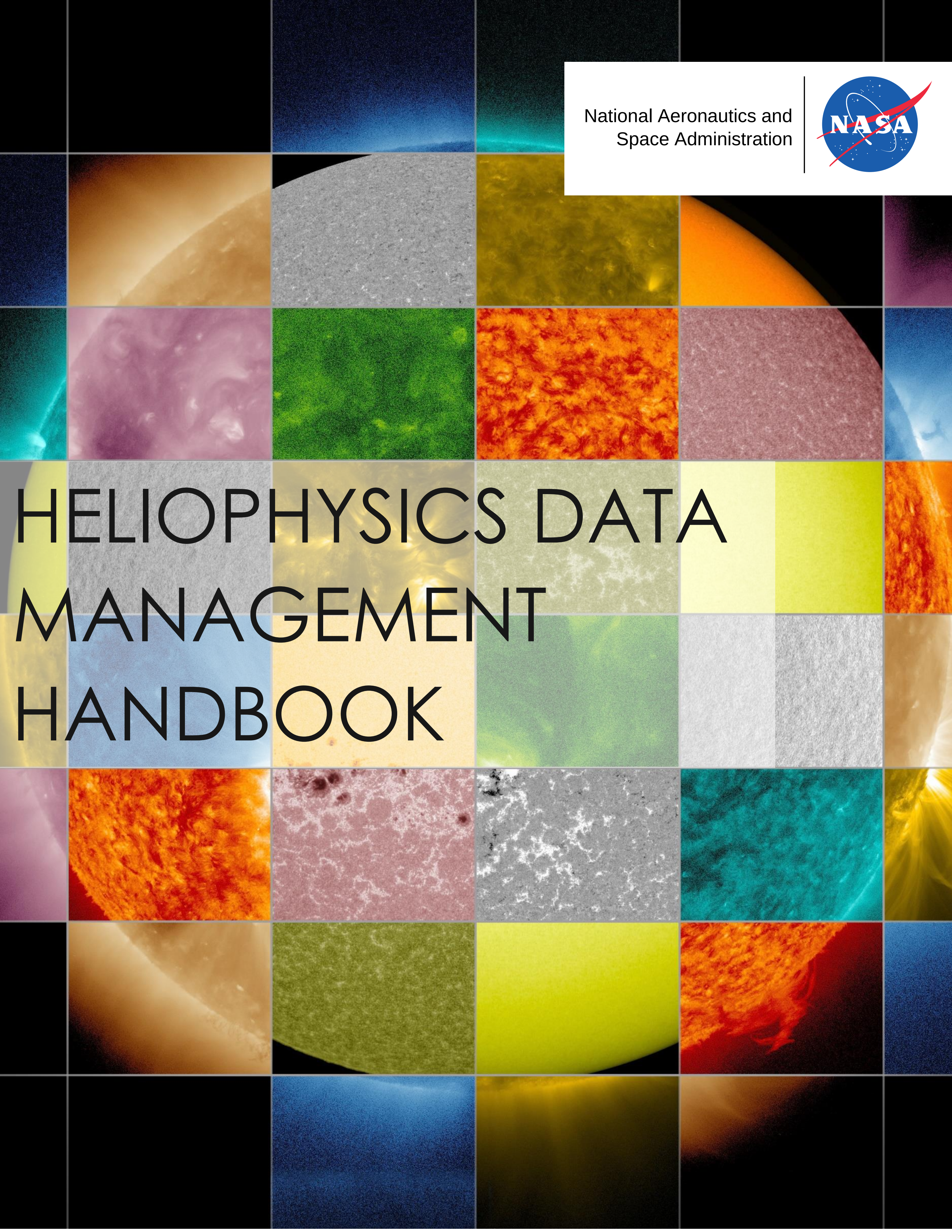
National Aeronautics and
Space Administration

HELIOPHYSICS DATA MANAGEMENT HANDBOOK



Preface

The NASA Science Mission Directorate Heliophysics Division (HPD) Data Management Handbook serves to elaborate on the governing principles and standards of the Heliophysics Digital Resource Library (HDRL) as established in the Heliophysics Science Data Management Policy.[1]

This is a living document, to be modified as needed as the HPD science program and Science Mission Directorate (SMD) science data requirements change. The contents of this document supplement the Heliophysics Science Data Management Policy and DO NOT supersede other NASA, SMD, and HPD policies. Information subject to specific laws, regulations, or policies that would prevent the release of this information—as specified in SMD Policy Document (SPD)-41a, *Scientific Information Policy for the Science Mission Directorate*[2]—are exempt from the guidance in this document.

This document will be posted publicly and will be updated on an as-needed basis. We welcome your feedback, as only through such interaction will the HDRL continue to be responsive to community needs.

Cover image: This collage of solar images from NASA's Solar Dynamics Observatory (SDO) shows how observations of the sun in different wavelengths helps highlight different aspects of the sun's surface and atmosphere. The collage also includes images from other SDO instruments that display magnetic and Doppler information. (Credit: NASA/SDO/Goddard Space Flight Center)

[1] Link to policy found on https://science.nasa.gov/heliophysics/data/
[2] https://science.nasa.gov/researchers/science-data/science-information-policy

Contents

Contents

Section 1. Purpose and Scope

1.1 Purpose

The intention of this document is to elaborate on the data management policies laid forth in the Heliophysics Science Data Management Policy. This document details how scientific information from the science programs sponsored by NASA's Heliophysics Division (HPD) is shared and managed, and its scope encompasses all phases of the mission and research data life cycles. This includes more detailed descriptions of HPD data management policies and standards, an outline of the Heliophysics data lifecycle for both HPD missions and HPD research projects, as well as an overview of HDRL's services and tools.

This document aims to be useful for existing and new HPD-funded data providers in delivering their data to the HDRL so that the data complies with NASA's Open Data Policy and is useful to the Heliophysics community. Both data management requirements and best practices are outlined within this document.

1.2 Scope

The document applies to all HPD-funded scientific activities regardless of the funding vehicle. This includes mission scientific information and research scientific information:

Mission Information

Information produced by NASA HPD-funded flight missions. Missions include strategic or flagship missions and investigations selected under Announcements of Opportunity (AOs), including those executed using a Principal Investigator-managed mission, a mission directed to a NASA implementing institution, or a mission implemented by a NASA partner.

Research Information

Information produced by investigations funded via research awards. This includes funding from investigations selected under a Notice of Funding Opportunity (NOFO) as set forth in the Code of Federal Regulations, Title 2 Grants and Agreements; NASA Research Announcements (NRAs) including those selected under Research Opportunities in Space and Earth Science (ROSES) NRAs; and other types of Broad Agency Announcements (BAAs), including Cooperative Agreement Notices (CANs).

1.3 Definitions

Per SPD-41, key terms in this document are defined as follows:

Archive: The process of storing data to ensure long term retention.

Accessible: As per the definition in the FAIR principles, data are retrievable by their identifier using standardized communications protocols.

Commercial Software: Software produced for the purposes of sale. This includes software that would be categorized as commercial-off-the-shelf (CoTS) software that NASA does not have a license to distribute.

Data: Scientific or technically relevant information that can be stored digitally and accessed electronically. This includes:

- Information produced by missions include observations, calibrations, coefficients, documentation, algorithms, and any ancillary information.
- Information needed to validate the scientific conclusions of peer-reviewed publications. This includes the data required to derive the findings communicated in figures, maps, and tables.
- This does not include laboratory notebooks, preliminary analyses, intermediate data products, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

Open Science and Data Management Plan (OSDMP): A document that describes whether and how data will be shared and preserved. An OSDMP should be compliant with NASA policies. An OSDMP includes any software that would enable the replication/reproduction of published results and any future research building on those results.

Findable: As per the definition of the FAIR principles, metadata and data should be easy to find for both humans and computers.

Information: Scientific knowledge produced as part of a research activity. This can include, but is not limited to, publications, data, and software.

Interoperable: As per the definition in the FAIR principles, data are able to work with other applications or workflows for analysis, storage, and processing.

HPD Mission: Missions include strategic[3] or flagship missions and flight investigations selected under Announcements of Opportunity (AOs). This can include investigations executed using a Principal Investigator-managed mission, a mission directed to a NASA implementing institution, or a mission implemented by a NASA partner.

Open Source Software (OSS): Software that can be accessed, used, modified, and shared by anyone, OSS is often distributed under licenses that comply with the definition of "Open Source" provided by the Open Source Initiative or meet the definition of "Free Software" provided by the Free Software Foundation.[4]

Persistent identifier: A long-lasting reference to a digital source. The digital object identifier (DOI) system is an example of a persistent identifier.

[3] Powering Science: NASA's Large Strategic Science Missions https://doi.org/10.17226%2F24857
[4] Free Software Foundation https://www.gnu.org/philosophy/free-sw.html

Permissive license: Guarantee the free use, modification, and redistribution of software while still permitting proprietary derivative works. Examples include the Apache License 2.0[5], the BSD 3-Clause "Revised" License[6], and the MIT License.[7]

Publication: Documents released through print, electronic, or alternative media. This includes peer reviewed manuscripts, technical reports, conference materials, and books. This does not include laboratory notebooks, preliminary analyses, drafts of scientific papers or preprints, plans for future research, peer review reports, or communications with colleagues.

Report: Documents produced through print, electronic, or alternative media containing Scientific and Technical information. These documents are usually not peer reviewed. This includes Technical Publication, Technical Memorandum, Contractor Report, Conference, Publication, Special Publication, and Technical Translation. This does not include interim research Grant reports.

Repository: An organized store of the data that makes data findable and accessible.

Research Activities: These activities provide information produced by investigations funded via research awards. This includes, but is not limited to, investigations selected under a Notice of Funding Opportunity (NOFO) as set forth in the Code of Federal Regulations, Title 2 Grants and Agreements; NASA Research Announcements (NRAs), including Research Opportunities in Space and Earth Science (ROSES) NRAs; and other types of Broad Agency Announcements (BAAs), including Cooperative Agreement Notices (CANs). Research and analysis activities also include other HPD-funded activities such as, but not limited to, experiments, investigations, using smallsats/CubeSats or sub-orbital platforms, field campaigns, or citizen science projects. This also includes research awards for scientific investigations that are Mission-funded via an Agency contract.

Research awards: These include grants, cooperative agreements, contracts, task orders, interagency transfers, direct internal NASA funding, and other applicable funding vehicles.

Restricted Software: Software that shall not be released due to an existing Federal law or guidance, NASA policy, or security concern. This includes software supporting security requirements described in NASA-STD-1006, Space System Protection Standard[8] or items covered under International Traffic in Arms Regulations (ITAR). For Mission software, projects should engage with the software release authority to determine their status. Examples of software that may be restricted are command related software, instrument control, authentication, or communication software.

Reusable: As per the definition in the FAIR principles, metadata and data should be well-described so that they can be replicated and/or combined in different settings. This includes releasing the data with a clear and accessible data usage policy.

---

[5] https://opensource.org/licenses/Apache-2.0
[6] https://opensource.org/licenses/BSD-3-Clause
[7] https://opensource.org/licenses/MIT
[8] https://standards.nasa.gov/standard/NASA/NASA-STD-1006

Scientific Utility: Information that is necessary to validate research findings or beneficial for future research activities.

Single use software: Software written for use in unique instances, such as making a plot for a paper, or manipulating data in a specific way.

Software: computer programs in both source and object code that provide users some degree of utility or produce a result or service.

Software project: An activity to develop software. A software project typically has a version control platform on which develop can occur collaboratively.

Source code: Human-readable set of statements written in a programming language that together compose software. Programmers write software in source code, often saved as a text file on a computer. The terms code and source code are often used interchangeably.

Section 2. HDRL Overview

HPD seeks to understand the nature and dynamical interactions of the Sun, the heliosphere, and the plasma environments of the Earth and planets based on data from the fleet of spacecraft, and appropriate ground-based assets collectively termed the Heliophysics System Observatory (HSO). This requires easy access to data and tools from a distributed set of active archives—including the Solar Data Analysis Center (SDAC) and Space Physics Data Facility (SPDF)—which make up the Heliophysics Digital Resource Library (HDRL).

The HDRL is a comprehensive repository housing the collective set of data, documentation, and tools resulting from the services of the HSO and Heliophysics research efforts.

The HDRL serves as the public face to access this data and related documentation, tools, and services—providing and assisting users in understanding data and metadata standards; archiving and serving all the data products for the long term; providing tools for data access and analysis; allowing users to run and use models to understand and predict processes in the Solar and Space Environments; and working with national and international partners to make data and services uniformly Findable, Accessible, Interoperable, and Reusable (FAIR)[9]. The HDRL maintains openly accessible data that are independently scientifically usable.

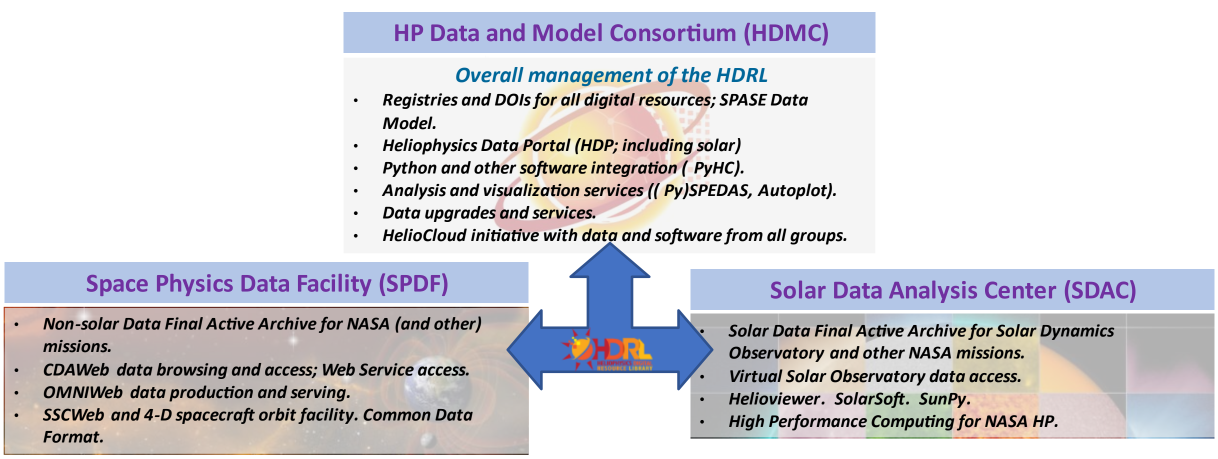


Figure 2-1 HDRL Organizational Chart

The core of the HDRL is the data that is produced by space flight missions and research activities. It is the responsibility of the HDRL to both store and serve this data, including during a mission's active phase. An active mission may have a duplicate set of data available through mission or

[9] https://www.go-fair.org/fair-principles/

instrument data centers, but the HDRL remains the primary repository of this data. The HDRL is the final location for all HPD-funded scientific data and documentation.

The HDRL is built from data systems driven by community needs and guidance and founded on community-based standards. Consistent with this approach, data providers and data users share responsibility for the quality and proper use of the data for research.

Section 3. Open Data Policies

3.1 Open Data

NASA is making a long-term commitment to building an inclusive open science community over the next decade. Open science is a movement to make scientific research more accessible and transparent to everyone at all levels. By making science more "open," researchers are given greater access to transparent research that can enable further scientific discoveries.

HPD seeks to make high-quality, high spatial and temporal resolution data publicly available as soon as possible and to adhere to the goal of early and continuing scientific data usability. This requires uniform descriptions of high-quality data products, adequate documentation, sustainable and open data formats, easy electronic access, appropriate analysis tools, and care in data preservation. These policies apply to all data sources, including space-based, airborne, ground-based, models, and in-situ sensors, and data supplied by citizens scientists, as well as supporting metadata, algorithms, source code, and documentation.

> **Why Open Data?**
>
> An open data policy is crucial for the acceleration of scientific discovery. Providing open data allows the products of NASA programs to be used by a wide community of researchers with differing backgrounds and perspectives. These communities include research and applications communities, private industry, academia, and the general public.
>
> By not impeding these researchers from achieving their scientific goals and encouraging others to engage in research, the HDRL strives to curate and serve the data products to answer fundamental science questions and further the goals of HPD.

The standards and best practices outlined in this document are designed to maximize the scientific potential of Heliophysics data, providing data that is not only available, but also usable and useful for a wider audience.

3.2 FAIR Data

Heliophysics data shall follow the **FAIR Guiding Principles** for scientific data management and stewardship. This means data should be Findable, Accessible, Interoperable, and Reusable.

The Space Physics Archive Search and Extract (SPASE) and ISTP metadata standards and FITS, CDF, and netCDF scientific data file formats are utilized by the HDRL to ensure that Heliophysics data follows the FAIR principles. Data producers are expected to adhere their data to one of the widely used data formats and work with the SPASE Metadata Working Team (SMWT)[10] in creating SPASE records for the data. See Section 6. for more details on data standards. Resources that provided guidelines and best practices for the application of metadata can be found in 6.6Appendix C.

The use of standards for data and metadata formats allows uniform data access through tools such as IDL, Python, MATLAB, or application interfaces that give researchers the ability to apply user or community developed tools for a wide array of research from single instrument to cross-mission

[10] SMWT website https://hdrl.gsfc.nasa.gov/smwt_home/smwt_index.html

analysis and visualization. A standard infrastructure – with uniformity of product documentation, formats, and access methods – allows scientists to access data from many missions from one web location or service.

Heliophysics data are to be reusable with a clear, open, and accessible data license (e.g., Creative Commons license). If there are no other restrictions on the data, the data is to be released with a **Creative Commons Zero license**[11].

> **FAIR Principles**
>
> **F**indable – metadata and data should be easy to find for both humans and computers
>
> **A**ccessible – data are retrievable by their identifier using standardized communications protocols
>
> **I**nteroperable – data are able to be worked with other applications or workflows for analysis, storage, and processing
>
> **R**eusable – metadata and data should be well-described so that they can be replicated and/or combined in different settings. This includes releasing the data with a clear and accessible data usage policy

3.3 Data Availability

To ensure HPD follows on its open data policies, the HDRL commits to the full and open sharing of heliophysics data obtained from NASA HPD-sponsored programs (e.g., observing satellites, sub-orbital platforms, field campaigns, and research) with all users as soon as such data become available, to the extent allowed by applicable law and existing NASA policies. This includes all data products, metadata, documentation, information on the instrument(s), models, and research results, as well as the source code used to generate, manipulate, and analyze them (including any calibrations, coefficients, algorithms, and ancillary data used to generate these products).[12]

In some cases, it is expected that some information will not be made public. The relevant laws, regulations, and policies that generate exceptions include but are not limited to:

- patent or intellectual property laws including the Bayh-Dole Act,
- the Export Administration Regulations (EAR),
- the Health Insurance Portability and Accountability Act (HIPAA),
- the International Traffic in Arms Regulations (ITAR),
- the Freedom of Information Act (FOIA),
- NASA STD 1006.1 Space System Protection Standard,
- NASA NPR 2810.7 Controlled Unclassified Information, and
- the Federal laws and regulations governing classified information or security
- requirements.

Timely availability of data is essential to maximizing its utility and value. Beyond the mission

[11] Creative Commons License: https://creativecommons.org/licenses/

[12] This does not include laboratory notebooks, preliminary analyses, intermediate data products, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

responsible for collecting it, timely science data and its associated metadata are necessary for validation, calibration, and use of models by not just the heliophysics community but also other groups and stakeholders with a need for space weather forecasting (e.g., human spaceflight).

HPD-funded data shall be delivered to the HDRL as they are produced and made public

Mission Data

A period not longer than ***six months*** after mission data have been downlinked may be allowed for activities such as initial calibration and validation of the data per the current SMD standard prior to the data reaching

Research Data

Scientifically useful research data shall become publicly available no later than the ***publication*** of the peer-reviewed manuscript that describes the scientific results, or at the ***conclusion*** of the research award

All users of HPD data shall not be discriminated against. Availability and sharing of data are essential to evolving the wider community's fundamental understanding of the Sun and its effects. For data products supplied from an international partner or another agency, the HDRL will restrict access only to the extent required by the appropriate agreements.

3.4 Data Usability

In general, but especially for long-term and non-specialist use, it is desirable to have data products that are ready-for-use, and thus in standard units, despiked, corrected for backgrounds, etc., and not dependent on specialized software packages. Lower-level products and the software and algorithms to use them should be archived, but these become increasingly difficult to use. One way of preserving data reduction routines is to implement them through analysis tools such as the Space Physics Environment Data Analysis Software (SPEDAS)[13] or SolarSoft[14]. Instrument-specific data reduction source code may also be preserved as documentation. Other products, such as browse plots and event lists, provide additional utility, and their archiving is strongly encouraged.

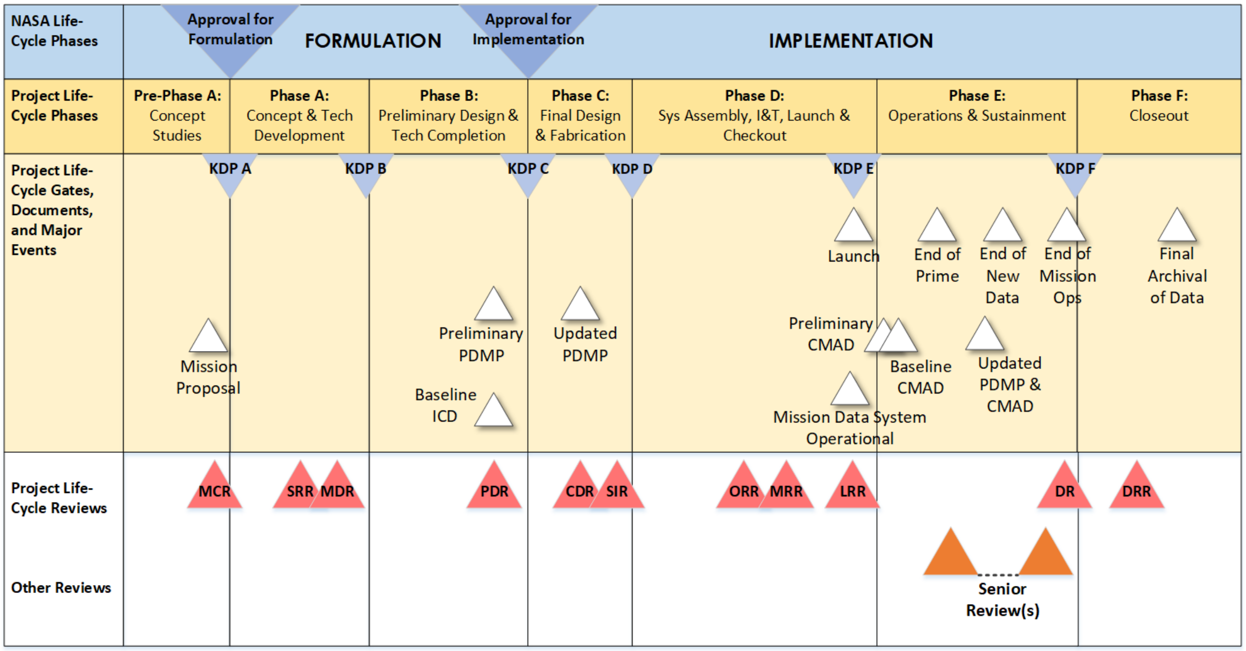
[13] https://spedas.org/
[14] https://www.lmsal.com/solarsoft/

Section 4. HDRL Mission Data Life Cycle

This section intends to help HPD missions understand how to manage their data throughout the mission life cycle. This includes details on the expectations of the missions in planning and providing their data to the HDRL as well as an overview of the important data-events in the mission timeline.

4.1 Overview of Data-Related Events throughout Mission Data Lifecycle

The data lifecycle of a mission is shown in Figure 4-1 and starts with the science objectives for a new mission or research project as captured in the Mission or Research Proposal. Understanding each data-related event is important in ensuring that missions can align their data products to the scientific goals of the mission and then deliver those data products to the wider community. A list of each data-related event for missions is provided following Figure 4-1:



Products	Pre-Phase A KDP A	Phase A KDP B		Phase B KDP C	Phase C KDP D		Phase D KDP E		Phase E KDP F	Phase F
	MCR	SRR	MDR	PDR	CDR	SIR	ORR	MRR/FRR	DR	DRR
Mission Proposal or Science Definition Report	Final									
PDMP				Preliminary	Update				Update	
CMAD							Preliminary		Baseline: Commissioning + 6 months Update at end of prime mission and at regularly intervals during extended operations (any data review and each Senior Review)	
ICD				Baseline						

CDR	Critical Design Review	MDR	Mission Definition Review
DR	Decommissioning Review	MRR	Mission Readiness Review
DRR	Disposal Readiness Review	ORR	Operational Readiness Review
ICD	Interface Control Document	PDR	Preliminary Design Review
KDP	Key Decision Point	SIR	System Integration Review
LRR	Launch Readiness Review	SRR	System Requirements Review
MCR	Mission Concept Review		

Figure 4-1 Heliophysics Mission Data Lifecycle and Product Maturity Matrix

The schedule of significant data-related events in the lifecycle of a HPD mission is as follows:

1. **Pre-Phase A:** A mission proposal, or similar pre-formulation document, defines the science investigation(s) to be conducted, the measurements required, and a science payload (with a fidelity that varies by document).
2. **Preliminary Design Review (PDR):** A preliminary Project Data Management Plan (PDMP) (see Section 4.2), prepared in consultation with the HDRL, will state the mission's data requirements and how the mission's data system will be implemented in order to meet the Mission's requirements. One of the requirements on the mission data system will be the production and open distribution of independently usable data, usually in conjunction with the HDRL.
3. **Critical Design Review (CDR):** The PDMP is updated.
4. **Mission Readiness Review/Flight Readiness Review (MRR/FRR):** The PDMP and CMAD will be updated as necessary.
5. **Launch:** The mission's data system as laid out in the PDMP will be operational and ready for production and distribution of the mission's science data once initial testing and calibration have been completed. Routine data production and, usually, continual data reduction improvements mark the active phase of the mission.
6. **Commissioning:** Six months after commission, a preliminary Calibration and Measurement Algorithm Document (CMAD) (see Section 4.3) will provide a detailed description for the processing of the telemetry into scientifically useful quantity. A baselined CMAD will be delivered twelve months after commissioning.
7. **Senior Review(s):** The mission continues to update its PDMP and CMAD, with significant focus on its plans for the final archiving of the data while working with the relevant archive(s) to refine plans for transition of mission data and tools to the archive(s).
8. **The end of the mission operations:** When science operations end, the mission shall verify that final, well-documented products have been transitioned to the HDRL in accordance with the PDMP.
9. **Final closeout of the mission:** All data and applicable source code have been delivered to the HDRL per the mission PDMP.

A more detailed description of the roles and expectations during these events is described in Section 4.4

4.2 Project Data Management Plan

All NASA HPD missions must document the implementation of their data management policies in the form of a PDMP. A PDMP is the interface document between NASA, the mission systems, and the instrument teams that describes the science and ancillary data associated with the mission and how the data will be managed. The PDMP defines the data, processing approaches and implementation, data and documentation products, data availability, and storage and archival strategies. It also defines the access method(s) for the heliophysics scientific community.

Missions are expected to work with the appropriate archive to ensure that the appropriate level of detail is captured in the PDMP. For Missions that plan to use multiple repositories to store their data (e.g., cross-division or international collaborations), the PDMP is the place to designate the primary archive to store their data.

PDMP Template

A PDMP template is provided for missions to capture how their data will be managed A baselined (signed) PDMP will help ensure that the scientific data is standardized and organized in accordance with the Heliophysics Science Data Management Policy, enabling long-term accessibility and utility for the wider heliophysics community. The HDRL uses the PDMP to help inform infrastructure and resource utilization and planning.

https://science.nasa.gov/heliophysics/data/

The PDMP shall be developed and available according to the timeline for the Science Data Management Plan within the Project Plan Control Plans Maturity Matrix in NASA Procedural Requirements (NPR) 7120.5[15], or at the discretion of the Program Scientist.

Typically, the PDMP will be available in draft form at the time of Preliminary Design Review for the mission and signed at the time of the Operational Readiness Review. The PDMP may be revised at the Flight Readiness Review/Mission Readiness Review as well as the next Senior Review or at the discretion of the Program Scientist. Refer to section 4.1 for an overview of important data events throughout the mission life cycle.

Each data provider will be expected to generate and make available metadata and other supporting material on the data products, spacecraft, and instrumentation appropriate to their investigation. The details of these will be defined during discussions with the Project and Program personnel during the drafting of the PDMP. The intent of such metadata and materials will be to make the data correctly and independently usable for science investigations.

Baselined PDMPs, as well as any updates made over the life of the mission, shall be submitted to the HDRL and be accessible by the public via the HDRL in a machine-readable format.

4.3 Calibration and Measurement Algorithm Document

CMAD Template

The HDRL provides a CMAD template for PI-led missions to follow. The CMAD is useful in understanding the production of data products and can aid in the development of additional data reduction tools.

https://science.nasa.gov/heliophysics/data/

Principal Investigator (PI)-led missions must provide a CMAD while directed missions must generate a CMAD for each instrument.

The CMAD is the interface document between NASA, the mission systems, and the instrument teams that describes the overall concept for calibrating a specific instrument. It describes how the mission will conduct preflight and inflight calibrations and provides a detailed description for the processing of the telemetry into scientifically useful quantities in geophysical units including calibrations, error analysis, and documentation issues.

[15] NPR 7120.5 https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=5F

The CMAD identifies sources of input, provides the physical theory and mathematical background underlying the use of those inputs to retrieve data products, and describes what has been accounted for during algorithm development.

The CMAD is developed and available according to the same timeline for the PDMP, or at the discretion of the Program Scientist. Typically, the CMAD will be available in draft form at the time of Preliminary Design Review for the mission and signed at the time of the Operational Readiness Review. The CMAD may be revised at the Flight Readiness Review/Mission Readiness Review as well as the next Senior Review or at the discretion of the Program Scientist.

Baselined CMADs, as well as any updates, shall be distributed to the HDRL and be accessible by the public via the HDRL in a machine-readable format.

4.4 Interface Control Document

Missions using the Solar Data Analysis Center (SDAC) for their solar data must generate an Interface Control Document (ICD). The ICD describes the details of the interface agreement between the two entities and their respective responsibilities for data delivery of the relevant datasets and webpages from the mission. This ICD also lists the datasets and products that the mission will deliver to SDAC along with their expected delivery schedule and availability at the SDAC.

ICD Template

The HDRL provides an ICD template for HPD-funded missions to follow. The ICD defines the mission's interface and associated performance requirements between the mission SOC and the SDAC.

https://science.nasa.gov/heliophysics/data/

4.5 Expectations Throughout Mission Data Lifecycle

Contact the relevant archive as early as possible

Either SDAC for space-based solar-physics missions or SPDF for non-solar Heliophysics missions.

It is imperative that missions begin a working relationship with the relevant, primary archive as early as possible. The HDRL needs to be aware and understand the expectation of what data will be served from the mission. Contacting the archives helps ensure that data will be released on schedule and provides the mission with a resource in how to handle their data products. Preferably, the mission should reach out to the HDRL prior to the proposal submission (pre-phase A). During Phase B, the mission and HDRL should discuss what additional services a mission expects the archives to provide following the end of mission operations. Initially, this may involve developing a notional science workflow to identify HDRL services needed by the data consumers. The result of this planning will then be formalized in the PDMP and CMAD. Services and data products can change during a mission’s lifetime, so expect the contents of an agreement between the HDRL and a mission to change over time, with appropriate and timely documentation.

During the Implementation phases prior to launch, missions are to provide data descriptions and sample data products to the relevant archive along with a baselined PDMP and CMAD so that the HDRL has the information to prepare the appropriate capacity and services to eventually host the data. All data that is provided by the mission must adhere to existing data and metadata standards (detailed in Section 6.) and must have supporting material (e.g., documentation, software) as needed to ensure independent data usability.

Data User Workflows

Missions should anticipate and define the workflows they expect the users of their data to follow, both during mission operations and more importantly, after.

Do the science workflows during and after mission operations need to be the same? If so, why?

Missions explicitly thinking about how they expect users to work on the mission science after mission operations are complete would be a useful exercise that would help in defining operational mission and HPD archive data systems and services. Operating and future missions may also want to think about their science in the context of workflows supportable in cloud environments.

A Mission Science Operations Center (SOC) – or equivalent – provides calibrated data products to the HDRL for the HDRL to distribute during the active phase of the mission. While the mission may serve data products from its own repository (such as a mission-specific website) that is consistent with the *Desirable Characteristics of Data Repositories for Federally Funded Research*[16], **the HDRL is ultimately the main archive of the data to the public.** Having the data in the HDRL makes it possible to implement easy-to-use multi-mission tools and ensures long-term archiving.

It is recommended that the mission provides a point of contact (POC) who works closely with the SOC to communicate with the applicable archive, especially for phases A through D (pre-launch). A POC will help better facilitate the eventual transfer of data from SOC to HDRL.

Missions using the SDAC will develop an ICD that describes the interactions and respective responsibilities for data delivery between the mission and the SDAC.

For missions using the Space Physics Data Facility (SPDF), an SPDF curation scientist should periodically attend mission science team meetings as well as mission telecons, typically meeting with the mission team 2-3 times per year.

Data is to be transferred from the SOC to the HDRL within a period that is no longer than 6 months for the distribution and open use of the data via the HDRL. If there are latency requirements for the data products, this will be communicated to the HDRL in the PDMP. For additional guidance, missions should refer to *Requirements for Data Delivery to the Heliophysics Archives (SPDF and SDAC) for New Missions*[17].

16 https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf
17 https://spdf.gsfc.nasa.gov/guidelines/archive_newdata_reqt.html

Data production is continually modified with new data, analysis methods, etc. as needed to provide the best possible data for scientific analysis. The use of data evolves as the mission matures; therefore, the data system and mission standards may change. The evolved mission data system will be captured in the PDMP as it is revised.

It is expected that missions think about how to archive the data so that the data is accessible and useful long after mission operations. Data producers are to ensure that all final, calibrated, and well-documented data products are provided to the relevant archive at closeout and share any relevant algorithms, scripts, or code that can assist in using the data. This includes ensuring the preservation of Level 0 data and documents describing how to calibrate and validate this data. Supporting material is essential for users to interpret the data long after mission closeout when the mission may no longer be able to provide support.

Data Preservation

The definitive dataset containing all of the science potential of the investigation, in that no irreversible transformations have been applied to the data, shall be archived along with tools for its reduction to science products and documented algorithms for this process. Relevant engineering and "housekeeping" data should also be preserved.

In the early stages of a mission's prime operational phase, data reduction routines may be continually updated and refined, and the best products may only be produced for time intervals of high interest. These early tools lead to the ability to routinely produce high-quality, carefully documented data that is served to the community using standard formats and delivery mechanisms. As the mission progresses, and especially in the extended phase, products mature and the argument for further data refinement becomes less compelling.

An overview of interactions and expectations for Data Producers and the HDRL throughout the Heliophysics data management lifecycle are laid out in Figure 4-2.

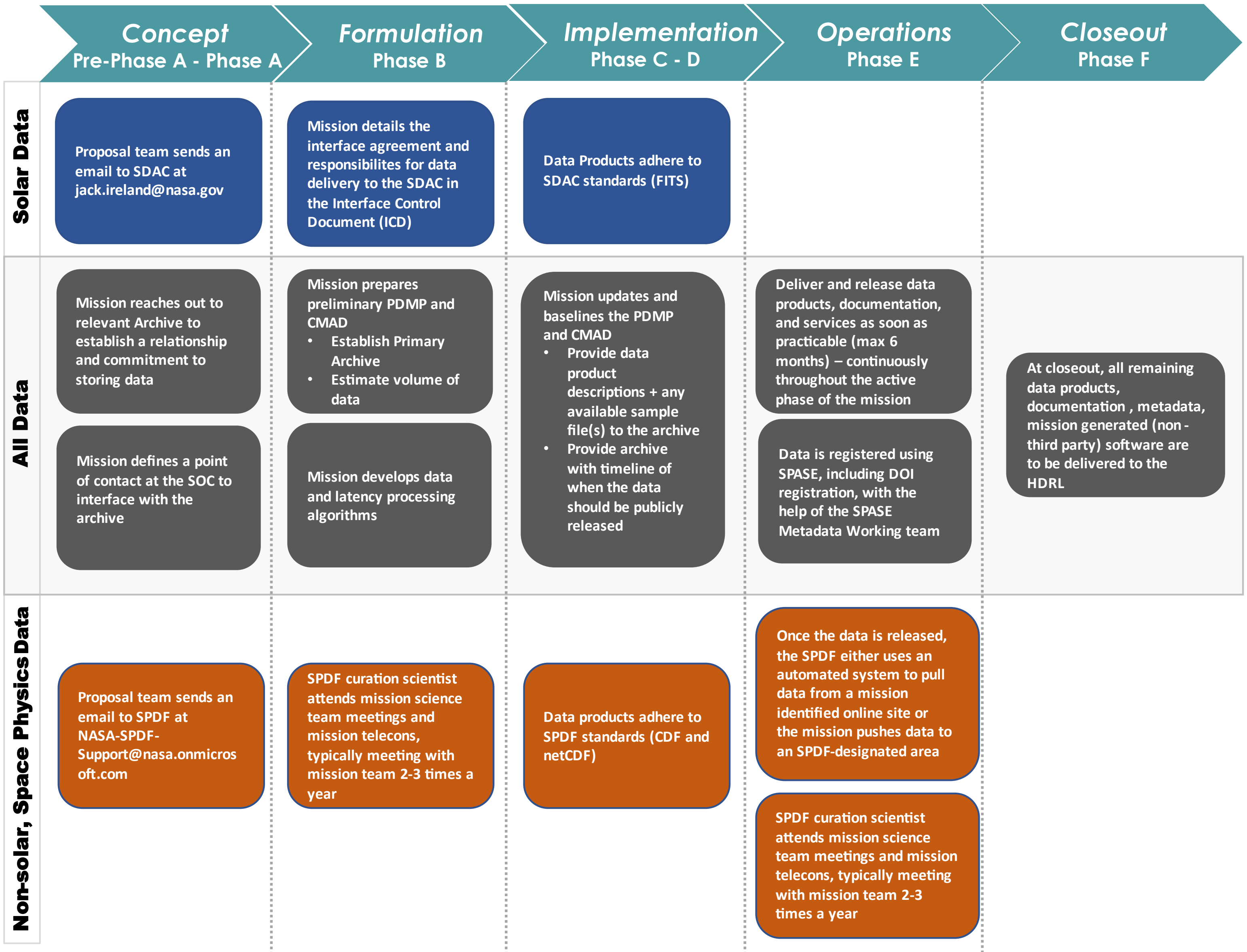


Figure 4-2 Mission and HDRL Interactions

Section 5. Research Data Life Cycle

Complementary to the missions are Research and Analysis (R&A) projects. These are projects funded from investigations selected under NRAs, commonly solicited through ROSES, and they provide data analysis and theory to understand the data from the missions as well as data from more rapid access to space (via sounding rockets, balloons, ground-based observations, CubeSat/smallsat programs, etc.).

5.1 Data Life Cycle for Research Projects

The HPD Research Project life cycle adapted from NPR 7120.8A[18] is shown in Figure 5-1.

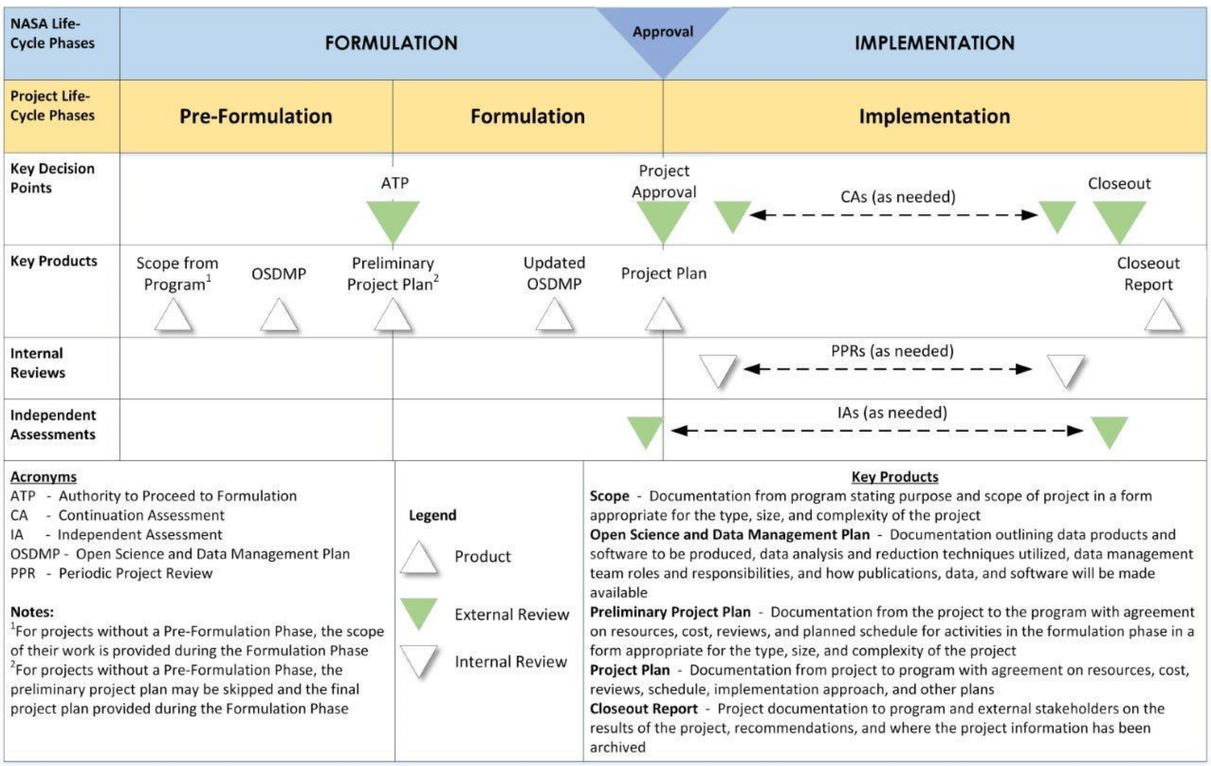


Figure 5-1 Research Project Life Cycle

R&A Projects are expected to prepare the Open Science and Data Management Plan (OSDMP) along with their Project Plan. Data products are produced during the implementation phase and are delivered according to agreements established in the OSDMP.

HPD research data shall become publicly available no later than the publication of the peer-reviewed article that describes it, or at the conclusion of the research award. All research data

[18] NPR 7120.8A https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=8A

must adhere to existing data and metadata standards (detailed in Section 6.) and must have supporting material (e.g., documentation, software) as needed to ensure independent data usability.

No later than the research data is publicly available, the research project must contact the relevant archive to organize the transfer of the data to the HDRL, either to SDAC for space-based solar-physics missions or to SPDF for non-solar Heliophysics missions.[19]

Figure 5-2 is an overview of the roles and expectations for the data producer, NASA HQ, and the Archives throughout a Heliophysics R&A project timeline. For each phase of the project, steps are outlined for each of the relevant parties to ensure that R&A projects are able to capture their data and the HDRL can ensure this data is archived and usable by the larger community.

5.2 Open Science and Data Management Plan

To broaden access to the results of NASA-funded research, all proposals must contain an Open Science and Data Management Plan (OSDMP), which will be evaluated as part of the Intrinsic Science Merit. Proposers should review Section 1.6 of B.1 The Heliophysics Research Program Overview for the required content of the OSDMP.

The OSDMP must describe:

1) the data products to be produced and the plans for sharing and archiving that data;
2) the plan for archiving and providing open access to any software developed;
3) roles and responsibilities of team members for data management;
4) expected documentation and/or publication of the results;
5) any open science activities associated with the project.

Projects are to update the OSDMP throughout the project implementation phase, as needed, to adjust for any changes in data plans. Baselined OSDMPs, as well as any updates, shall be distributed to the HDRL and be accessible by the public via the HDRL in a machine-readable format.

OSDMP Template

Proposers are strongly encouraged to refer to the Heliophysics Division's OSDMP template. A baselined OSDMP will help ensure that the scientific data and software produced by the research activity is standardized and organized in accordance with the Heliophysics Science Data Management Policy, enabling the long-term accessibility and utility for the wider heliophysics community. The HDRL will use the OSDMP to help inform infrastructure and resource utilization and planning.

https://science.nasa.gov/researchers/templates-heliophysic-division-appendix-b-roses-proposals

[19] https://spdf.gsfc.nasa.gov/guidelines/archive_newdata_reqt.html

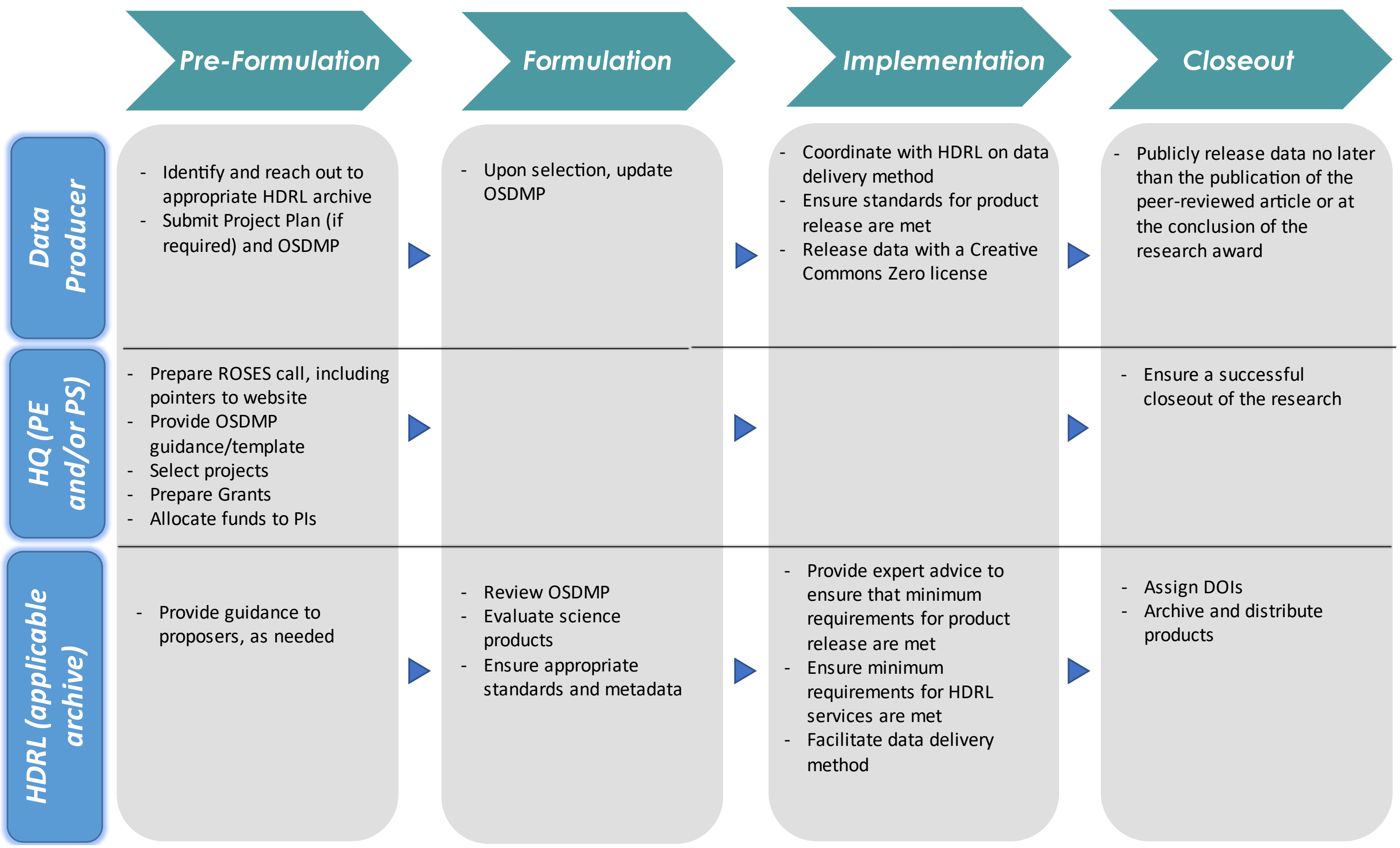


Figure 5-2 Heliophysics R&A Project Data Lifecycle Roles and Responsibilities

5.3 Citizen Science

Citizen Science is defined as a form of open collaboration in which individuals or organizations participate voluntarily in the scientific process in various ways. "Citizen Science Projects" are science projects that rely on volunteers to analyze or produce data. These volunteers are welcome at all scientific levels, making these projects important for inclusion and outreach, which are critical in achieving SMD's open science goals. Therefore, data utilized or created by citizen science projects shall be introduced to ensure these volunteers are encouraged in contributing to HPD's scientific principles and goals. HPD provides multiple opportunities for citizen science participation[20] and periodically solicits citizen science projects via the Heliophysics Citizen Science Investigations (H-CSI) program. In accordance with HPD policy, data collected as part of HPD crowdsourcing projects or citizen science projects shall be made public. All proposed and funded citizen science projects must develop an OSDMP.

Citizen scientists, those who are recruited by the project, shall work with the PI and the research team to provide necessary metadata and documentation. The PI and the research team are to specify the data collection requirements and citizen scientists are expected to follow these requirements. HPD-funded citizen science projects are encouraged to use the data formats and metadata standards described in Section 6. to increase interoperability with other Heliophysics observation data. Additional citizen science guidance and best practices can be found in the *NASA Citizen Science Program Handbook*.[21]

---

[20] https://science.nasa.gov/heliophysics/programs/citizen-science/
[21] https://docs.google.com/document/d/11R9xt6Hg0Zs_ZdfyGLC6W1PjQS_uYFlA_0pptwlD9Pg/edit

Section 6. Standards

This section provides expectations and overviews on the standards and data formats commonly used within the Heliophysics data environment.

6.1 Space Physics Archive Search and Extract (SPASE) Data Model

Central to the success of the HDRL is a uniform set of terminology to describe products and their sources. This allows the HDRL to make registry of data products that is useful for the search and discovery of relevant data. To foster interoperability between the various partners in the HDRL, NASA HPD has sponsored the SPASE collaborative, consisting of scientists and software designers from a number of U.S. and international institutions, to develop the SPASE Data Model that will allow uniform descriptions of products and services.

SPASE Tutorials
Data providers are to work with the Space Physics Archive Search and Extract (SPASE) Metadata Working Team (SMWT) in creating SPASE records for the data collections. SPASE records are used to populate the HDRL as a discipline-wide data registry and will be used to create Digital Object Identifiers (DOIs) for the data collections, that can be cited in science papers. For tutorials and more information, visit the SPASE school website https://school.spase-group.org/

The SPASE Data Model is a set of terms and values along with the relationships between them that allow describing all the resources in a Heliophysics data environment[22]. The intent of the SPASE Data Model is to provide the means to describe resources, most importantly scientifically useful data products, in a uniform way so they may be easily registered, found accessed, and used. The SPASE Data Model unifies descriptions of data from observatories or simulation models using the same language to describe all products, streamlining finding accessing, and using underlying variegated fields thus speeding up and improving research. Combined with other standards, like the data formats outlined in Section 6.2, SPASE facilitates an integrated approach in which any required data can be quickly found and accessed, allowing the ability to use a multiplicity of data sets for a given project.

The SPASE Data Model is consistent with the FAIR principles. SPASE descriptions allow users to easily interact with the data and request subsets of data using application programming interfaces (APIs) using any appropriate programming language (Java, Python, etc.).

In addition, SPASE Product IDs provide unique identifiers of Heliophysics data products and hence can be used to create Digital Object Identifiers (DOIs) for the data collections. The SPASE

[22] SPASE group https://spase-group.org/index.html

records are used to create DOIs for the data collections that can be cited in scientific papers. The SPASE group intends to register all relevant Heliophysics data resources, including space, ground, and model-based data.

As requirements arise that cannot be met by the current SPASE model, from new missions or other sources, the SPASE consortium works with any members of the community to address these needs. For existing missions, the funding for providing SPASE metadata will come through the HDRL. New missions may include the relatively small, required funding for basic SPASE descriptions, as described above, in their budgets.

Digital Object Identifiers (DOIs)

DOIs are unique, persistent identifiers that do not change over time. DOIs have long been routinely used for assigning unique identifiers to journal articles or books, but they are now the *de facto* standard for registering datasets.

The SPASE group offers a service that makes acquiring a "Digital Object Identifier" easy for data providers.

For missions, the HDRL has already compiled product descriptions for nearly all current and past NASA missions, so all that is needed in order to mint DOIs for datasets is to get agreement on the publication information, namely, the *title¸ creator(s), publisher(s), and publication year.*

https://hdrl.gsfc.nasa.gov/DOI_Initiative_October_2020.pdf

6.2 Data Formats

In accordance with the FAIR principles, it is NASA's policy to optimize data for ease of accessibility and use, including preparing data for long-term stewardship and to be machine readable for automated processing. Consequently, HPD data is to be made in available in convenient, modifiable, and open formats as well as have robust, standards-compliant metadata that clearly and explicitly describe the data. HPD data providers are to provide their data in one of the following self-describing formats:

Name	Relevant Archive	Links	Description
Flexible Image Transport System (FITS)	*SDAC*	https://fits.gsfc.nasa.gov/	Typically used for HPD *solar images*, FITS is a file format endorsed by the International Astronomical Union designed to store, transmit, and manipulate scientific images and associated astronomical data for long-term archival usage.
Common Data Format (CDF)	*SPDF*	https://cdf.gsfc.nasa.gov	Typically used for HPD *space physics data*, CDF is a self-describing format maintained by SPDF used for the storage of scalar and multidimensional data. When using CDF, follow the ISTP/IACG Guidelines on the SPDF website and dataset and file naming guidelines.
Network Common Data Format (netCDF4-Classic)	*SPDF*	https://www.unidata .ucar.edu/software/netcdf/	Typically used for major *data sets closer to the Earth in the Ionosphere-Mesosphere-Thermosphere realm*, netCDF is a self-describing, portable, scalable, appendable, sharable, and archivable data model. NetCDF has two main data models, the classic (netCDF-classic) and the enhanced model (netCDF4). NetCDF4 is an extension of the classic model and introduces a third binary format that uses Hierarchical Data Format (HDF5). NetCDF4 offers more data representation features and data types (such as groups, compound types, unsigned interger types, parallel I/O, etc.) at the expense of some additional complexity. For netCDFs, use netCDF4-Classic plus the string data variable type, and use time as the unlimited dimension, but don't include groups, unsigned 64-bit integers, or user-defined variable types. Follow the ISTP/IACG Guidelines on the SPDF website and dataset and file naming guidelines.

- **Note that the SDAC uses FITS, while SPDF utilizes both CDF and netCDF**

It is crucial that HPD data is self-describing and machine-readable so that users can easily access and understand the data such that the data is independently usable long after the data has been archived. Data that are difficult to read by the community, or have non-standard components in any way, are barriers to scientific discovery. The HDRL uses these data formats to bring uniformity to its repository, allowing for a greater usage of the data across different sources.

Data files shall contain fully compliant metadata and organization following the Space Physics Guidelines for CDF [and netCDF][23] or FITS standards[24] as appropriate.

Supporting information of scientific utility, such as documentation and graphics, may use commonly accepted formats including, but not limited to, PDF, PNG, and JPEG. Projects may also work with the applicable archive to obtain a waiver for unsupported file formats to ensure the information is not lost.

6.3 Code and Software

Code and software developed by HPD data producers are often used to convert lower-level data products to a higher-level. While the CMAD for missions covers at a fundamental level how these conversions are made, it is important for all data producers to share their code and software to assist in making these conversions in the future. By properly documenting and openly sharing code and software produced by the data producer, the data products can be reusable and better follow the FAIR principles.

Additionally, many data producers develop data analysis software (including web applications) that help them browse, understand, and analyze their data. However, relying on this software after a mission is completed is not suitable for widespread and future use and long-term archiving. Therefore, this software should be made available and archived.

Along with the NASA-generated data products, data producers are to provide and make available any algorithms or first-party software used to generate these products. This means:

HPD-funded software shall be released under a permissive license that has broad acceptance in the community and shall be made available in a publicly accessible repository that is widely accepted by the community.

- *If software would be developed but not made available under a permissive license, there must be an explanation why.*
- *If commercial software is used, then HPD-funded source code will be released and made available.*

[23] https://spdf.gsfc.nasa.gov/sp_use_of_cdf.html
[24] https://fits.gsfc.nasa.gov/fits_conventions.html

Mission software is software developed or used to support HPD missions. This is software typically developed under NPR 7150.2 *NASA Software Engineering Requirements*[25] or other requirements tailored for the missions. Mission software is often restricted software and shall not be shared openly unless approved for release, though software may also be unrestricted. For missions, the classification and examples of software is described in NPR 7150.2, Appendix D.

Research software is software produced by investigations funded via research awards. This software should be developed and released as described in the project's software management plan. Scientifically useful software developed under a research award shall be released no later than the time of publication or at the end of the research award.

Open-source software can be classified as documentation and archived and is to be citable using a DOI.

Examples of permissive licenses include the Apache License Version 2.0[26], BSD[27], and the MIT License[28].

Sharing of well-documented code is important to show how higher-level data products were generated. Software is constantly changing, so sharing code makes it easier to account for these changes and any repeat data product generation. In addition, sharing code facilitates knowledge sharing and allows others to utilize the code for their own scientific purpose.

Unrestricted software that is developed using HPD-funding is expected to be shared openly. This software must be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community. These platforms improve the efficiency in the development of the software, enables reuse of the software, and shares openly the software developed with public funding. Furthermore, the sharing of mission software supports reproducibility for the scientific results. GitHub[29] and Zenodo[30] or equivalent tools and archives may be used for archiving and dissemination of software.

All open-source projects on GitHub, Zenodo, or equivalent platform should contain documentation including but not limited to a README that is the instruction manual that explains to users why the project is useful, how to get started, and how to build and use the software. Additionally, HPD-funded software projects that allow others to contribute must include a code of conduct and guidelines for how to make those contributions.

Restricted software (e.g., license restrictions or patent protections) should be discussed with the program officer, as restricted software does not have to be released.

[25] https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7150&s=2
[26] https://www.apache.org/licenses/LICENSE-2.0
[27] https://opensource.org/licenses/BSD-3-Clause
[28] https://opensource.org/licenses/MIT
[29] https://github.com/
[30] https://zenodo.org/

The following table provides guidelines for sharing different types of research software[31].

Research Software	
Software Type	**Guideline**
Single Use Software	Software written for use in unique instances, such as making a plot for a paper or manipulating data in a specific way. Single-use software developed from SMD-funding shall be released no later than with the publication it supports. It can be released as supplementary material to the publication or shared in a data repository that produces a permanent identifier that is linked to from the publication.
Libraries	Generic tools, often with a larger user base, implementing well-known algorithms, providing statistical analysis or visualization, etc., that are incorporated in other software categories. Libraries developed using SMD funding shall be released no later than the publication it supports. Libraries should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source libraries. Major versions or releases of the software should be archived in a NASA designated repository.
Analysis Software	Generalized software (not low-level libraries) used to manipulate measurements or model results to visualize or gain understanding. This software often evolves from single-use utility software and may incorporate libraries. Analysis software developed from SMD-funding shall be released no later than the publication it supports. Analysis software should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source analysis software. Major versions or releases of the software should be archived in a NASA designated repository.

[31] https://www.nap.edu/catalog/25217/open-source-software-policy-options-for-nasa-earth-and-space-sciences

Research Software	
Model and Simulation Software	Software that either implements solutions to mathematical equations given input data and boundary conditions or infers models from data. Includes first-principles models, data-assimilation tools, empirical models, machine learning, mission planning and engineering tools, among others. They often use libraries. Model and Simulation software developed from SMD funding shall be released no later than the publication it supports. Model and Simulation software should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contributed to existing, open source software. Major versions or releases of the software should be archived in a NASA designated repository.
Data Processing	Software for processing uncalibrated sensor measurements into calibrated sensor data and derived data products. This software type applies calibration coefficients, corrections or algorithms, which may be dependent on forward modeling, simulated observations, equations, and data filtering. It may include modeling and simulation software and libraries. Data processing software developed from SMD funding shall be released no later than the publication it supports. Data processing software should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source software. Major versions or releases of the software should be archived in a NASA designated repository.
Commercial Software	Software produced for the purposes of sale. This includes software that would be classified as commercial-off-the-shelf (CoTS) and software that NASA does not have a license to distribute. Commercial software is not required to be released. SMD-funded research software developed in a commercial language (a programming language that requires a license to compile or run software) is not considered commercial software and shall be shared in a similar manner as other types of software developed in open languages.

6.4 Models and Simulations

Scientifically useful data from models and simulations must adhere to the timeline for sharing scientific data. As with all other scientific data resulting from SMD-funded scientific activities, model and simulation output must be made publicly available and adhere to existing metadata and DOI standards. Mission-specific outputs may be archived at the relevant HDRL archive. Model and simulation outputs for R&A projects may reside at whatever publicly accessible archive the

PI chooses, given that the archive provides a persistent DOI and is consistent with the *Desirable Characteristics of Data Repositories for Federally Funded Research*.

Simulation and model output may be registered by the Community Coordinated Modeling Center (CCMC). CCMC is a jointly funded federal interagency organization that provides modeling capabilities to the heliophysics community and to groups interested in space weather forecasting[32]. The CCMC uses a standardized model on-boarding process to ensure any given model installation is successful and useful for the community. [33]

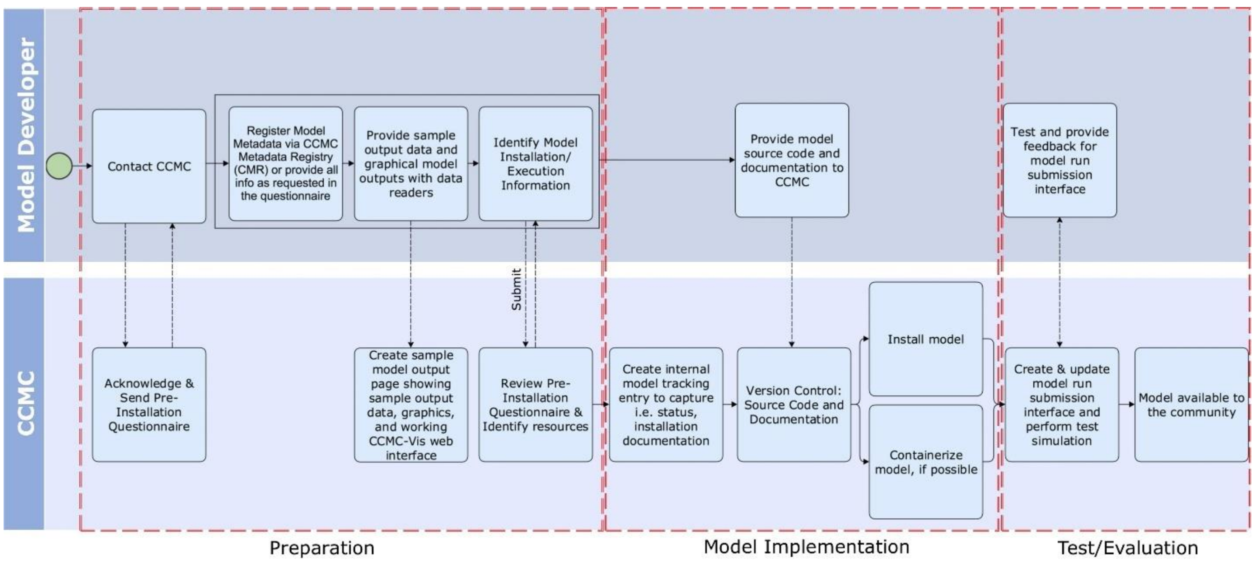


Figure 6-1 CCMC Model On-boarding Process

For simulation output, metadata regarding the model and model run(s) are required and shall conform to the requirements set by the CCMC. Proposers should work with the CCMC to produce model output readers and interpolators compatible with the CCMC Kamodo Open-Source Project[34].

6.5 Documentation

To ensure independent usability and reproducibility, data products and their purpose shall be clearly described. Submissions to the HDRL are expected to have additional descriptions of the instrument, data collections, and processing necessary to make the data fully understandable for long-term and correct use by non-experts.

In addition, the steps in creating each level of data product shall be clearly documented and fully described to ensure that data products are understandable and reproducible after the mission ends. For more information on the levels of data, see the Appendix.

Documentation shall be in formats that are widely-used and actively maintained (e.g., ASCII text and PDF), with free and open-source definitions, and free and open-source file reading/writing

32 https://ccmc.gsfc.nasa.gov/faq/
33 https://ccmc.gsfc.nasa.gov/model-onboarding/
34 https://github.com/nasa/Kamodo

software support. Source files that compile to documentation can be provided (e.g., .tex files) but the compiled, readable documentation must also be provided in standard formats.

6.6 Publications

All HPD-funded publications—scientific and technical documents funded by HPD or reporting on HPD-funded research which is released through print, electronic, or alternative media—shall be made publicly accessible via a NASA designated repository at the time of their publication.[35] These include peer reviewed manuscripts, technical reports, conference materials, and books. This does not include internal reports, laboratory notebooks, preliminary analyses, drafts of scientific papers or preprints, plans for future research, peer review reports, or communications with colleagues.

An openly accessible version of the as-accepted, peer-reviewed manuscripts shall be available via PubSpace at the time of publication. There are two options for how to comply with this requirement: Either (1) the manuscript is individually uploaded to NASA PubSpace by the time of publication, or (2) it is published in a journal that makes it openly available at the time of publication and also it is indexed by either the Clearinghouse for the Open Research of the United States (CHORUS) publishing group or the Astrophysics Data System (ADS). More information on Public Access can be found on the Scientific and Technical Information (STI) website (https://sti.nasa.gov/research-access/). The process for NASA civil servant and contractor (with a NASA Identity) submissions to PubSpace is shown in Figure 6-2. The PubSpace submission process for NASA-external grantees, contractors without a NASA Identity, and Cooperative Agreement holders is shown in Figure 6-3. Further guidance on how to make the publication publicly accessible is available at https://sti.nasa.gov/submit-to-pubspace/. Publishing the manuscript as Open Access and posting a version on a community recognized preprint server are encouraged.

[35] Office of Science and Technology Policy *Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*, https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

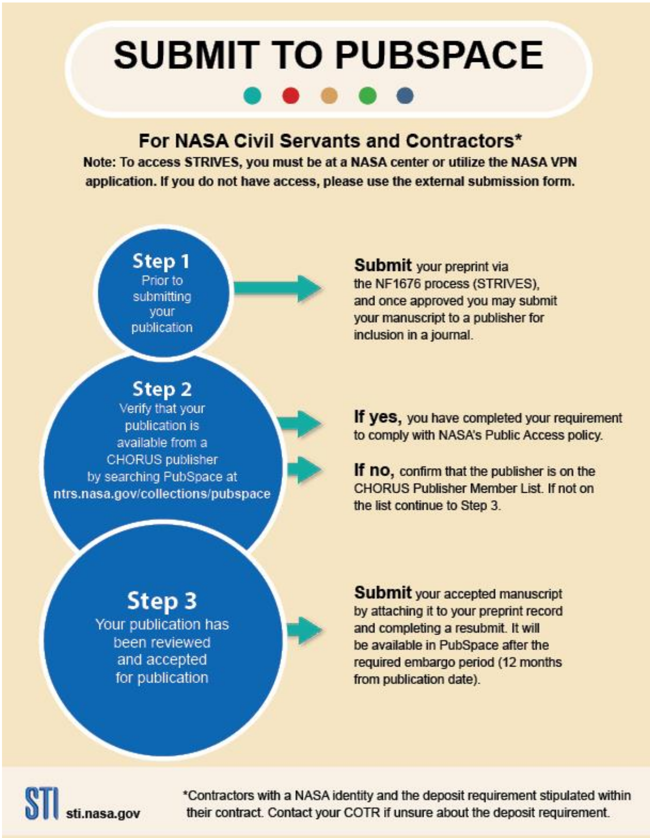


Figure 6-2 PubSpace Submission Process for NASA Civil Servants and Contractors with a NASA Identity

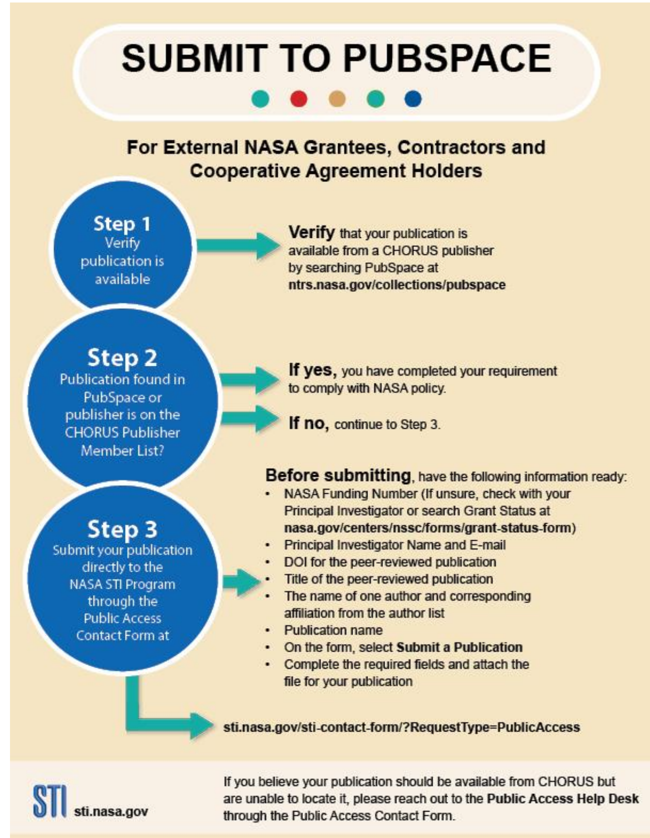


Figure 6-3 PubSpace Submission Process for External NASA Grantees, Contractors without a NASA Identity, and Cooperative Agreement Holders

Appendix A. Community Services and Tools

For any services or tools to be consistently useful, they need to become findable or searchable by the HDRL, not just grant-funded projects that may or may not be renewed in the next grant cycle. To this end, several projects have been identified by various groups within the HDRL or in Senior Reviews and endorsed by NASA HQ as being of lasting value. Some of these have been incorporated into the HDRL and CCMC, and others are supported through grants or contracts with the HDRL Project. These services provide both tools for individual researchers and the means for missions to avoid duplication of effort in designing data systems. The following tools are increasingly used in both those ways. This section provides a current snapshot of the various services available to the heliophysics community but is not intended to be an exhaustive list. Tools directly supported by the HDRL are denoted with an asterisk (*). The HDRL will develop recommendations, standards, and best practices enabling others to register their tool and services such that they are findable and usable by the heliophysics community via the HDRL.

Services/Tools sponsored by HDRL

CDAWeb/CDAS Data Access and Display: Access to most of the active archive of old and new NASA mission products via both a web interface that allow file download and graphical display (https://cdaweb.gsfc.nasa.gov) and via web services of both "SOAP" and "RESTful" varieties. (https://cdaweb.gsfc.nasa.gov/WebServices/)

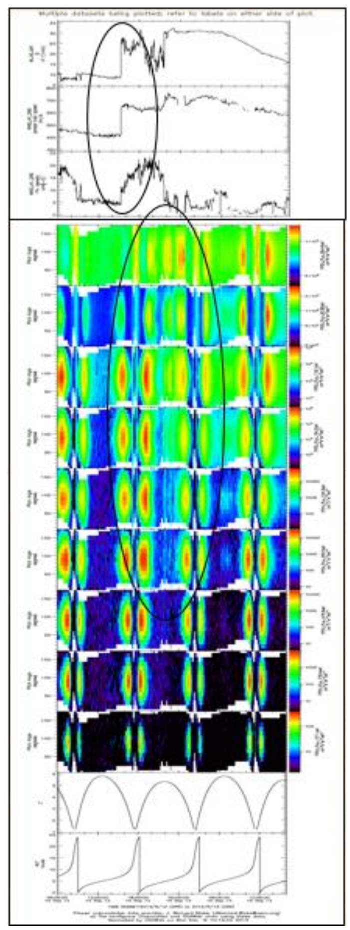


Figure 6-4 The CDAWeb data browsing system makes it easy to display data quickly and download desired parameters and time ranges across missions

SSCWeb & 4-D Orbit Viewer: Views in both planar projections with options to plot multiple spacecraft orbits and to apply region and other query restrictions (https://sscweb.gsfc.nasa.gov); and in 3-D space and time for interactive views. (https://sscweb.gsfc.nasa.gov/tipsod/)

OMNIweb: Cross-normalized solar wind fields and plasma dataset mapped the nose of the bowshock, and geomagnetic and solar activity indices. (https://omniweb.gsfc.nasa.gov/)

CDF Maintenance and Upgrades: SPDF provides continual upgrades in conjunction with user needs to assure that the standard CDF data format continues to provide efficient and complete access to heliophysics data. (See https://cdf.gsfc.nasa.gov)

Heliophysics Data Portal: Find and access nearly all heliophysics data products using the uniform SPASE data model. Searches can be by time range, spacecraft, instrument, data type, keyword, SPASE ID, cadence, location, and other descriptors. The Heliophysics Data Portal (HDP) is adding digital object identifier (DOI) information as we gather it. The HDP provides direct links to data repositories, and "get data/plots" buttons provide access from within the tool for many datasets. (See https://heliophysicsdata.sci.gsfc.nasa.gov/ and the "Help" button there.)

Virtual Solar Observatory: Web browser and API access to most solar physics data. Searches can be made by time range, mission, instrument, observables, nicknames (e.g., "H-alpha"), and spectral ranges. VSO is integrated into SolarSoft to provide API access to data via Interactive Data Language (IDL). (See http://docs.virtualsolar.org, https://sdac.virtualsolar.org/cgi/search)

SolarSoft: A mostly IDL-based collection of routines that provide everything from basic data access and processing of Level 0 data, to general plotting tools, to advanced analysis tools tailored to the needs of particular solar missions. This is gradually being complemented by the routines in SunPy, but it still serves as the major workhorse for solar data analysis. (See http://www.lmsal.com/solarsoft/; http://www.mssl.ucl.ac.uk/surf/sswdoc/solarsoft/ssw_install_howto.html)

Helioviewer (and JHelioviewer): A general visualization tool for registered solar images from many missions, including some ground-based. Data products include multi-wavelength SDO, SOHO, Hinode, and other images in multiple wavelengths, with easy overplotting and movie-making. Many videos are already archived on YouTube. The European version, JHelioviewer, runs on desktops/laptops and adds such things as potential-field modeled magnetic field lines. (See https://www.helioviewer.org and https://www.jhelioviewer.org) Both applications include links to the Heliophysics Events Knowledgebase (https://www.lmsal.com/hek/).

HelioCloud: A next generation environment for conducting heliophysics research based on a NASA instance of Amazon Web Services (AWS) Cloud web services. This project has the dual goals of increasing research team collaboration and unlocking the science which may be done with high end computing and big data. Together these goals will help support the need for Open Science in heliophysics. (See http://heliocloud.org/)

CCMC Services/Databases and Activities Relevant to HDRL

*Kamodo**: an open source CCMC tool for access, interpolation, and visualization of space weather models and data in python. Kamodo allows model developers to represent simulation results as mathematical functions which may be manipulated directly by end users. Kamodo handles unit conversion transparently and supports interactive science discovery through jupyter notebooks with minimal coding and is accessible through python.

(See: https://github.com/nasa/Kamodo)

*Kameleon**: A software suite that is being developed at the CCMC to address the difficulty in analyzing and disseminating the varying output formats of space weather model data. Through the employment of a comprehensive data format standardization methodology, Kameleon allows heterogeneous model output to be stored uniformly in a common science data format. The converted files contain both the original model output as well as additional metadata elements to create platform independent and self-descriptive data files. To facilitate model data dissemination, data reuse, and code reuse – the Kameleon access and interpolation library provides direct access to both the model data as well as the embedded metadata.

(See: https://ccmc.gsfc.nasa.gov/Kameleon/)

*CCMC Metadata Registry (CMR)**: Provides detailed SPASE descriptions of simulation models, model runs, data sets from model simulation runs, and observation/instrument data sets. (See: https://kauai.ccmc.gsfc.nasa.gov/CMR/view/metadata)

General Services/Tools Utilized by HDRL

SPASE Descriptions, Registration, and Tools: The SPASE Data Model needs to be continually updated, and new or revised product descriptions are needed frequently; the SPASE group provides these, partly as a contract-funded activity for continuity and community service, and partly as an open consortium that continues to refine and update the model. SPASE descriptions include DOIs for data reference and "product and parameter keys" to aid the use of APIs. (See http://spase-group.org, and Roberts, et al., 2018.)

Space Physics Environment Data Analysis Software (SPEDAS): Roughly speaking, *SolarSoft* for non-solar physics. A set of IDL routines, along with a version that does not require IDL, for loading, plotting, analyzing, and integrating data from many ground and space-based observatories. The latter include all the CDAWeb accessible data, with enhancements for some analysis routines (e.g., visualization of some 3D distributions) and continual expansion and improvement. A Python version of SPEDAS is in the works. (See http://spedas.org/wiki/index.php?title=Main_Page; Angelopoulos et al., 2018.)

Autoplot: A Java application that reads and displays/plots all formats used in Heliophysics plus many others. Autoplot can be used to plot data for a server or as a standalone Webstart application. It can be used to form a data access layer, and can make "png walks" to provide rapid data surveys. A variety of output formats are also possible. (See http://autoplot.org)

HAPI: The Heliophysics Application Programmer's Interface is a generic API with a full and mature specification that provides uniform access to a wide range of heliophysics data, including time series of scalars, vectors, spectrograms, and more complex matrix fields. It has implementations at CDAWeb, integrated Space Weather Analysis System (iSWA) at CCMC, Laboratory for Atmospheric and Space Physics (LASP) Interactive Solar Irradiance Datacenter (LISIRD), and a growing number of other places. Clients exist for IDL, Matlab, and Python. It is being built and maintained by an informal (but funded) group at a variety of institutions. (See https://github.com/hapi-server)

Python in Heliophysics Community (PyHC): There is an increasing use of Python for data access and analysis, particularly among younger researchers. The HDRL is unifying these efforts and has initiated funding of some efforts. This will be a community-directed effort that builds on the models of other open source communities. (See https://heliopython.org)

	Stand-alone Tool	Web Service/App	Standard	Code
CDAWeb/CDAS Data Access and Display		x		
SSCWeb & 4-D Orbit Viewer		x		
CDF Maintenance and Upgrades			x	x
Heliophysics Data Portal		x		
Virtual Solar Observatory		x		
SolarSoft	x			
Helioviewer (and JHelioviewer)		x		
HelioCloud		x		
Heliophysics Events Knowledgebase		x		
SPASE Descriptions, Registration, and Tools	x		x	
Space Physics Environment Data Analysis Software (SPEDAS) / (Py)SPEDAS	x			x
Autoplot		x		
Heliophysics Application Programmer's Interface (HAPI)		x	x	
Python in Heliophysics Community (PyHC)			x	x
Support for various maintenance/upgrade tasks				
Data/model comparison activities; model and output registries				
Visual System for Browsing, Analysis and Retrieval of Data (VisBARD)	x	x		x
OMNIWeb		x		
integrated Space Weather Analysis System (iSWA)		x		
CCMC Run-on-Request (ROR) system		x		
Space Weather Database Of Notifications, Knowledge, Information (DONKI)		x		
Comprehensive Assessment of Models and Events using Library Tools (CAMEL) Framework		x		
Kamodo*				x
Kameleon*	x			

Appendix B. Levels of Data

The following table describes levels of data for SMD missions as defined in SPD-41a. It should be considered a guideline, and ***HPD missions should describe their specific data levels in their respective PDMPs***. Missions may have only a subset of these levels of data available or might have further sub-divisions of these levels, which should be indicated by a letter (e.g., Level 1b data would represent a processing stage between Level 1 and Level 2).

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 have been processed to remove instrument or sensor effects. This level of data is typically in physical units that correspond to an observable or physical quantity.
3	Data that have been mapped onto 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 are 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	Data produced on the ground for testing the Mission, either prior to launch or during operations. When not excepted, scientifically useful data produced as part of ground testing should be made accessible in accordance with available resources. This type of data may only be available upon request.
Calibration	Data generated for the calibration of scientific data. Data may be generated on-board or on the ground. These data are typically used during processing in removing instrumental effects or creating higher level data products.

Appendix C. Metadata Resources

The consistent use of metadata helps ensure that data is organized, findable, and accessible. Below is a list of resources that provide guidelines and best practices for the application of metadata.

Recommended dataset and file naming guidelines:
https://spdf.gsfc.nasa.gov/guidelines/filenaming_recommendations.html

Dataset URI template standard:
https://github.com/hapi-server/uri-templates/wiki/Specification

ISTP internal metadata:
https://github.com/IHDE-Alliance/ISTP_metadata/tree/main/v1.0.0

Appendix D. Acronyms

ACRONYM	DEFINITION
API	Application Programming Interface
AWS	Amazon Web Services
CAMEL	Comprehensive Assessment of Models and Events using Library Tools
CCMC	Community Coordinated Modeling Center
CDF	Common Data Format
CDR	Critical Design Review
CHORUS	Clearinghouse for the Open Research of the United States
CMAD	Calibration And Measurement Algorithm Document
CMR	CCMC Metadata Registry
CUI	Controlled Unclassified Information
DMP	Data Management Plan
DOI	Digital Object Identifier
DONKI	Database Of Notifications, Knowledge, Information
DR	Decommissioning Review
DRR	Disposal Readiness Review
EAR	Export Administration Regulations
FAIR	Findable, Accessible, Interoperable, And Reusable
FITS	Flexible Image Transport System
FOIA	Freedom of Information Act
FRR	Flight Readiness Review
HAPI	Heliophysics Application Programmer's Interface
H-CSI	Heliophysics Citizen Science Investigations
HDF5	Hierarchical Data Format
HDP	Heliophysics Data Portal
HDRL	Heliophysics Digital Resource Library
HPD	Heliophysics Division
HIPAA	Health Insurance Portability and Accountability Act
HSO	Heliophysics System Observatory
ICD	Interface Control Document
IDL	Interactive Data Language
ITAR	International Traffic in Arms Regulations
LASP	Laboratory for Atmospheric and Space Physics
LISIRD	LASP Interactive Solar Irradiance Datacenter
LRR	Launch Readiness Review
MCR	Mission Concept Review
MDR	Mission Definition Review
MRR	Mission Readiness Review
NOFO	Notice of Funding Opportunity
NPR	NASA Procedural Requirements
ORR	Operational Readiness Review

ACRONYM	DEFINITION
OSDMP	Open Science and Data Management Plan
OSS	Open Source Software
PDMP	Project Data Management Plan
PDR	Preliminary Design Review
PI	Principal Investigator
POC	point of contact
R&A	Research and Analysis
ROR	Run-on-Request
ROSES	Research Opportunities in Space and Earth Science
SBIR	Small Business Innovation Research
SDAC	Solar Data Analysis Center
SIR	System Integration Review
SMD	Science Mission Directorate
SMWT	SPASE Metadata Working Team
SOC	Science Operations Center
SPASE	Space Physics Archive Search and Extract
SPEDAS	Space Physics Environment Data Analysis Software
SPD	SMD Policy Document
SPDF	Space Physics Data Facility
SRR	System Requirements Review
STI	Scientific and Technical Information
STTR	Small Business Technology Transfer