InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **Case Study 08 Final Report: Mars Global Surveyor Data Records in the Planetary Data System**

Status:	Final (public)
Version:	1.0
Submission Date:	September 2005
Release Date:	September 2007
Author:	The InterPARES 2 Project
Writer(s):	William Underwood Information and Telecommunications Laboratory, Georgia Tech Research Institute
Project Unit:	Focus 2
URL:	http://www.interpares.org/display_file.cfm?doc= ip2_cs08_final_report.pdf

TABLE OF CONTENTS

List of Figures

Abstract

This case study was designed to collect information to answer InterPARES 2 core research questions regarding the creation, management and preservation of e-science data records. It was also designed to collect information that could be used to address research issues in the Description, Policy, and Modeling Cross-domains. Finally, it was designed to collect information to perform a walkthrough of the InterPARES 2 Chain of Preservation model of records creation, maintenance, appraisal and preservation of authentic e-records.

The focus of the case study is the data records of the Mars Global Surveyor mission and the Planetary Data System (PDS). The MGS was the first mission of the National Aeronautics and Space Administration (NASA) Mars Surveyor Program. The MSG spacecraft was launched in 1996. Last contact with the spacecraft was in November 2006. The MGS spacecraft contains instruments that send observational data back to the Jet Propulsion Laboratory Spacecraft Control Center where it is stored as experimental data records in a Project Database. Spacecraft Operations Planning Computers at the Principal Investigator's site, and at seven discipline nodes—Atmospheres, Geosciences, Imaging, Navigation Ancillary Information Facility, Plasma Physics, Rings and Small Bodies—are used to create reduced data records.

A data product is a labeled reduced data record resulting from a scientific observation, usually stored in one file. Templates are used to create product labels that identify, describe and define the structure of the data. An example of a data product is a planetary image, a spectrum table or a time series table. An archive volume is data products plus documentation and ancillary files. Volumes are transferred to the Science Data Validation Team for peer review. If judged complete, reliable and PDS-compliant, they are cataloged and stored in the PDS archive for access by planetary scientists.

The designers of the PDS adopted self-describing data files as a preservation strategy. The labels of self-describing files describe the file format of attached data as well as the context in which the data was created. The labels are interpreted using PDS Label Library Light. The PDS Object Access Library uses this interpretation to access the data, and the accessed data is displayed using NASAView.

The PDS has been operational since 1989 and it has not been necessary to update (convert or migrate) any of the data products to other data formats. When the hardware and/or operating system become obsolete, only Label Library Light, the Object Access Library and the NASA Viewer need to be recompiled or translated to the new platform. The result is that technology obsolescence requires only the migration of an interpreter, a data access library, and a viewer rather than the creation of this software for each of the file structures.

The PDS is referred to as an "active archive," whereas the National Space Science Data Center's (NSSDC's) repository is referred to as a "deep archive." The PDS is the entrance for Planetary Science data into the NSSDC archives for long-term preservation.

A. Introduction

Background

The Planetary Data System Case Study is a case study of the InterPARES 2 Science Focus Group. It was designed to collect information to answer InterPARES 2 core research questions regarding the creation, management and preservation of e-science data records. The focus was on the data records of the Mars Global Surveyor mission. It was also designed to collect information that could be used to address research issues in the InterPARES 2 Description, Policy, and Modeling Cross-domains. Finally, it was also designed to collect information to perform a walkthrough of the InterPARES 2 activity model of records creation, maintenance, appraisal, and preservation of authentic e-records.

Purpose

The purpose of this report is to provide the results of a case study of scientific data records from a NASA spacecraft mission in order to provide answers to the InterPARES 2 research questions.

Scope

The next section gives a brief overview of the subjects of the case study, the Mars Global Surveyor mission and the Planetary Data System. Section C describes the methodology of the case study. Section D describes the context of the records. Section E provides narrative answers to the core InterPARES research questions. Section F provides answers to the cross-domain questions and describes the potential use of the case study data in refinement and validation of InterPARES 2 Chain of Preservation (COP) activity model.

B. Mars Global Surveyor Mission and the Planetary Data System

The Mars Global Surveyor (MGS) was the first mission of the National Aeronautics and Space Administration (NASA) Mars Surveyor Program. The MGS spacecraft is illustrated in Figure 1. The MSG spacecraft was launched in 1996. Last contact with the spacecraft was in November 2006. Subsequent missions in the Mars Surveyor Program were the 2001 Mars Odyssey mission and the 2003 Mars Exploration Rovers.

The Mars Global Surveyor is an orbiter carrying the following instruments:

- Mars Orbiter Camera (MOC),
- Mars Orbiter Laser Altimeter (MOLA),
- Thermal Emission Spectrometer (TES),
- Magnetometer/Electron Reflectometer (MAG/ER),
- Radio Science Subsystem (RSS), and
- Accelerometer.

These instruments are illustrated in Figure 2.

Figure 1. The Mars Global Surveyor Spacecraft

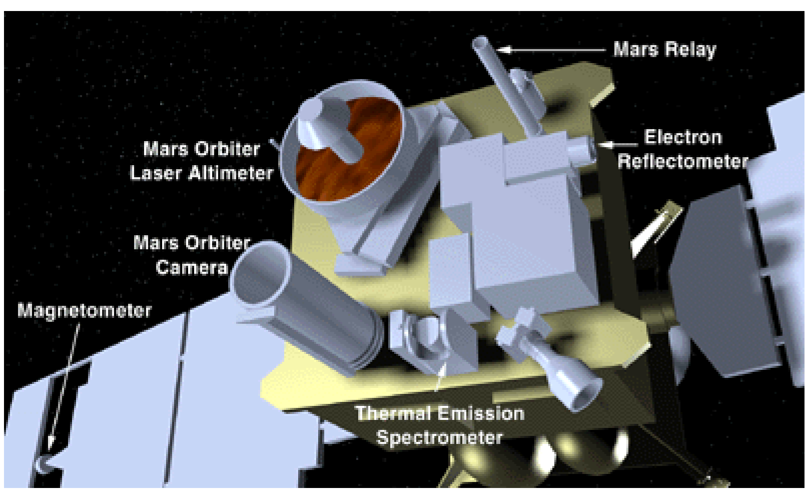


Figure 2. Mars Global Surveyor Instruments

Let us consider just one of the Mars Global Surveyor instruments, the Mars Orbital Camera (MOC). The Principal Investigator for the Mars Orbital Camera is Mike Malin. The camera was designed to obtain images of the surface and atmosphere of Mars. The camera housing, which is shown in Figure 3 is 80 by 40 centimeters. It contains three cameras. A .4 degree narrow field of view camera with a resolution of 1.5 meters per pixel, and two 140 degree wide field of view cameras with blue and red bandpass filters.

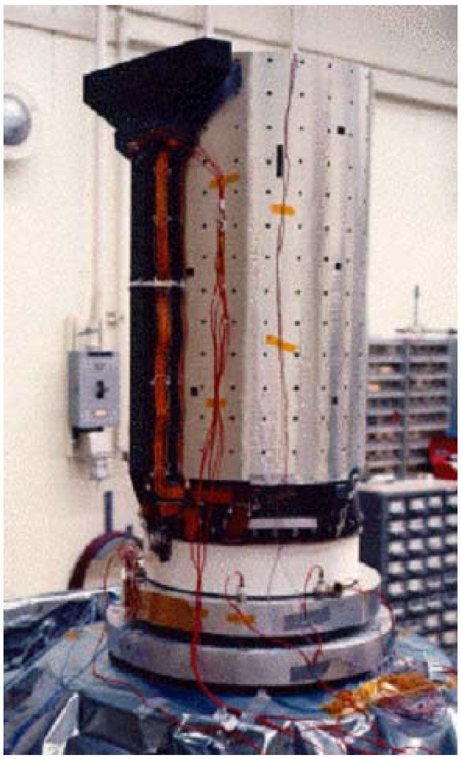


Figure 3. The Mars Orbital Camera

Collection of scientific data in the solar system requires a telecommunication network. Spacecraft, such as the Mars Global Surveyor and the Mars Exploration Rover, contain instruments that send observation data into a network of spacecraft tracking other spacecraft, as well as ground-based transmitters and receivers for communicating with spacecraft in deep space. This telecommunication network, called the Deep Space Network (DSN), is illustrated in Figure 4.

This data is sent to the Jet Propulsion Laboratory (JPL) Spacecraft Control Center. The MGS ground data system includes a centralized mission operations component at Jet Propulsion Laboratory (JPL) and a distributed component located at home institutions of the Principal Investigators, the Radio Science Team Leader, Interdisciplinary Scientists, Team Members, Co-Investigators, and Participating Scientists. Science Operations Planning Computer (SOPC) workstations located at Principal Investigator, Team Leader, and Interdisciplinary Scientist home institutions are electronically connected via NASCOM and Ethernet links to a Project Data Base (PDB) at JPL. Personnel at JPL and those with SOPC workstations use the Project Data Base to coordinate observation planning. Level 0 science data in packetized form, SPICE files, and relevant engineering data are placed in the PDB for access by Principal Investigators and the Radio Science Team Leader. These files are transferred to the relevant home institutions for examination and generation of reduced data records.

Each Principal Investigator and the Radio Science Team Leader is responsible for getting the products to his respective team members and to interdisciplinary scientists and participating scientists, under the guidelines of data use outlined in the MGS Science Data Management Plan.

Standard products form the core of the archives produced by MGS Project Teams and are transferred to the PDS for distribution to the science community. These products and associated raw data, SPICE files, and ancillary information are placed on archive volumes for validation and transfer to the PDS [Arvidson et al 1999, p. 3].

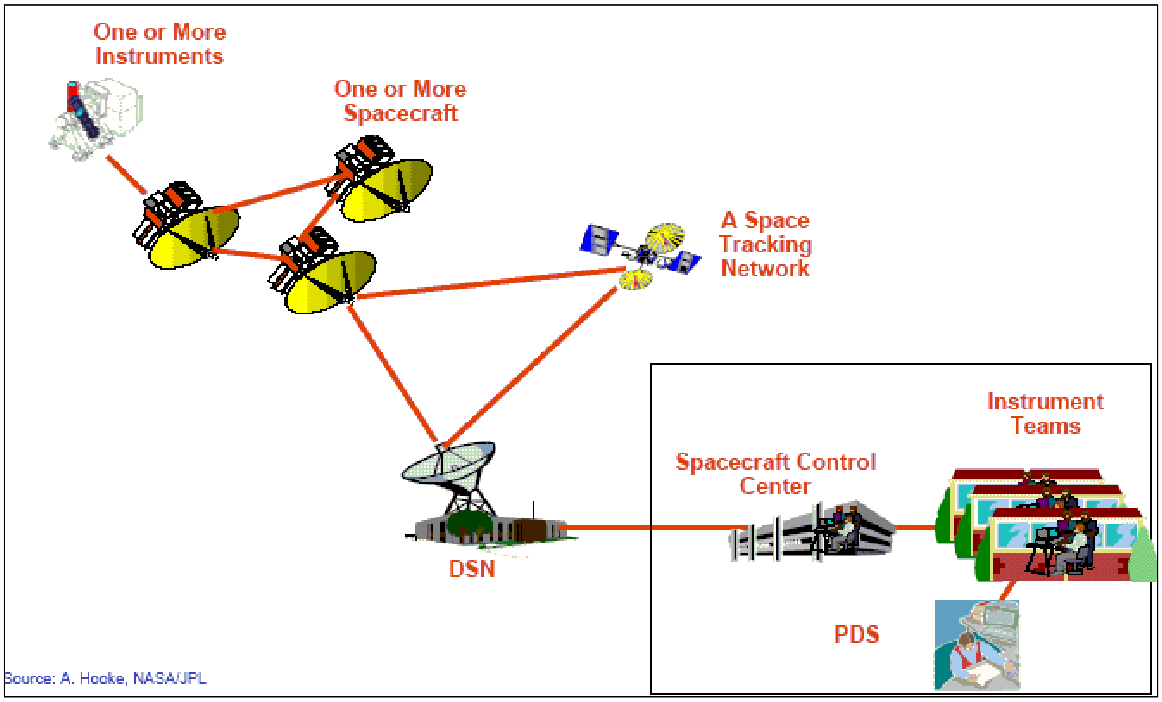


Figure 4. The Deep Space Network

The SPICE System provides scientists with ancillary observation geometry data and related tools that are useful in the planning and interpretation of science instrument observations returned from planetary spacecraft. SPICE stands for Spacecraft, Planet, Instrument, C-Matrix, and Events.

The Planetary Data System archives and distributes scientific data from NASA planetary missions. NASA's Office of Space Science sponsors the PDS. Its purpose is to ensure the long-term usability of NASA science data and to stimulate advanced research.

The PDS is a distributed data archives. There is a central node and seven other discipline nodes—Atmospheres, Geosciences, Imaging, Navigation Ancillary Information Facility (NAIF), PPI, Rings and Small Bodies.

Figure 5 shows at a high-level the Planetary Science Data Model. The primary classes in the data model include Mission, Spacecraft, Instrument and Target. The objects in these classes are not digital objects. Other classes include Document, Software, Image, Time Series and Spectrum. Objects in these classes are digital objects.

A dataset is an accumulation of data products, supplemental data, software, and documentation that will completely document and support the use of these data products. For example, the images of Jupiter taken by both Voyager spacecraft comprise a single dataset. The digital data products include those shown in Fig. 5 (image, time series, spectrum) and others. Other classes of digital objects are documents and software. Figure 6 shows the major subsystems of the Planetary Data System.

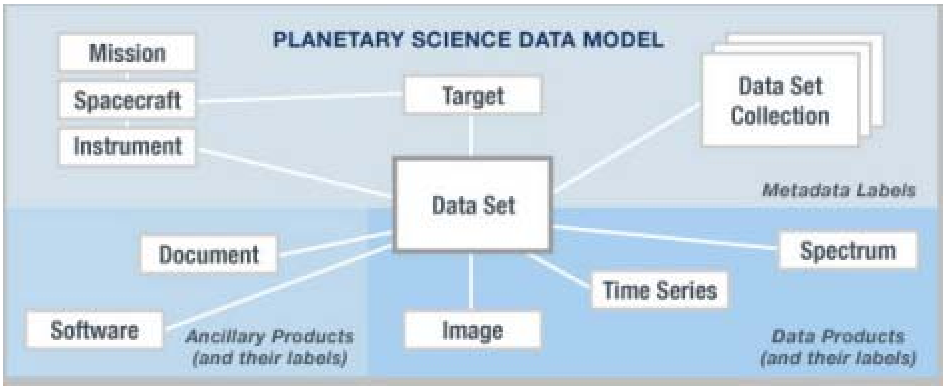


Figure 5. Planetary Science Data Model

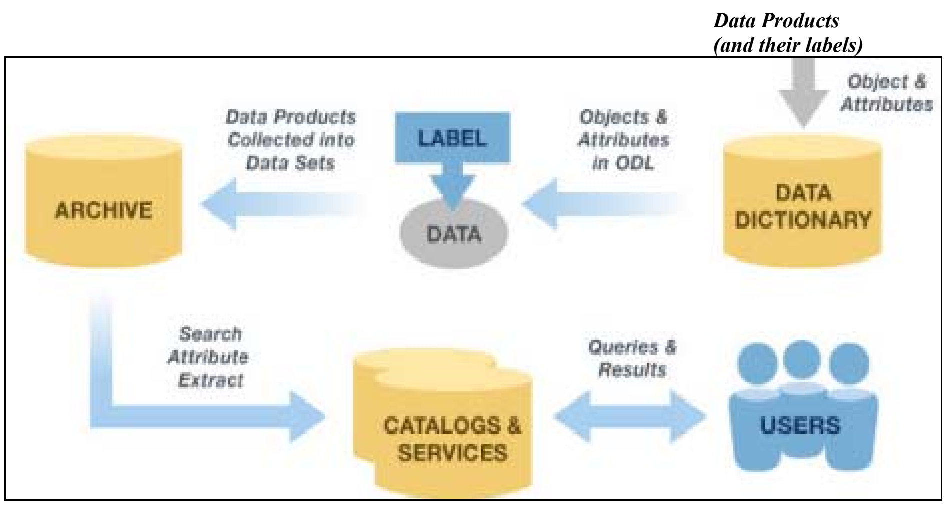


Figure 6. Planetary Data System

The data product labels and ancillary product labels are included on data volumes. When data volumes are transferred to the PDS, they are automatically cataloged by mission, target name, target type, instrument, instrument type and dataset identifier. Standard terminology is maintained in the Planetary Science Data Dictionary (PSDD) [PDS 2002], which is jointly maintained by the PDS and the multi-mission ground data system. The metadata values for new data products are compared with the PSDD and existing values used wherever possible.

C. Statement of Methodology

Due to extensive documentation of the Mars Global Surveyor mission and the Planetary Data System, that documentation was gathered and reviewed to answer the core questions posed to researchers, and any applicable domain and cross-domain questions. It is believed that the documentation also supports an understanding of records creation, maintenance, appraisal and preservation activities to a level of detail to support a walkthrough of the InterPARES 2 Chain of Preservation (COP) model. In addition, examples of experimental data records and reduced data records from the Mars Orbital Camera (MOC) were collected. A viewer for records in the Planetary Data System was also obtained. These examples support a walkthrough of the COP activity model as well as a diplomatic analysis of the scientific data records.

D. Description of Context

The context of digital record creation and management is described in five contexts: Provenancial, Juridical-Administrative, Procedural, Documentary and Technological.

Provenancial Context

The provenancial context of the data records includes the name of the creating body, its mandate, structure, and functions. Indicators of provenancial context include organizational charts, annual reports, the classification scheme, etc. The sponsor of the Mars Global Surveyor Mission/Project, and the records creator, was NASA - Office of Space Science. Figure 7 shows a high-level NASA organizational Chart. The Office of Space Science corresponds to the box labeled Science.

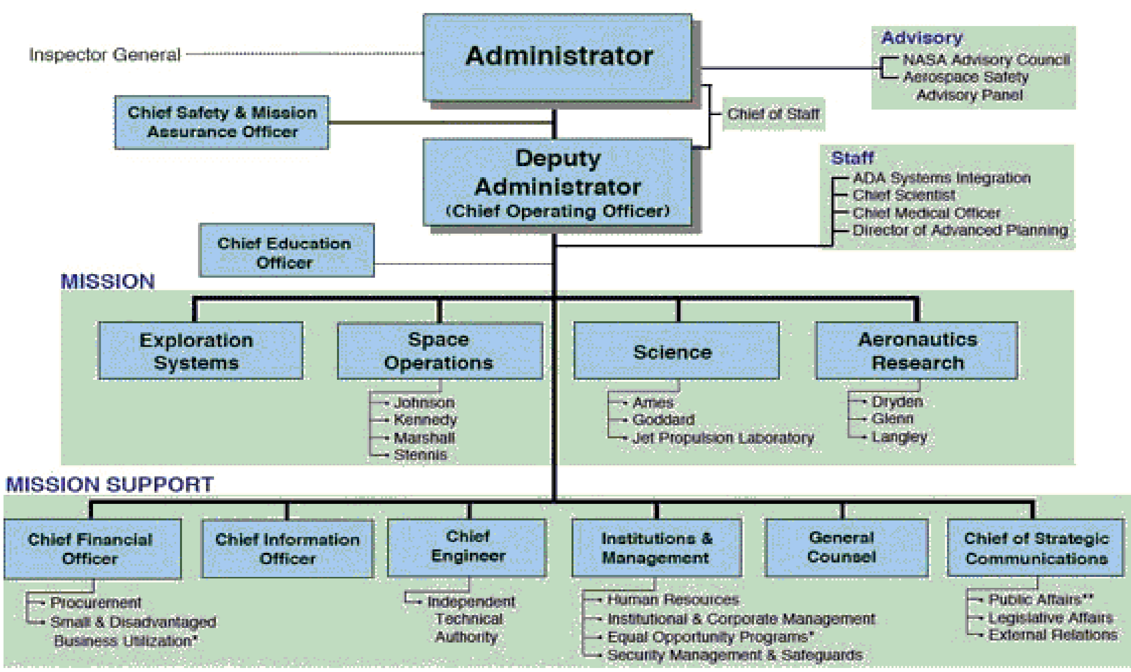


Figure 7. NASA Organizational Chart

Figure 8 shows the MARS Exploration Program Organization chart, one mission of which is the Mars Global Surveyor Project. Thomas E. Thorpe is the Mars Surveyor Operations Project Manager at NASA's Jet Propulsion Laboratory. He is responsible for the flight operations of all science Investigators and the facilities necessary to conduct instrument operation, data acquisition, telemetry processing and science data analysis and archiving to achieve the science objectives for the Mars Surveyor series of missions.

Arden L. Albee has served as Project Scientist for the Mars Observer and Mars Global Surveyor Missions since 1984. He is Professor of Geology and Planetary Science and Dean of Graduate Studies at the California Institute of Technology.

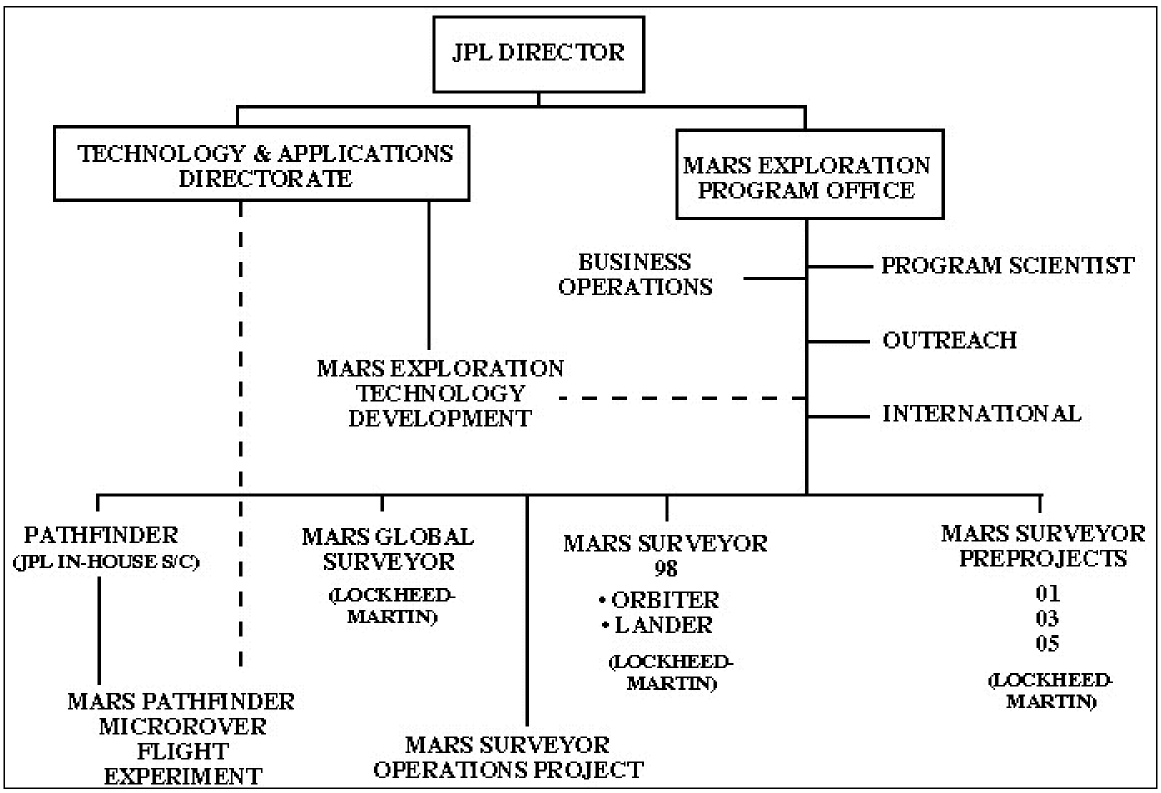


Figure 8. Mars Exploration Program Organization Chart [Shirley 1996]

The MGS Project Archive Generation, Validation and Transfer Plan sets forth the roles and responsibilities for persons and organizations creating, validating and transferring the MGS data records to the PDS.

“The Project Scientist and the Project Science Group (Project Scientist, Team Leaders, Principal Investigators, and Interdisciplinary Scientists) provide an oversight function for implementation of the Archive Generation, Validation, and Transfer Plan. The MGS Interdisciplinary Scientist for Data and Archives, with input from the SDVT and the MGS Project Science Group, will advise the Project with regard to archiving and will work with MGS and the PDS to help ensure that detailed plans are in place for generation of Planetary Data System-compatible products and associated documentation, and that archive volumes are generated, validated, and transferred to the Planetary Data System." [Arvidson et al 1995, pp. 4-5]

Juridical-Administrative Context

The Juridical-administrative context is the legal and organizational system in which the creating body operates. Laws and regulations indicate juridical-administrative context.

NASA regulations are published in the U.S. Code of Federal Regulations, Title 14 - Aeronautics and Space, Chapter V, National Aeronautics and Space Administration, Parts 1200-1299.[1]

"The National Aeronautics and Space Administration was established by the National Aeronautics and Space Act of 1958 (72 Stat. 426, 42 U.S.C. 2451 et seq.), as amended." "It is the purpose of the National Aeronautics and Space Administration to carry out aeronautical and space activities of the United States." "NASA is authorized to conduct research for the solution of problems of flight within and outside the Earth's atmosphere; to develop, construct, test, and operate aeronautical and space vehicles for research purposes; to operate a space transportation system including the space shuttle, upper stages, space program, space station, and related equipment; and to perform such other activities as may be required for the exploration of space." "NASA is headed by an Administrator, who is appointed from civilian life by the President by and with the advice and consent of the Senate. The Administrator is responsible, under the supervision and direction of the President, for exercising all powers and discharging all duties of NASA. The Deputy Administrator of NASA is also appointed by the President from civilian life by and with the advice and consent of the Senate. The Deputy Administrator acts with or for the Administrator within the full scope of the Administrator's responsibilities." [14 CFR § 1201, Subpart 1**]**

"NASA's basic organization consists of the Headquarters, eight field installations, the Jet Propulsion Laboratory (a Government-owned, contractor-operated facility), and several component installations which report to Directors of Field Installations. Responsibility for overall planning, coordination, and control of NASA programs is vested in NASA Headquarters." [14 CFR § 1201, Subpart 2**]**

Procedural Context

The procedural context is the business procedure in the course of which the digital entity is created. Indicators include workflow rules, codes of administrative procedure, classification schemes, etc.

The *PDS Data Preparation Workbook* (DPW) serves as a guide for the organization and preparation of datasets intended for submission to the Planetary Data System [JPL 1995a]. For active projects, archive planning consists of identifying the data to be archived, developing a detailed archiving schedule, and defining an end-to-end data flow through the ground system. NASA requires a Project Data Management Plan (PDMP) for all new projects. This plan provides a general description of the project data processing, cataloging, and communication

[1] http://www.access.gpo.gov/nara/cfr/waisidx_04/14cfrv5_04.html.

plan [JPL]. The Archive Policy and Data Transfer Plan (APDTP) provides a detailed description of the production and delivery plans for archive products for a project. A Data Product Software Interface Specification (SIS) is a document that describes the format and size of the individual data products. Figure 9 shows the process of MGS archive generation, validation and transfer [Arvidson 1999, p. 8].

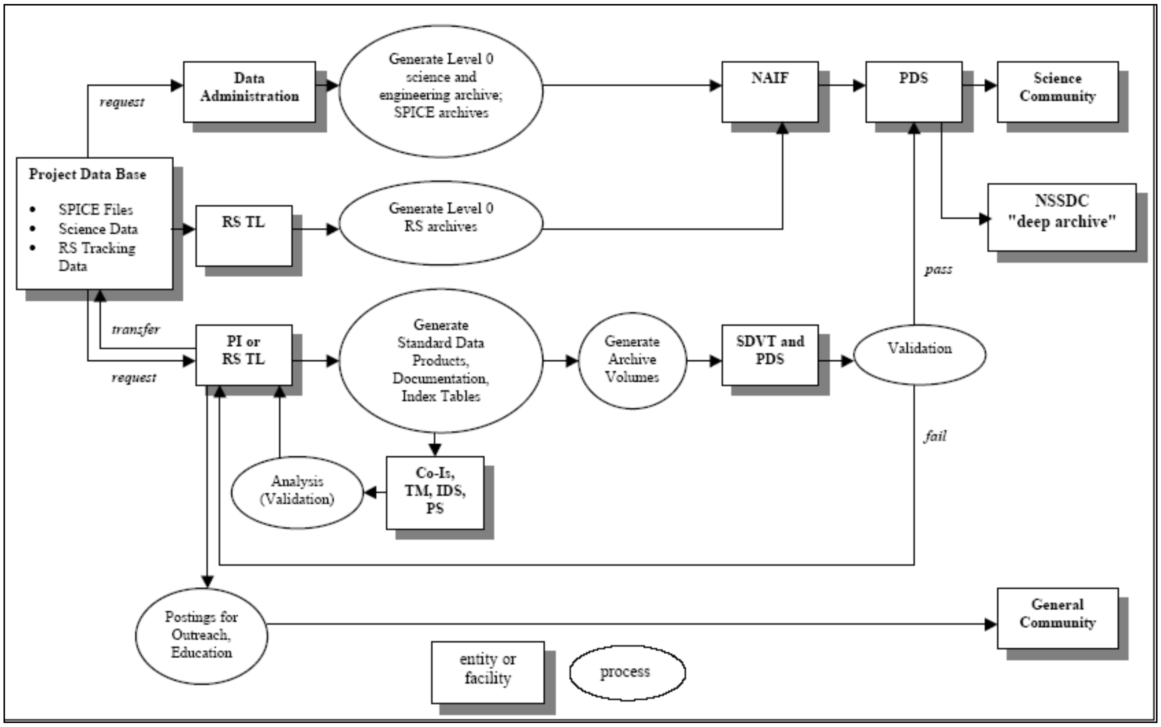


Figure 9. Process of MGS Archive Generation, Validation and Transfer.

The Project Database (PDB) contains data received from spacecraft instruments. Spacecraft Operations Planning Computers at the Principal Investigator's site, and at seven discipline nodes—Atmospheres, Geosciences, Imaging, Navigation Ancillary Information Facility (NAIF), PPI, Rings and Small Bodies—are used to process the data. PDS templates are used to create labels describing the context and structure of the data.

Experimental (or reduced) data records are transferred to the Science Data Validation Team for Peer Review. If judged complete, reliable and PDS-compliant, they are cataloged and stored in the PDS for access by planetary scientists. Planetary Scientists can find and retrieve MGS datasets using the PDS catalog.

The PDS is referred to as an "active archive," whereas the National Space Science Data Center's (NSSDC's) Repository is referred to as a "deep archive." The NSSDC and the PDS have a Memorandum of Understanding whereby the PDS is the entrance for Planetary Science data to the NSSDC long-term archives and the PDS will provide the NSSDC with copies of all PDS products for distribution as well as long-term digital product storage [MOU 2006].

Documentary Context

The documentary context of the MGS scientific data records is the fonds to which the digital entities belong and its internal structure. Indicators of documentary context include classification schemes, record inventories, indexes, registers, etc. In some organizations, business procedures are integrated with documentary procedures.

PDS datasets are organized by mission, spacecraft instrument and target (sun, planet, moon, comet, etc.). A PDS index table is contained in each dataset volume. Each row contains a product ID and the set of attributes describing that product. The attributes are typically the same ones found in the product's label.

Technological Context

The technological context of the MGS scientific data records is the characteristics of the digital environment in which the records are created and maintained.

The Project Database (PDB) contains data received from spacecraft instruments. Spacecraft Operations Planning Computers at the Principal Investigator's site, and at seven discipline nodes—Atmospheres, Geosciences, Imaging, Navigation Ancillary Information Facility (NAIF), PPI, Rings and Small Bodies—are used to process the data. PDS templates are used to create labels describing the context and structure of the data. The software tools used to create product labels and data products operates on Solaris and Linux operating systems.

Experimental (or reduced) data records are transferred to the Science Data Validation Team for Peer Review. If judged complete, reliable and PDS-compliant, they are cataloged and stored in the PDS for access by planetary scientists.

Planetary Scientists can find and retrieve MGS datasets using the PDS catalog. Software known as Label Library Light (L3) and Object Access Library (OAL) can be used to access the MGS data products for further analysis. A viewer known as NASAView can be used to view data such as MGS images.

E. Narrative Answers to the 23 Core Research Questions

Answers are given in this section to the twenty-three InterPARES 2 core research questions to be addressed by a case study.

1. What activities of the creator have you investigated?

The activities investigated include:
- generation of standard data products, documentation and index tables by the MGS MOC Science Team;
- generation of archive volumes;
- validation of the archive volumes by the SDVT and PDS;
- transfer to the PDS;
- access by the planetary science community; and
- transfer to the NSSDC deep archives.

2. Which of these activities generate the digital entities that are the objects of your case study?

The activities that generate the digital entities are:
- generation of standard data products;
- generation of documentation;
- generation of index tables;
- generation of archive volumes;
- validation of archive volumes; and
- transfer of archive volumes to PDS.

3. For what purpose(s) are the digital entities you have examined created?

The objectives of the Mars Global Surveyor Mars Orbital Camera (MOC) are to:
1. Obtain global, synoptic views of the Martian surface and atmosphere to study meteorological, climatological, and related surface changes during the course of the mission;
2. Examine and monitor surface and atmospheric features at moderate resolution for changes on time scales of hours, days, weeks, months, and years, and at a spatial scale that permits the details of their morphological character to be discriminated; and
3. Systematically examine local areas at extremely high resolution to quantify surface/atmosphere interactions and geological processes that operate on short time scales and at extremely small spatial scales. [JPL 1995c, p. D-16]

The PDS has three main user groups -- Planetary and Space Scientists, Mission Planners, and Education and Public Outreach. Each user group places different requirements on the system. The Scientists require high-resolution data, both raw and processed data products, rapid access to the latest data, detailed documentation, and transformation and data analysis support. The

Mission Planners required processed or derived data products, access to historic datasets, and detailed documentation. The Educators and Public require small amounts of highly derived products (maps, plots, animations, etc.), current and historic datasets, and easy to understand documentation.

4. What form do these digital entities take?

The kinds of digital entities created are described below. In the PDS, they are known as PDS structure objects. Structure objects outline the format in which the science data appear in PDS labels. An explanation of each PDS structure object is included in the PDS Standards Reference [JPL 2003]. For each object, there is text that describes the object, outlines its uses, and illustrates one or more examples.

ARRAY—"The ARRAY object is provided to describe dimensioned arrays of homogeneous objects. Note that an ARRAY can contain only a single object, which can itself be another ARRAY or COLLECTION if required. A maximum of 6 axes is allowed in an ARRAY."

BIT_COLUMN—"The bit_column object identifies a bit string embedded in a column. Bit_columns defined within columns are analogous to columns defined within rows."

BIT_ELEMENT—"The bit_element object identifies a bit string embedded in an element."

CATALOG—"The CATALOG object is used within a VOLUME object to reference completed PDS high level catalog templates. These provide additional information related to the datasets on the volume."

COLLECTION—"The COLLECTION object allows the ordered grouping of heterogeneous objects into a named collection. The COLLECTION object may contain a mixture of different object types including other COLLECTIONS."

COLUMN—"The COLUMN object identifies a single column in a data object. Columns must not contain embedded COLUMN objects."

CONTAINER—"The container object is a method of grouping a set of sub-objects (such as columns) that repeat within a data objects (such as a table). Use of the container object allows repeating groups to be defined within a data structure."

DIRECTORY—"The Directory object is used to define a hierarchical file organization on a media such as tapes or CD-ROMs. It identifies all directories and subdirectories below the root level. Subdirectories are identified by embedding DIRECTORY objects. Files within the directories and subdirectories are sequentially identified by using FILE objects with a sequence_number value corresponding to their position on the media."

DOCUMENT—"The DOCUMENT object is used to identify a particular document provided on a volume to support a dataset or dataset collection. A document can be made up of one or many files in a single format."

ELEMENT—"The ELEMENT object provides a means of defining a lowest level component of a data object that is stored in an integral multiple of 8-bit bytes. Element objects may be embedded in COLLECTION and ARRAY data objects."

FILE—"The file object is used to define the format of a file, to reference external files, and to indicate boundaries between label records and data records in data files with attached labels. In the PDS, the file object may be used in two ways: 1) As a container, or envelope, for label files. All label files contain an implicit file object that starts at the top of the label and ends where the label ends. In these cases, the PDS recommends against using the NAME keyword to reference the file name. 2) As an explicit object, used when a file reference is needed in a label, in which case the optional file_name data element is used to identify the file being referenced. The keywords in the file object always describe the file being referenced, not the file in which they are contained, i.e., if used in a detached label file, they describe the detached data file, not the label file itself."

HEADER—"The HEADER object is used to identify and define the attributes of commonly used header data structures for non-PDS formats such as VICAR or FITS. These structures are usually system or software specific and are described in detail in a referenced description text file."

HISTOGRAM—"The histogram object is a sequence of numeric values that provides the number of occurrences of a data value or a range of data values in a data object."

IMAGE—"An image object is a regular array of sample values. Image objects are normally processed with special display tools to produce a visual representation of the sample values. This is done by assigning brightness levels or display colors to the various sample values. Images are composed of LINES and SAMPLES. IMAGE objects may be associated with other objects, including HISTOGRAMs, PALETTEs, HISTORY, and TABLEs which contain statistics, display parameters, engineering values, or other ancillary data."

INDEX_TABLE—The INDEX_TABLE object is a specific type of TABLE object that provides information about the data stored on an archive volume. The INDEX table contains one row for each data file (or data product label file in the case where detached labels are used) on the volume. The table is formatted so that it may be read directly by many data management systems on various host computers. All fields (columns) are separated by commas, and character fields are enclosed by double quotation marks. Each record ends in a carriage return/line feed sequence. This allows the table to be treated as a fixed length record file on hosts that support this file type, and as a normal text file on other hosts.

PALETTE—"The PALETTE object is a sub-class of the table object. It contains entries that represent color assignments for SAMPLE values contained in an IMAGE."

QUBE—"The QUBE object is a multidimensional array (called the core) of sample values in multiple dimensions."

SERIES—"The series object is a sub-class of the table object. It is used for storing a sequence of measurements organized in a specific way (e.g., ascending time, radial distances)."
SPECTRUM—"The spectrum object is a form of table used for storing spectral measurements. The spectrum is assumed to have a number of measurements of the observation target taken in different spectral bands."

SPICE KERNEL—"The spice kernel object defines a single kernel from a collection of SPICE kernels. SPICE kernels provide ancillary data needed to support the planning and subsequent analysis of space science observations."

TABLE—"The TABLE object is a uniform collection of rows containing ASCII and/or binary values stored in columns."

TEXT—"The TEXT object provides general description of a file of plain text. It is recommended that text objects contain no special formatting characters, with the exception of the carriage return/line feed sequence and the page break."

VOLUME—"The volume object describes a physical unit used to store or distribute data products (e.g., a magnetic tape, CD-ROM disk, On-Line Magnetic disk or floppy disk) which contains directories and files. The directories and files may include a catalog of data products, an index table, documentation, software, calibration and geometry information as well as the actual science data (including for each data product a label and primary and supplemental data)."

4a. What are the key formal elements, attributes, and behavior (if any) of the digital entities?

The Planetary Science Data Dictionary (PSDD) contains the set of definitions for all attribute names that are valid for use in resource descriptions across the PDS. The PDS data nomenclature standards define the rules for constructing Data Element and Data Object names [JPL 2003].

A keyword (standard data element name) is an element of the Planetary Science Data Dictionary (PSDD) that defines a named property of an object. The keyword plus its value is an attribute.

A label (product label) is a resource description stored in a file. If the label is in the same file as the resource, it is called an attached label. If it is in a separate file, it is called a detached label.

The Object Description Language (ODL) is used to create labels (data descriptions) for data files and other objects such as software and documents. The PDS labels contain the key attributes of the digital objects. Figure 10 shows a PDS Label for a MOC image of the Martian surface.

Each line, other than comment lines, in the label contains an attribute. The attribute name (or keyword) is on the left of the equal sign and the attribute value on the right. The author of this data product is indicated by PRODUCER_ID = MGS_MOC_TEAM. The Subject or Matter is the RATIONALE_DESCRIPTION = "Monitor gullies and bright features in S05-01463 and E11-03412". The date of compilation is the PRODUCT_CREATON_TIME = 2005-12-28T22:01:28.

```
PDS_VERSION_ID = PDS3
FILE_NAME = "/tmp/S0902603.img"
RECORD_TYPE = FIXED_LENGTH
RECORD_BYTES = 2048
FILE_RECORDS = 6657
LABEL_RECORDS = 1
^IMAGE = 2
SPACECRAFT_NAME = MARS_GLOBAL_SURVEYOR
MISSION_PHASE_NAME = "SUPPORT"
TARGET_NAME = MARS
INSTRUMENT_ID = "MOC-NA"
PRODUCER_ID = MGS_MOC_TEAM
DATA_SET_ID = "MGS-M-MOC-NA/WA-2-SDP-L0-V1.0"
PRODUCT_CREATION_TIME = 2005-12-28T22:01:28
SOFTWARE_NAME = "makepds 1.11"
UPLOAD_ID = "UNK"
PRODUCT_ID = "S09/02603"
START_TIME = 2005-08-26T03:29:46.82
IMAGE_TIME = 2005-08-26T03:29:46.82
STOP_TIME = 2005-08-26T03:29:50.03
SPACECRAFT_CLOCK_START_COUNT = "809494240:80"
SPACECRAFT_CLOCK_STOP_COUNT = "N/A"
FOCAL_PLANE_TEMPERATURE = 275.7 <K>
GAIN_MODE_ID = "0A"
OFFSET_MODE_ID = "34"
LINE_EXPOSURE_DURATION = 0.482100 <MILLISECONDS>
DOWNTRACK_SUMMING = 1
CROSSTRACK_SUMMING = 1
EDIT_MODE_ID = "0"
RATIONALE_DESC = "Monitor gullies and bright features
in S05-01463 and E11-03412"
DATA_QUALITY_DESC = "OK"
ORBIT_NUMBER = 28896
OBJECT = IMAGE
LINES = 6656
LINE_SAMPLES = 2048
LINE_PREFIX_BYTES = 0
LINE_SUFFIX_BYTES = 0
SAMPLE_TYPE = UNSIGNED_INTEGER
SAMPLE_BITS = 8
SAMPLE_BIT_MASK = 2#11111111#
CHECKSUM = 16#FF208C26#
END_OBJECT = IMAGE
END
```

Figure 10. A PDS label for a MOC image of the Martian surface.

The label also describes the structure or format of the data. For this data product, the file format is described as 6657 records, each record of length 2048 bytes. There is one label record and the image begins at the second record. There are 6656 lines (or rows) in the image, and 2048 samples (or pixels) in each row. There are 8-bits in each pixel. The checksum is the integrity metadata.

The behaviors of a digital object consist of the various operations that can be performed on the object. For instance, an image object is a regular array of sample values. Image objects are normally processed with special display tools to produce a visual representation of the sample values. This operation on the digital object to produce a visual representation is a behavior of an image object. Other behaviors of these digital objects consist of the processing and analytic tools that can be used to create other objects, e.g., a tool to produce an image histogram from an image.

4b. What are the digital components of which they consist and their specifications?

This question was answered above in the descriptions of the types of digital objects created. For example, "The ELEMENT object provides a means of defining a lowest level component of a data object that is stored in an integral multiple of 8-bit bytes. Element objects may be embedded in COLLECTION and ARRAY data objects."

4c. What is the relationship between the intellectual aspects and the technical components?

The Planetary Science Data Model and data dictionary represent the intellectual concepts and relationships of the Planetary Science domain as well as the technological components of the data products and PDS Archive. Missions are associated with spacecraft, and spacecraft with Instruments and targets. Data products are created that are a labeled grouping of data resulting from measurements by an instrument and stored in a file. An example of a data product is a planetary image, a spectrum table, or a time series table. A product label identifies, describes, and defines the structure of the data. Data products are aggregated into volumes and volumes into datasets. A volume has a directory structure whose primary directories are shown in Figure 11. It is at this level of data objects, data object structure, volumes and directories that we encounter the technical components.

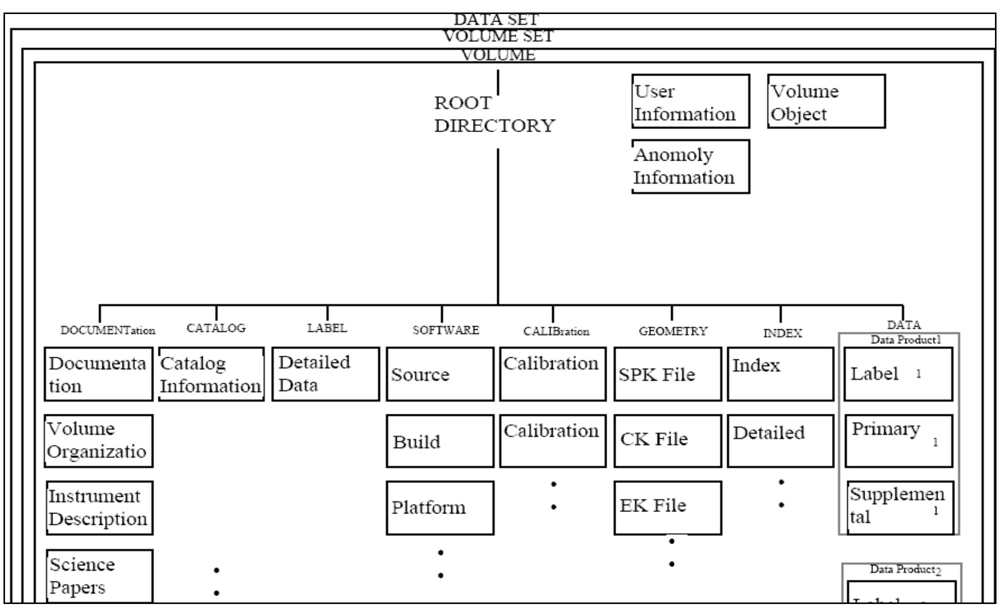


Figure 11. The Directory and File Structure of a Volume

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

The PRODUCT_ID data element represents a permanent, unique identifier assigned to a data product by its producer. In the PDS, the value assigned to PRODUCT_ID must be unique within its dataset.

The PDS Standards Reference [JPL 2003] also specifies the rules for dataset and volume names and ids. Each PDS dataset must have a unique DATA_SET_NAME and a unique DATA_SET_ID, both formed from up to seven components. The components are:

- Instrument host
- Target
- Instrument
- Data processing level number
- Dataset type (optional)
- Description (optional)
- Version number

Within the DATA_SET_ID, abbreviations are separated by hyphens. For instance, for the DATA_SET_ID

MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0

MGS is the instrument host,
M (Mars) is the target,
MOC-NA/WA (Mars Orbiter Camera-Narrow Angle/Wide Angle) is the instrument,
2 is the data processing level number,
DSDP (Distributed Standard Data Product) is the Dataset type,
LO is the Description, and
V1.0 is the version.

Within datasets, there are unique volume ids. Within volumes, the file names are unique. For instance, data file names have the form

MMMNNNOO.IMG

Where MMM is the mission phase, NNN is the mission phase index (For pre-mapping data, NNN is the orbit number), and OO is a unique index within that orbit.

4e. In the organization of the digital entities, what kind of aggregation levels exists, if any?

Data products, software and documents are aggregated into volumes, volumes aggregated into datasets, and datasets are associated with instruments and Targets, targets and instruments are associated with spacecraft and spacecraft with a mission.

4f. What determines the way in which the digital entities are organized?

The PDS data model determines the way in which the data objects are organized. That data model was determined by an identification of the classes of objects in the planetary science domain, classes of data products and technologies for aggregating those data products.

5. How are those digital entities created?

5a. What is the nature of the system(s) with which they are created? (e.g., functionality, software, hardware, peripherals etc.)

A Project Data Base (PDB) maintains the digital versions of experimental data records (EDRs), SPICE files and Standard Data products that are acquired from the spacecraft or created. Science Operations Planning Computers (SOPC) interface to the PDB. The SPICE System is used with the unprocessed data records (UDRs) to produce experimental data records. The software used on the SOPC includes a routine for associating a label template with a data object. Volume Verifier software is used by the Science Data Validation Team (SDVT) to validate standard data products, that is, to determine whether they are PDS compliant.

5b. Does the system manage the complete range of digital entities created in the identified activity or activities for the organization (or part of it) in which they operate?

In addition to the Space Operation Planning computers, there are additional computers, e.g., at Goddard Space Flight Center (GSFC), that are used for processing and analysis of science data. These are part of the system for producing standard and specialized science data products.

6. From what precise process(es) or procedure(s), or part thereof, do the digital entities result?

The digital scientific data records originate from the MGS spacecraft in the form of instrument measurements (observations) of a target. The Command and Data Handling Subsystem (C&DHS) records science and engineering data for later transmission or transmits this data in real-time [JPL 1995c].

This data is transmitted through the Deep Space Network to Mission Ground Control where it is accumulated in a Project Data Base. The process of MGS archive generation, validation and transfer shown in Fig. 9 then occurs.

In addition to the analysis of the data by the project science team, the scientific data is provided to planetary scientists for scientific analysis and potential subsequent scientific publication.

7. To what other digital or non-digital entities are they connected in either a conceptual or a technical way? Is such connection documented or captured?

The primary classes in the Planetary Science Data Model are Mission, Spacecraft, Instrument and Target. The objects in these classes are not digital objects. Instruments are associated with spacecraft and targets, and spacecraft are associated with missions. Other classes include Document, Software, Image, Time Series and Spectrum. Objects in these classes are digital objects. These digital objects are aggregated into volumes and volumes are aggregated into datasets. Datasets are associated with instruments and targets.

8. What are the documentary and technological processes or procedures that the creator follows to identify, retrieve, and access the digital entities?

The documentary and technological processes used to identify the digital entities has already been discussed in the answer to question 4d. The PRODUCT_ID data element represents a permanent, unique identifier assigned to a data product by its producer. In the PDS, the value assigned to PRODUCT_ID must be unique within its dataset.

Ingest is the PDS function that receives, stores and catalogs resources. When volumes and datasets are transferred to the PDS, the catalog information for the dataset and the Volume descriptions are used to automatically update the PDS catalog. Volumes are indexed by mission, target name, target type, instrument and instrument type and volume identifier attributes.

Search is the PDS function that supports the location of resources based on their cataloged attributes. Retrieve is the PDS function that enables the user to download resources to the user's computer. There is also a browse function that allows users to view resources online through a Web browser.

NASAView is the display tool for the PDS archive that runs of multiple platforms. This application was constructed using Label Library Light (L3) [Hughes 1998], Object Access Library (OAL) [Davis and Monk 1997] and the XVT Portability Toolkit (tm). Label Library Light reads and parses a data product label and creates an in-memory representation of the label information. The object access library uses the parse tree and accesses the actual PDS object. The XVT Development Tool supplies the cross-platform GUI for NASAView.

OAL allows the inclusion of standard routines for processing specialized representations. For example, the decompression routines for JPEG, GIF, and other standard compression schemes have been included. PDS labels have also been created to describe non-PDS data formats such as FITS and VICAR labeled images. Using these “detached” PDS labels to describe non-PDS formatted data, standard OAL function calls can be used to access the data.

Figure 12 shows an image of the Martian surface captured with the Mars Orbital Camera and displayed with NASAView. This image (S0902603.img) shows a new gully deposit in a crater in Terre Sirenum (at 36.6° S, 168.8° W) that may provide evidence that water flowed on Mars in this decade. This image was captured 26 August 2005. The light-toned gully deposit near the center of the displayed mage does not appear in a MOC image (E11-03412.img) of the same area captured 22 December 2001. “The new materials covers the entire gully floor from the point

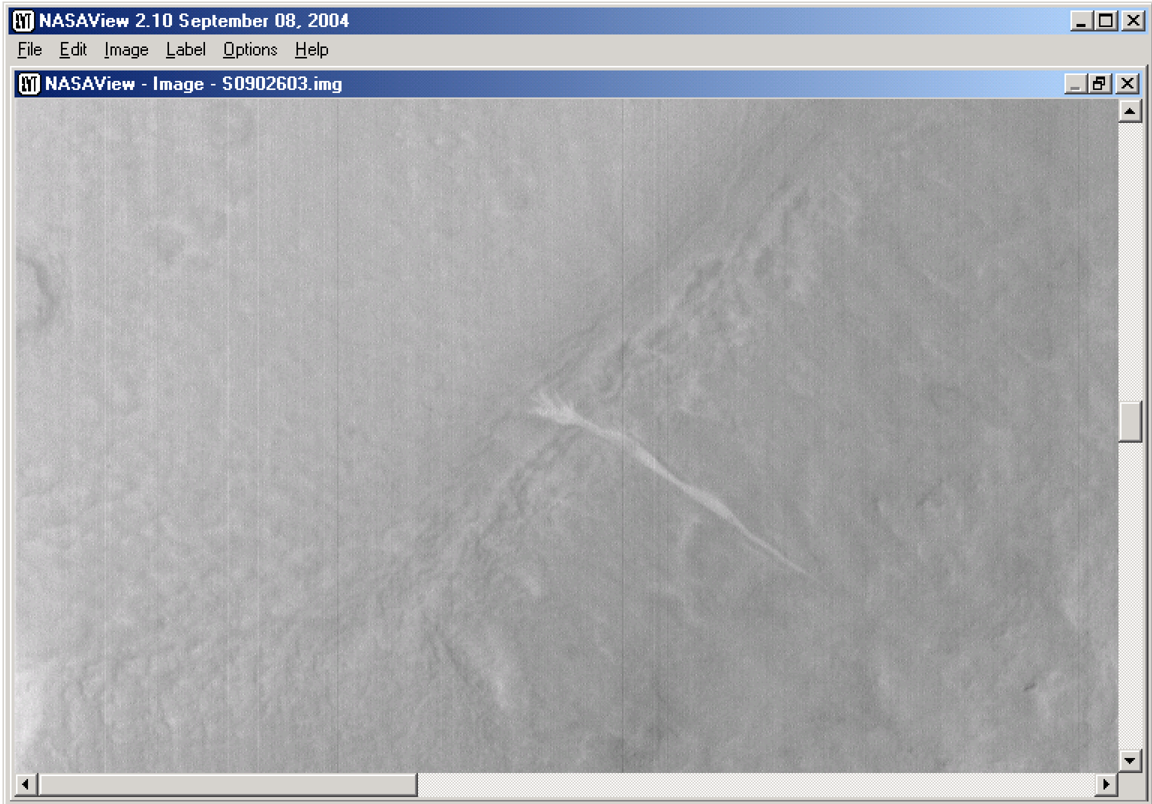


Figure 12. PDS Image S0902603.img Displayed with NASAView

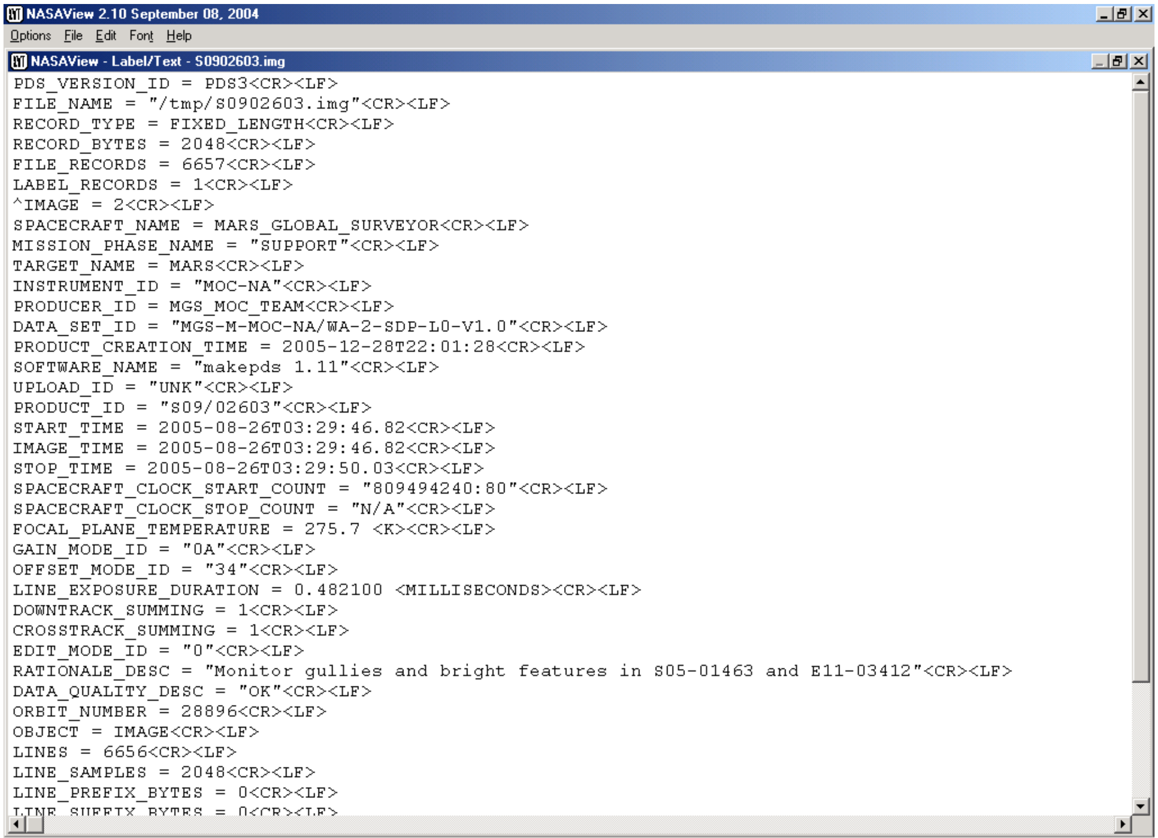
where the gully emerges from beneath a mantled slope, down to where the spot at which the channel meets the crater floor. At this break in slope the gully material, as it was emplaced, spread out in five or six fingers."[2]

If one selects Label from the Menu bar in NASAView, the label for the data product is displayed as shown in Figure 13. This is the same label as that shown in Figure 10.

9. Are those processes and procedures documented? How? In what form?

The identification procedures and standards are described in the PDS Standards Reference. Documentation of the Ingest and Search can be found at the Planetary Data Systems Homepage (http://pds.jpl.nasa.gov/). An access procedure to the labels is described by Hughes et al. [1998]. The Object Access Library is described by Davis and Monk [1997].

[2] http://www.msss.com/mars_images/moc/2006/12/06/gullies/sirenum_crater/index.html.




```
PDS_VERSION_ID = PDS3<CR><LF>
FILE_NAME = "/tmp/S0902603.img"<CR><LF>
RECORD_TYPE = FIXED_LENGTH<CR><LF>
RECORD_BYTES = 2048<CR><LF>
FILE_RECORDS = 6657<CR><LF>
LABEL_RECORDS = 1<CR><LF>
^IMAGE = 2<CR><LF>
SPACECRAFT_NAME = MARS_GLOBAL_SURVEYOR<CR><LF>
MISSION_PHASE_NAME = "SUPPORT"<CR><LF>
TARGET_NAME = MARS<CR><LF>
INSTRUMENT_ID = "MOC-NA"<CR><LF>
PRODUCER_ID = MGS_MOC_TEAM<CR><LF>
DATA_SET_ID = "MGS-M-MOC-NA/WA-2-SDP-L0-V1.0"<CR><LF>
PRODUCT_CREATION_TIME = 2005-12-28T22:01:28<CR><LF>
SOFTWARE_NAME = "makepds 1.11"<CR><LF>
UPLOAD_ID = "UNK"<CR><LF>
PRODUCT_ID = "S09/02603"<CR><LF>
START_TIME = 2005-08-26T03:29:46.82<CR><LF>
IMAGE_TIME = 2005-08-26T03:29:46.82<CR><LF>
STOP_TIME = 2005-08-26T03:29:50.03<CR><LF>
SPACECRAFT_CLOCK_START_COUNT = "809494240:80"<CR><LF>
SPACECRAFT_CLOCK_STOP_COUNT = "N/A"<CR><LF>
FOCAL_PLANE_TEMPERATURE = 275.7 <K><CR><LF>
GAIN_MODE_ID = "0A"<CR><LF>
OFFSET_MODE_ID = "34"<CR><LF>
LINE_EXPOSURE_DURATION = 0.482100 <MILLISECONDS><CR><LF>
DOWNTRACK_SUMMING = 1<CR><LF>
CROSSTRACK_SUMMING = 1<CR><LF>
EDIT_MODE_ID = "0"<CR><LF>
RATIONALE_DESC = "Monitor gullies and bright features in S05-01463 and E11-03412"<CR><LF>
DATA_QUALITY_DESC = "OK"<CR><LF>
ORBIT_NUMBER = 28896<CR><LF>
OBJECT = IMAGE<CR><LF>
LINES = 6656<CR><LF>
LINE_SAMPLES = 2048<CR><LF>
LINE_PREFIX_BYTES = 0<CR><LF>
LINE_SUFFIX_BYTES = 0<CR><LF>
```

Figure 13. Label for S0902603.img Displayed with NASAView.

10. What measures does the creator take to ensure the quality, reliability and authenticity of the digital entities and their documentation?

The *PDS Data Preparation Workbook* (DPW) serves as a guide for the organization and preparation of datasets intended for submission to the Planetary Data System (PDS) [JPL 1995a]. For active projects, archive planning consists of identifying the data to be archived, developing a detailed archiving schedule, and defining an end-to-end data flow through the ground system. NASA requires a Project Data Management Plan (PDMP) for all new projects. This plan provides a general description of the project data processing, cataloging, and communication plan [JPL]. The Archive Policy and Data Transfer Plan (APDTP) provides a detailed description of the production and delivery plans for archive products for a project. A Data Product Software Interface Specification (SIS) is a document that describes the format and size of the individual data products.

All data incorporated into the PDS archives must undergo a peer review. The purpose of the review is to determine that:

- the data are accurate, complete and reliable;
- the data are suitable for archiving; and
- the PDS standards have been followed. [JPL 2003]

The primary validation tool of the PDS is the Volume Verifier. The Central Node data engineers run this program on each product delivered from a project. It validates the format and content of all product labels, and validates the integrity of data files using checksums.

Access to PDS resources is controlled through user authentication. Access to restricted areas of the PDS is determined by the user's assigned role.

11. Does the creator think that the authenticity of his digital entities is assured, and if so, why?

Project team members, PDS managers and engineers and other planetary scientists do not traditionally use the term “authentic” to characterize the data products that they create, maintain and use. They are concerned that the data records are complete, reliable, accurate and that the integrity of the data record is assured.

Given the definition of authentic digital record as “a digital record that is what it purports to be and is free from tampering or corruption,” one can conclude that, due to the emphasis on completeness and reliability of the planetary science data records, the peer review, role-based authentication of access to archived data products and data integrity checks, the scientific data records are maintained authentic. However, it is unlikely that the creators or maintainers of the records would customarily use that term to characterize their quality.

12. How does the creator use the digital entities under examination?

The Mars Global Surveyor science community includes Principal Investigators, Interdisciplinary Scientists, Co-Investigators, and will include Participating Scientists. It also includes a Radio Science Team, which utilizes both spacecraft and ground equipment to conduct experiments [JPL 1995b].

Principal Investigators and the Radio Science Team Leader are responsible for the acquisition, reduction, and analysis of data from their instruments. They are also responsible for generation of standard data products and archive volumes for their instruments.

Interdisciplinary Scientists will enhance science analyses by participating in reducing data for specific instruments, and/or by conducting analyses involving multi- instrument datasets to address specific science objectives.

Participating Scientists will be added during the Mars Global Surveyor operations and data analysis phase of the mission. Participating Scientists will fulfill Co-Investigator, Team Member, or Interdisciplinary Scientist functions during the Mission.

It is expected that the Interdisciplinary Scientists will chair Science Working Groups. These Groups will be composed of relevant Mars Global Surveyor scientists and will focus on coordinated data analyses. Other Mars Global Surveyor scientists will, from time to time, chair other groups to facilitate coordinated analyses. For planning purposes, it is assumed that there will be five groups to cover science (atmospheres, polar processes, atmosphere-surface interactions, geodesy and geophysics) and one to cover data and archives.

The Project Science Group will be chaired by the Project Scientist and will include Principal Investigators, and Interdisciplinary Scientists and the Radio Science Team Leader as members. The Project Science Group also includes the Mars Global Surveyor Program Scientist from NASA Headquarters and, by invitation, the principal interface for the Mars Balloon Relay Experiment. The Project Science Group will establish science and measurement strategies and also establish and coordinate cross-instrument and cross-discipline data analyses. The Science Working Groups are subgroups of the Project Science Group. The Project Scientist will track agreements for cooperative analyses among Mars Global Surveyor scientists and will, with the help of the Project Science Group, adjudicate any disputes that may arise.

13. How are changes to the digital entities made and recorded?

The lifecycle of a resource is the set of distinct changes that a resource undergoes from its creation to its final place, if any, on the System and in the Archive. The status of a resource is the attributes that represent the state of a resource at a given point in time. Any set of attributes can be defined for this purpose and thus be "tracked" by System. However, it is natural to use the STATUS keywords, which define the state of a resource with respect to a given lifecycle. For example, the ARCHIVE_STATUS of a dataset tells which phase of the archive lifecycle the dataset is in. The ARCHIVE_STATUS_NOTE element is available to describe the ARCHIVE_STATUS value in finer detail. The ARCHIVE_STATUS_DATE element provides the date that the archive status will in the future, or has in the past, changed.

The Track function is the System function that reports on the status of a resource, starting from the time it is submitted to the PDS and continuing throughout its lifecycle.

14. Do external users have access to the digital entities in question? If so, how, and what kind of uses do they make of the entities?

Because of the widespread scientific and public interest in new results from Mars and the strong commitment of Mars Global Surveyor scientists to releasing data on a timely basis, clear release policies have been established [JPL 1995b].

Generation and validation of products and archive volumes may require up to six months from time of receipt of raw data at instrument processing facilities in order to produce and validate useful archive volumes. After the six-month generation/validation period, the relevant archive volumes and posted products will be transferred to the Planetary Data System, which will make them available to the general science community.

To ensure rapid dissemination of new and significant information, each Team Leader and 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 of data receipt. Postings on Internet are used as a cost-effective way for widespread dissemination of these special products. The posted data will include images, derived spectra, topographic information, and other forms of data that illustrate new and significant results. (See Mars Global Surveyor Homepage http://mars.jpl.nasa.gov/mgs/)

15. Are there specific job competencies (or responsibilities) with respect to the creation, maintenance, and/or use of the digital entities? If yes, what are they?

The MGS Project Archive Generation, Validation and Transfer Plan [Arvidson et al. 1995, pp. 4-5] sets forth the roles and responsibilities for persons and organizations creating, validating and transferring the MGS data records to the PDS.

"The Project Scientist and the Project Science Group (Project Scientist, Team Leaders, Principal Investigators, and Interdisciplinary Scientists) provide an oversight function for implementation of the Archive Generation, Validation, and Transfer Plan. The MGS Interdisciplinary Scientist for Data and Archives, with input from the SDVT and the MGS Project Science Group, will advise the Project with regard to archiving and will work with MGS and the PDS to help ensure that detailed plans are in place for generation of Planetary Data System-compatible products and associated documentation, and that archive volumes are generated, validated, and transferred to the Planetary Data System."

"MGS Data Administration will compile archive volumes containing engineering, level 0 science packets from spacecraft instruments, DSN monitor data, and SPICE files. The Radio Science Team Leader will produce the Radio Science level 0 archive collection. Principal Investigators are responsible for generation of reduced science data records, documentation, algorithms/software to generate levels 1 to 3 products, and archive volumes containing standard products and supporting information. The Radio Science Team Leader will not supply reduction software, but will instead provide documentation to explain how the processing is carried out. … The author of each archive collection (and associated volumes) is responsible for publishing a Software Interface Specification document that delineates the format and content of the respective volumes. These SISs are due in final form by the time of orbit insertion. Generally, the relevant SIS is included on the archive volume."

"Each Principal Investigator and the Radio Science Team Leader will also be responsible for posting a subset of reduced data (and relevant documentation) on a system accessible via the Internet for public access."

"The Science Data Validation Team (SDVT) will be responsible for reviewing the initial archive volumes."

The PDS Management Council is the governing body of the PDS, consisting of the project manager, project scientist, project engineer, mission interface lead, and managers of all the PDS discipline nodes.

16. Are the access rights (to objects and/or systems) connected to the job competence of the responsible person? If yes, what are they?

Access to restricted areas of the PDS is determined by the user's assigned role. In the PDS, roles may include Subscriber, Data Engineer, System Administrator, etc. This type of access control is called role-based authentication.

17. Among its digital entities, which ones does the creator consider to be records and why?

The measurements received from the instruments at the mission ground control system are *experimental data records* (EDRs) [JPL 1995, pp. 5-45]. Processed EDRs are referred to as *reduced data records* (RDRs) [JPL 2005, pp. 7-14].

Engineering data records (EDRs) consist of raw data in the form of packets containing time-ordered sequences of science data obtained by a given instrument together with engineering information that allows instrument teams to check operations of its instruments.

The MGS mission and PDS documentation also refer to the digital objects created as data products**.** A data product is a resource consisting of a PDS compliant label and one or more files containing data. The label provides a resource description of the data files.

Data products consist of EDRs, SPICE kernels, Standard Data Products, and Special Data Products (SPDPs). SPDPs are defined as those science data products produced during the course of science analysis.

18. Does the creator keep the digital entities that are currently being examined? That is, are these digital entities part of a recordkeeping system? If so, what are its features?

Volumes of created data products are submitted to the Planetary Data System, which keeps the digital objects that are created. The PDS is referred to as an active archive, but it resembles a recordkeeping system, albeit of the records of scientific activities, rather than business activities.

The PDS was also designed to be an archive for long-term preservation of planetary science data records. The peer review process resembles the appraisal process for selection of records for long-term preservation. The datasets have already been arranged and described at the time they are transferred to the PDS. Therefore, the PDS could also be viewed as an archive for long-term preservation, rather than a recordkeeping system for active records.

18a. Do the recordkeeping system(s) (or processes) routinely capture all digital entities within the scope of the activity it covers?

The purpose of the Planetary Data System is to archive and provide access to the scientific data records of the NASA Planetary Science spacecraft missions. The PDS receives, catalogs, stores, and provides access to EDRs, SPICE kernels, Standard Data Products, and Special Data Products (SPDPs) created from data received at JPL during these missions. Therefore, the PDS seems to be routinely capturing all the digital entities within the scope of the activity it covers.

18b. From what applications do the recordkeeping system(s) inherit or capture the digital entities and the related metadata (e.g., e-mail, tracking systems, workflow systems, office systems, databases, etc.)?

Instrument measurements are sent as data packets from spacecraft through the Deep Space Network to computers at the Mission Ground Station at JPL. Science Operations Planning Computer (SOPC) workstations located at Principal Investigator, Team Leader, and

Interdisciplinary Scientist home institutions are electronically connected via NASCOM and Ethernet links to a project database (PDB) at JPL. The workstations are used to create standard data products, documentation and index tables. These are packaged into archive volumes and sent to the Science Data Validation Team (SDVT) for validation. The SDVT transfers the archive volumes to the PDS where there is additional validation.

18c. Are the digital entities organized in a way that reflects the creation processes? What is the schema, if any, for organizing the digital entities?

The schema for organizing and describing these data objects are the PDS data Model, the PDS templates used to create data products labels and the directory and file structures of volumes.

<table>
<tr><th>DIRECTORY/FILE</th><th>CONTENTS</th></tr>
<tr><td>-------------------------</td><td>--</td></tr>
<tr><td><root></td><td></td></tr>
<tr><td>|-INDEX.HTM</td><td>Starting point for web-browsing CD contents.</td></tr>
<tr><td>|-AAREADME.TXT</td><td>The file you are reading (ASCII Text).</td></tr>
<tr><td>|-ERRATA.TXT</td><td>Description of known anomalies and errors present on the volume set (optional file).</td></tr>
<tr><td>|-VOLDESC.CAT</td><td>A description of the contents of this volume in a format readable by both humans and computers.</td></tr>
<tr><td>|-<CATALOG></td><td>Catalog Directory</td></tr>
<tr><td>| |-CATINFO.TXT</td><td>Describes Contents of the Catalog directory</td></tr>
<tr><td>| |-DATASET.CAT</td><td>Dataset description.</td></tr>
<tr><td>| |-DSMAP.CAT</td><td>Map Projection description.</td></tr>
<tr><td>| |-INSTHOST.CAT</td><td>Spacecraft description.</td></tr>
<tr><td>| |-MISSION.CAT</td><td>Mission description.</td></tr>
<tr><td>| |-PERSON.CAT</td><td>Contributors to this dataset.</td></tr>
<tr><td>| |-REFS.CAT</td><td>References</td></tr>
<tr><td>| |-MOCINST.CAT</td><td>MOC instrument description.</td></tr>
<tr><td>|</td><td></td></tr>
<tr><td>|-<DOCUMENT></td><td>Documentation Directory. The files in this directory provide detailed information regarding the MOC DSDP archive.</td></tr>
<tr><td>| |-DOCINFO.TXT</td><td>Description of files in the DOCUMENT directory</td></tr>
<tr><td>| |-VOLINFO.TXT</td><td>Documentation regarding the contents of this CD Volume Set.</td></tr>
<tr><td>| |-VOLINFO.LBL</td><td>PDS Label file describing the VOLINFO documents</td></tr>
<tr><td>| |-MOCSIS.TXT</td><td>Mars Observing Camera Software Interface Specification document.</td></tr>
<tr><td>|-<INDEX></td><td>Directory for the image index files.</td></tr>
<tr><td>| |-INDXINFO.TXT</td><td>Description of files in <INDEX> directory.</td></tr>
<tr><td>| |-IMGINDX.TAB</td><td>Image Index table.</td></tr>
<tr><td>| |-CUMINDX.TAB</td><td>Cumulative Image Index table for all volumes.</td></tr>
<tr><td>| |-IMGINDX.LBL</td><td>PDS label for IMGINDX.TAB.</td></tr>
<tr><td>|</td><td></td></tr>
<tr><td>|-<data directories></td><td>Data directory names indicating the mission phase (first three characters) and a unique-within-phase index (last three characters). For pre-mapping data, this has the form MMMNNN, where NNN is the orbit number. Files within each directory have the form MMMNNNOO.IMG, where MMM and NNN are as above and OO is a unique index within that orbit.</td></tr>
<tr><td>|-<BROWSE></td><td>The files in this directory provide HTML browsing of the files contained on this volume.</td></tr>
<tr><td>|-MAP.HTM</td><td>Map-based browser page.</td></tr>
<tr><td>|-TAB.HTM</td><td>Table-based browser page.</td></tr>
</table>

Figure 14. Directory structure of a PDS CD Volume

The files in a volume are organized starting at the root directory. Below the root directory is a directory tree containing data, documentation, and index files. In Figure 14, directory names are indicated by angle brackets (<...>), upper-case letters indicate an actual directory or file name, and lower-case letters indicate the general form of a set of directory or file names.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

When datasets are transferred to the PDS, they are automatically cataloged by mission, target name, target type, instrument, instrument type and dataset identifier. Figure 15 shows the Dataset Search interface of the PDS.[3]

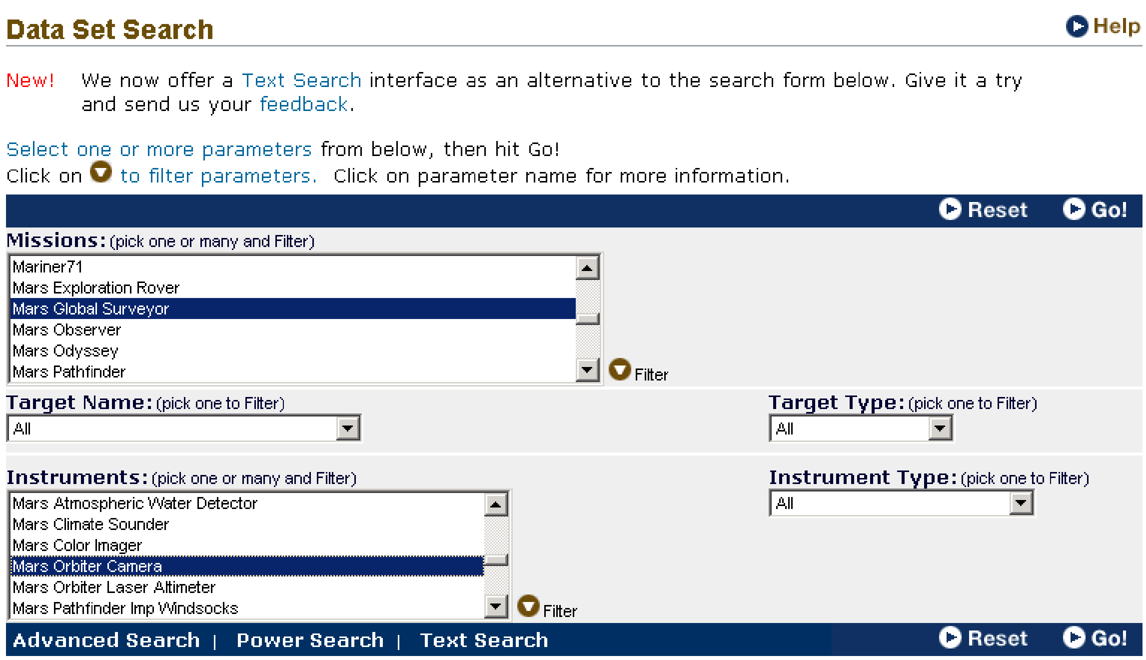


Figure 15. PDS Dataset Search

If one selects the subject of the case study, the Mars Global Explorer Mission and the Mars Orbital Camera Instrument, and then selects Go, the page shown in Figure 16 is displayed.

If one selects the dataset that is being used in this walkthrough, MGS-M-MOC-NA/WA-2-DSDP-LO-V1.0 (Mars Orbital Camera, Decompressed Standard Data Products), the page shown in Figure 17 is displayed.

This Web page allows one to view the description of the dataset—Mission, Dataset, Instrument and Target Information. If one selects Dataset Information, one will see displayed the descriptive information for the dataset.

3 http://pds.jpl.nasa.gov/data_services/.

Search Results (2 data sets found) Help

Data Set	Instrument Host	Information About the Data Set	Data Products & Related Files	Other Resources
1. Mars Global Surveyor Imaging (MOC) Mars Mapping Phase, Decompressed Standard Data Products	MGS	MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0	Search for Products with Atlas	• Atlas • Imaging Online Archives • EN Backup Volumes Online • Atlas • MGS Home Page
2. MOC COMPRESSED IMAGES	MGS	MGS-M-MOC-NA/WA-2-SDP-L0-V1.0	Search for Products with Atlas	• Atlas • Imaging Online Archives • EN Backup Volumes Online • Atlas • MGS Home Page

Page 1 | New Search

Figure 16. Datasets for the MGS Mars Orbital Camera.

Mars Global Surveyor Imaging (MOC) Mars Mapping Phase, Decompressed Standard Data Products

Citation	Citation TBD
Access/Download Data Set	Search for Products with the Atlas
Data set abstract	This data set contains portions of the MOC Decompressed Standard Data Product (DSDP) Archive, a collection of decompressed images from the Mars Orbiter Camera on the Mars Global Surveyor spacecraft. Images are stored with PDS labels, but are otherwise unprocessed and uncalibrated.

Additional Information	
Mission Information	MARS GLOBAL SURVEYOR
Dataset Information	MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0
Instrument Host Information	MGS
Instrument Information	MOC
Target Information	MARS
Other Resources	Atlas Imaging Online Archives EN Backup Volumes Online Atlas MGS Home Page

Figure 17. Dataset Descriptions

As shown in Figure 18, a second method of accessing MGS Planetary Science Data from the Data Services Web page is to select PDS Explorer, which allows one to browse PDS datasets by Target, Mission, Instrument and Dataset ID.

If one selects for the MGS Mission the dataset ID used in this walkthrough, the Web page shown in Figure 19 is displayed. If one selects the mgsc_001 volume, the Web page shown in Figure 20 is displayed.

18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

The PDS logs accesses to restricted areas of the system. User id, date, time and operation are logged.

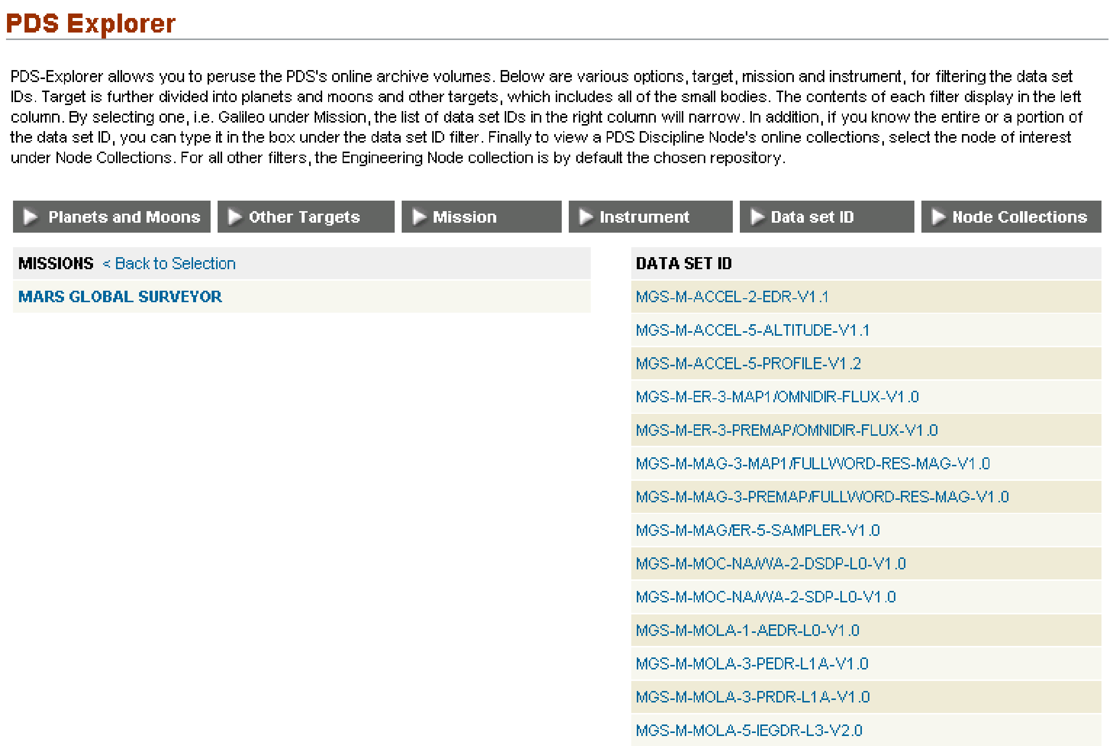


Figure 18. PDS Explorer for Browsing PDS Datasets

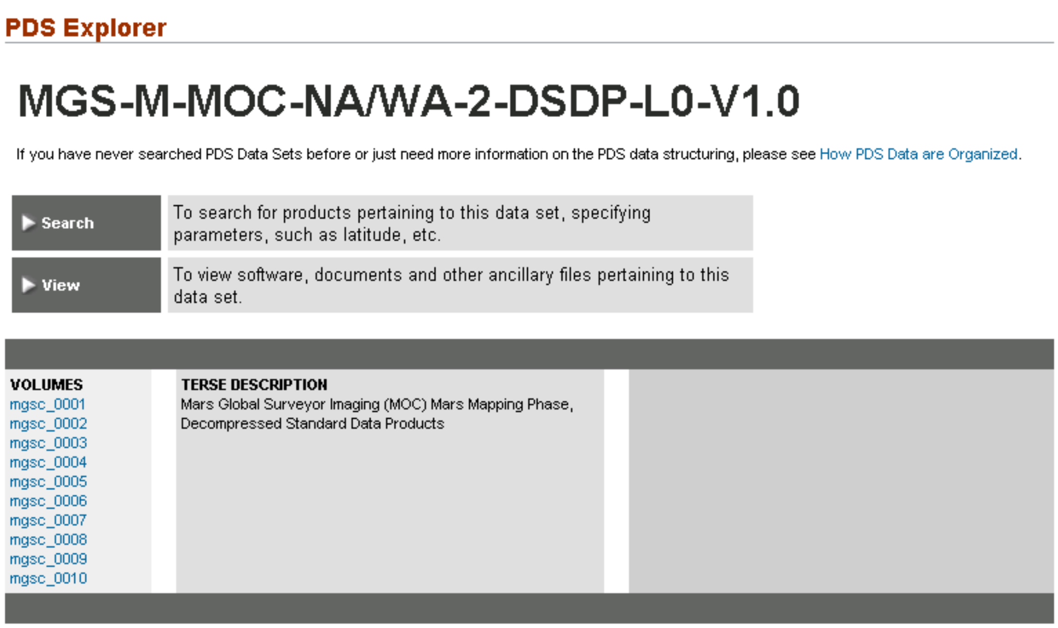


Figure 19. Volumes in the MGS-M-MOC-NA/WA-2-DSPS-L0-V1.0 Dataset

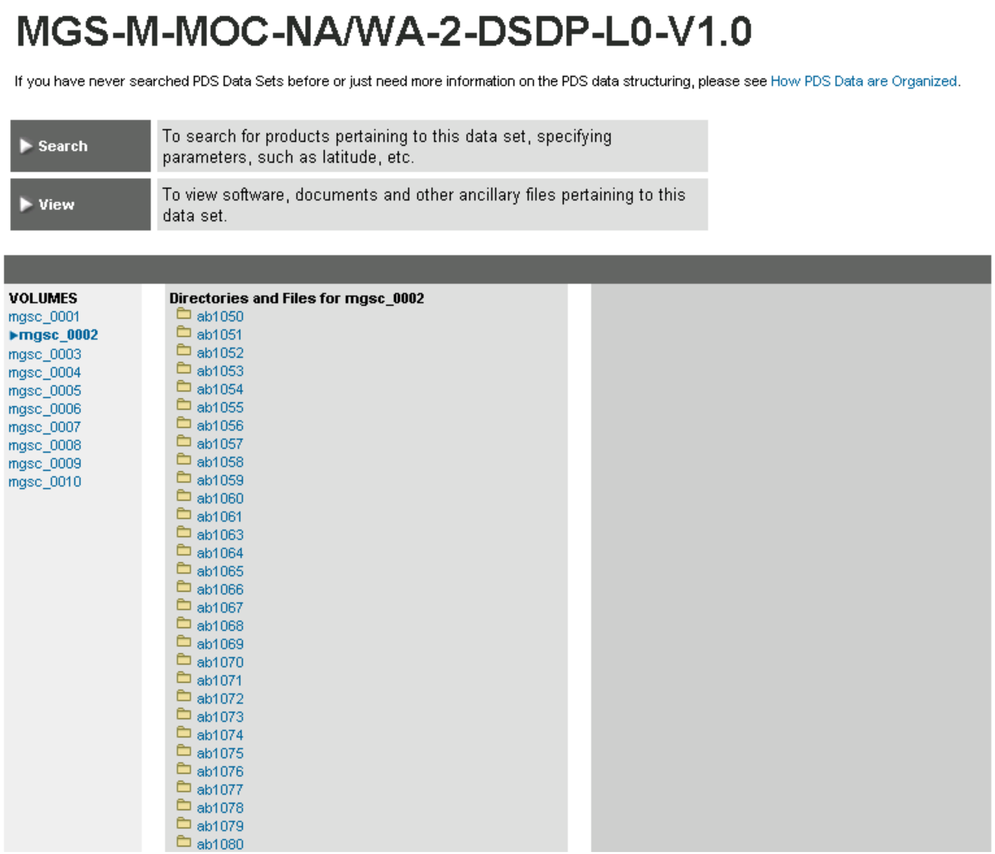


Figure 20. Files in the mgsc_0001 Volume

19. How does the creator maintain its digital entities through technological change?

19a. What preservation strategies and/or methods are implemented and how?

The designers of the PDS adopted self-describing data files as a preservation strategy. Self-describing files describe the file format of attached data as well as the context in which the data was created, e.g., instrument, location, time, project, etc. Self-describing data files eliminate the need for converting scientific data to other formats. Data description labels are interpreted using PDS Label Library Light. The PDS Object Access Library uses this interpretation to access the data, and the accessed data is displayed using NASAView. Self-describing files are a type of abstraction mechanism. Classes such as TABLE, HISTOGRAM OR IMAGE have attributes that enable one to describe specific objects in that class.

The PDS has been operational since 1989 and it has not been necessary to update (convert or migrate) any of the data products to other data formats. When hardware and/or operating system become obsolete, only PDS Label Library Light (L3), Object Access Library (OAL) and the NASAViewer need to be recompiled or translated to the new platform.

Moore et al. describe an archival preservation strategy called Collection-based Persistent Object Preservation (POP) [Moore et al 2000a, 2000b]. POP achieves its power through abstraction mechanisms, one of which is self-describing files. Another abstraction mechanism is storage abstraction. This abstraction is used to define the fundamental operations on storage needed to support manipulation and access to data files. There is a mapping from the storage abstraction to the operations of a particular vendor product. By adding drivers for a new vendor storage product as they are created, it is possible to manage storage indefinitely into the future. When a storage product becomes obsolete, the archival objects in storage can be migrated transparently to a new storage technology. The archival repository continues to operate at the level of the storage abstraction.

San Diego Supercomputer Center is investigating Persistent Object Preservation through abstraction mechanisms provided by data grid technologies. The PDS preservation strategy would not be a Persistent Object Preservation strategy unless it incorporated other abstraction mechanisms. It does so by applying the abstraction mechanisms of data grids to its storage repositories [Hughes et al 2003].

One of the Preservation Technological Requirements of the PDS was a robust media that would survive 50 or more years, yet be affordable to store and use. PDS has experimented with several types of storage techniques. They found the optical CD-ROM and CD-Write Once technologies provide them with reliable storage and access. Data saved on CD-ROM could contain approximately ten digital tapes worth of data and have a shelf life of 50 to 100 years, compared to 15 years for tape. This technique provided large savings over tape usage in controlled storage environments, labor to copy/re-write, and risk of loss. [McMahon 1994]

19b. Are these strategies or methods determined by the type of digital entities (in a technical sense) or by other criteria? If the latter, what criteria?

This strategy is determined in part by having scientific data whose file structure can be described in the Object Description Language. This file structure description can be prefixed to the file or associated with it. An interpreter for this file structure description language can thus apply to many file structures. The result is that technology obsolescence requires only the migration of the interpreter rather than the creation of access software for each of the file structures.

As discussed earlier, this strategy has been applied to arrays, qubes, images, histograms, series, tables, spectrum, SPICE kernels, catalogs, directories, text documents, and others. The question arises as to whether this strategy applies to e-records created by word processing software that include embedded markup for elements of intellectual and physical form. The answer is yes, if the word processing documents are represented in XML with XSL stylesheets. This has not been implemented in the PDS.

F. Cross-domain Research Questions

Three cross-domain teams also have research questions. They are the Description Cross Domain, the Policy Cross Domain and the Modeling (System Analysis) Cross Domain.

Description Cross-domain

The Description cross-domain identified two questions that should be addressed by each case study in order to gather data for the purposes of comparative analysis.

20. What descriptive or other metadata schema or standards are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

The *Planetary Science Data Dictionary* (PSDD) [PDS 2002] is being used in the creation, maintenance, use and preservation of the Planetary Data System. The PSDD contains definitions of the standard data element names and objects.

In the PDS, metadata refer specifically to the resource description. As stated earlier, a resource description is the set of attributes describing a resource. In the PDS, the standard data element names of all attributes must be defined in the Planetary Science Data Dictionary (PSDD).

21. What is the source of these descriptive or other metadata schema or standards (institutional convention, professional body, international standard, individual practice, etc.)?

The Planetary Science Data Dictionary is a NASA institutional standard for Planetary Science Metadata [PDS 2002]. The PDS procedures for assigning standardized names to data elements follow closely the NBS Guide on Data Entity Naming Conventions [NBS].

Policy Cross-domain

The Policy Cross Domain identified two questions that should be addressed by each case study in order to gather data for the purposes of comparative analysis.

22. To what extent do policies, procedures, and standards currently control records creation, maintenance, preservation and use in the context of your activity? Do these policies, procedures, and standards need to be modified or augmented?

The *PDS Standards Reference* [JPL 2003] defines all PDS standards for data preparation. The PDS Standards Reference is used in conjunction with the *PDS Data Preparation Workbook* [JPL 1995a] and the *Planetary Science Data Dictionary* [PDS 2002]. The *Planetary Science Data Dictionary* (PSDD) contains definitions of the standard data element names and objects. The *PDS Data Preparation Workbook* describes the end-to-end process for submitting data to the

PDS and gives instructions for preparing datasets. In addition, a glossary of terms used throughout the documentation is included as an appendix to the Workbook.

23. What legal, moral (e.g., control over artistic expression) or ethical obligations, concerns or issues exist regarding the creation, maintenance, preservation and use of the records in the context of your focus group's activity?

Professional ethics in the Planetary Sciences and Institutional ethics in NASA dictate that Principal Investigators, Project Science Group and the PDS project manager, project scientist, project engineer, mission interface lead, and managers of all the PDS discipline nodes be competent and trustworthy.

Modeling Cross-domain

The InterPARES Chain of Preservation (COP) activity model is a generic model of the process of creating, maintaining, selecting and preserving authentic electronic records. The model is intended to provide a framework for making and carrying out archival decisions.

Walkthroughs using case data are an effective way to test whether a model, design, program code, or user interface achieve what is intended and to improve the quality of the product [Yourdon 1990, Freedman and Weinberg 1990]. A walkthrough is a peer group review of any information system product. A walkthrough of an activity model, such as the chain of preservation model, is concerned with the functionality of the system. To demonstrate that the COP activity model applies to specific cases of electronic records and to refine and validate the model, it is intended to conduct a walkthrough (General Study 12). The method used in the walkthrough is to iteratively step through each of the lowest-level activities in the model:

1. Reviewing the activity definition and the input, output and control definitions.
2. Identifying data elements of labels on input and output arrows.
3. Defining the transformation of inputs to outputs.
4. Determining values of the data elements that are related to the specific body of records in the case study.
5. Recording the results and any problems or issues that arise and suggesting possible solutions.

To conduct a walkthrough of the model, it is necessary to have specific actual objects and information from a case. In this case study, PDS documents have been collected that characterize describe the Planetary Data System. Copies of the Mars Global Surveyor Science Data Management Plan [JPL 1995b] and the Mars Global Surveyor Project Archive Generation, Validation, and Transfer Plan [Arvidson 1999] were also obtained. Examples of data products and indexes, catalogs, and documents in a dataset volume have also been collected. A copy of NASAView has also been obtained.[4]

Because a walkthrough of an activity model provides a concrete example of the application of the model, the walkthrough may also provide archival institutions with an example of how they might apply the model in their own institutions.

[4] One can obtain a copy of NASAView at http://pdsproto.jpl.nasa.gov/Distribution/license.html.

G. References

Planetary Data Systems Homepage http://pds.jpl.nasa.gov/
PDS MGS MOC Image Collection http://ida.wr.usgs.gov/

[Arvidson et al 1999] R. Arvidson, E. Guinness, S. Slavney and R. J. Springer. Mars Global Surveyor Project Archive Generation, Validation, and Transfer Plan. Document # 542-312. February 11, 1999.
http://pds-geosciences.wustl.edu/missions/mgs/docs/mgsarchive.pdf

[Arvidson et al 2002] R. E. Arvidson, S. Slavney and S. Nelson. MARS EXPLORATION PROGRAM DATA MANAGEMENT PLAN. Rev. 3.0. Dept. of Earth and Planetary Sciences Washington University St. Louis, MO March 20, 2002.
http://pds-geosciences.wustl.edu/missions/mep/mepdmp.pdf

[Commission 1995] Commission on Physical Sciences, Mathematics, and Applications Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government: Working Papers, National Academies Press, 1995.
http://books.nap.edu/books/NI000157/html/index.html

[Davis 1990] R. L. Davis. Specification for the Object Description Language, NASA, Jet Propulsion Laboratory, Pasadena, CA, 1990. [Also published by U of Colorado Press, Boulder]

[Davis & Monk 1997] R. Davis and S. Monk. Planetary Data System Object Access Library User's Guide**,** OAL Version 1.2. Laboratory for Atmospheric and Space Physics, University of Colorado, December 1997.
http://pds-rings.arc.nasa.gov/toolkits/oal_131/oalguide.pdf

[Freedman and Weinberg 1990] E. Freedman and G. Weinberg. *Handbook of Walkthroughs, Inspections and Technical Reviews*, 3rd Ed., New York: Dorsett Home Publishing, 1990.

[Hughes 1993] J. S. Hughes and Y. P. Li. "The Planetary Data System Data Model", Proceedings of Twelfth IEEE Symposium on Mass Storage Systems, April 25-29, 1993.

[Hughes 1998] S. Hughes, D. Bernath, S. Monk. Planetary Data System Label Library Light (L3) - Version 1.1, User's Guide, February 3, 1998.

[Hughes et al 2003] J. S. Hughes, D. J. Crichton, S. C. Kelly, S. K. LaVoie and R.J. Walker, A data grid framework for sharing data across distributed heterogeneous repositories, 2003, presented at American Geophysical Union (Fall Meeting), San Francisco, California.

[JPL] PDS Guidelines for Project Data Management Plans, JPL Document D-5111. Jet Propulsion Laboratory, Pasadena.

[JPL 1995a] Planetary Data System Data Preparation Workbook, Version 3.1, JPL D-7669, Part 1, Feb. 1995, Jet Propulsion Laboratory, Pasadena.

[JPL 1995b] Mars Global Surveyor Science Data Management Plan. JPL D-12529, Jet Propulsion Laboratory, California Institute of Technology, February 1995. http://www.msss.com/mars_images/moc/mgssdmp.html#RTFToC26

[JPL 1995c] Mars Global Surveyor: Investigation Description and Science Requirements Document, 542-300 Jet Propulsion Laboratory, Pasadena, February 1995 http://mars.jpl.nasa.gov/mgs/pdf/300.PDF

[JPL 2003] Planetary Data System Standards Reference, Version 3.6, JPL D-7669, Part 2, Aug. 2003, Jet Propulsion Laboratory, Pasadena. http://pds.jpl.nasa.gov/documents/sr/

[King 1998] J. H. King White paper on NASA science data retention. NSSDC, NASA Goddard Space Flight Center, June 1998. http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html

[McMahon 1996] S. K. McMahon. Overview of the Planetary Data System, *Planetary Space Science,* v. 44, #1, 1996, pp. 3-12.

[Moore et al 2000a] R. Moore, C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, W. Schroeder, and A. Gupta. Collection-Based Persistent Digital Archives-Part 1, *D-LIB Magazine*, March 2000.

[Moore et al 2000b] R Moore, C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, W. Schroeder, and A. Gupta. Collection-Based Persistent Digital Archives - Part 2, *D-LIB Magazine*, April 2000.

[MOU 2006] Memorandum of Understanding between the Planetary Data System and the National Space Science Data Center, November 29, 2006.

[NAIF 2003] Navigation and Ancillary Information Facility. SPICE Overview, August 2003. http://centauri.larc.nasa.gov/newfrontiers/SPICE_Overview.pdf

[NASA] NASA Procedures and Guidelines (NPG) 1000.3.

[NBS] Guide on Data Entity Naming Conventions, NBS Special Publication 500-149.

[PDS 2002] Planetary Data System Project and Advanced Multimission Operations System . Planetary Science Data Dictionary Document. JPL D-7116, Rev. E, Jet Propulsion Laboratory, California Institute of Technology, August 28, 2002. http://pds.jpl.nasa.gov/documents/psdd/psdd.pdf

[Pearce-Moses] R. Pearce-Moses. A Glossary of Archival and Records Terminology. http://www.archivists.org/glossary/index.asp

[Shirley 1996] D. Shirley. "Mars on $300 A Day:" The Mars Exploration Program. (second of two parts) The Martian Chronicle, Issue 7, Nov 1996. http://mpfwww.jpl.nasa.gov/MPF/martianchronicle/martianchron7/

[Yourdon 1990] E. Yourdon. *Structured Walkthroughs*, 4th Ed., Englewood Cliffs, NJ: Yourdon Press, 1989.

H. Glossary

active archive
An archive of space science datasets that is in active use by the space scientists.

archive
An archive consists of one or more datasets 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
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 provider
The organization responsible for delivering mission archives to the PDS (see mission archive).

dataset
A collection of related data products. The data products typically (but not always) come from a series of observations from an instrument and have been processed in the same manner

deep archive
The collection of PDS archive volumes held by the National Space Science Data Center (NSSDC).

***experimental data record* (EDR)**
NASA Level 0 data that consist of time-ordered sequences of raw science data obtained by a given instrument, together with engineering information that allows instrument teams to check their instruments.

mirroring
Setting up a duplicate repository on the network, in case the original repository fails or becomes overloaded.

mission archive
A dataset, along with all its ancillary files, organized in a PDS-compliant volume structure.

namespace
A set of names that are all unique relative to each other. Typically, there is a single "controlling authority" that controls the names in the namespace. For example, a mission or discipline node may control all the names in a local data dictionary

packetized data record
These data records consist of time-ordered packetized telemetry received from the spacecraft, with duplications removed, together with ancillary information needed to understand what is contained in a given packet.

***persistent object preservation* (POP)**
a technique to ensure electronic records remain accessible by making them self-describing in a way that is independent of specific hardware and software. [Pearce-Moses]

***reduced data record* (RDR)**
Science data that have been processed from raw data to NASA Level 1 or higher.

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. Standard Data Products include both EDRs and Reduced Data Records (RDRs). RDRs are typically generated from EDRs and SPICE data at investigator facilities. Standard Data Products are produced primarily by Instrument Teams, although some may be generated by Interdisciplinary Scientists addressing specific scientific problems.

Acronyms and Abbreviations

EDR – Experimental Data Record
MGS – Mars Global Surveyor
NAIF – Navigation and Ancillary Information Facility
PDB – Project Data Base
SDVT – Science Data Validation Team
SIS – Software Interface Specification
SOPC – Science Operations Planning Computer
SPICE – Spacecraft, Planet, Instrument, C-matrix, Events
UDR – Unprocessed Data Record