



EXPLORE SCIENCE

HPD Open Science and Open Software

Matt McClure (matthew.r.mcclure@nasa.gov)

Susanna Finn (susanna.c.finn@nasa.gov)

Alvin Robles (alvin.s.robles@nasa.gov)

Steven Crawford (steven.m.crawford@nasa.gov)

September 21, 2022

A decorative graphic on the left side of the slide, featuring a curved white border. Inside the border, there is a space-themed illustration with a bright yellow sun at the bottom left, a blue and white Earth at the bottom, a grey moon, a reddish-brown Mars, a brown Saturn with its rings, and a blue and green nebula in the background.

Outline

- What is HDRL?
- Why change the data policy for Open Science?
- The new SPD-41a: Scientific Information Policy
- The Heliophysics Data Policy
- HPD Data Management Handbook
- Next steps



Heliophysics Digital Resource Library (HDRL)

Vision: Where the Heliophysics System Observatory Comes Together

The HDRL enables the scientific analysis goals of the Heliophysics System Observatory:

- *provisioning and curation of scientific big data* from many sources, PB volumes; (the Foundation: data, metadata, standards)
- *support for data analysis and modeling* in multiple computational environments;
- the design and implementation of a *collaborative open science infrastructure*.

Individual missions can do great science

Unlocking groundbreaking systems science requires the HDRL



HDRL ORGANIZATIONAL CHART

HP Data and Model Consortium / HDMC

**D. Aaron Roberts (PS), Brian Thomas (DPS),
Tressa Helvey-Kasulke (PM)**

Overall management of the HDRL.

Registries and DOIs for all digital resources; SPASE Data Model.

Heliophysics Data Portal (HDP; including solar)

Python and other software integration (PyHC).

Analysis and visualization services ((Py)SPEDAS, Autoplot).

Data upgrades and services.

HelioCloud initiative with data and software from all groups.

All activities within the various HDRL components are interrelated.

Space Physics Data Facility / SPDF

Robert Candey (PS), Lan Jian (DPS)

Non-solar Data Final Active Archive for NASA (and other) missions.

CDAWeb data browsing and access; Web Service access.

OMNIWeb data production and serving.

SSCWeb and 4-D spacecraft orbit facility. Common Data Format.

Solar Data Analysis Center / SDAC

Jack Ireland (PS)

Solar Data Final Active Archive for Solar Dynamics Observatory and other NASA missions.

Virtual Solar Observatory data access.

Helioviewer. SolarSoft. SunPy.

High Performance Computing for NASA HP.

Collaborators

Community Coordinated Modeling Center

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

Center for HelioAnalytics

Science Research Projects, Outreach, Open Science, User Testing

HelioCloud

HelioCloud Vision: Connect people, data, papers, and software in the Heliophysics community

What?

- ❖ Publicly accessible cloud based high value data (Goal: 8-10 Pb by 2025)

Why?

- ❖ Close to the compute - unlock big data science for everyone!
- ❖ Accelerate Collaboration - Everyone can work and share in the same environment

Who?

- ❖ SDAC (SDO/AIA, SOHO, Stereo A/B)
- ❖ SPDF (All High Level Data)
- ❖ CCMC (Model Outputs)
- ❖ APL (Research Team Data, SuperMAG)

Accomplishments:

- ❖ PyHC Summer School (May 30 - Jun 3) - HelioCloud supported 150 simultaneous users
- ❖ Over 70 active NASA researchers
- ❖ Coordinating with the AWS Space Act support from OSSIC to move more data to HelioCloud

Why do we have a new policy?

1. Open Science!

- **F**indable
- **A**ccessible
- **I**nclusive
- **R**eproducible

2. New White House Guidance:

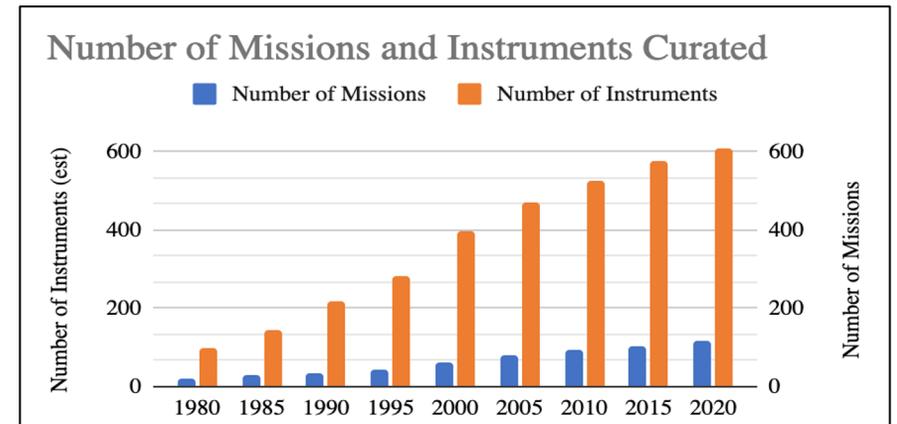
- [New Guidance to Ensure Federally Funded Research Data Equitably Benefits All of America](#)
- [OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay](#)

3. Community Feedback

- 2021 Heliophysics Infrastructure Workshop
- Request for Information on the Heliophysics Data Environment (HPDE)

4. Data growth and variety:

	Now	2023
SDAC	1.3 PB	~20PB
SPDF	500 TB	700TB
HelioCloud	100 TB	1.5PB



OSTP Guidance

On August 25, OSTP issued Guidance to Make Federally Funded Research Freely Available Without Delay. This had three high level of recommendations of:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data

This memo has been integrated into the latest version of SDP-41a. The most significant change was removing the embargo period to peer reviewed publications. Other requirements were already consistent between the memo and SDP-41a.

RELEASE: [OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay](#)

BLOG: [Breakthroughs for All: Delivering Equitable Access to America's Research](#)

MEMO: Memo to the Heads of Executive Departments and Agencies on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research can be found [here](#).

CONGRESSIONAL REPORT: Economic Landscape of Federal Public Access Policy can be found [here](#).

SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Goal 1: Develop and Implement Capabilities to Enable Open Science

1.1

Develop and implement a **consistent open data and software policy** tailored for SMD

1.2

Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**

1.3

Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions

1.4

Increase transparency into how science data are being used through a free and open unified journal server

Goal 2: Continuous Evolution of Data and Computing Systems

2.1

Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques

2.2

Integrate investment decisions in High-End Computing with the strategic needs of the research communities

2.3

Invest in capabilities to use commercial cloud environments for open science

2.4

Invest in the tools and training necessary to enable breakthrough science through application of AI/ML

Goal 3: Harness the Community and Strategic Partnerships for Innovation

3.1

Develop **community of practice and standards group**

3.2

Partner with **academic, commercial, governmental and international organizations**

3.3

Promote opportunities for continuous learning as the field evolves through collaboration

SPD-41: Scientific Information Policy

SPD-41 was released in August 2021.

SPD-41 brings together existing NASA and Federal guidance.

- SPD-41: The Science Information Policy - <https://go.usa.gov/xtNTJ>
- Science Information Policy Website - <https://go.usa.gov/xtNTt>



SPD-41a was released in November with proposed additions. An RFI was released to the community and closed on **March 4, 2022.**

What is the **current** policy in SPD-41?

Data

Scientific data shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research.

Mission data shall be openly available with no period of exclusive access.

Software

Research software should be publicly available no later than the publication of the research and assigned a permissive software license.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months.

Mission publications shall additionally be made publicly available at the time of their publication.

What are the **new** proposed changes in the SPD-41a draft?

Data

Scientific data should be FAIR and shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research, and be citable.

Mission data shall be openly available with no period of exclusive access.

Software

Research software shall be publicly available no later than the publication of the research, assigned a permissive software license, and be citable.

Mission software shall additionally be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months. Publishing as open access is supported and posting preprints is encouraged.

Mission publications shall additionally be made publicly available at the time of their publication.

Science workshops and meetings shall be open to broad participation and documented in public repositories.

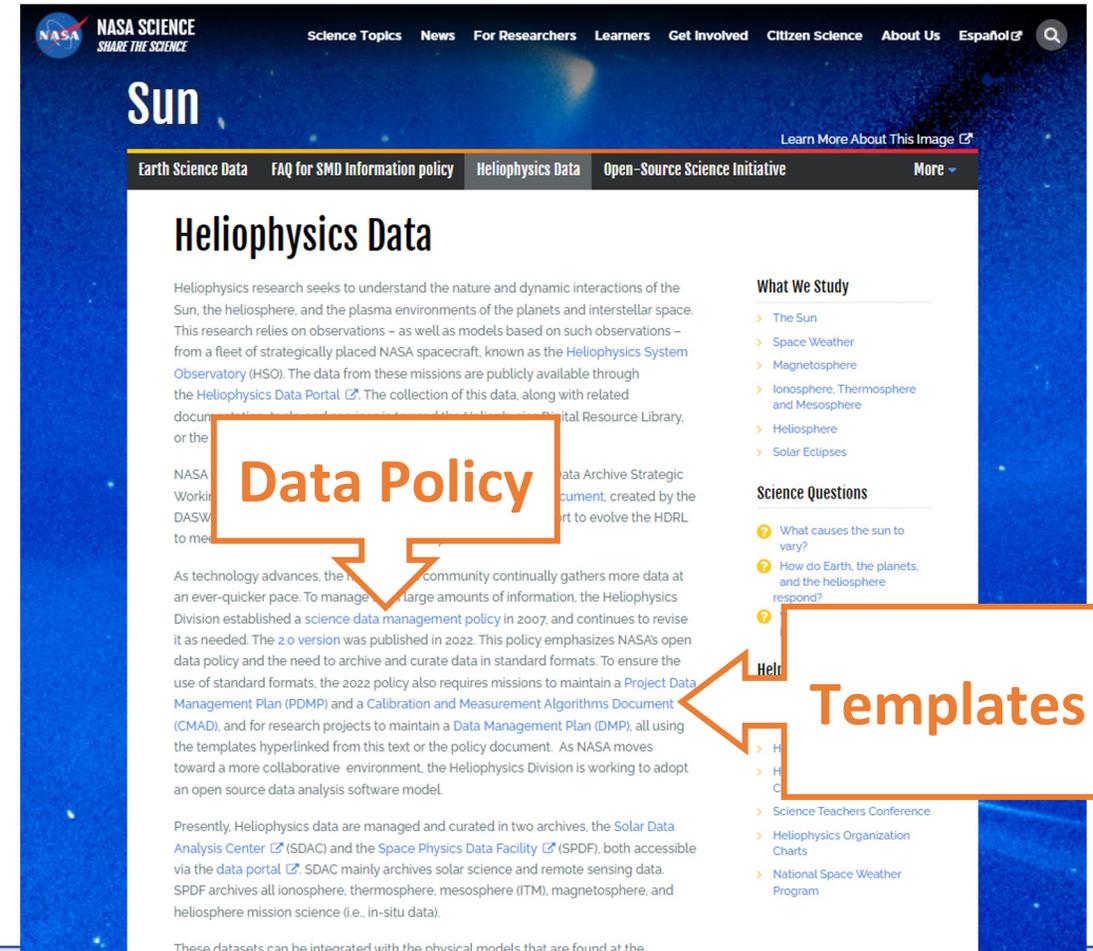
Open science activities will be considered in reviews of proposals.

HPD Science Data Management Policy

The previous HPD Science Data Management Policy release (v1.2) was in October 2016

HPD Archives Strategic Working Group worked with Archives personnel at GSFC to develop v2.0 (released February 14, 2022)

- Accounts for changes in requirements and organization of the HPD Archives
- Aligns with SPD-41
- Encourages adherence to FAIR Guiding Principles for scientific data management and stewardship.
- Facilitates long-term curation of HPD-funded data



<https://science.nasa.gov/heliophysics/heliophysics-data>

Information protected by other NASA policies, Executive Orders, or legislation are exempt from the open science requirements.

HPD Policy and Open Science

Heliophysics benefits from the contributions of the open science community and [is committed to promoting open science](#) across its initiatives and funded programs.

The new [Heliophysics Data Policy](#) [Feb 2022] sets two Division policies:

- Codifies “making high-quality, high spatial and temporal resolution data publicly available as soon as practical”
- Prescribes full and open sharing of Heliophysics Division-funded software under permissive license with broad community acceptance and made available in public repositories

HPD participates as an active member in the SMD Open Source Science Working Group and Councils and supports key initiatives including:

- [Transform to Open Science](#) (TOPS), SMD initiative
- HPD funded activities for Python in the Heliophysics Community (PyHC)
- SMD Year of Open Science - Initiative to inspire open science engagement through events and activities



HPD Data Policy Key Points/Highlights

- Missions are required to work with the Archives early in the mission, prior to adoption of the Project Data Management Plan (PDMP), in the preparation of their data products.
- Archiving of products in specific formats* (e.g., CDF, FITS) with standard metadata and adequate documentation are now required. All data (including real-time) are now required to be placed in the Heliophysics archives and registered upon production.
- Complete (including descriptions of variables) SPASE product Registry entries are required.* Missions encouraged to use HDRL services to streamline the development of tools for data discovery, access, and use.
- For **research activities** that *do not* result in publication of a peer-reviewed article, scientifically useful data and advanced data products associated with the award that have not already been made public shall be made publicly available *at the conclusion of the research award*.*
- Directs missions to use new PDMP and Calibration and Measurement Algorithms Document (CMAD) templates and clarifies timelines for their development and submission.

* HPD-specific supplement to SPD-41

All ROSES efforts must share their data.

Applicability (in alignment with SPD-41)

Information produced by NASA Heliophysics missions. Missions include strategic or flagship missions and investigations selected under Announcements of Opportunity (AOs), including those selected under the Stand-Alone Missions of Opportunity Notice (SALMON) and Cooperative Agreement Notices (CANs).

Information produced by investigations funded via research awards. This includes funding from investigations selected under NASA Research Announcements (NRAs), including those selected under the ROSES NRA. This class will be referred to as “Research” information.

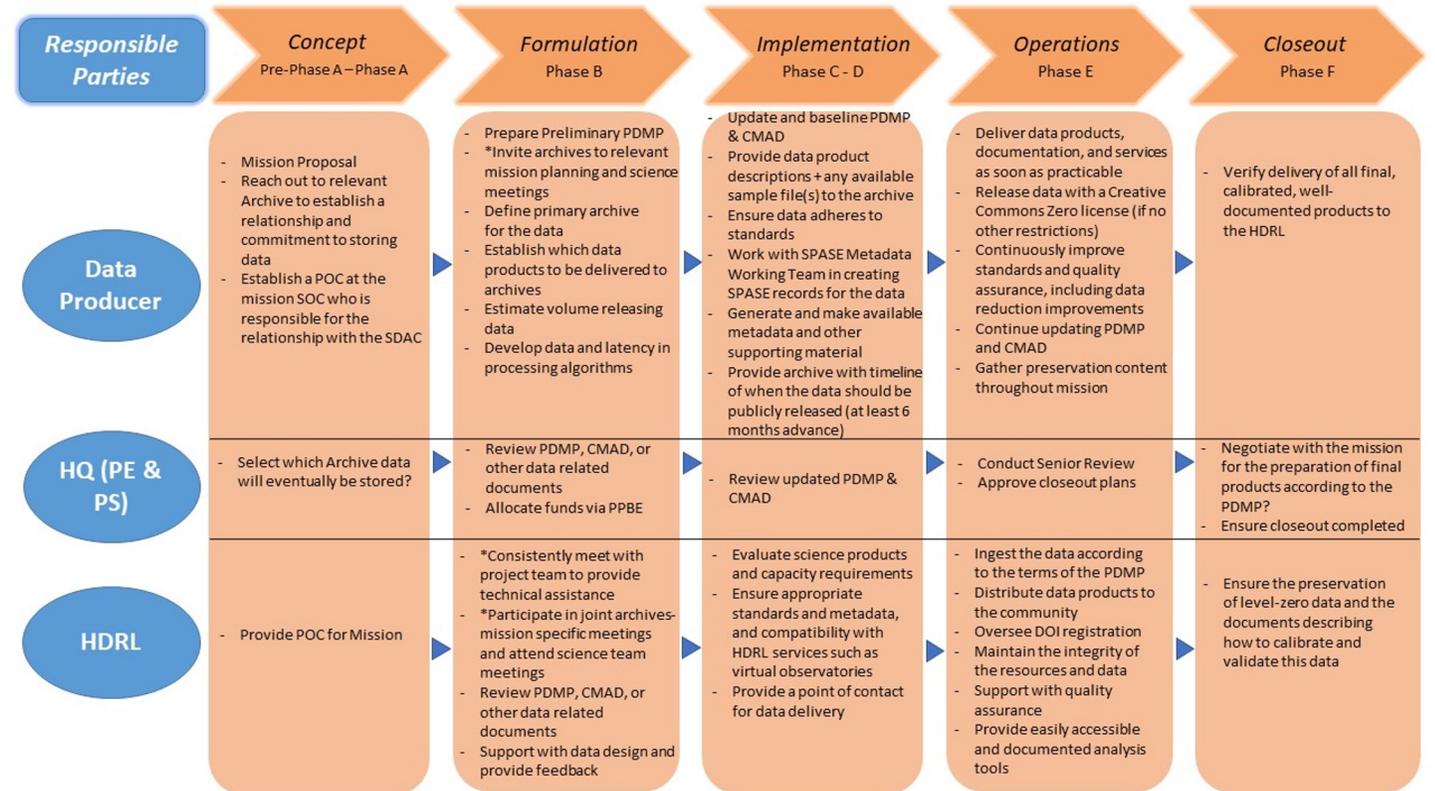
- This also includes investigations funded via research sub-awards for research made as part of Mission-funded activities or cooperative agreements.
- Research awards can include grants, cooperative agreements, contracts, task orders, interagency transfers, direct internal NASA funding, and other applicable funding vehicles.

Information produced by other HPD-funded activities such as, but not limited to, experiments, investigations using sub-orbital platforms, field campaigns, or citizen science projects.

New missions and investigations shall follow all parts of this policy. Existing missions and investigations should adopt all parts of this policy consistent with available resources.

HPD Data Management Handbook

- The HPD Data Management Handbook is a companion to the Data Policy and will provide guidance, best practices, and processes to help **data providers** meet the requirements of the Data Policy
- The handbook:
 - explains the motivation behind adherence to Open Science and FAIR principles
 - breaks down the Mission Data Life Cycle as well as the Research Data Life Cycle, detailing the roles, responsibilities, and expectations within each
 - specifies and explains the data and metadata standards used by the HDRL
 - highlights tools and services provided by the HDRL and CCMC
- A separate HDRL User's Guide will be developed



Supporting Heliophysics Open Science with ROSES-22

- B.12 Heliophysics Data Environment Enhancements (HDEE)
 - upgrade existing Heliophysics data products to **improve the quality, utility, and accessibility of datasets** relevant to Heliophysics research.
 - POC Reiner Friedel reinhard.h.friedel@nasa.gov
- F.8 Supplemental Open-Source Software (SOSS)
 - Supplemental open-source software awards are used to encourage the conversion of legacy software into modern code to be released under a generally accepted, open source license. The supplement would **add a software component to an existing "parent" research and analysis award.**
 - POC Steven Crawford steven.m.crawford@nasa.gov

Supporting Heliophysics Open Science with ROSES-22

- B.16 Heliophysics AI/ML Ready Data (HARD)
 - **create AI/ML ready data sets and products**, which allow the use of AI/ML methodologies and approaches to address specific science problems
 - POC Katya Verner ekaterina.m.verner@nasa.gov
- B.20 Heliophysics Tools and Methods (H-TM)
 - advance the goal of a robust, vital, and cohesive Python environment for Heliophysics (PyHC) with **open source code**.
 - POC Reiner Friedel reinhard.h.friedel@nasa.gov
- B.21 Heliophysics Citizen Science Investigations (H-CSI)
 - **expand participation** of citizen scientists in heliophysics research.
 - POC Susanna Finn susanna.c.finn@nasa.gov



Next Steps:

1. We want the Heliophysics Community's Feedback!
2. Continually evaluate the implementation of new policy and HDRL architectures!
3. Enable future of Heliophysics by preparing infrastructure for tomorrow!

ADDITIONAL SLIDES



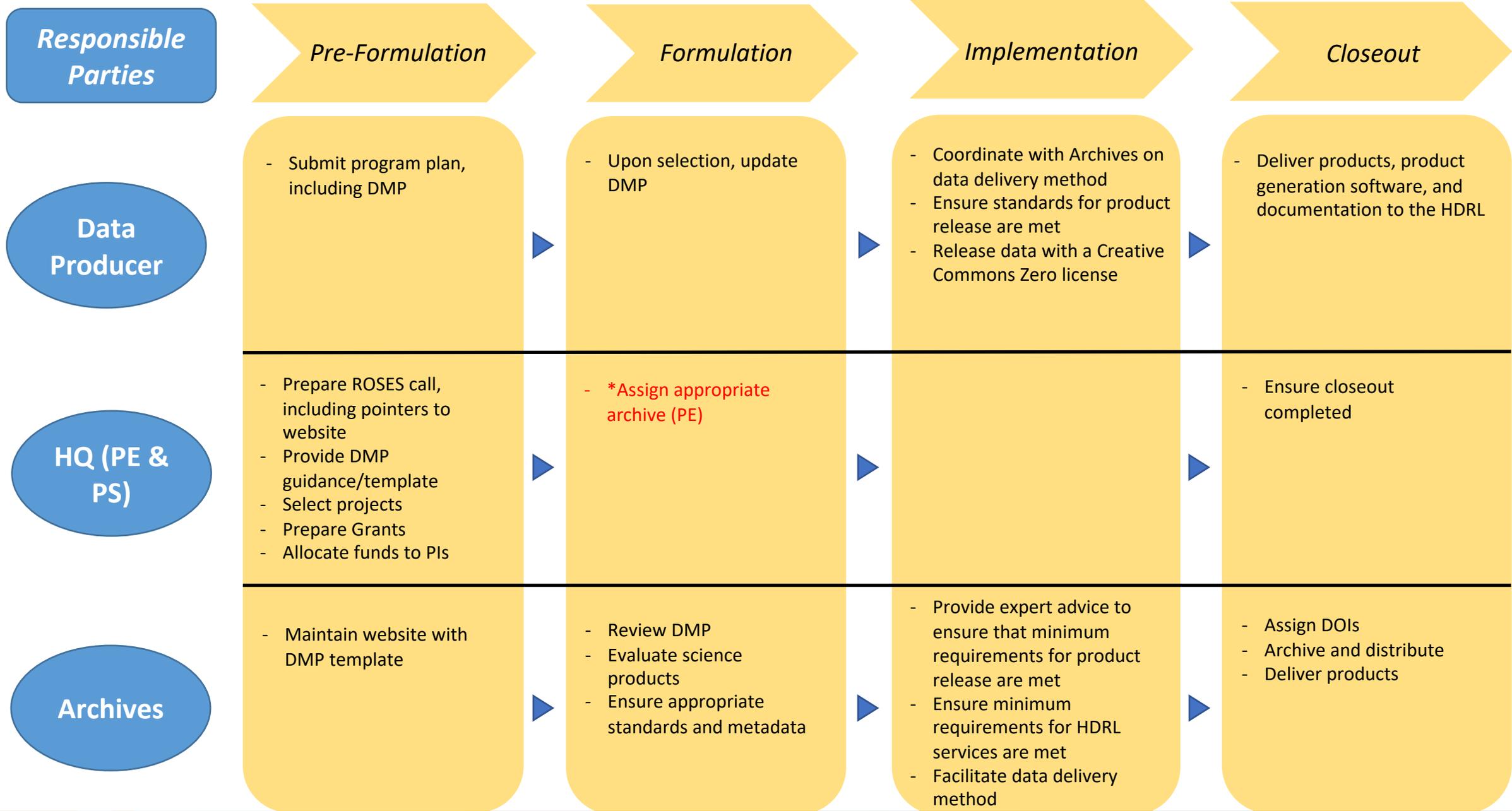
HPD Data Archives Strategic Working Group

New ways of probing solar and space plasmas have led to a rapid growth in data size and complexity

Understanding the intricacies of these systems requires cross mission and model comparison, machine learning and other large-scale processing, which has implications for data archiving and modeling needs:

- Greatly increased storage, retrieval, and computation services, including collaborative analysis, and data cleaning for ML analyses
- Increased discoverability and usability of data and model results
- Added data curation requirements, such as better user and data provider training and feedback, more complete metadata and provenance, dataset registry and DOIs, quality control
- Open science and software initiatives

DASWG formed in 2020 to address these needs for the future.



Data Levels per SPD-41a

Data Level	Description
0	Unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. This level of data may only be available upon request
1	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information. This level of data will be in a standard format that is accessible
2	Data that has been processed to remove instrument or sensor effects.
3	Data that has been mapped on to a uniform space-time grid, resampled, or combined to produce a set of data with greater completeness and consistency
4	Products delivered as part of a Mission derived from data. This could include model outputs, analysis of results, catalogs, or databases derived from Mission data
5	Products contributed from the community derived from Mission data
Auxiliary	Technical data generated by support or other systems as part of the Mission. This could include environmental sensors, spacecraft telemetry, or other technical information. When not excepted, scientifically useful data that is produced as part of auxiliary systems should be made accessible in accordance with available resources. This level of data may only be available upon request
Ground Test	Data that are produced on the ground for testing the Mission either prior to launch or during operations. When not excepted, scientifically useful data that is produced as part of ground testing should be made accessible in accordance with available resources. This level of data may only be available upon request.



Heliophysics RFI Responses

“This RFI seeks to gather information and community feedback on the current and future needs for NASA’s Heliophysics data, and associated archives, tools, models, and resources. This information will be used to inform NASA’s restructuring, expansion, and evolution of the Heliophysics data archiving infrastructure, currently known as the Heliophysics Data Environment (HPDE) located at <https://hpde.gsfc.nasa.gov>.”

Released 15 Oct 2021, Responses due 28 Feb 2022

28 responses, 19 relevant

Top repeated themes seen in the responses include:

- Curation
- Interface/site design
- Standards
- Guidance and best practices

Current HPD Archives Structure

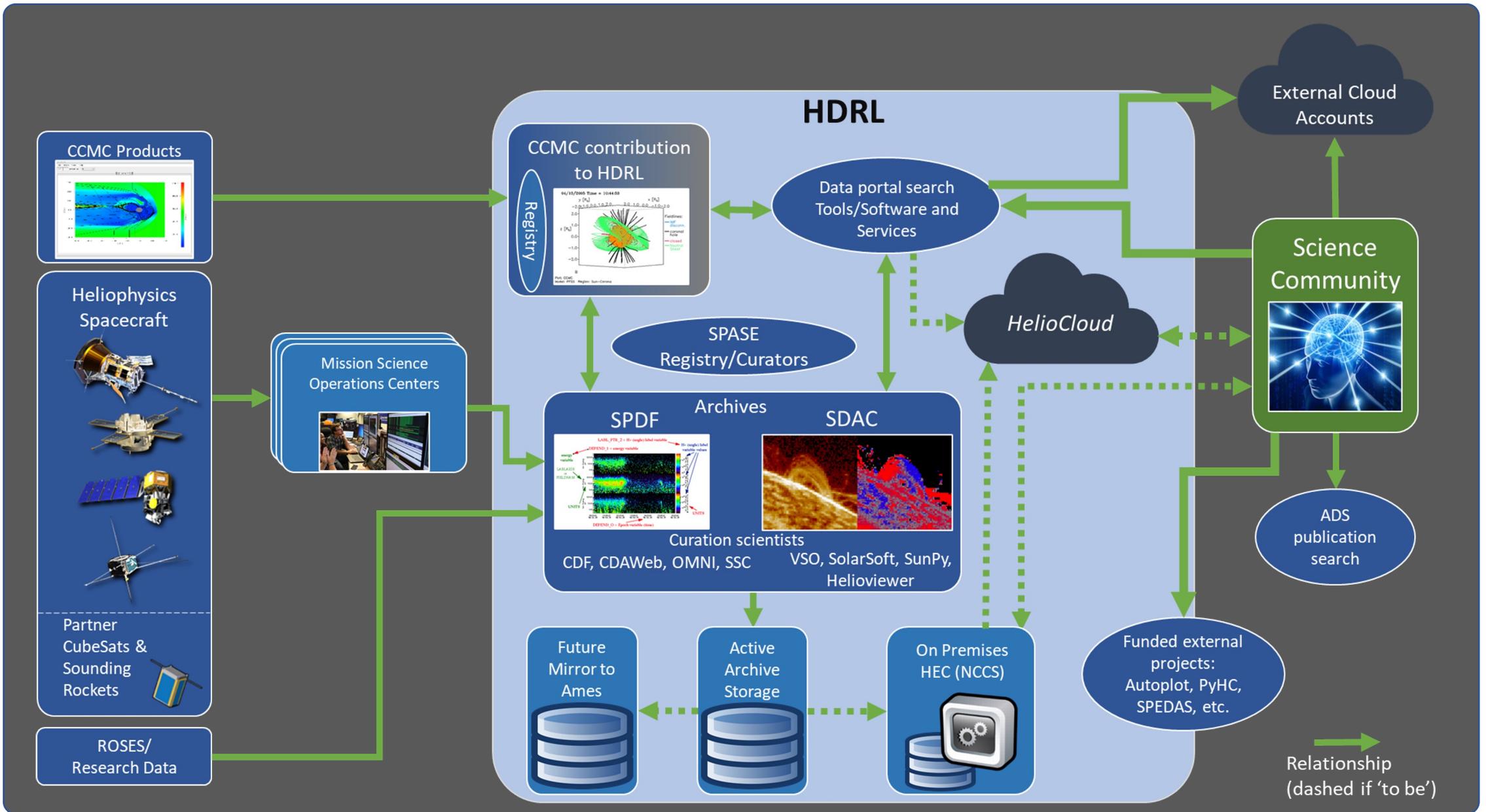
Both datasets also available through the Heliophysics Data Portal

The screenshot displays the Heliophysics Data Portal interface. At the top, it features the NASA logo and the text "GODDARD SPACE FLIGHT CENTER Space Physics Data Facility". Navigation links include "Goddard Home" and "Visit NASA.gov". The main header reads "Heliophysics Data Portal 'Find it. Browse it. Get it.'" with a "SPASE inside" logo. A secondary navigation bar contains "Help", "Geo Orbits", "Helio Orbits", "SPASE Registry", "ADS Abstracts", and "Feedback".

The interface is divided into several sections:

- Text Restriction:** A search box with an "Add" button.
- Time Span Restriction:** A section with an information icon, a format "YYYY-MM-dd or YYYY-DDD", and "from:" and "to:" input fields with an "Add" button.
- Element Restriction:** A list of filter categories, each with an information icon: Resource type, Measurement type, Observatory Group, Observatory, Instrument, Observed region, Spectral range, Cadence, Repository Name, Access rights, and Format.
- Current Product Restrictions:** A section stating "No restrictions are currently set." with a "View Current List" link and a "Sort by" dropdown menu set to "Observatory".
- Results Table:** A table showing search results. The first row indicates "Showing 1 - 20 of 2938 Results". The table has two main columns: "# Products (& SPASE descriptions)" and "Access Links".

#	Products (& SPASE descriptions)	Access Links
1	ACE 27-day Survey Plots	<ul style="list-style-type: none">· Polar-Wind-Geotail 'gif-walk' site Get Images/Plots
2	ACE Cosmic Ray Isotope Spectrometer (CRIS) 1-Hour Level 2 Data	<ul style="list-style-type: none">· FTPS from SPDF (not with most browsers)· HTTPS from SPDF· CDAWeb· CDAWeb HAPI Server· ACE Science Center FTP area· ACE Science Center· ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page Get Data/Plots
3	ACE Cosmic Ray Isotope Spectrometer (CRIS) Daily Level 2 Data	<ul style="list-style-type: none">· FTPS from SPDF (not with most browsers)· HTTPS from SPDF· CDAWeb· CDAWeb HAPI Server· ACE Science Center ftp area· ACE Science Center· ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page Get Data/Plots
4	ACE Daily Survey Plots	<ul style="list-style-type: none">· Polar-Wind-Geotail 'gif-walk' site Get Images/Plots

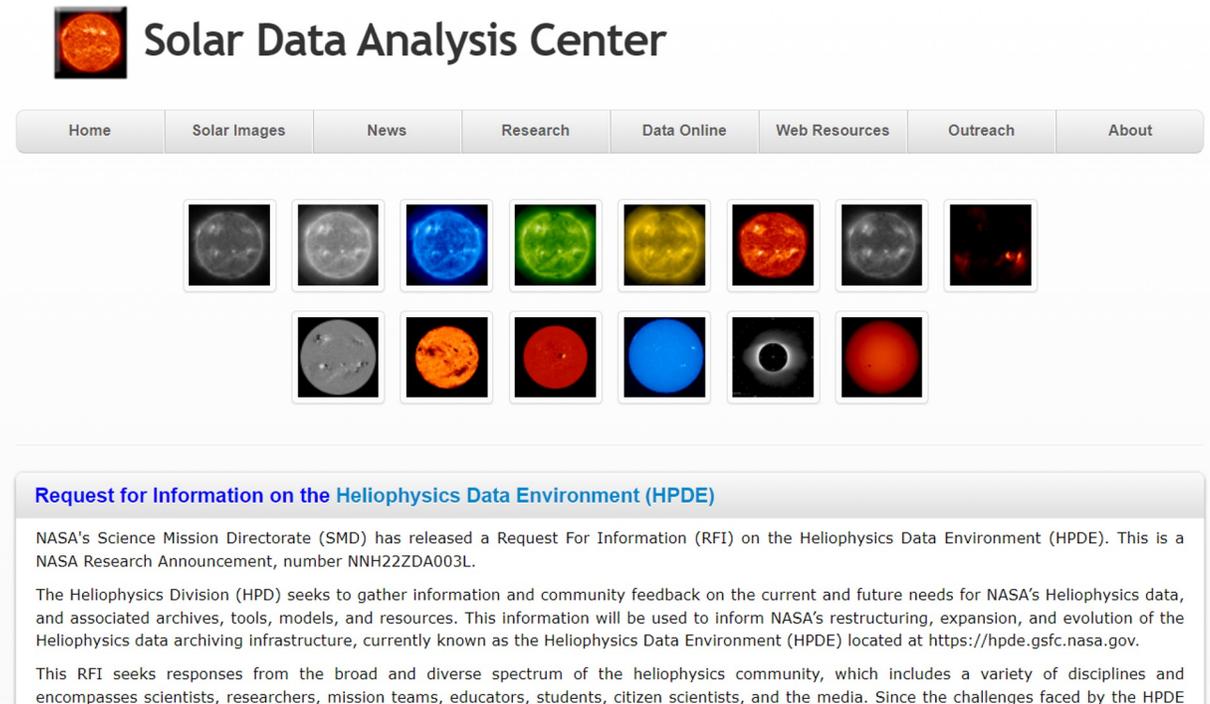


Current HPD Archives Structure

Heliophysics Mission Data Archived in 2 separate archives:

Space physics (non-solar) data archive: SPDF, Solar data archive: SDAC

Each archive has a website with access to data and tools

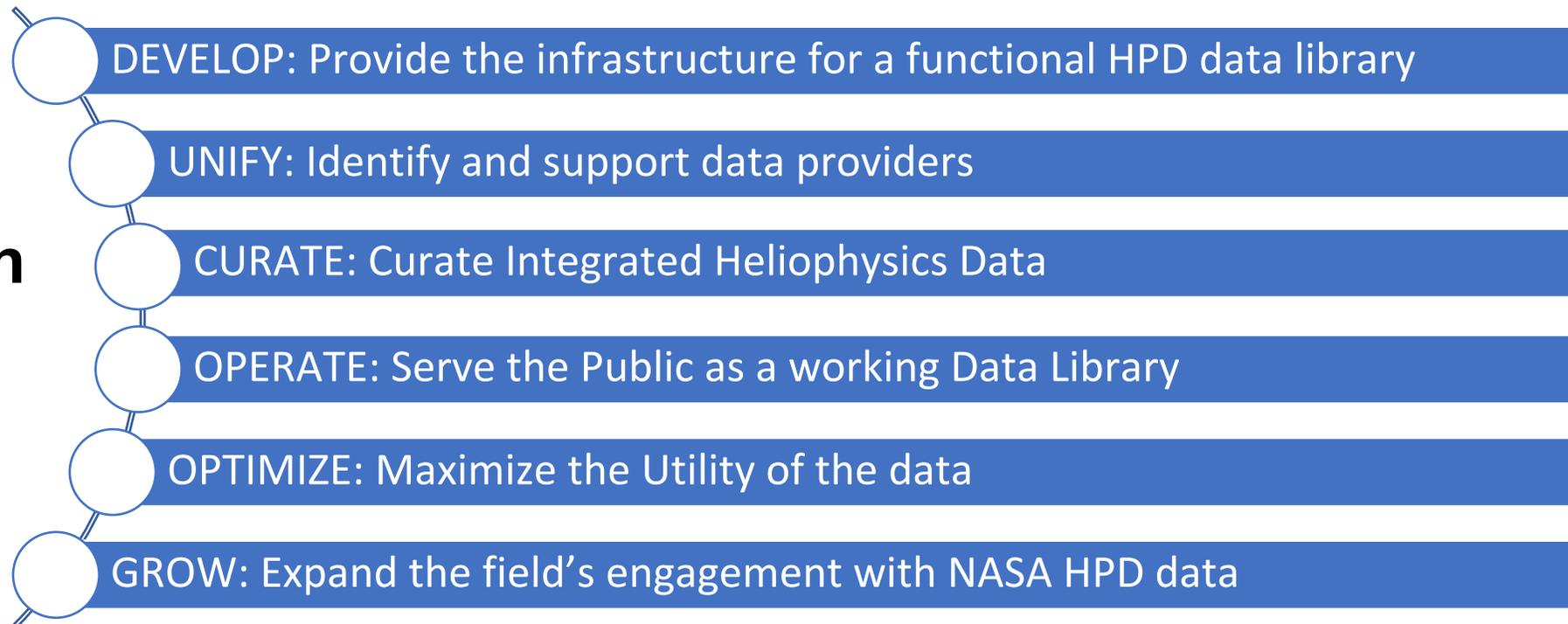


Heliophysics Archiving Strategy

VISION: Democratize the Science and Data of Heliophysics

MISSION: In line with US government strategy, HDRL is committed to being the premier resource for all NASA HPD data needs. Moving beyond a traditional repository and toward a functional, collaborative data library, the NASA HPD archives will maximize the utility of the data of the HSO, sustainability of the archives, and access for the public to this data.

Mission Goals:



Archives Strategy by Goal

Theme	Goal	Objective
1. Develop	Provide the infrastructure for a functional Heliophysics Digital Resource Library (HDRL)	1.1 Establish standards, protocols, procedures, documentation, equipment/software, and technology as needed
		1.2 Maintain said standards, protocols, documentation, equipment/software, technology, simulation services, and analysis services as needed
		1.3 Upgrade hardware and software to match the technology as it evolves
		1.4 Ensure adequate staffing, training, and regular review of required capabilities for HPD Data Library Curators
		1.5 Heliophysics Data Archive working group
		1.6 Ensure comprehensive datasets including CubeSats, SmallSats, sounding rockets, and balloons
		1.7 Coordinate with and across SMD regarding archival efforts
2. Unify	Identify, connect, and unify support for data providers	2.1 Identify, build, maintain relationships with new, existing, and former providers
		2.2 Provide technical assistance to missions and the data provider community to build capacity and ensure adherence to standards
		2.3 Engage and support synthetic data providers, model output providers, and model developers
3. Curate	Curate integrated heliophysics data	3.1 Ensure the long-term preservation of heliophysics datasets (i.e., 100+ years)
		3.2 Acquire/receive the data via a standardized process
		3.3 Adhere to FAIR principles and processes for existing and new datasets.
		3.4 Ingest the data according to the terms of the PDMP
		3.5 Establish and regularly update quality assurance protocols for curated data
		3.6 Maintain a standard process for a deep “backup”/cold storage
4. Operate	Serve the Public as a working <i>Digital Resource Library</i>	4.1 Staff research librarians to scope and enhance data products
		4.2 Design and host a public-facing website/interface for viewing <u>all</u> the data
		4.3 Design guides to understand how to access and use the data
		4.4 Provide support to stakeholders
		4.5 Produce and maintain data access via an application programming interface
5. Optimize	Maximize the Utility of the data	5.1 Continue to host community-provided models and provide access to modeling services
		5.2 Conduct in-house analysis
		5.3 Partner with HPD Citizen Science
		5.4 Baseline and continually track metrics
		5.5 Utilize data-informed analytics to identify opportunities enhancing services for Archives Stakeholders
		5.6 Establish and maintain access to and analysis services for heliophysics data products
6. Grow	Expand the field’s engagement with	6.1 Communicate the goals and objectives of the HDRL for existing and new stakeholders