeb database because the database is the largest non-solar database in the world and the data served by CDAWeb have passed the ISTP metadata compliance checks and undergone some quality control.
- SPDF has engaged worldwide users by providing timely responses to all kinds of questions, giving tutorials in summer schools, etc. Many users, including missions, have used SPDF's orbit viewers to find spacecraft conjunctions for mission planning and conduct space weather forecasting and research.
- SPDF actively solicits feedback from the community in many ways. Our science and technical staff actively participate in and present at domestic and international science meetings (DASH/IHDEA, AGU, etc.), both to advertise our data and services in presentations and for direct contact with Heliophysics users and data providers. SPDF scientists interact directly in meetings and telecons with missions/projects, PIs, project scientists and in project science working groups. We interact to understand and meet specific project requirements and to explain how SPDF capabilities can simultaneously better enable mission science while improving community cross-disciplinary data access and long-term archiving of appropriate science products. SPDF actively solicits direct feedback and questions to our data system curators and discipline scientists from researchers and other users to SPDF.

2.3. DISH

The nature of DISH's work critically depends on collaboration with a large variety of organizations. These are shown in Table 2.3. below.

Table 2.3. DISH key interfaces and partnerships

Partner(s)	Nature of collaboration
RDA, Research Software Alliance (ReSA), DataCite, Zenodo, Sci-X, COPDESS	To measure and increase the efficiency of citation tracking for data and software.
CodeMeta, LASP, ReSA, DataCite, Zenodo	Improved citation information for software and increase the interlinking of software with data and publications.
DataCite, ESIP, CodeMeta, Metadata Game Changers	Metadata interoperability with international systems.
IHDEA, UAT, IVOA	Heliophysics-specific improvements for metadata interoperability.
OCSDO, HSD missions, research software engineers, publishers	Open science guidelines for Heliophysics.
IHDEA, Madrigal, George Mason University, JHU APL, NASA GSFC 619 (mEditor team), ESIP	Modernize SPASE workflows, tooling and pipelines.
OCSDO	Improvements for Heliophysics science research on the Science Discovery Engine.
IHDEA, JHU APL, ESAC, ESA DataLabs, 2i2c, Zenodo, SMD partner repositories	Improve interoperability of DISH services and associated standards.
Inter-agency Heliophysics Data Working Group (IHDWG), with NSF, NOAA, USGS	US Interagency working group for standards and services in Heliophysics data environments.
CCMC	Facilitate model verification via comparison with observational data.

3. Current State of HDRL Data Holdings

3.1. HDRL Metrics

HDRL's data holdings are distributed between the on-premise infrastructures of the SPDF and SDAC with copies of mission data hosted on HSD's HelioCloud alongside user-uploaded research and non-NASA mission datasets. A previous effort led by the OCSDO generated a metrics report for fiscal year 2024 for all NASA science data repositories with many of the requested metrics (Murphy 2026). Relevant metrics extracted from that report are included in Table 3.1 below. Updated metrics for fiscal years 2025 and 2026 are not yet available, but we expect these numbers to increase. The current status of HDRL's compliance with the desirable characteristics relevant to metadata and data documentation is either "Fully Met" or "Partially Met" for all three HDRL members as detailed in row 4 of Ringuette et al. (2026). See Section 4.1. for details.

Table 3.1. Selected Fiscal Year 2024 Metrics

	Total	On-Premise	Cloud
Number of Datasets	11,500	~8,000	3,500
Dataset Volume (TB)	3,151	3,085	1,321
Growth (TB/yr)	470	370	350
2030 Projected Volume (TB)	35,000	35,000	2,000
Unique User IPs	~8,000,000	~8,000,000	–
Data Download Volume (TB/yr)	–	~750	–
Datasets with API Access	–	3,000	3,500

*Access metrics on the cloud were not yet available for FY 2024. Development of this capability is ongoing with results planned for inclusion in FY26 metrics. Total metrics for the last two rows were not included in the cited report. See Murphy (2026) for details.

3.2. SDAC

The SDAC ingests data regularly and frequently from all operating solar physics missions - SOHO, STEREO, Hinode, IRIS, Parker Solar Probe, Solar Orbiter, PUNCH, and PADRE. The SDAC also holds data from previous missions, yielding a total volume of about 2.4 PB. Data volumes at the SDAC increase by tens of TB a year. The SDAC also acts as a data distribution site for SDO data, holding a rolling, separate 1.25PB cache of SDO data. Larger volumes of SDO data are not yet held, however about 140 TB of SDO Level 0 data has been transferred to the SDAC since November 2025.

SDAC data holdings are copied to IMCC (Integrated Modeling Computing Center, formerly NCCS) and on to tape. Almost all the data is in the FITS format. Use of FITS confers many advantages to NASA's Heliophysics community. The format is used internationally, is free and open source, has a long heritage, and has a wide variety of free and open source tools available to read and write FITS files. Most operating missions (except PUNCH) use the Solarsoft library, housed at the SDAC, to store and distribute mission-specific data preparation and analysis software. Usage of the software is explained via mission-developed data analysis guides (in multiple formats) and online tools that are hosted at mission/instrument team institution websites. Missions are typically not required to deliver final data products and documentation to the SDAC until a mission ends. Therefore, final data products and associated documentation are typically described in a mission's final data delivery plan.

SDAC captures some information concerning data usage via the VSO API. Statistics for 2025 show over five million VSO requests, with Python-based clients outnumbering IDL-based clients by a factor four.

3.3. SPDF

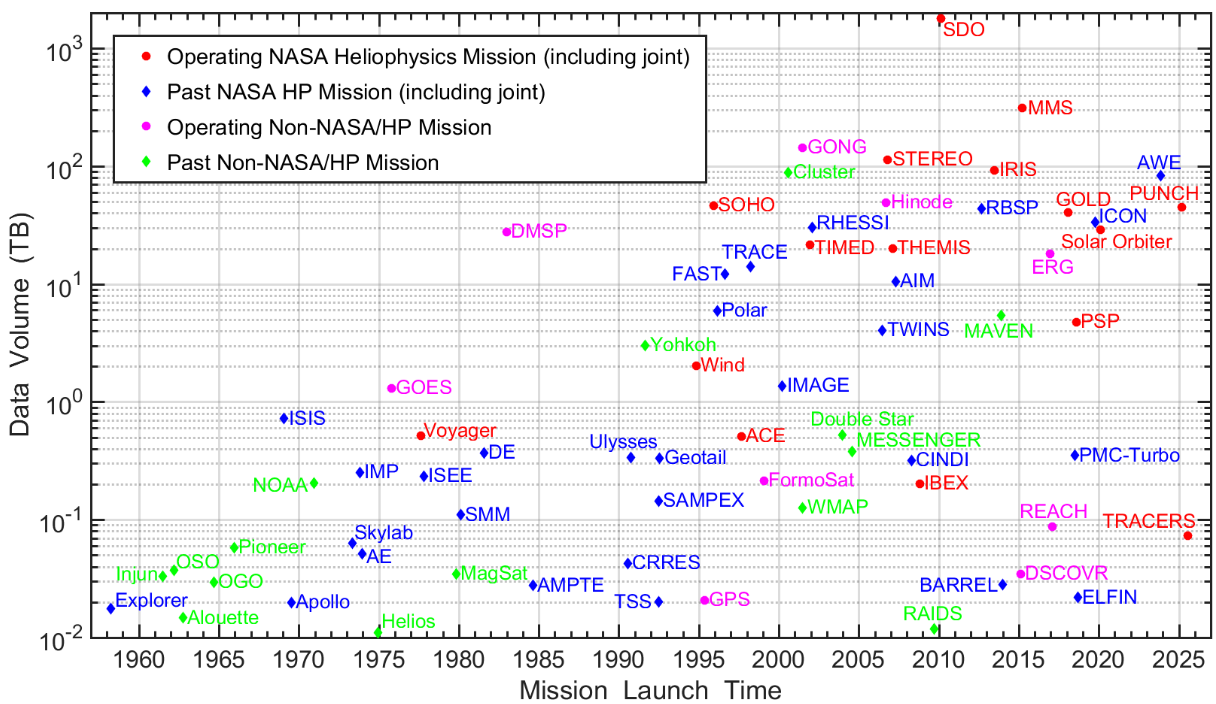


***Figure 3.3**. **Data volume of major missions archived at SPDF and SDAC** versus their launch time. Only missions with more than 10 GB of data are plotted. The expected total data volumes of SDO and DKIST (NSF project) are 40 and 200 PB, respectively, too large to be included in the plot. The missions are sorted into four groups: red (blue) for operating (past) NASA Heliophysics missions including joint ones with other agencies, magenta (green) for operating (past) non-NASA-Heliophysics missions.*

SPDF ingests about 15 TB of data and 0.85 million files per month, with some fraction replacing older-version data. At present, SPDF archives about 110 million files and 8000 datasets, 3000 of which are CDF datasets. The data archive at SPDF increased from about 400 TB to about 800 TB in five years, so may reach about 1.6 PB in the next five years. The SPDF data is also copied to IMCC for on-premises high-performance computing, and science data is copied to HelioCloud for data analysis and computing in the cloud.

Figure 3.3 shows the current data volume of major missions archived at SPDF and SDAC. Among them, four missions (SOHO, STEREO, PSP, Solar Orbiter) provide both remote-sensing solar data and in-situ space physics data, so their data are archived at both SPDF and SDAC. Ten missions are only archived at SDAC including 1.7 PB from SDO. The other 56 missions have data archived at SPDF. The MMS mission has the largest volume of data (~40 TB) and most data files (~20 million). The data volume from recent missions are generally larger than earlier missions as the mission instruments advance with time and some of the recent missions include multiple probes (e.g., Cluster, THEMIS, MMS).

In addition to science data, SPDF has archived about 300 orbit datasets in SSC. About 350 plot types and 17.5 million pre-generated plots from about 30 missions are stored alongside the digital data of corresponding missions. These are typically summary or quick-look plots and served at Plot Walk, and helpful for checking the space environment encountered by the missions. Documentation in various formats are also archived, including processing and analysis source code.

3.4. DISH

DISH holdings include the SPASE records which describe both space- and solar-physics HDRL datasets (and associated metadata) and a cloud-based “analysis cache” of high level data.

The SPASE records have associated landing pages and a significant number are registered with DOIs (see the DataCite Metadata Dashboard, currently 3076 DOIs). Registration and curation of these records is an ongoing activity which includes addressing a sizable backlog of older datasets.

The cloud “analysis cache” holdings are copies of data available from both SDAC and SPDF and come to 1.2 PB. This cache is limited by space and data availability. Some NASA solar mission data which were not at SDAC are not yet available (ex STEREO). Furthermore SDO data products exceed the resourcing available and so a reduced set of holdings is present. Some solar data and service catalogs, such as for the HEK are being added with a limit of 1.5 PB being available. The current cache limit is set by an agreement with AWS which defrays the cost of cloud storage. Additional data product storage in the cloud would have to be paid for by NASA. Depending on the available resourcing for this capability, the cache may be expanded to hold a greater number of solar mission data products including an expanded SDO dataset.

4. Metadata, Discoverability, and Interoperability

4.1. How HDRL Ensures Compliance with SPD-41a

HDRL participated in the 2024 SMD Data Repository Workshop to develop a list of desirable characteristics for NASA data archives in response to the direction in NASA SPD-41a (Ringuette, Downs & Morgan 2024). Table 4.1. shows a summary of HDRL's degree of alignment with these characteristics (more detail can be found at Ringuette et al. 2026). Fifteen of the twenty-two characteristics are fully met for all three members of HDRL, demonstrating the high quality of HDRL members' service and compliance. The "Public Review" characteristic is "Partially Met" for all three HDRL members due to a need for increased community feedback mechanisms. The single characteristic that is not met for all three is the "Certified" characteristic from Appendix D of SPD-41a. The certification process is time and resource intensive and is beyond the scope of HDRL's current resources.

Table 4.1 Summary of Desirable Characteristics for HDRL

	SDAC	SPDF	DISH
Fully Met	17	18	19
Partially Met	4	3	2
Not Met	1	1	1

4.2. Assessment of the Metadata Quality and Completeness for Data at HDRL

The creation of high quality and complete metadata critically depends on the automated and manual curation efforts of the SPDF and the SDAC. Those metadata are supplemented by DISH with additional information necessary to support discovery and Open Science practices such as citation with DOIs. As of the end of May 2026, 2881 datasets have DOIs out of the roughly 8000 datasets archived by HDRL[7,8].

No standard tool exists to measure metadata quality and completeness, although a wide variety of available tools measure FAIR[9]. No tool exists to measure FAIRness for SPASE, ISTP, or FITS metadata, so we instead selected from currently available tools to objectively assess our metadata in the Schema.org and DataCite formats. Multiple FAIR assessments of our entire collection of DataCite metadata were completed using Habermann (2025), a tool used in a series of blog posts describing, calculating, and comparing FAIR scores across all DataCite repositories (e.g., Habermann & Robinson, 2024). FAIR assessments of the Schema.org

[7] https://science.data.nasa.gov/about/repository-metrics
[8] https://metadata.datacite.org/heliophy.spdf
[9] E.g., https://catalogue.fair-impact.eu/resources

metadata for an example dataset were conducted at two times using the F-UJI tool (Devaraju & Huber, 2023).

HDRL's DataCite repository was included in an external FAIR analysis in May 2025, when our previous pipeline was in place (Haberman & Robinson, 2025), and then again in January and May 2026 after our new pipeline (Boquet et al. 2026) was implemented in operations (Habermann, 2026a, Habermann, 2026b). The observed improvement was a significant increase in the FAIR score to 37%, which now places us as being the most compliant among measured SMD repositories (Figure 4.2).

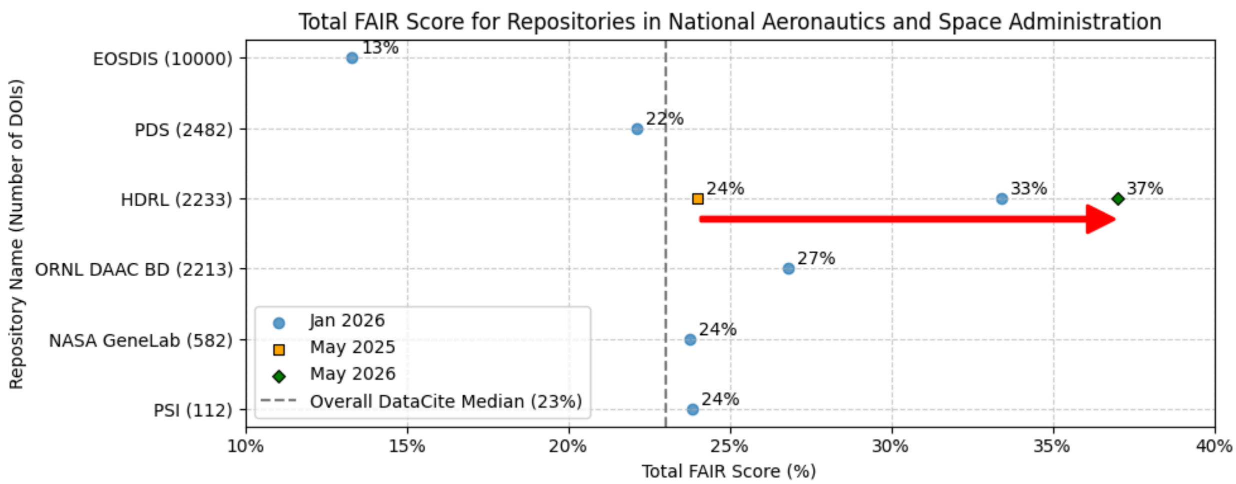


Figure 4.2. Total Score Distribution by Repository ID *in the NASA DataCite repository. Repositories are shown on the vertical axis with the calculated FAIR scores plotted horizontally. The dashed vertical line indicates the median FAIR score of 23% across all NASA DataCite repositories.* ***The FAIR score for all HDRL data repositories is the highest*** *among all repositories as shown by the red arrow. Plot from (Habermann, 2026b).*

We have implemented a similarly improved pipeline for the metadata embedded in our landing pages in support of increased discovery by internet search engines and generic AI tools. Using the F-UJI tool as an external measurement of FAIR for these metadata, we found an increase of about 30% in our FAIR score for an example dataset due to changes made across all datasets (from 58% as measured in March 2025 to nearly 90% measured in May 2026).

4.3. Description of Discoverability Mechanisms and Interoperability Standards

HDRL supports discoverability through a rich spectrum of services. Alignment of those metadata and data services with existing, widely used standards increases the interoperability and effectiveness of those services for the international science community.

At the most general level, HDRL enables discoverability by generic internet search engines and similarly generic AI searches by embedding rich, valid, structured data into the landing pages of each dataset, which are aligned with both the relevant ESIP standard and schema.org's requirements. This approach ensures metadata interoperability between Heliophysics, Earth Science, and other sciences in support of recent Solar and Space Physics Decadal Survey Recommendation 5-12. For Heliophysics-specific searches, HDRL provides discoverability through the accessibility-compliant HelioData website and API or a growing selection of our datasets and missions with SPASE records, with the complete index of SPASE records supported by the HDP. Alignment of these metadata with SPASE as the current de facto standard across Heliophysics internationally provides interoperability in support of discovery across the diverse areas of Heliophysics. In support of Recommendation 5-10, HDRL is additionally collaborating with external groups to create a pipeline from HAPI servers into SPASE to add the several thousand ground-based datasets described there into these discoverability services.

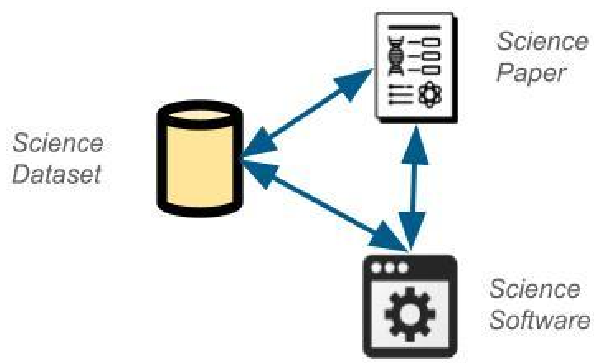


Figure 4.3. Interlinked research artifacts. *Science papers, software and datasets are research artifacts which are being interlinked by HDRL. Finding any one of these artifacts will lead to easy discovery of the others and accelerate the research process.*

Modern research no longer solely relies on access to data and the related analyses in publications, but additionally increasingly depends on software requiring connections between these resource types (Figure 4.3). HDRL's discovery services will soon provide connections from data to related publications and software to decrease the time it takes for researchers to get to the science in support of the recommendations of the 2024 Decadal Survey (e.g., Recommendation 5-15). These connections will be possible using a recently developed AI curation pipeline for publications developed through a collaboration with the Science Explorer and the growing index of software available through the NASA-funded HSSI. In the future, increased connections between these critical resource types will be made possible by continued collaboration with the Science Explorer and HSSI and a completely redesigned SPASE metadata system, called SPASE 3.0, which will provide increased support for FAIR and discovery through a purposefully interoperable community-driven design (Thomas & Ringuette 2024, Koval et al. 2025).

VSO uses the internationally established SOAP and TAP protocols to access and search for data. All data available via the VSO is completely free and open for use. The VSO, originally written in 2002 in Perl, is currently being completely rewritten in Python, and is moving to the Model-View-Controller model of application development. The website is also being completely overhauled. VSO 2.0 will move to using the REST API model. The redevelopment takes advantage of modern programming paradigms for web applications, and will make the VSO easier to maintain and develop.

4.4. Adherence to Open-Science Practices

NASA's Open Science website indicates that the four main components of Open Science are Collaboration, Reproducibility, Availability/FAIR, and Transparency (or CRAFT). All data available at HDRL (and via the VSO) is free from any restriction on use.

Collaboration: Wherever possible, HDRL staff openly collaborate with international groups in related sciences, international cross-science groups, other Heliophysics repositories, and other groups within NASA to improve data infrastructure. HDRL also supports the practice of collaboration in the research community through standing up and managing instances of HSD's HelioCloud for community-led workshops by request, such as the NASA Heliophysics Summer School and PyHC's Summer School, and by hosting the research of NASA-funded Heliophysics researchers, resulting in a growing number of peer-reviewed publications.

Reproducibility*:* HDRL is working to reduce the amount of manual labor and resulting metadata inconsistencies through building updated, automatable, and reproducible pipelines to accelerate curation processes, including the ongoing adoption of NASA-developed mEditor, a content management system for metadata. HDRL supports the practice of reproducibility for NASA-funded researchers through HSD's HelioCloud (Thomas et al. 2022). For broader impact, additional instances of HelioCloud are increasingly available at other institutions (e.g., APL) with interlinked data discovery services.

The VSO implements data search reproducibility via the VSO cart concept, wherein the search parameters can be referenced by a unique ID. This concept will be extended in VSO 2.0 so that both the search and results are uniquely identified and stored, and therefore reproducible. Data selections and views made via helioviewer.org are reproducible using an automatically generated shareable and unique link; JHelioviewer users can download the state of the application. Therefore, both applications enable reproducibility.

Availability/FAIR*:* HDRL support FAIR in a variety of ways, including assigning DOIs to hosted resources in support of findability, providing access to those resources through standardized communication protocols (both generic and discipline-specific), striving for increased metadata interoperability as described above, and promoting the creation of domain-relevant standards for Heliophysics through IHDEA (reusability). Work remains to include DOIs in existing data access services. Our support of these practices for the research community will increase as our Zenodo-hosted Heliophysics Research Repository comes online later this year.

The VSO holds metadata on available data sources and the FITS files at those sources. The metadata of VSO 2.0 will be made publicly available and accessible via OpenAPI-compliant APIs. This will provide a transparent accounting of VSO data connections and allow third-parties to build their own applications. Similarly, SDAC Watch metadata describing SDAC data holdings will be made publicly available when its API is released (Section 1.1.1).

Transparency: HDRL staff support transparency by presenting ongoing efforts in relevant conferences, workshops, and mission meetings, in addition to a recent HDRL Strategy Meeting designed to collectively share progress and gather input from HDRL staff and grantees. In support of researchers' adoption of transparency, HDRL ensures that dataset citations are easily available on the helio.data.nasa.gov website as well as CDAWeb Notes and supports similar efforts for software by the HSSI for use in peer-reviewed publications.

VSO has built a dashboard that publicly reports the results of a daily test of the ability of the VSO to find and download data from all its sources. These results can be used to diagnose VSO issues but also give users a snapshot view of the reachability of data sources. VSO also has a social media presence which is used to disseminate information on upgrades, planned outages, degraded service, etc. Helioviewer.org maintains a blog, accessible via the interface but running on an independent system through which service and other notifications are disseminated. Finally, the SDAC webpage is being rebuilt to give more information on data held at the SDAC, help on data analysis, and other tools and services targeted to Heliophysics research.

SPDF provides the mission data inventory plots, statistics of data ingestion and usage to the public. Besides giving reports in major meetings such as DASH/IHDEA and AGU, SPDF sends out the notices of system maintenance, data releases, CDF upgrades regularly in the homepages of SPDF, CDAWeb, SSCWeb, and other associated websites. For important changes (e.g., service outages, new OMNI data products), we notify the worldwide Heliophysics community through SPA newsletters and SPDF emails which people can subscribe to.

4.5. FITS

The vast majority of data produced by NASA solar missions are written in the FITS format. FITS is an open and international data standard for astronomical image data that has been endorsed by NASA and the International Astronomical Union. The standard, and tools to read and write FITS files are all completely free and open source. Tools to read and write FITS files are available in the two primary data analysis libraries, Solarsoft and SunPy (2020). The IAU FITS working group maintains the FITS standard. By using the FITS standard, NASA's solar physics community leverages a well known and broadly supported format with a long heritage and broad knowledge base. SDAC has implemented the SDAC Watch project to measure FITS file compliance for every file held at the SDAC. FITS file compliance results will be passed back to the missions in the next six months and made publicly available. The SDAC Watch API, although not yet publicly available, will follow the OpenAPI standard.

The SOLARNET project was an European-led project to better describe solar physics data within the FITS format. Its further development has been adopted by IHDEA. SDAC staff participate in its further development and are working on testing existing FITS files against the SOLARNET standard to assess existing compliance and identify gaps in the SOLARNET standard.

4.6. ISTP Metadata

ISTP Metadata Guidelines were originally developed in the 1990s as part of the ISTP science initiative. It is now adopted by most of the Heliophysics missions and supported by numerous display and analysis tools. The ISTP Metadata Guidelines are like a blueprint of a house before adding the data values. A dataset using ISTP metadata by definition forms a logically complete and self-sufficient whole (data and descriptions).

SPDF leverages the standardized self-describing data formats (CDF, netCDF), the global attributes and variable attributes for datasets, time conventions, and filenaming conventions. The markdown version of the ISTP Metadata Guidelines is on GitHub, with non-ISTP global and variable attributes used by missions listed in the appendices. We will release it in version-controlled form soon with a DOI on Zenodo. Upcoming activities of development include:

- Formalize a steering committee to define governance and update processes
- Improve support for FAIR principles and compatibility with SPASE data model
- Add a standard syntax for unit of measurements (VOUnits recommended by IHDEA)
- Add standard descriptions for tensors, including vectors and components (external vocabularies)
- Add standard descriptions of multi-dimensional data with dependent dimensions, modeled data (multi-dimensional data on irregular grid), waveform bursts data, and quality/status flags
- Add simplified descriptions of multi-dimensional data independent of plotting choices.

4.7. DISH Discoverability Mechanisms

In addition to the HelioData website/API and the paper linking efforts mentioned above, DISH supports discoverability of cloud-based data using the “cloudcatalog” software and standard (Antunes et al. 2025). Cloudcatalog is built on top of HAPI and is used to find and retrieve scientific data which are stored in designated HelioCloud S3 buckets. Over 1.2 PB of cloud-hosted science data are currently discoverable using cloudcatalog.

5. Processes for Data Ingest and Automation

5.1. SDAC

The SDAC typically establishes contact with a mission regarding data products at least by Phase B, and an ICD is written between the mission and the SDAC establishing the basis for an ongoing relationship. The ICD comes into effect in Phase E when observational data becomes available. SDAC pulls the agreed-upon data from remote institutes using secure, encrypted mechanisms, and makes that data available via the VSO and for direct download. Figure 1.1.2.1 shows how data arrives to the user. The SDAC's historical role was an "active archive" and the *secondary* location for the community to obtain data, a mission institute being considered as the primary location. This relation to the missions has changed recently, with PUNCH, PADRE, and CODEX data being publicly available only from SDAC.

The primary responsibility to provide FITS files that comply with the FITS standard and accurately represent the observation lies with the mission. Inspection by the scientific community typically surfaces compliance and scientific errors in the data. Compliance with the FITS standard is not required for the SDAC to store a mission FITS file. The SDAC Watch application counts files held at the SDAC and tests each against the FITS standard. In the next year, SDAC will make available mission-specific reports on SDAC data holdings. This will include number and type of files per data product, data volumes, and FITS compliance assessments. This will enable SDAC and missions to identify gaps in the SDAC archive and improve the quality of every FITS file.

As a result of SPD-41a, NASA-funded researchers (such as via ROSES) are required to make any data products created as a result of their research publicly available. Small datasets can be hosted in the appropriate Zenodo community; large ones are hosted by SDAC. SDAC now has two multi-TB archives of such data available at https://umbra.nascom.nasa.gov/contributed/ . For these datasets, the proposers contacted the SDAC, and details on the transfer and accompanying documentation are settled. SDAC staff test the research-generated FITS files for compliance. Once the FITS files are compliant and transferred, the data are made available online at SDAC.

The VSO maintains and updates a registry of data sources. The Heliophysics community produces new data sources infrequently, so this process is handled manually. Once a source has been added, metadata from each FITS file from the source is extracted automatically and used by the VSO to enable data search and download via the existing VSO API and clients.

5.2. SPDF

SPDF has set up data submission guidelines, with the basic steps from data preparation to public serving displayed in Figure 5.2.1. Whenever the mission team or other data provider is ready, we check the sample data files and review them against ISTP Metadata Guidelines and in the test CDAWeb site. Only when the sample data passes our validation, the data providers are expected to produce the data files in bulk. The data provider submits a list of datasets and source locations (mission portal, HTTPS/FTPS server, API), and SPDF configures an auto-ingest script to pull updated data files every day, or the data provider pushes the data to an SPDF-designated ingest area. For a small mission or project, this could be a one-time-only submission. For most missions, this is a routine automated process, and we run over 2500 automated scripts a day looking for new data. One primary and one secondary SPDF curation scientist are assigned to each major mission to make sure the data archiving is timely, accurate, and complete.

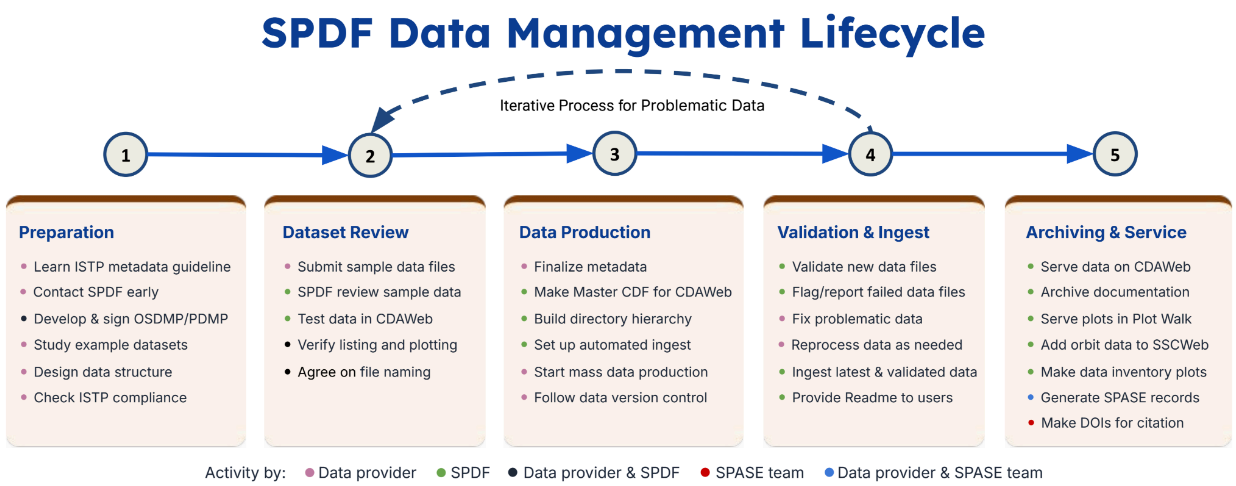


Figure 5.2.1. The SPDF data management lifecycle *from preparation to archiving and service. The purple (green) dots indicate the activities by the data provider (SPDF). The black dots mark the activities by the data provider and SPDF. The red dots indicate the activities by the SPASE team, and the blue dots mark the activities by the data provider and SPASE team.*

SPDF ingests new datasets from many missions, along with regular updates to existing datasets using automated pipeline and directory-based workflows, including conversion to CDF form as needed. For older datasets, such as from NSSDC legacy migration efforts, we restore smaller datasets under the regular SPDF budget and work with others on data upgrade proposals for bigger projects. We work closely with data providers to get the data in a usable and descriptive form. To ease this effort, we have developed tools for creating and describing the structure and metadata following both ISTP and SPASE guidelines. By using standardized metadata, CDAWeb and other services are able to automatically open and use the data with appropriate units and plot types. We recently created SPDF services FAQ and quick start guides to help. In addition to science data ingests, SSCWeb regularly ingests orbit data from mission data, NAIF SPICE kernels, and NORAD two-line elements (TLEs).

One of the most important functions SPDF provides are our data quality review and validation processes. For data files coming into the CDAWeb system, every file is opened and validated to make sure it has valid time stamps and valid variables. Invalid files are reported to the data provider and in most cases replaced by them in a timely manner. In some cases, data providers would not be aware of problematic data files if not for our validation measures. In addition, the CDAWeb and SSCWeb systems are set up to automatically notify the staff when dataset problems occur so we can correct metadata or develop software solutions as quickly as possible. Data files failing the validation against the pre-determined master metadata (often caused by variable dimension size changes, addition or removal of variables) are flagged, and the curation scientist will contact the data provider for more information or necessary fixes.

Our process of ingesting data into a test data browsing and access system ensures datasets are cleaned and tested before going public. Inspection and use by a combination of SPDF data curation scientists and researchers in the community ensures that data are usable and reliable, while there is still time to get updates from original data providers. This is similar to the many eyes approach of the best open source projects.

Although time consuming, upfront interactions (even before launch) with data providers makes for a much better product, especially for long term archiving, since we provide an external and knowledgeable review with an eye to the product’s use by future non-experts and describing datasets so they are independently useful and complete for the long-term.

Nearly as important to long-term archiving is sufficient and high quality metadata and documentation. We work with data providers to collect and review this information and make it accessible within the data files and on the HTTPS site. Where needed, we can add additional metadata for a dataset to a Master CDF that overrides and improves on the metadata in the original data files, and add links to the documentation and webpages of mission and instrument teams.

New or revised SPASE records are pulled from the Github repos into the Heliophysics Data Portal to provide search functionality for datasets and other Heliophysics resources, in support of the SPASE effort discussed below.

5.3. DISH

DISH undertakes metadata operations activities in the following areas: Heliophysics resource registration for discovery, linking HDRL resources to other external resources, curation of a cloud-hosted analysis cache and the registration and curation of NASA-funded researcher contributed datasets and related resources (late CY26).

5.3.1. Heliophysics Resource Registration

Both space- and solar-physics resources may be registered after ingestion into the respective HDRL archives is accomplished. Registration of resources is done using SPASE metadata and minting of associated DOIs (see Figure 5.3.1). Resources which are registered using DOIs and SPASE include scientific mission datasets and quicklook/visualization data products along with their associated resources such as mission, instrument, and researcher metadata. Registration is handled by the SMWT - SPASE Metadata Working Team which is led by DISH. Both SPDF and SDAC also contribute some curation support to the SMWT.

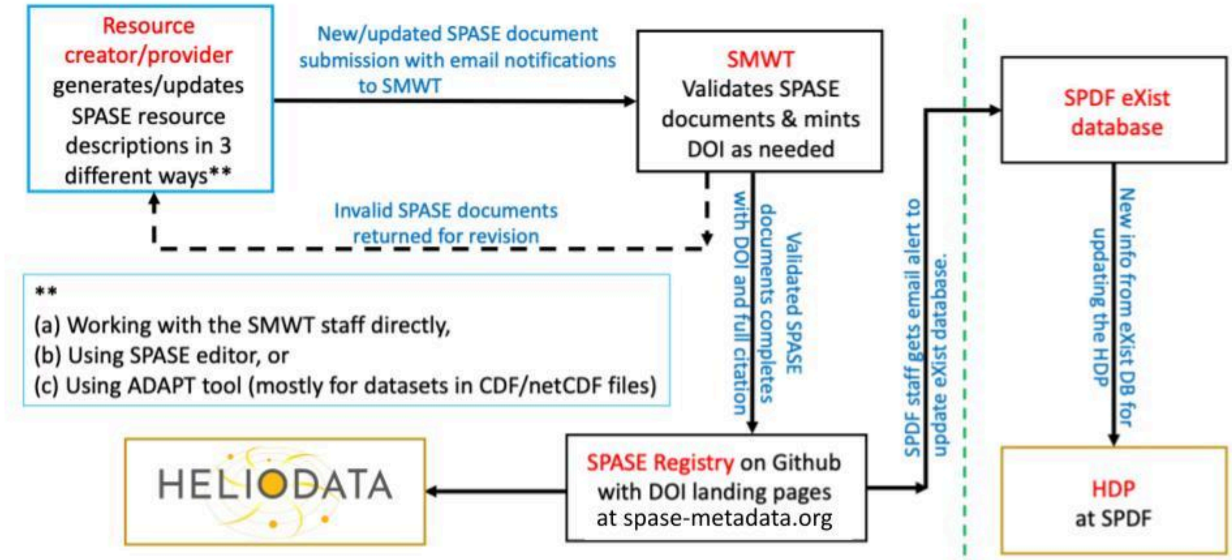


Figure 5.3.1. SPASE registration workflow diagram*. SMWT - SPASE Metadata Working Team curates SPASE records using the workflow shown. SPASE metadata support three public facing resources including SPASE landing pages, the Heliophysics Data Portal (HDP) and HelioData. SPDF staff support the updating and maintenance of the HDP.*

Several projects are led by DISH staff to actively improve the registration workflows. Projects in this area include the adoption of a new Content Management System (CMS) based on "MEditor"[10] which will streamline and further automate the curation workflow shown in Figure 5.3.1. Several development projects involve improving the creation of SPASE records. These include a new modernized pipeline for drafting SPASE records from ISTP metadata, a new automated pipeline to create SPASE records for all datasets served through HAPI (Weigel et al. 2021), and a including the upcoming NASA-funded HAPI server for Madrigal data. This directly supports recommendations from the recent Solar and Space Physics Decadal Survey (National Academies of Sciences, Engineering, and Medicine 2025) to enable research combining space and ground-based datasets. Another development project includes the creation of a "SPASE Metadata Extractor" based on the successful HSSI Metadata Extractor to further accelerate drafting, curation and validation of all SPASE records using AI.

5.3.2. Linking Datasets to Related Resources

This area of metadata operations involves the linking of HDRL datasets with other research artifacts (see Figure 4.3.). A service developed and hosted by DISH connects HDRL datasets to

[10] Adapted with help from Goddard Earth Sciences Data and Information Services Center (GES DISC)

publications utilizing various advanced Natural Language Processing (NLP) and Machine Learning (ML) technologies (including LLMs) and has recently been operationalized for making connections between HDRL datasets and Heliophysics peer-reviewed publications. As a result, there are now searchable links between these resource types at Sci-X and HDRL. The current operational effort is to further automate and scale up the process of linking datasets to papers with the expectation that over 50,000 paper-dataset links will be registered by the end of the current calendar year. In CY27 we expect to transition the discovery of papers by associated datasets from the current service to HelioData, the VSO, and Helioviewer using a new paper linking API.

DISH is also working to interlink software with NASA datasets. A NASA-funded project led by the HSSI and supported by DISH is using AI capabilities to extract connections between software, the associated peer-reviewed publications, and HDRL datasets from GitHub repositories. A sample of connections between these resources is now searchable on the HSSI website and is planned to be available on the HelioData website and Sci-X. Once recently discovered gaps in the citation pipeline are resolved (Stall et al. 2025), further connections between data, publications, and software will be sourced from citations in peer-reviewed literature. Additional support for interlinking between these resource types is planned for SPASE 3.0, the next generation SPASE model.

5.3.3. Researcher Contributed Resource Ingest

DISH will soon be launching the Heliophysics Research Repository as an archive to host NASA funded Heliophysics research artifacts (datasets and software). This repository is being constructed as a specialized Zenodo community which features automated metadata curation for required metadata fields. DISH will utilize these metadata to list these research datasets, software, or publications on the HelioData website as related resources alongside mission data. Irreducible datasets too large for submission (>200 GB) may be supported through the SPDF and the SDAC. Coordination with NASA HPD to make this resource known to the community and to integrate it into Heliophysics funded opportunities is a FY27 objective.

5.3.4. Cloud Cache

DISH supports a cache of HDRL and related scientific data in the cloud. The current cache is hosted in AWS S3 buckets and comprises both space- and solar-physics data along with some contributed AI/ML datasets. The current volume of data is 1.2 PB with an upper limit of 1.5 PB being available. These limits are significantly less than the total volume of available HDRL data and only ‘high level’ datasets (typically level 1 or higher for solar and level 2 or higher for space physics) appear in the cache. All high level datasets resident at the SPDF and SDAC are present in the cache with the exception of SDO for which a specialized dataset was produced for the AIA and HMI instruments (only). This SDO dataset has reduced time resolution and time span compared to the full AIA and HMI datasets.

The cloud cache is maintained currently in a heterogeneous fashion. In some cases it is automated via a periodic pipeline (high-level SPDF datasets) which pulls new data, in other cases the archive uploads the data in coordination with cloud operations (high-level SDAC datasets), in other cases cloud operations obtains and uploads the data (HEK event list, SDO EVE data). Regularizing the updating of the cache is an upcoming objective for the team. In all of these cases once the data are uploaded they are registered using the cloudcatalog software and are thereby made discoverable. As the cache space is limited, logging of the use of these data is being performed and will be utilized for analysis to determine which datasets are not being sufficiently utilized in order that they may be deleted in favor of new data (CY27 objective).

6. Challenges to Core Mission

Change is a constant and each year HDRL must adapt to handle the associated challenges to its core mission (see Section 1 for definition). Significant challenges are broken out thematically below.

6.1. Data Related Challenges

First, and obviously, HDRL must handle new sources of mission data, including smallsats, rockets and balloons. The associated curation and maintenance of an ever expanding number and volume of archived resources for its use by stakeholders is a related challenge. Specifically

- **Large volumes of SDO data are not present at the SDAC**. The SDO data resides at the SDO JSOC, and is about 28 PB at present. The data must be transferred and stored at the SDAC. Further, a new data management system needs to be developed to replace the JSOC's homegrown data management system.

- **Missions design mission-specific data systems and services for their operational period, not for the post-operational period.** These systems and services cannot be run indefinitely when the responsibility for data curation and access is transferred to HDRL. Mission-developed tools and documentation are often closely related to data analysis but are heterogeneous in format and are typically scattered across multiple mission and instrument websites.

- **Different agencies have different metadata standards**. For example, NOAA's SOLAR-1 mission uses the ISO geographic metadata standard and Climate and Forecast (CF) metadata conventions. To serve its data at CDAWeb, SPDF staff have to convert the metadata following the ISTP guideline and change the setting of the time variables. Tools are needed to ease the data format conversion and enable the interoperability of data in different services.

- **Significant curation backlog.** Approximately 50% of the preserved datasets at HDRL are fully described and integrated into the system. There is a large backlog of older datasets (going back to the early 60's) which need significant curation work in order to be brought up to standards.

- **Communicating data curation status.** The data validation and ingest pipeline could be more visible to the data providers after the data submission, therefore SPDF is developing a way to systematically record and update the data ingest status.

6.2. Analysis Related Challenges

Evolving techniques and methodologies for doing science are another source of change. Analysis no longer amounts to only single instrument science and HDRL must enable system science with multiple sources of data including big data analysis which requires substantial computational capabilities next to the data. Another example includes the application of AI/ML (e.g. Surya foundational model, Sujit et al. 2025). Specific challenges include:

- **Increasing complexity for data analysis.** With the advancement of space instrumentation and time-dependent operating modes, more complicated datasets including multiple dimensions and dynamic dimension sizes are being made making it difficult to illustrate the relationship among these variables using a traditional 2-D display. Some missions consist of many spacecraft (e.g., nine satellites for HelioSwarm) and it is not yet known which science products are effective and sufficient in describing the physics from the combined measurements. These changes in science data products will drive changes in HDRL systems and services.

- **Planned new data analysis capabilities under threat**. HDRL has substantial assets co-located with the IMCC (formerly NCCS), and has plans to utilize their extant capabilities to facilitate new data analysis capabilities for the Heliophysics community. However, earlier in 2026, it was communicated to HDRL that ESD is refocusing the use of IMCC to be more closely aligned to ESD priorities. It is unclear at present what this means for the HDRL presence at IMCC.

- **Big data analysis infrastructure**. To enable this analysis, computational capabilities must be co-located with data storage. The open infrastructure of HelioCloud provides democratic access to much, but not all, of HDRL's science data. In order to do so on premise, big data analysis needs to be leveraged. Present plans to evolve the storage at IMCC into a big data analysis facility are imperiled for the same reasons as the prior item.

- **Upgrading and modernizing tools.** This affects all HDRL tools and services. For example, while CDAWeb is fast and simple to use, it needs higher resolution plotting, more interactive interface, better handling of large mission constellations, and plotting of extended multi-dimensional data. This modernization effort will leverage contemporary web technologies and design patterns. Key challenges include balancing interactivity with performance and accessibility, especially selecting amongst thousands of datasets and sometimes hundreds of variables per dataset.

6.3. Policy and Compliance Related Challenges

New requirements flowing from policy also add to the challenges which HDRL must solve. These policies may levy requirements such as for better metadata curation (Open Data, Open

Science) or inclusion of new sources of data (researcher contributed datasets; SPD-41a) or new types of services (Heliophysics Decadal Survey). Significant challenges include:

- **Supplying requested metrics in the age of AI bots.** The recent HQ/OCSDO emphasis on science data repository metrics motivates a focus on developing metrics measurement systems for all three members of HDRL (see Table 3.1 for metrics concerning the cloud). However, AI bots present a modern challenge for HDRL's infrastructure, polluting our access metrics and straining (and even crashing!) our data access services.

- **Supporting the decadal survey objectives.** The decadal calls on US agencies (NSF and NASA) to collaborate to develop several objectives including a HelioSystems Laboratory and better enabling combined ground-based and space-based science. Currently NSF lacks a central entity in Heliophysics (several parts of NSF correspond to HDRL's scope of space, atmospheric and solar physics) which makes it difficult for HDRL management to meaningfully collaborate.

- **External dataset integration.** Some data (many are ground-based) from the projects supported by NSF or other agencies are added to SPASE/DOI registry and findable at HelioData per the data providers' request, but a large number of datasets are still missing. They could be added to the SPASE/DOI registry and made discoverable through HelioData, although they are not needed to be archived at SDAC/SPDF if they have permanent archives outside of NASA.

6.4. Technology Related Challenges

The technological landscape is ever changing. New capabilities, changing costs along with evolving technologies provide another source of challenges. Significant challenges include:

- **Cost effective storage** - either on-premise or in cloud - is currently expensive. On-premise hardware refreshes are required at cadences determined by equipment durability, data volumes, and usage patterns.

- **Leveraging AI for HDRL operations and development.** LLMs may provide significant productivity gains for HDRL, however staff need to be upskilled to take advantage of these resources. HDRL will have to develop the appropriate safeguards against AI hallucination which may mitigate potential productivity gains and will also have to procure appropriate services to defray costs of using these capabilities.

- **Hardware and software obsolescence.** SPDF and SDAC handle software and hardware obsolescence by planned migration to new hardware every 5-7 years, using open source software like Bacula backup software that won't lock us out, and migrating to new software libraries as needed, such as the conversion of Java-based SKTEditor and TIPSOD orbit viewer packages to browser-based tools.

- **Disaster recovery and mitigation planning.** SPDF maintains multiple copies of the data in two buildings and offsite on tape.The main storage servers use the copy-on-write ZFS file system which provides checksums of every block, hourly snapshots for tracking changes, and an efficient hourly mirror to the read-only copy in the other building. The data are backed up on LTO-9 and soon LTO-10 tape for storage in an offsite Iron Mountain facility twice a year (courtesy of a NSSDCA-led contract). In order to store the incremental backups outside the building between full backups, we plan to also copy the incrementals to AWS Glacier storage. The data are also synced to IMCC for on-premises high-performance computing, and science data is copied to HelioCloud for data analysis and computing in the cloud. The operational servers have redundant twins in the other building. SDAC data is held using the enterprise-level NetApp/ONTAP data system which maintains data integrity and offers powerful and easy to use tools for data management. SDAC data is also copied to be colocated with IMCC computing systems. A third copy is made to tape. Finally, given that many missions maintain their own data systems, additional copies of data are located at mission/instrument institutes (Figure 1.1.2.1).

6.5. Staffing Related Challenges

A perennial challenge is finding people to do the work of HDRL. This work requires a specialized set of skills. Ideally, HDRL staff should be well-versed in the relevant science, understand mission and instrument operations, and have familiarity with data-related research and infrastructure.

- **There is intense competition for staff with the private sector.** The same specialized skills which HDRL needs are also highly coveted in industry. It may be appreciated that the public sector does not pay as well as Facebook, Google, Amazon, Microsoft and other similar companies. A partial remedy for this is to hire interns and early career staff but we still have a gap when seeking more experienced staff.

- **Lack of deputies.** HDRL is unable to hire civil servant deputies for the lack of available billets (Robert Candey will retire this CY, leaving all HDRL members without deputies). ***Lack of deputies results in a significant risk of loss of institutional knowledge and a critical leadership gap*** should any of the project scientists for SDAC, SPDF, or DISH leave. This gap also constrains needed and required HDRL engagements with other US agency peer organizations and at an international level.

References

Antunes et al. (2025). “CloudCatalog: an API plus Tools for Lazy Indexing of Millions of Cloud-Stored Data Files”. Data, Analysis and Software in Heliophysics (DASH), San Antonio. Zenodo. https://doi.org/10.5281/zenodo.17398630

Boquet, Z., Ringuette, R., Cecconi, B., Koval, A., & Kasulke, T. (2026). “Improving and Assessing FAIR for HDRL's Metadata: Mapping to Schema.org and DataCite” (ESIP DSC). Zenodo. https://doi.org/10.5281/zenodo.18966019

Freeland, S.L., Handy, B. N. (1998) Data Analysis with the SolarSoft System, Sol. Phys., 182, 497. https://doi.org/10.1023/A:1005038224881

Müller, D., Nicula, B., Felix, S., Verstringe, F., Bourgoignie, B., Csillaghy, A., Berghmans, D., Jiggens, P., García-Ortiz, J. P. ,Ireland, J., Zahniy, S., Fleck, B. (2017) JHelioviewer Time-dependent 3D visualisation of solar and heliospheric data, A&A, 606 A10. https://doi.org/10.1051/0004-6361/201730893

National Academies of Sciences, Engineering, and Medicine. 2025. *The Next Decade of Discovery in Solar and Space Physics: Exploring and Safeguarding Humanity's Home in Space*. Washington, DC: The National Academies Press. https://doi.org/10.17226/27938

Roberts, D., Thomas, B., Fung, S., Weigel, R., Kepko, L., Viall, N., Walker, R., Rawafi, N., Ireland, J., Jian, L., and Candey, R. (2023). The Heliophysics Digital Resource Library: An Open, Flexible, Easily Used Infrastructure for Data, Software, Registration, Documentation, and Services. Bulletin of the AAS. https://doi.org/10.3847/25c2cfeb.99007c8b.

Stall, S., Ringuette, R., Vrouwenvelder, K., & Sedora, B. (2025, March 28). COPDESS Data and Software Citation Deep-Dive Seminar. Zenodo. https://doi.org/10.5281/zenodo.15103616

The SunPy Community, Barnes, W. T., Bobra, M. G., Christe, S. D., Freij, N., Hayes, L. A., Ireland, J., Mumford, S., Perez-Suarez, D., Ryan, D. F., Shih, A. Y., Chanda, P., Glogowski, K., Hewett, R., Hughitt, V. K., Hill, A., Hiware, K., Inglis, A., Kirk, M. S. F., Konge, S., Mason, J. P., Maloney, S. A., Murray, S. A., Panda, A., Park, J., Pereira, T. M. D., Reardon, K., Savage, S., Sipőcz, B. M., Stansby, D., Jain, Y., Taylor, G., Yadav, T., Rajul, Dang, T. K. (2020). The SunPy Project: Open Source Development and Status of the Version 1.0 Core Package. Ap.J., 890:68, 2020. doi:10.3847/1538-4357/ab4f7a.

Thomas, B., Antunes, A., Yeakel, K., Bradford, J., Winter, E., Mo, W., Smith, B., Vandegriff, J., & Roberts, A. (2022). HelioCloud: An Open Cloud-based Platform for Heliophysics Research. *Bulletin of the AAS*, *54*(7). https://baas.aas.org/pub/2022n7i103p03

Thomas, B., & Ringuette, R. (2024). “Report on Effort to Create a SPASE 3.0”, International Heliophysics Data Environment Alliance (IHDEA) meeting, Madrid, Spain. Zenodo. https://doi.org/10.5281/zenodo.13910320

Ringuette, R., Thomas, B. A., Ireland, J., Jian, L., & Candey, R. (2026). “HDRL's 2026 Status on the Desirable Characteristics of Data Repositories for Federally Funding Research [Data set]”. Zenodo. https://doi.org/10.5281/zenodo.20672271

Sujit et al. (2025), “Surya: Foundation Model for Heliophysics using SDO Data”, AGU Fall Meeting 2025, held in New Orleans, LA, 15-19 December 2025, Session: Informatics / AI Foundation Models for Earth, Space, and Planetary Sciences III Oral, id. IN43A-04. doi: https://arxiv.org/abs/2508.14112

Weigel, R. S., Vandegriff, J., Faden, J., King, T., Roberts, D. A., Harris, B., et al. (2021). “HAPI: An API standard for accessing Heliophysics time series data”. *Journal of Geophysical Research: Space Physics*, 126, e2021JA029534. https://doi.org/10.1029/2021JA029534

Acronyms

ACE - Advanced Composition Explorer
AIA - Atmospheric Imaging Assembly
AGU - American Geophysical Union
AMDA - Automated Multi-Dataset Analysis
AMS - American Meteorological Society
API - Application Programming Interface
AWS - Amazon Web Services
CCMC - Community Coordinate Modeling Center
CDF - Common Data Format
CEDAR - Coupling, Energetics, and Dynamics of Atmospheric Regions
CODEX - COronal Diagnostic EXperiment
COHOWeb - COordinated Heliospheric Observations Web
COPDESS - Coalition for Publishing Data in the Earth and Space Sciences
DCAT - Data Catalog
DASH - Data, Analysis, and Software for Heliophysics
DISH - Data Integration Services for Heliophysics
DOI - Digital Object Identifier
DONKI - Database of Notifications, Knowledge, Information
DMSP - Defense Meteorological Satellite Program
EIT - Extreme Ultraviolet Imaging Telescope
ESA - European Space Agency
ESAC - European Space Astronomy Centre
ESIP - Earth Science Informatics Partners
EVE - Extreme Ultraviolet Variability Experiment
FAIR - Findable, Accessible, Interoperable, Reproducible
FUV - Far ultraviolet
GEM - Geospace Environment Modeling
GIF - Graphics Interchange Format
GOES - Geostationary Operational Environmental Satellite
GPU - graphics processing unit
GUVI - Global Ultraviolet Imager
HAPI - Heliophysics API
HAO - High Altitude Observatory
HDMC - Heliophysics Data and Model Consortium
HDRL - Heliophysics Digital Resource Library
HMI - Helioseismic Magnetic Imager
HPD - NASA Heliophysics Division
HSSI - Heliophysics Software Search Interface
HTTPS - Hypertext Transfer Protocol Secure
IAS - Institute d'Astrophysique
ICD - Interface Control Document
IDL - Interactive Data Language
IHDEA - International Heliophysics Data Environment Alliance

IHDWG - Interagency Data environments Working Group
IMAGE - Imager for Magnetopause-to-Aurora Global Exploration
IMCC - Integrated Modeling Computing Center
ISTP - International Solar-Terrestrial Physics
JAXA - Japan Aerospace Exploration Agency
JGR - Journal of Geophysical research
JSON - JavaScript Object Notation
KASI - Korea Astronomy and Space Science Institute
LASCO - Large Angle and Spectrometric Coronagraph
LLM - Large Language Model
LMSAL - Lockheed Martin Solar & Astrophysics Laboratory
LTO - Linear Tape Open
MAVEN - Mars Atmosphere and Volatile Evolution
MCP - Model Context Protocol
ML - Machine Learning
MLSO - Mauna Loa Solar Observatory
MSFC - Marshall Space Flight Center
NAIF - Navigation and Ancillary Information Facility
NCCS - National Center for Climate Simulation
NLP - Natural Language Processing
NOAA - National Oceanic and Atmospheric Administration
NORAD - North American Aerospace Defense Command
NRL - Naval Research Laboratory
NSF - National Science Foundation
NSO - National Solar Observatory
NSSDC - National Space Science Data Center
ONTAP - Open Network Technology for Appliance Products
PADRE - Solar Polarization and Directivity X-Ray Experiment
PFSS - potential flux source surface
PDF - Portable Document Format
PNG - Portable Network Graphics
PS - PostScript
PS - Project Scientist
PSP - Parker Solar Probe
PUNCH - Polarimetry to UNify the Corona and Heliosphere
RDA - Research Data Alliance
ReSA - Research Software Alliance
REST - REpresentational State Transfer
ROB - Royal Observatory of Belgium
S3 - Simple Storage Service
Sci-X - Science eXplorer
SDAC - Solar Data Analysis Center
SDE - Science Discovery Engine
SDO - Solar Dynamics Observatory

SHINE - Solar Heliospheric and INterplanetary Environment
SMWT - SPASE Metadata Working Team
SOAP - Simple Object Access Protocol
SOHO - Solar and Heliospheric Observatory
SOLAR-1 - Space weather Observations at L1 to Advance Readiness – 1
SPASE - Space Physics Archive Search and Extract
SPICE - Spacecraft, Planet, Instrument, C-matrix, Events
SPD - Solar Physics Division
SPDF - Space Physics Data Facility
SSC - Satellite Situation Center
STEREO - Solar and TErrestrial RElations Observatory
SWPC - Space Weather Prediction Center
SwRI - Southwest Research Institute
TAP - Table Access Protocol
TESS - Triennial Earth-Sun Summit
TIMED - Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics
TIPSOD - Tool for Interactive Plotting, Sonification, and 3D Orbit Display
UAT - Unified Astronomy Thesaurus
UCAR - University Corporation for Atmospheric Research
URL - Uniform Resource Locator
VSO - Virtual Solar Observatory
WSA - Wang-Sheeley-Arge
XML - Extensible Markup Language
ZFS - Zettabyte File System