

MARS EXPLORATION PROGRAM DATA MANAGEMENT PLAN

**Raymond E. Arvidson
Susan H. Slavney
Jennifer G. Ward**

**McDonnell Center for the Space Sciences
Dept. of Earth and Planetary Sciences
Washington University
St. Louis, MO 63130**

**Rev. 4.0
June 15, 2011**

Mars Exploration Program Data Management Plan

Approved by:

Thomas Morgan **Date**
Program Manager
Planetary Data System

Reta Beebe **Date**
Chief Scientist
Planetary Data System

Michael Meyer **Date**
Lead Scientist
Mars Exploration Program, NASA Headquarters

Douglas McCuiston **Date**
Program Director
Mars Exploration Program, NASA Headquarters

Edwin Grayzeck **Date**
National Space Science Data Center

Richard Zurek **Date**
Program Chief Scientist
Mars Exploration Program Office, Jet Propulsion Laboratory

Fuk Li **Date**
Program Manager
Mars Exploration Program Office, Jet Propulsion Laboratory

Table of Contents

Acronyms	iv
1 Purpose and Scope of Document	1
2 Overview of the Mars Exploration Program	1
3 Overview of the Planetary Data System	1
4 Roles and Responsibilities	2
4.1 Mars Exploration Program	2
4.2 Planetary Data System.....	3
4.3 National Space Science Data Center	3
4.4 Jet Propulsion Laboratory Media Relations Office	3
4.5 Principal Investigator Home Institution Media Relations Offices.....	3
5 Data Products.....	3
5.1 Experiment Data Records.....	4
5.2 Standard Data Products	4
5.3 Ancillary Files	4
5.3.1 SPICE Files.....	4
5.3.2 Engineering and Other Data.....	4
5.4 Public Information Products.....	5
5.5 Data Set Documentation.....	5
5.6 Software	5
6 Policies for Release of Data and Public Information	5
6.1 Data Rights and Release Policy.....	5
6.2 Public Information Release Policy	7
6.3 Data Release Acknowledgement Policy.....	8
7 Documentation Concerning Returned Samples	8
8 Archive Planning, Generation, Validation, and Delivery	8
8.1 Archive Media.....	8
8.2 Planning and Documentation	8
8.3 Delivery Testing.....	9
8.4 Generation	9
8.5 Peer Review.....	9
8.6 Validation and Delivery	10
9 Applicable Documents.....	11
10 Definitions of Terms	12
11 Tables.....	13
12 Figures	21

Tables

Table 1 – Example Outline for Project Archive Generation, Validation and Distribution Plan	13
Table 2 – NASA Definition of Processing Levels for Science Data Sets.....	14
Table 3 – Annotated Example Outline for Data Product SIS for Instrument-Related Data	15
Table 4 – Example Outline for Archive Volume SIS for Instrument-Related Data	17
Table 5 – Example of PDS Archive Volume Structure and Contents	20

Figures

Figure 1 – Planetary Data System Organizational Structure.	21
Figure 2 - Flow of Archive Volume Generation, Validation and Distribution Functions for Mars Exploration Program Projects.....	22

Acronyms

CD-ROM	Compact disc read only memory
CODMAC	Committee on Data Management and Computation
DAWG	Data and Archive Working Group
DSN	Deep Space Network
EDR	Experiment Data Record
JPL	Jet Propulsion Laboratory
MEP	Mars Exploration Program
MGS	Mars Global Surveyor
NAIF	Navigation and Ancillary Information Facility
NASA	National Aeronautics and Space Administration
NSSDC	National Space Science Data Center
PDS	Planetary Data System
PSG	Project Science Group
RDR	Reduced Data Records
SIS	Software Interface Specification
SPICE	Spacecraft, Planet, Instrument, C-matrix, Events

1 Purpose and Scope of Document

The Mars Exploration Program (MEP) Data Management Plan is a high-level plan for timely generation, validation, and delivery of data products from MEP projects to the Planetary Data System (PDS) and the National Space Science Data Center (NSSDC) in complete, well-documented, permanent archives. The Plan also describes policies and procedures for distributing the data and information to the general public in a timely fashion, consistent with the MEP Program Plan.

As stated in that Program Plan, each MEP mission shall develop a Science Data Management Plan covering all of that mission's science investigations, whether or not associated with specific instrument teams. This plan will define details of policy and implementation for science data release within and by the teams, particularly for data release to the general community and public and for data archival in approved archive centers.

This document describes how to develop the archive portion of the Science Data Management Plan; it is for use by personnel associated with MEP projects to understand programmatic data release policies and how data products are to be archived. The intent is to have each Project write an Archive Plan that is responsive to this document. Principal Investigators associated with MEP projects will use this document to understand what is expected of them in terms of delivery of a complete data archive, and how the PDS can help them prepare their data products. PDS and NSSDC personnel will use this document to understand the types and volumes of data products and archives they can expect to receive from MEP projects, in order to allocate resources accordingly.

2 Overview of the Mars Exploration Program

MEP is a multiyear coordinated exploration of Mars (<http://mars.jpl.nasa.gov/>). The focus is on understanding current and past climates, searching for evidence of past and extant life, and determining the extent to which human expeditions to the planet are feasible.

The long-term scientific payoff of MEP will come about if the observations and resultant data feed from one project to the next and if documented archives are used by the scientific community to increase knowledge about the planet. The Project has been extraordinarily successful to date. Current and planned missions, including separate and joint missions by NASA and the European Space Agency, foretell a continuing need to properly archive and make accessible Mars mission data.

3 Overview of the Planetary Data System

The NASA PDS is an active archive that provides high quality, usable planetary science data products to the science community (<http://pds.jpl.nasa.gov/>). This system evolved in response to science community requests for improved availability of planetary data from NASA projects through increased scientific involvement and oversight. PDS provides access to data archives for scientists, educators, and the public.

The first objective of the PDS is to publish and disseminate documented data sets of use in scientific analyses. PDS-produced products are reviewed by scientists and data engineers to ensure that the data and the related materials are appropriate and usable. PDS data sets are typically published as archives, collections of reviewed data and documentation, ancillary information (such as calibration data), software, and any other tools needed to understand and use the data. PDS provides access to data by a system of online and compact disc archives.

The second objective of the PDS is to work with projects to help them design, generate, and validate their data products as archives. The advantage of having projects deliver well-documented products is that use by the scientific and public communities is maximized.

The third PDS objective is to develop and maintain archive data standards to ensure future usability. PDS has developed a set of standards for describing and storing data so that future scientists unfamiliar with the original experiments can analyze the data using a variety of computer platforms with no additional support beyond that provided with the product. These standards address structure of the data and description of archive contents through consistent use of common terms (PDS Standards Reference:

<http://pds.jpl.nasa.gov/tools/standards-reference.shtml>; Planetary Science Data Dictionary:
<http://pds.jpl.nasa.gov/tools/dictionary.shtml>).

The fourth PDS objective is to provide expert scientific help to the user community. Most of the archive products may be accessed or ordered automatically by users, but PDS also provides teams of scientists to work with users to select and understand data. In some cases, special processing can be done to generate user-specific products. In addition, PDS-funded scientists receive suggestions from the science community about new and/or improved software, data sets, and archives.

To accomplish the four objectives, the PDS is structured as a distributed system composed of eight teams, called nodes (Figure 1). The science nodes include the Atmospheres, Geosciences, Imaging, Planetary Plasma Interactions, Rings, and Small Bodies Nodes. Each has an advisory group of discipline scientists to provide guidance and priorities for PDS. The support nodes are the Navigation and Ancillary Information Facility (NAIF) Node, specializing in observation geometry and sequence of events information, and the Engineering Node, at the Jet Propulsion Laboratory.

4 Roles and Responsibilities

This section summarizes the roles and responsibilities for personnel and organizations involved in the generation, validation, transfer, and distribution of MEP archives.

4.1 Mars Exploration Program

Each MEP project is responsible for the production and delivery to PDS of documented archives that use PDS standards. The investigation teams selected for or added later in flight are responsible for collecting the necessary scientific, engineering, and ancillary information prior to depositing data and related information in the appropriate NASA data archive. Principal Investigators and Team Leaders are responsible, under contractual obligation, for working with PDS personnel on planning, generating, and validating archive data volumes based on data from their instruments and spacecraft subsystems, or from data analysis activities, as appropriate. Each investigation lead is responsible for validation and transfer of those data products and of supporting documentation to the PDS. All science data and designated data products, with supporting documentation, will be archived in an approved archive center (PDS for NASA-funded investigations) following a short period for verification and validation of not more than 6 months from the time of receipt by the investigation teams of usable scientific and ancillary data. (There are no investigator proprietary data rights; see Section 6.)

For joint missions between NASA and another space agency, data from all investigations will be shared and will be archived in the PDS for NASA and also in the archive center chosen by the partnering agency. In the case of non-NASA missions, the organization sponsoring the full mission will have entered into an appropriate agreement with NASA providing for sharing and archiving of flight data necessary for the completion of the selected NASA-supported investigation/investigators. The Project Science Data Management Plan and embedded Archive Plan will reflect these agreements between agencies.

The Project Science Group (PSG) or equivalent working team will coordinate plans for archiving through such entities as a Data and Archive Working Group (DAWG). DAWG membership will include an archive representative from each instrument team and from each PDS node that will receive the data. The DAWG is typically co-led by a designated Project archive representative and a member of the lead PDS node for the Project.

Each MEP project will generate an archive plan document that describes the specific generation, validation, and transfer to PDS of the project's data products. The plan will specify the contents of archives, the personnel or organization responsible for generating each archive, estimates of the amount of data to be delivered, and a schedule for delivery. The Project archive plans will be based on the policies and procedures specified in the Mars Exploration Program Data Management Plan (this document). An outline of a project archive plan is provided in Table 1. Each MEP project is responsible for distributing its data products internally, i.e. among scientists associated with the project. Note that missions preparing and/or returning sample caches have additional requirements (e.g., see Section 7).

The Project Office of each MEP project is responsible for ensuring that the policies and implementation plans concerning release of information and interactions with the public and press are followed. The Project also is responsible for primary interactions with NASA Headquarters, including ensuring that Headquarters personnel are aware of information releases before they occur, and focusing feedback from Headquarters back to the appropriate parties. The Project has the responsibility for primary interactions with the JPL Media Relations Office, serving as coordinating entity for media relations under the purview of the MEP Program Office at JPL. The Project Scientist has primary responsibility for ensuring that scientific findings are made available to the media.

4.2 Planetary Data System

The PDS is the primary organization within NASA responsible for archiving and distributing planetary data. The PDS is thus the designated point of contact for MEP projects on archive-related issues. The PDS is also the interface between MEP projects and the NSSDC.

The PDS Geosciences Node Director has been selected to take the lead in working with each MEP project to ensure that overall plans are in place and that individual projects are meeting archiving commitments. Relevant PDS personnel will participate in archive planning efforts through DAWGs for each project to ensure that: (a) archives are planned, generated, and reviewed using PDS standards, (b) archives are constructed in ways that facilitate cross-mission and cross-instrument data analyses, (c) PDS provides user services that allow search and access to data in ways that transcend particular instruments or data sets.

The PDS is responsible for distributing MEP project archives to the broad science community after the data have been released by the project. Distribution is done electronically via the PDS websites. The PDS maintains copies of project archives, and the relevant PDS discipline nodes are available to provide information and expert assistance to users of the data.

4.3 National Space Science Data Center

The NSSDC's primary role is to maintain a "deep archive" of data for long-term preservation, for filling any large orders requested by the science community, and for filling orders requested of the NSSDC by the general public. The PDS will deliver copies of MEP project archive volumes to NSSDC according to agreements in place between the PDS and NSSDC.

4.4 Jet Propulsion Laboratory Media Relations Office

The JPL Media Relations Office is chartered to be the prime point of contact with the public and the media for activities under the purview of the MEP Program Office at JPL. The overall objective of the Media Relations Office is to maximize the positive impact of information releases from the project. This office helps organize and implement press releases and press conferences. They facilitate release of data and information meant for consumption by the press and public, including video feeds of interviews.

4.5 Principal Investigator Home Institution Media Relations Offices

Each Principal Investigator may release findings from his or her instrument to the press, through the home institution media relations office, if those releases are coordinated with the MEP Project Scientist and the JPL Media Relations Office.

5 Data Products

The science-related data products to be produced during MEP projects are described in this section; these include standard products, engineering and other ancillary products, and documentation that accompanies data sets as they are delivered to the PDS. For reference, Table 2 defines numerical processing levels, using the common assignments developed by the NASA community. Special data products should also be described to the extent possible to aid estimation of the volume and type of products that may be archived at a future time.

5.1 Experiment Data Records

Experiment Data Records (EDRs) are NASA Level 0 products (see Table 2) consisting of time-ordered sequences of raw science data. EDRs are sorted by instrument and are usually accompanied by engineering data which can be monitored for proper instrument operation. The specific contents of EDRs will be instrument-dependent. PDS does not archive telemetry data, except those related to radio science.

5.2 Standard Data Products

Standard Data Products include both EDRs and pre-defined Reduced Data Records (RDRs). RDRs are typically generated from EDRs and SPICE data (Section 5.3.1) at investigator facilities. Standard Data Products are defined during the proposal and selection process and are contractually promised by the Principal Investigators as part of the investigation. Standard Data Products are produced in a systematic way during the course of the project, primarily by Investigation Teams, although some may be generated by Interdisciplinary Scientists addressing specific scientific problems. Standard Data Products are sometimes called operational products or pipeline products, since they are produced routinely with fairly well-defined procedures. During MEP projects it is expected that increased knowledge of instrument calibrations and operating nuances, together with increased knowledge of Mars, and better understanding of how to reduce raw data, will result in updates to some Standard Data Products during the course of the mission. The documentation for Standard Data Products should be thorough enough to understand quantitatively the processing history.

Special Data Products are RDRs which have not been pre-defined and/or are not produced routinely.

5.3 Ancillary Files

5.3.1 SPICE Files

SPICE is an acronym used to describe a suite of elemental ancillary data sets, often called kernels:

- S Spacecraft Ephemeris, with files containing spacecraft location as a function of time
- P Planetary/Satellite ephemerides and associated target body physical and cartographic constants
- I Instrument information, including mounting alignment and field-of-view size/shape/orientation
- C Orientation of a spacecraft's primary coordinate systems and possibly angular rates as well
- E Event information, including nominal sequences, real-time commanding, unscheduled events, and experimenter's notebook comments

Most SPICE kernels will be generated by flight operations teams with some instrument information and sequence data obtained from instrument teams. When appropriate, there will be both predict and reconstructed SPICE files. Reconstructed SPICE kernels will typically be generated within a couple of weeks after data acquisition and made available to instrument teams. SPICE kernels will be used together with a toolkit of software modules (to be supplied by JPL's Navigation and Ancillary Information Facility) at investigator institutions to generate the derived ancillary data needed to help plan observations and to process EDRs. SPICE kernels and NAIF toolkit software are applicable to landers and rovers as well as to orbiting spacecraft.

5.3.2 Engineering and Other Data

Engineering data, while not strictly science data, can be used to aid the interpretation of science data and as such will be archived. Other data that may be useful include sequence of events files, log files, DSN monitor files, etc. The engineering and ancillary file content of Project archives will be negotiated between

each Project and the PDS and described in the Project Archive Plan. These products will be archived by NAIF. In some cases, engineering data may be bundled with science archives, e.g. rover wheel currents and rover yaw, pitch, and roll data would be needed to infer soil properties during traverses and trenching operations. Thus, these engineering data might be included in archives focused on imaging and *in-situ* data acquired during rover operations.

5.4 Public Information Products

Public Information Products consist of selected data products that have been annotated for the media and delivered to relevant NASA or home institution offices for use in public relations. It is expected that these products will be produced mainly at investigator institutions, although materials for press distribution may be produced at JPL.

5.5 Data Set Documentation

Documentation of data acquisition and processing histories is crucial to successful long-term use of Project data. The SPICE E Kernel will contain information explaining why certain observations were acquired, including whether they were part of a coordinated sequence. Information on spacecraft events that affected the sequences and data products will also be included. The reconstructed (as opposed to predicted) SPICE files will be transferred to a long-term archive along with raw and reduced science data.

Information needed to fully understand EDRs, SPICE files, and reduced data will also be generated for long-term use of MEP Project data. Documentation to be provided by each instrument team will typically include:

- Calibration Requirements Document
- Calibration Report
- Science Reports
- User's Guides that describe data products from particular instruments and how they were generated
- Instrument-specific sequence of events

Data sets archived with PDS are described by entries in the PDS Catalog. Data set providers are required to supply descriptive information for the PDS Catalog. Specifically, the Project is responsible for descriptions of the mission and the instrument hosts (spacecraft), and instrument teams are responsible for descriptions of their instruments and data sets.

5.6 Software

It is expected that standard data products delivered by Principal Investigators will be in formats that are accessible by other users with commonly available software tools. PDS discourages including software in an archive.

6 Policies for Release of Data and Public Information

6.1 Data Rights and Release Policy

Because of the expected widespread scientific and public interest in new results from Mars and the strong MEP commitment to releasing data on a timely basis, it is important to establish a clear release policy for archives and other data sets (e.g. quick-look products).

This document highlights aspects of data release related to archive and release outside the Project. MEP policy also addresses science data access within the Project and the investigation teams:

1. All team members, including co-investigators, collaborators and other identified associated scientists shall have full and immediate access to scientific data obtained by the team's

investigation. The PI or Team Leader will establish rules and procedures consistent with the project policies defined by the Project Science Group (or its equivalent) for publication of results from the investigation.

2. Project investigators shall share data, interpretations, pre-publication manuscripts, and presentations amongst themselves. The PSG/Mission PI will establish “Rules of the Road” and implement mechanisms and forums by which this sharing will take place.

A key element of the data release policy is the need for a reasonable interval of time to generate and validate standard data products and archive volumes before release to the general community. Based on numerous past experiences, up to a six-month period is necessary to produce useful RDRs for data from most instruments. EDRs and RDRs should be produced, validated, and released as soon as possible, with the six-month period considered to be a maximum (although certain data sets such as global maps may require more time to generate).

Specific release policies are:

- I. The generation/validation time period for standard data products is defined to be the period from receipt of science packets containing raw data at instrument processing facilities until release of archive volumes to the PDS. The archive volumes will include, as appropriate, standard data products, SPICE files, relevant software, and documentation describing the generation of the products. During the generation/validation period, Investigation Teams are expected to use raw data from flight instruments or spacecraft subsystems to generate standard products and archives, and to validate the products and volumes. Efforts involving all of the project investigators are expected to be underway during this period, including product validation. Generation and validation of products and archives may typically require up to six months from the time of receipt of raw data at instrument processing facilities in order to produce and validate useful archive volumes. More rapid release is highly desired.
- II. To ensure rapid dissemination of new and significant information, each Principal Investigator will also release a significant subset of data earlier as a form of public outreach and education. These releases will typically be available within a week (or sooner) of data receipt. Postings on the Internet (e.g., World Wide Web) can be used as a cost-effective way for widespread dissemination of these special products. The posted data may include images, derived spectra, topographic information, and other forms of data that illustrate new and significant results. Postings will include documentation. These data releases will conform to the Public Information Release Policy (section 6.2 below).
- III. During the generation/validation period, use and analysis of raw and derived products from a particular instrument, use of the results of unpublished papers derived from such analysis, or use of data posted on the Internet for public outreach and education should be done only with the agreement of the relevant Principal Investigator(s) unless required by MEP (see VII below).
- IV. After the six-month or less generation/validation period, the relevant archive volumes will be transferred to the PDS, which will make them available to the general science community.
- V. Individual Investigation Teams are strongly encouraged to make data available as soon as possible, and they may decide to make a number of versions available to the PDS earlier than the typical six-month period. They may not release data from other investigations without concurrence of the respective PI or Team Leader. The Project Scientist/Mission PI will adjudicate conflicts.
- VI. Items I. to V. above pertain to nominal mission operations. In the event of anomalies, data should be archived using stated policies, unless observational, personnel, and/or financial constraints force the need for a longer period between data receipt and transfer of archives to the PDS. Exceptions to the 6-month schedule must be identified and approved by the Project and MEP.
- VII. MEP reserves the right to require a Principal Investigator to acquire and/or release in a timely manner, possibly immediately, a limited set of data for purposes of mission planning and

ensuring mission safety, for both current and future missions, and/or MEP public engagement. This includes the right to require joint release of data and in all cases with appropriate attributions.

6.2 Public Information Release Policy

Public information release includes press conferences and written material concerning both mission operations and scientific analyses. Specific policy statements for Public Information Release for Mars Exploration Program projects are:

- I. Information concerning spacecraft and instrument anomalies may be released only by the Project Office, in coordination with NASA Headquarters and Principal Investigators.
- II. Information concerning significant scientific results may be released during press conferences, press releases organized by the Project Office and the Jet Propulsion Laboratory's Media Relations Office, and on the World Wide Web, in coordination with NASA Headquarters and relevant members of the project's science teams.
- III. Information concerning scientific results from a given instrument may also be released by the Principal Investigator for the instrument. For example, such releases may be organized by the home institution of the Principal Investigator. These releases can be in any medium, but significant results must have prior coordination through the Project Office, the MEP Program Office at JPL, and NASA Headquarters.
- IV. Information concerning scientific results from instrument studies may be released by an Interdisciplinary Scientist, Participating Scientist, or Co-Investigator before it is deposited in the Planetary Data System, only with the concurrence of the Principal Investigator whose instrument data are used.
- V. In cases II through IV the Project Office must receive, in advance of the release, a copy of the release material (e.g., images, spectra, captions, summary of results), a schedule for the release, and a statement of the mechanisms for release. The intent is not to require concurrence for the release, but only to make sure that the Project Office, the Public Information Office, the JPL MEP Program Office, and NASA Headquarters are informed of the releases before they happen.
- VI. MEP reserves the right to require a Principal Investigator to acquire and/or release immediately a limited set of data for purposes of public or educational outreach.

The Project Scientist (or Mission PI for competed missions) has primary responsibility for ensuring that scientific findings are made available to the media. In addition, the Project Scientist or Mission PI or his/her designates, working with the appropriate NASA center's Media Relations Office, will:

- I. Help instrument teams prepare plans for generation and release of public-oriented data and results.
- II. Ensure that PSG members prepare release materials with enough lead time so that NASA Headquarters receives the materials about a week before the general release to the media.
- III. Coordinate the preparation of materials (press releases, interviews, press conference scripts, video file productions, etc.) with the NASA center media relations representative.
- IV. Identify themes or subjects for each media interaction opportunity months in advance, and to schedule the preparation work.
- V. Decide what the best media interaction vehicle should be for the information to be presented.
- VI. Help enforce the graphical standards of the Media Relations Office for best television viewing.

6.3 Data Release Acknowledgement Policy

In any release of data or information as described in sections 6.1 and 6.2 above, appropriate credit will be given to all entities involved, including the Principal Investigator, the Principal Investigator's home institution, JPL, and NASA. The credit will appear on public web sites that contain data releases and on documentation that accompanies the data products.

7 Documentation Concerning Returned Samples

Samples of rocks and soils may be returned from Mars as part of MEP. It is expected that the samples will be archived in an appropriate curatorial facility or facilities, where initial analyses will be conducted. This facility or facilities will also serve as a source of sample material for the science community. The Science Data Management Plan for those missions which have as their objective the preparation and caching of samples for return to Earth or the actual return of such sample caches from Mars to Earth will contain or will reference a Sample Handling, Curation and Distribution Plan, which is not discussed further here.

It is crucial that observations that provide context (e.g., geological, geochemical, environmental) for returned samples also be archived. Such data must be archived in ways that can be easily linked to sample information developed by the curatorial facilities. To facilitate such links, it is expected that the PDS, MEP, and curatorial facility personnel will work together to establish data bases and procedures to connect mission and sample archives.

8 Archive Planning, Generation, Validation, and Delivery

Standard products form the core of the archives to be produced by MEP projects and released to the PDS for distribution to the science community. These products, SPICE files, software, and engineering and other ancillary information will be validated for transfer to the PDS.

The following section discusses the processes and schedules for generation and validation of standard products and archives by MEP projects. Figure 2 shows the flow of data through the stages of archives generation, validation, transfer to the Planetary Data System, and distribution to the science community. Also shown is posting of timely results on the Internet for education and outreach.

8.1 Archive Media

An *archive* consists of one or more data sets along with all the documentation and ancillary information needed to understand and use the data. Thus defined, an archive is a logical construct independent of the medium on which it is stored.

An *archive volume* is a unit of media on which an archive is stored; for example, one CD-ROM. When an archive spans multiple volumes, it is called an *archive volume set*. Usually the documentation and some ancillary files are repeated on each volume of the set, so that a single volume can be used alone.

Whole or partial archives may be placed online for convenient distribution via the Internet. However, to ensure long-term viability, archives are also stored offline on physical archive volumes. The choice of storage medium will depend on available technology, but it should be a stable, industry-standard medium intended for long-term storage. PDS periodically refreshes archives in long-term storage and reevaluates the choices of storage media.

8.2 Planning and Documentation

Preparation for generating project data archives involves designing the data products and the archive volumes on which they will be stored, and developing the systematic generation, validation, and distribution procedures for these archive volumes. These plans will be documented in an Archive Volume Generation, Validation and Distribution Plan created by each project. These documents will cover all the standard data products generated by the project (e.g., Arvidson et al., 1999; Arvidson, 2003; Arvidson et al., 2006a; Arvidson et al., 2006b; Arvidson, 2008). For reference, Table 1 shows a suggested outline for an Archive Plan.

The format and content of data products are described in Data Product Software Interface Specification (SIS) documents, created by the instrument Principal Investigators or ancillary data product's cognizant engineer. Typically a separate SIS is created for each type of data product, including raw, derived and ancillary products. Table 3 is an example of a Data Product SIS outline, including annotations concerning the specific content recommended for each section. (Note that the example outline given is that of a completely self-contained Data Product SIS. If the pertinent information is available from other sources, e.g. data product labels and catalog descriptions, then sections of the Data Product SIS may simply refer the reader to those sources.)

The format and content of the archive volumes are described in an Archive Volume SIS for each data set, which is also generated by the Principal Investigator for the instrument or a cognizant engineer, working with relevant PDS personnel as needed. Table 4 is an example outline for an Archive Volume SIS. Table 5 shows a typical directory structure and contents for a PDS archive.

8.3 Delivery Testing

Well in advance of the first delivery of data from an MEP project to the PDS, delivery tests are carried out using sample data products. The purpose of the tests is to smooth the path for actual deliveries. Typically between one and three tests are conducted, covering these aspects of data delivery:

1. Delivery path from project to PDS
2. Communication protocols between data providers and PDS recipients
3. Verification of safe delivery, e.g. via checksums
4. Verification of complete delivery, e.g. via file manifests
5. Verification that delivered archive corresponds to archive documentation
6. Verification that delivered archive complies with PDS standards
7. Verification that test procedures can handle expected delivery volumes.

Testing is overseen by the Data and Archive Working Group. Each science team or operations entity that will deliver to PDS participates in testing. The personnel involved in the tests are the same personnel who will be involved in the actual delivery and receipt of data during operations. A test plan is written to specify the tests to be conducted, the roles and responsibilities, and the schedule. After each test a report is submitted to the DAWG with recommendations for changes, if any, to be made to the delivery procedures.

8.4 Generation

Generation of science data packets, EDRs, RDRs, and other data sets for a given MEP project is the responsibility of that project, as discussed in Section 4. Archive volumes of raw (level 0) data are not required by PDS, except for radio science data.

The DAWG or its equivalent for each project will provide oversight on implementation of the Project Archive Plan and will ensure timely generation, validation, and delivery of archives to the PDS.

The Navigational and Ancillary Information Facility (NAIF) will generate an archive volume set of all SPICE data.

8.5 Peer Review

All data formally incorporated into the PDS archives must undergo a peer review. The purpose of the review is to determine that a particular data set is scientifically useful, well-documented, and accessible. Any problems/deficiencies identified during the review must be resolved before the data are accepted into the PDS archives. A peer review committee includes scientists who have some interest and expertise in the

kind of data being reviewed as well as PDS personnel whose job is to see that the data set is PDS-compatible.

Typically, once any peer review liens have been resolved, the data are considered acceptable for PDS archiving. As the final data products are delivered to PDS during operations, they will undergo additional validation by the science teams and PDS. The science teams will validate the science aspects of their data products ahead of each delivery, and once delivered, the PDS will validate the final files for PDS compliance.

8.6 Validation and Delivery

Validation of engineering and science packets and SPICE files will be an intrinsic part of standard product generation. Further science validation of standard products will be done in part during analysis of the data.

A key additional requirement is the validation of archives for integrity of scientific content, file structures (e.g., do the files conform to Software Interface Specification documents?), directory structures, compliance with PDS standards, and integrity of the physical media used to transfer the data products. Validation will be coordinated through the DAWG or its equivalent and will follow the relevant Project Archive Plan. Interested members of the scientific community may also be asked to comment on plans for archive generation, validation, and delivery.

During the validation period, problems due to obvious errors in processing, missing files, and inadequate documentation will be referred back to the data suppliers for correction. Minor errors may simply be documented on the volumes. A fundamental aspect of the release schedule is that SPICE files and algorithms or software for generating derived data products are released at the same time as reduced data records generated from the raw data.

The DAWG will periodically report results of validation to the Project Science Group (or equivalent) for the specific project. If the volumes are approved for release by the Project, then the Principal Investigator or relevant Project personnel would transfer the archives to the PDS, based on the release schedule specified by the Project's Archive Generation, Validation, and Distribution Plan.

The PDS is responsible for delivering a copy of each archive volume to the NSSDC.

9 Applicable Documents

Arvidson, R., E. Guinness, S. Slavney, and R. J. Springer, 1999, *Mars Global Surveyor Project Archive Generation, Validation, and Transfer Plan*, MGS Project Document 542-312, <http://pds-geosciences.wustl.edu/missions/mgs/docs/mgsarchive.pdf>.

Arvidson, R. E., 2003, *2001 Mars Odyssey Orbiter Archive Generation, Validation, and Transfer Plan*, Odyssey Project Document 722-302, Rev. 1, http://pds-geosciences.wustl.edu/missions/odyssey/docs/ody_archive.pdf.

Arvidson, R., S. J. Noland, and S. Slavney, 2006a, *Mars Reconnaissance Orbiter (MRO) Project Archive Generation, Validation, and Transfer Plan*, MRO Project Document 31-468, Version 1.1, <http://pds-geosciences.wustl.edu/missions/mro/docs/MROarchive.pdf>.

Arvidson, R., S. Slavney, and J. A. Crisp, 2006b, *Mars Exploration Rover (MER) Project Archive Generation, Validation, and Transfer Plan*, MER Project Document 420-1-200, Rev. C, http://pds-geosciences.wustl.edu/missions/mer/docs/mer_archive.pdf.

Arvidson, R., 2008, *Phoenix Project Archive Generation, Validation, and Transfer Plan*, Phoenix Project Document 214-096, Initial Release, http://pds-geosciences.wustl.edu/missions/phoenix/Phoenix_Archive_Plan.pdf

Data Management and Computation, Volume 1, Issues and Recommendations, 1982, National Academy Press, 167 p., http://www.nap.edu/openbook.php?record_id=12366&page=R1.

Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences, 1986, National Academy Press, 111 p., http://books.nap.edu/openbook.php?record_id=12343&page=R1.

Mars Exploration Program (MEP) Program Plan, JPL D-66674.

Planetary Science Data Dictionary Document, August 17, 2010, Planetary Data System, JPL D-7116, Version 1.80, <http://pds.jpl.nasa.gov/tools/dictionary.shtml>.

Planetary Data System Data Archive Preparation Guide, April 1, 2010, Version 1.4, JPL D-31224, <http://pds.jpl.nasa.gov/tools/archiving.shtml>.

Planetary Data System Standards Reference, February 27, 2009, Version 3.8, JPL D-7669, Part-2, <http://pds.jpl.nasa.gov/tools/standards-reference.shtml>.

10 Definitions of Terms

Archive – An archive consists of one or more data sets along with all the documentation and ancillary information needed to understand and use the data. An archive is a logical construct independent of the medium on which it is stored.

Archive Volume, Archive Volume Set – A volume is a unit of media on which data products are stored; for example, one CD-ROM. An *archive volume* is a volume containing all or part of an archive; that is, data products plus documentation and ancillary files. When an archive spans multiple volumes, they are called an *archive volume set*. Usually the documentation and some ancillary files are repeated on each volume of the set, so that a single volume can be used alone.

Data Product – A labeled grouping of data resulting from a scientific observation, usually stored in one file. A product label identifies, describes, and defines the structure of the data. An example of a data product is a planetary image, a spectrum table, or a time series table.

Data Set – An accumulation of data products, typically from a single instrument at a particular level of processing. A data set together with supporting documentation and ancillary files is an archive.

Experiment Data Records – NASA Level 0 data for a given instrument; raw data.

Reduced Data Records – Science data that have been processed from raw data to NASA Level 1 or higher. See Table 2 for definitions of processing levels.

Standard Data Product – A data product that has been defined during the proposal and selection process and that is contractually promised by the Principal Investigator as part of the investigation. Standard data products are generated in a predefined way, using well-understood procedures, and processed in "pipeline" fashion.

Special Data Products - RDRs which have not been pre-defined and/or are not produced routinely.

11 Tables

Table 1 – Example Outline for Project Archive Generation, Validation and Distribution Plan

1. Introduction
 - 1.1 Purpose
 - 1.2 Scope
 - 1.3 Contents
 - 1.4 Applicable Documents and Constraints
2. Archive Generation, Validation, Transfer, and Distribution
 - 2.1 The Mission
 - 2.2 Data Flow
 - 2.3 Data Volume
 - 2.4 Data Validation and Peer Review
 - 2.5 Data Delivery Schedule
 - 2.6 Online Access to Archives
 - 2.6.1 PDS Online Services
 - 2.6.2 The Analyst's Notebook
3. Roles and Responsibilities
 - 3.1 Project Responsibilities
 - 3.2 Planetary Data System Responsibilities
 - 3.3 National Space Science Data Center Responsibilities
4. Figures
 - 4.1 Mission Timeline
 - 4.2 Data flow from downlink to Planetary Data System
5. Tables
 - 5.1 Payload
 - 5.2 Definitions of Processing Levels for Science Data Sets
 - 5.3 Components of Archives
 - 5.4 Standard and Special Data Products
 - 5.5 Archive Generation, Validation, and Release Schedule

Table 2 – NASA Definition of Processing Levels for Science Data Sets

Level	Type	Data Processing Level Description
1	Raw Data	Telemetry data with data embedded.
2	Edited Data	Corrected for telemetry errors and split or decommutated into a data set for a given instrument. Sometimes called Experimental Data Record. Data are also tagged with time and location of acquisition. Corresponds to NASA Level 0 data.
3	Calibrated Data	Edited data that are still in units produced by instrument, but that have been corrected so that values are expressed in or are proportional to some physical unit such as radiance. No resampling, so edited data can be reconstructed. NASA Level 1A.
4	Resampled Data	Data that have been resampled in the time or space domains in such a way that the original edited data cannot be reconstructed. Could be calibrated in addition to being resampled. NASA Level 1B.
5	Derived Data	Derived results, as maps, reports, graphics, etc. NASA Levels 2 through 5.
6	Ancillary Data	Nonscience data needed to generate calibrated or resampled data sets. Consists of instrument gains, offsets, pointing information for scan platforms, etc.
7	Correlative Data	Other science data needed to interpret space-based data sets. May include groundbased data observations such as soil type or ocean buoy measurements of wind drift.
8	User Description	Description of why the data were required, any peculiarities associated with the data sets, and enough documentation to allow secondary user to extract information from the data.
N	N	Not Applicable.

Table 3 – Annotated Example Outline for Data Product SIS for Instrument-Related Data

1. Purpose and Scope of Document

The purpose of this document is to provide users of the data product with a detailed description of the product and a description of how it was generated, including data sources and destinations. The document is intended to provide enough information to enable users to read and understand the data product. The users for whom this document is intended are the scientists who will analyze the data, including those associated with the project and those in the general planetary science community. This section should include a one-sentence description of the data product.

2. Applicable Documents

Planetary Data System Standards Reference, February 27, 2009, Version 3.8, JPL D-7669, Part 2.

Planetary Data System Archive Preparation Guide, April 1, 2010, Version 1.4, JPL D-31224.

Science Data Management Plan for the project

Archive Generation, Validation and Distribution Plan for the project

Archive Volume SIS for the data product

Paper(s) describing the instrument

3. Relationships with Other Interfaces

State what products, software, and documents would be affected by a change in this data product. Refer to documents that describe the other interfaces.

4. Data Product Characteristics and Environment

4.1 Instrument Overview

Describe how the instrument acquires data; include a general description of operating modes as they relate to the data product.

4.2 Data Product Overview

Describe the type of data product (image, spectrum, etc.); what it measures (radiance, voltage, time, etc.); how observations are organized into data products; how a data product is stored (binary or ASCII, how grouped into files). Details should be specified in the following sections.

4.3 Data Processing

This section should provide general information about the data product content, format, size, and production rate. Details about data format should be specified later in section 6.

4.3.1 Data Processing Level

Describe the product in terms of its NASA and/or CODMAC processing levels.

4.3.2 Data Product Generation

Describe how and by whom data products are generated. Describe any calibrations, corrections, or compressions that have been applied to the product. Specify software, algorithms, calibration files or procedures. If a product has been compressed or otherwise processed in a reversible manner, indicate the software, algorithms, and/or ancillary data needed to reverse the processing, and describe how they should be applied. State whether multiple versions of the products will be generated.

4.3.3 Data Flow

Describe the sources, destinations, and transfer procedures for data products. State the size of an individual data product and the total size of all the data products generated over the course of each mission phase. State the time span covered by a product, if applicable, and the rate at

which products are generated and delivered. If more than one version of a product may be generated, state the number of expected versions and the rate at which they will be generated.

4.3.4 Labeling and Identification

Indicate how an individual product is identified and labeled, including PDS labels and any other labels or header information. Details about label and header formats should be specified later in section 6. Each individual product should have a unique identifier; describe how this identifier is assigned. If multiple versions of a product are possible, state how they will be distinguished.

4.4. Standards Used in Generating Data Products

4.4.1 PDS Standards

State that the data products comply with Planetary Data System standards for file formats and labels. Refer to the PDS Standards Reference.

4.4.2 Time Standards

4.4.3 Coordinate Systems

4.4.4 Data Storage Conventions

Address any data storage issues, such as byte order, machine dependence, and compression.

4.5 Data Validation

Describe validation procedures applied to data products to ensure that their contents and format are free of errors. This may be a brief overview if the details are described in another document.

5. Detailed Data Product Specifications

Describe the physical organization of the data product so that the user can access and analyze the data.

5.1 Data Product Structure and Organization

5.2 Data Format Descriptions

5.3 Label and Header Descriptions

Include PDS labels and other labels and headers that describe the data.

6. Applicable Software

Describe any applicable software used to examine, display, or analyze the data products. State who produces the software and who uses it.

6.1 Utility Programs

Include image display tools, plotting tools, etc., that allow a user to examine the data.

6.2 Applicable PDS Software Tools

List the PDS software tools that may be useful in examining the data products.

6.3 Software Distribution and Update Procedures

Describe how the software is distributed and updated, and how the user can obtain it.

7. Appendices

7.1 Glossary

7.2 Acronyms

7.3 Definitions of Data Processing Levels

7.4 Example PDS Label

Table 4 – Example Outline for Archive Volume SIS for Instrument-Related Data

1. Introduction

1.1. Purpose and Scope

This Software Interface Specification is intended to be used by those who wish to understand the format and content of this PDS archive. Typically these individuals are software engineers, data analysts, and planetary scientists.

1.2. Content Overview

Briefly describe the data products in the archive and what instrument and mission they come from. Say who is responsible for generating the products and assembling the archive.

1.3. Applicable Documents and Constraints

This Archive Volume SIS is intended to be consistent with the following documents.

1. Mars Exploration Program Data Management Plan [give current citation] [Only for data from Mars Program missions]
2. Project Archive Plan [give complete citation]
3. Data Product SIS [give complete citation]
4. *Planetary Data System Archive Preparation Guide*, April 1, 2010, Version 1.4, JPL D-31224.
5. *Planetary Data System Standards Reference*, February 27, 2009, Version 3.8, JPL D-7669, Part 2.

1.4. Relationships with Other Interfaces

State what products, software, and documents, if any, would be affected by a change in this SIS.

2. Archive Volume Contents

2.1. Root Directory Contents

Files in the Root Directory include an overview of the archive, a description of the volume for the PDS Catalog, and a list of errata and other notes about the archive.

2.2. Data Directory Contents and Naming

Describe the contents of the data directories. Explain how data are aggregated into subdirectories, explain the naming scheme for directories and files. If some information is already provided in the Data Product SIS, refer to the appropriate section of that document.

2.3. Index Directory Contents

Tables in the Index directory are provided to help the user locate products on the archive volume and on previously released volumes in this archive. An index table has one row for each product on the volume. List the columns in the index table(s).

2.4. Document Directory Contents

The Document directory contains documentation to help the user understand and use the archive. Usually the Data Product SIS and Archive Volume SIS are included in the Document directory.

2.5. Catalog Directory Contents

Files in the Catalog directory also contain documentation, but they are formatted in a way to enable them to be entered in the PDS Catalog for searching. They include descriptions of the data set(s), instrument(s), instrument host(s) (e.g. the spacecraft), mission, personnel contact information, and references.

2.6. Label Directory Contents (optional)

The optional Label directory contains PDS label fragments that are referenced by the data products' PDS labels. For example, rather than repeating a lengthy table definition in many labels, the definition may be captured in a file in the Label directory to which the labels refer.

2.7. Software Directory Contents (optional)

The optional Software directory contains utilities or application programs to aid the user in viewing or extracting data from the data products. Use of this directory is not common. Data providers are encouraged to create products that do not require any special software. If it is

desirable to provide software to the user, PDS prefers to make it available on a PDS web site but not to archive it due to the difficulties in maintaining software over the long term.

2.8. Calib Directory Contents (optional)

The optional Calib directory contains calibration data and/or reports that were used to create the products in the archive, or that can be used to create higher-level products from the products in the archive.

2.9. Geometry Directory Contents (optional)

The optional Geometry directory contains files needed to understand observation geometry.

2.10. Browse Directory Contents (optional)

The optional Browse directory contains reduced-size easily viewed versions of data products to help the user identify products of interest. Browse products are not intended to be used for data analysis.

2.11. Extras Directory Contents (optional)

The optional Extras directory contains documentation, utility programs or other materials that the user may find helpful, but that are beyond the scope of the required elements of the archive. The contents of this directory are exempt from PDS requirements for labeling, etc. The Extras directory is intended for value-added material that is not crucial for understanding the data.

3. Archive Volume Format

3.1. File Formats

3.1.1. Document File Format

Describe formats of documentation files in the archive, e.g. ASCII, PDF, HTML. See PDS guidelines for documents.

3.1.2. Table File Format (if any)

Describe formats of table files in the archive, either with fixed-length fields or variable-length fields. See PDS guidelines for tables.

3.1.3. PDS Label Format

Describe format of PDS labels in the archive, either attached or detached. Say whether a local Data Dictionary is used.

3.1.4. Software File Format (if any)

Describe file formats, languages used, machines for which compiled executables are provided.

3.1.5. Catalog File Format

Catalog files (suffix .CAT) exist in the Root and Catalog directories. Like PDS labels, they are text files formatted as keyword=value statements.

3.1.6. Science Data File Formats

Describe data file format(s). For example, type of data (image, spectrum, table); ASCII text or binary; integer or floating point; big-endian or little-endian byte order; how data are aggregated into one file (one file per orbit, per day, etc.). Refer to Data Product SIS for details.

4. Archive Volume Generation

4.1. Data Transfer and Validation Methods

Describe details of how, where and by whom the archive is generated and validated. Discuss how data products are delivered to PDS. If these details exist in an Interface Control Document, refer to it and include it in the list of Applicable Documents in section 1.3.

4.2. Data Product Sizes and Delivery Rates

Give a summary of expected sizes and production rates for each type of product in the archive. For example, give the number and size of products generated per day, per delivery to PDS, and total for the entire primary mission.

4.3. **Interface Media Characteristics**

This archive is intended to be made available online as its primary method of distribution. When it is necessary to store the archive on removable media such as a DVD, the archive volume will conform to ISO 9660 standards.

4.4. **Backup and Duplicates**

Backup copies of the archive will be stored by the data provider at [TBD] until the final versions of the products have been archived by the PDS. The PDS will store backup copies of the archive at the [TBD] PDS Node and at the National Space Science Data Center (NSSDC).

4.5. **Labeling and Identification**

Give the unique volume identifier(s) for the archive.

5. **Support Staff and Cognizant Persons**

List names and affiliations of the P.I. and other personnel, including PDS personnel, associated with the generation of the archive.

Table 5 – Example of PDS Archive Volume Structure and Contents

The following is an example of the directory structure and contents of a typical PDS archive volume or volume set. Asterisks (*) indicate directories and files that are required for PDS archives. The archive may be stored on one or more physical volumes. In the event that multiple volumes are used to store a particular data set, typically all the directories are present on each volume in the volume set.

Directory	Contents
Root	*AAREADME.TXT: Introduction and general information for the user. ERRATA.TXT: Summary of anomalies found on this volume and on previous volumes in the set, if any. (Required file if there are comments or anomalies.) *VOLDESC.CAT: Description of the volume for the PDS Catalog.
*DOCUMENT	*DOCINFO.TXT: Description of files in the document directory. *VOLINFO.TXT: Overview of the mission, instruments, and the archiving process. Archive Volume SIS Data Product SISs Science Papers
CATALOG	CATINFO.TXT: Description of files in the catalog directory. *CATALOG.CAT: Descriptions of mission, spacecraft, instrument and data sets for the PDS Catalog.
*INDEX	*INDXINFO.TXT: Description of files in the index directory. *INDEX.TAB: Table listing the data products on this volume and pertinent information about each one. *INDEX.LBL: PDS label describing the format of INDEX.TAB.
LABEL	LABINFO.TXT: Description of files in the label directory. Additional PDS labels and include files that are not packaged with the data products.
CALIB	CALINFO.TXT: Description of files in the calib directory Calibration files or reports used to create the data products or needed for processing the data products.
*DATA	Data products with PDS labels. A data product's label may be attached (embedded at the beginning of the file) or detached (stored in a separate file in the same directory). Typically the DATA directory is divided into subdirectories.

12 Figures

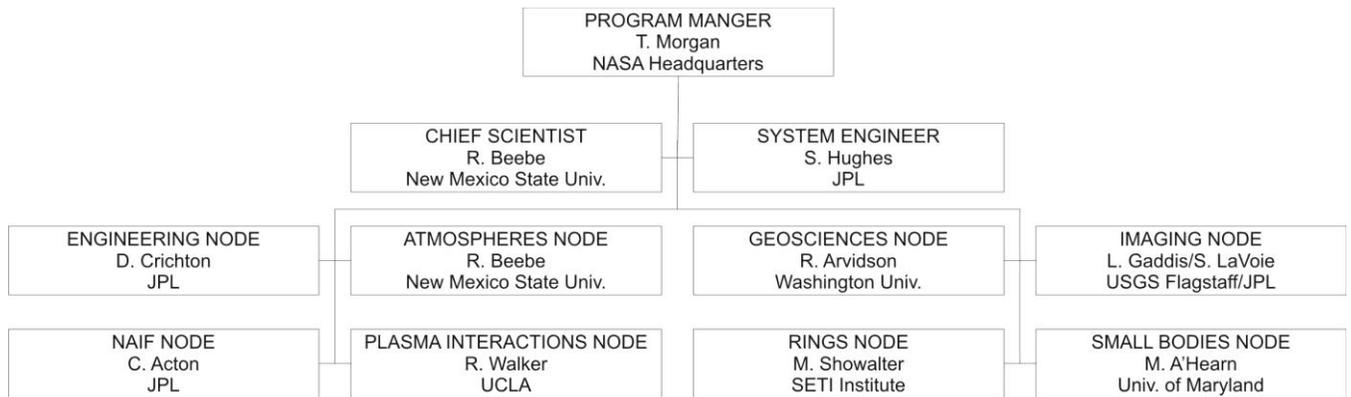


Figure 1 – Planetary Data System Organizational Structure.

Not shown are the extensive set of geographically distributed subnodes and data nodes, the Radio Science support function at Stanford University led by R. Simpson, and the User Centered Design support function at NASA Ames Research Center led by J. Trimble.

Overview of Mars Exploration Program Project Archive Volume Generation, Validation and Distribution

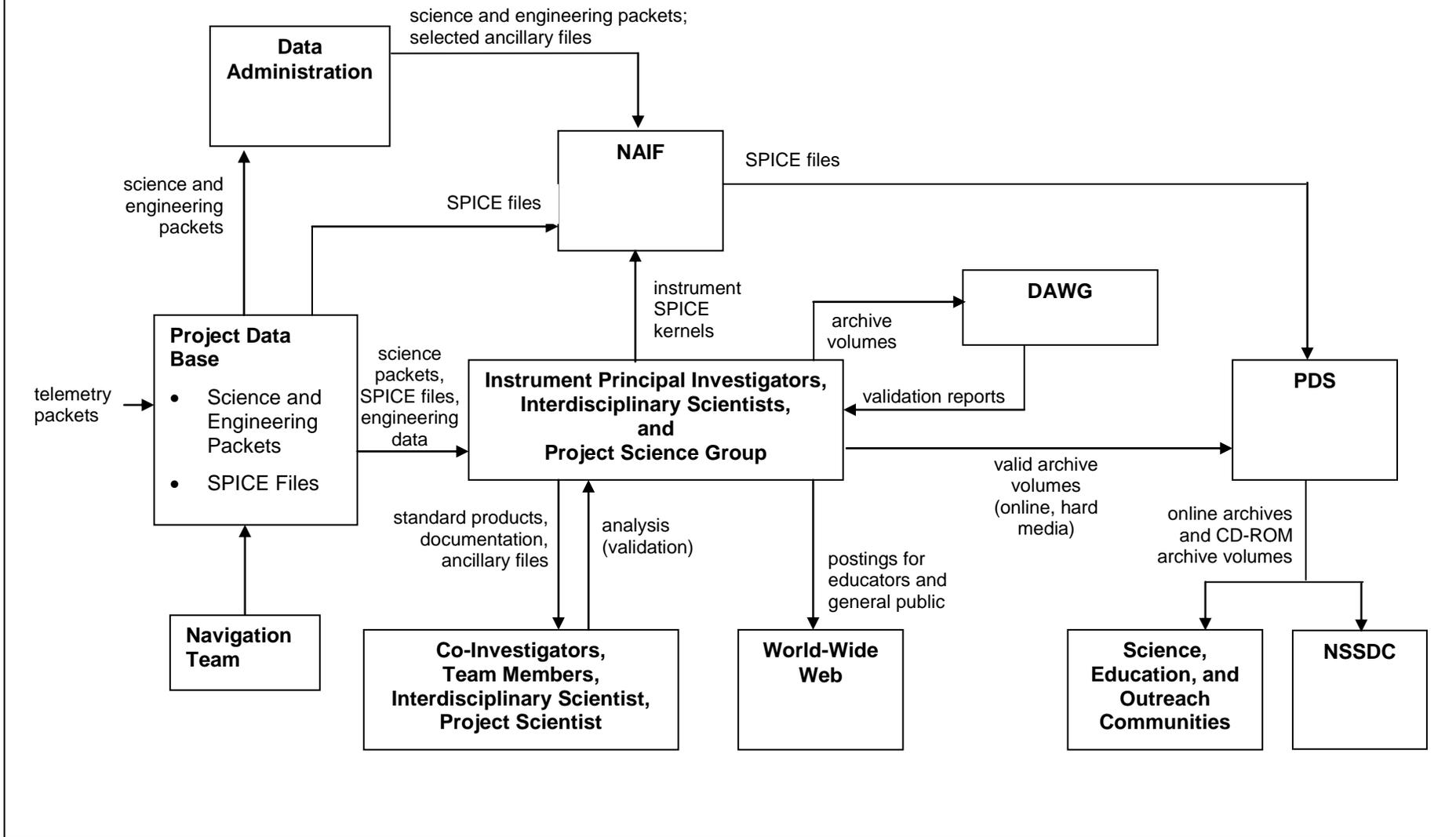


Figure 2 - Flow of Archive Volume Generation, Validation and Distribution Functions for Mars Exploration Program Projects