InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **General Study 12 Final Report: Validation of the InterPARES 2 Project Chain of Preservation Model Using Case Study Data**

Status:	Final (public)
Version:	1.0
Submission Date:	July 2007
Release Date:	February 2008
Author:	The InterPARES 2 Project
Writer(s):	William Underwood, Georgia Tech Research Institute
	Kevin Glick, Yale University
	Mark Wolfe, University at Albany, State University of New York
Project Unit:	Focus 2
URL:	http://www.interpares.org/display_file.cfm?doc= ip2_gs12_final_report.pdf

Table of Contents

List of Figures

ABSTRACT

The InterPARES 2 Project Chain of Preservation (COP) model is a generic model of the process of creating, maintaining, selecting and preserving authentic digital records. It is an integration of the UBC-MAS Project's Model of Electronic Recordkeeping[1] and the InterPARES 1 Project's models of Selection (Appraisal) and Preservation of Electronic Records.[2]

The COP model was developed via functional decomposition. It was not generalized from specific archival objects and information used to manage those objects. The COP model is prescriptive as well as descriptive. It prescribes criteria for determining whether digital records can be presumed to be authentic and a method for applying these criteria. The question arises: How can we ensure that the model actually applies to the creation, maintenance, selection and preservation of actual e-records?

Walkthroughs using case study 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. A walkthrough is a peer group review of any information system product. A walkthrough of a model, such as the COP model, is concerned with the functionality of the system. The purpose of this paper is to demonstrate that the COP model applies to specific record-making, recordkeeping and preservation systems for digital records and to refine and validate the model by conducting a walkthrough of the model using case study data.

Case study 08 (Mars Global Surveyor Data Records in the Planetary Data System) is a case study of the InterPARES 2 Project's Focus 2 (Science) Task Force. It was designed to collect information about the Planetary Data System (PDS) Space Science Data Archive 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 (MGS) mission. It was also designed to collect information to perform a walkthrough of the InterPARES 2 COP model.

NASA refers to the Planetary Data System as an active archive. Copies of the scientific datasets are transferred to the National Space Science Data Center (NSSDC) for long-term preservation. The NSSDC is referred to as a deep archives. The PDS activities of data preparation and management of datasets in the PDS Archive are similar to the activities *Managing Records Creation* and *Manage Records in a Recordkeeping System*. The management at the NSSDC of scientific datasets from the PDS and from other space science disciplines appears to be similar to the activity *Select and Preserve Records*. However, the Planetary Data System case study and the walkthrough using this case study data revealed that selection (appraisal) and preservation were central aspects of the PDS design and operation.

One of the novel aspects of the Planetary Data System is the choice and implementation of a preservation strategy that obviates the need to convert to data products to other file formats. This strategy is to prefix the scientific data records (and supplementary documentation) with labels

[1] See http://www.interpares.org/UBCProject/index.htm.
[2] See http://www.interpares.org/book/index.cfm, appendices 4 and 5.

that describe the context in which the scientific data was collected as well as a description of the format of the data. This label is described in a language called the Object Description Language. There are software tools to interpret the labels, including the description of the structure of the data, and other tools that use this description to read and/or display the data.

An interesting aspect of the walkthrough using MGS/PDS data is that the Description activity seems to take place during record-making rather than as in the COP model after transfer to the organization responsible for long-term preservation. Another interesting aspect of this case study is that parts of appraisal and validation activities take place in the PDS before recordkeeping activities; while in the COP model they take place after recordkeeping. A possible explanation for this is the PDS management decision to actively involve scientists in the archiving process. The scientists who create and use the data products are better able to describe and appraise them than archivists (or scientists) far removed from the mission and data creation. Because of the expense of space science missions, the investment dictates early description, appraisal and validation of the data sets.

Of the sixty-eight lowest-level activities in the COP model, data from the PDS and MGS case study was found to correspond to forty-six of those activities. There is no corresponding data in the PDS case study for seven of the COP activities. No data were collected for fifteen of the NSSDC activities that would correspond to the long-term preservation activities of the COP model. The walkthrough of the COP model using PDS and MGS case study data demonstrates that there is an interpretation of the record-making, recordkeeping and some of the preservation activities of the COP model in the domain of scientific data archives. That is to say, the COP model is satisfiable in this domain. A more thorough validation of the COP model would require walkthrough of the COP preservation activities using case study data with regard to the NSSDC preservation activities and walkthroughs using case study data from other archival domains.

Information was not found for PDS activities corresponding to COP activities for classifying or registering incoming or outgoing scientific data records or other supporting documents. This is because the PDS does not keep records for all mission activities, but only scientific data sets. This does not invalidate the COP model, but just emphasizes that it is a more general model of recordkeeping activities than the activities of scientific data recordkeeping.

Information was not found in the case study corresponding to the three activities in the decomposition of COP activity A4.2.2.2, *Assess Authenticity of Records*. This is not surprising as the criteria and method of assessment of the authenticity of digital records was a new research contribution of InterPARES 1. However, it is demonstrated in InterPARES 2 general study 06 (Validation of the Benchmark Requirements for Presuming the Authenticity of Electronic Records) that the assessment activity can be carried out with data from the MGS/PDS case study resulting in a high degree of belief in a presumption of authenticity of the records maintained in the PDS.

1. Introduction

1.1 Background

The InterPARES 2 Chain of Preservation (COP) model is a generic model of the process of creating, maintaining, selecting and preserving authentic digital records [InterPARES2 2005]. The set of data elements that are created, maintained and used by the activities of the COP model has been created [InterPARES2 2004].[3] The COP model and data elements are intended to provide a framework for making and carrying out archival decisions regarding the preservation of authentic digital records.

Case study 08 (Scientific Data Records from a NASA Spacecraft Mission) is a case study of the InterPARES 2 Project's Focus 2 (Science) Task Force [Underwood 2005]. It was designed to collect information about the Planetary Data System (PDS) Planetary Science Data Archive in order 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. Further, it was designed to collect information that could be used to address research issues in the InterPARES 2 Description and Policy Cross-domains. Finally, it was also designed to collect information to perform a walkthrough of the COP model of records creation, maintenance, appraisal and preservation of digital records.

The National Space Science Data Center (NSSDC) at Goddard Space Flight Center is responsible for top-level data management functions that span all NASA Space Science programs. It is also NASA's deep archives for scientific data from Space Science Missions. NASA has established a number of Discipline Data Center (DDC) active archives to augment the NSSDC by facilitating data access in those disciplines. The Planetary Data System is one of these. As an active archive within the NASA archive environment, the PDS has primary responsibility for the collection of lunar and planetary data, the definition of its content, its validation and catalog management [MOU 2006].

1.2 Purpose

Because the COP model was developed via functional decomposition rather than generalized from specific archival objects and information used to manage those objects, the following question then arises: How can a user be sure that the model actually applies to the creation, maintenance, selection and preservation of digital records?

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 1989, Freedman and Weinberg 1990]. A walkthrough is a peer group review of any information system product. A walkthrough of a model, such as the COP model, is concerned with the functionality of the system. Walkthroughs can also be used to determine whether an model or design meets functional requirements. The purpose of this paper is to demonstrate that

[3] Note: The version of the COP model used in this report was an early draft version from 2005.

the COP model applies to specific record-making, recordkeeping and preservation systems for digital records and to refine and validate the model by conducting a walkthrough of the model using case study data.

1.3 Scope

In section 2, the walkthrough team and method are described. In section 3, the results of the walkthrough are discussed. In section 4, conclusions drawn from the walkthrough are presented.

2. The Approach

2.1 Walkthrough Team

The walkthrough team consists of a *presenter*, who "puts on the table" the model being reviewed; *reviewers*, who have a good understanding of the model, ask questions of the case study expert to identify data corresponding to inputs and outputs of the activities, raise issues and suggest solutions to problems; a *case study expert*, who answers questions posed by the reviewers about data from the case study; and a *secretary*, who records the discussed facts and issues and distributes the minutes.

2.2 Walkthrough Method

The method used in the walkthrough is to iteratively step through each of the lowest-level activities in the model:

1. Reviewing the activity definitions and the input, output and control definitions.
2. Identifying activities in the case study that correspond to activities of the model.
3. Identifying objects from the case study that correspond to labels on controls, inputs and outputs of the activities.
4. Recording the results and any problems or issues that arise and suggesting possible solutions.

3. Walkthrough of the COP Model

The Walkthrough Team began by creating a list of all the lowest-level COP activities, their definitions, and their controls, inputs and outputs. There are sixty-eight lowest-level activities in the COP model. This document was used to record the instantiations of the COP model with data from the Planetary Data System Case Study as we walked through the COP Activity diagrams. The paragraphs below are the contents of that document.

A.1.1.1 Analyze the Records Creator

Definition: To assess the information concerning the records creator 's mission, organizational structure, activities, and existing technological, financial and human resources, and records related needs and risks to help identify the requirements for the chain of preservation framework.
Input: Information about Records Creator
Output: Analysis of Records Creator

Discussion: During the case study, mission statements, and organization charts were collected for NASA Headquarters, the Sciences and Exploration Directorate at Goddard Space Flight Center, the Space and Earth Sciences Directorate at the Jet Propulsion Laboratory, and the Mars Global Surveyor Mission/Project. Information was also collected on the juridical-administrative context, the documentary context, procedural context, and technological context of the Planetary Data System and Mars Global Surveyor science data records. All of these are inputs to the COP activity *Analyze Records Creator*.

There have been several studies of NASA as a Scientific Data Records Creator. These include studies by the National Academy of Science's Committee on Data Management and Computation [Arvidson 1986], The Government Accounting Office [GAO 1990], and the National Academy of Science's Commission on Physical Sciences, Mathematics and Applications [Commission 1995].

A1.1.2 Analyze Creator's Existing Records

Definition: To assess creator's existing records and information about those records to determine framework requirements.
Input: Information about Records Creator, Information about Creator's Existing Records, Creators Existing Records
Output: Analysis of Records

Discussion: "In 1982, the National Academy of Sciences chartered the Committee on Data Management and Computation (CODMAC), which identified serious problems in the way data was managed by NASA. Historically, much planetary data was not delivered to any archive facility. Frequently, data that was stored was difficult to locate or use because the documentation was inadequate for scientists outside the original investigation teams. In addition, in the years since early planetary missions, their tapes containing data were becoming physically unreadable." [McMahon 1996]

"CODMAC I [Bernstein, 1982] offered these principles to result in more scientific return from the data: 1) scientific involvement; 2) scientific oversight; 3) data availability including usable formats, ancillary data, timely distribution, validated data, and proposer documentation; 4) proper facilities; 5) structured, transportable, adequately documented software; 6) data storage in permanent and retrievable form; and 7) adequate data system funding." [McMahon 1996]

A1.1.3 Establish Management Policies

Definition: To develop management regime policies for establishing overall framework design requirements.
Input: Analysis of Records Creator, Analysis of Records
Output: Management Policies

Discussion: "Under CODMAC III [RUSSELL, 1988], the committee reviewed the NASA progress in addressing the issues previously defined. Its report included these recommendations: implement the principles of CODMAC I; write explicit data management plans for all missions; provide sufficient resources for data archiving; enforce proper archiving requirements on projects and principle investigators; build a secure archive; manage scientific data management units by discipline, with NSSDC as the deep archive; establish catalogs and directories; assess storage media and develop guidelines for their use; establish data advisory committees on data retention and preservation; promote and support use of data archives."

"In response to the CODMAC request, the NASA Solar System Exploration Division established the Planetary Data System as an active archive. PDS became operational in 1989. It currently satisfies all the CODMAC recommendations for those tasks within its responsibility. Data management plans are the responsibility of each flight project, although PDS provides help in the definition of archive products." [McMahon 1996]

A1.1.4 Establish Design Requirements

Definition: To identify the rules guiding the chain of preservation framework on the basis of the analysis of the records creator and its existing records.
Input: Analysis of Records Creator, Analysis of Records, Management Policies
Output: Design Requirements, Framework Policies

Discussion: In 1983, a workshop was held at Goddard Space Flight Center in which Planetary Scientists outlined what was to become the Planetary Data System [Kieffer 1984]. 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.

A1.2.1 Design Record-making System

This is not a lowest-level activity in the COP model, but is included here because the outputs of its lower-level activities are merged into a single output of this activity.

Output: Record-making System Design (Record-making Access Privileges, Record-making Metadata Schemes, Integrated Business and Documentary Procedures, Record-making Technological Requirements, Records Forms)

A1.2.1.1 Develop Records Forms and Record-making Metadata Schemes

Definition: To specify the documentary forms of records and, for each type of record, the metadata elements or attributes that need to be explicitly expressed and linked to every record.
Input: Design Requirements
Output: Record-making Metadata Schema, Records Forms

Discussion: The PDS Data Model [Hughes 1998a] and PDS Data Dictionary [PDS 2002] are the Record-making Metadata Schemes. The primary classes in the Planetary Science Data Model are Mission, Spacecraft, Instrument and Target. Missions are associated with Spacecraft, and Spacecraft with Instruments and Targets. Datasets (which are aggregations of data records, documents and software) are associated with instruments and targets.

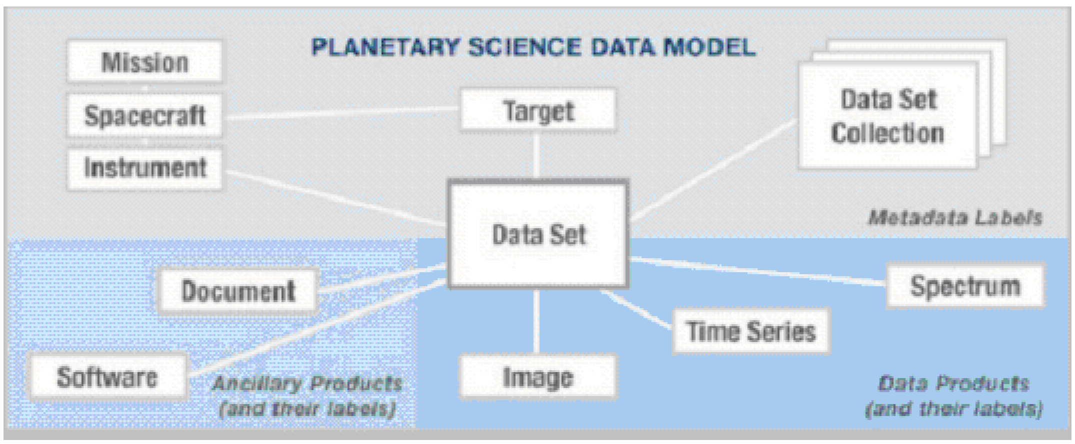


Figure 1. Planetary Science Data Model

The kinds of digital objects created are described below. In 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 a 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. It

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)."

A1.2.1.2 Establish Access Privileges for Record-making[4]

Definition: To define record-making access privileges for each user of the record-making system.
Input: Design Requirements, Records Forms
Output: Record-making Access Privileges

Discussion: Datasets are created on Space Operations Planning Computers (SOPCs) at the Principal Investigator's site and at seven Discipline Nodes. Security measures, which include access control, are specified in an "Operations Facility and Control Plan."

A1.2.1.3 Design Integrated Business and Documentary Procedures

Definition: To develop procedures for carrying out business, linked to a plan for organization of the creator's records.
Input: Design Requirements, Record-making Access Privileges
Output: Integrated Business and Documentary Procedures

Discussion: 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 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 2 shows the process of MGS archive generation, validation and transfer to the PDS [Arvidson 1999, p. 8].

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.

[4] Perhaps a better name for the activity A1.2.1.2 would be *Design Access Control System for Record-making* as the access privileges for individual users would be assigned after the Record-making System had been implemented.

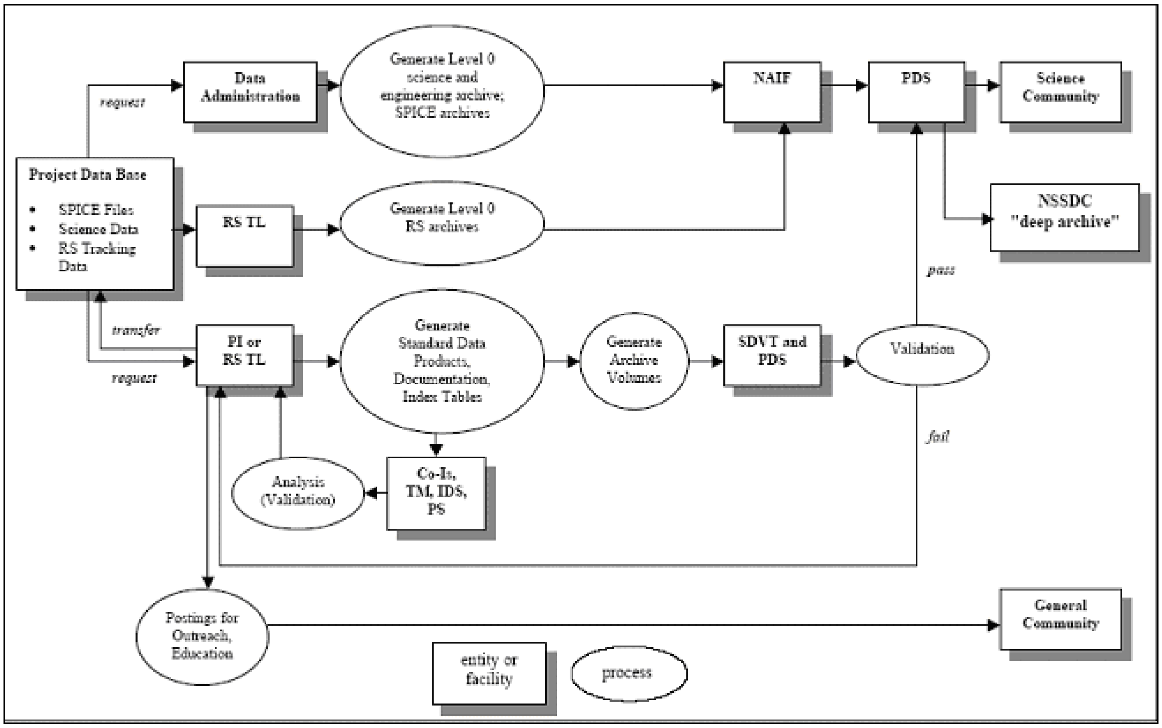


Figure 2. Dataset Generation, Validation and Transfer

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].

A1.2.1.4 Determine Record-making Technological Requirements

Definition: To specify the hardware and software needed for the record-making system.
Controls: Record-making Metadata Schema, Records Forms, Record-making Access Privileges, Integrated Business and Documentary Procedures
Input: Information Concerning Available Technology
Output: Record-making Technological Requirements

Discussion: Among the record-making technical requirements were that the data product and volume creation tools should operate on Solaris, Linux and DOS/Windows operating systems and should be written in the C programming language.

A1.2.2 Design Recordkeeping System

This is not a lowest-level activity in the COP model, but is included here to record the fact that the outputs of it lower-level activities are merged into a single output of this activity *Design Recordkeeping System*.

Output: Recordkeeping System Design (Recordkeeping Metadata Schemes, Registration Scheme, Classification Scheme, Retention Schedule, Procedures for Maintaining Authentic Records, Recordkeeping Access Privileges, Recordkeeping Retrieval System, Recordkeeping Technological Requirements)[5]

A1.2.2.1 Develop Recordkeeping Schemes

Definition: To establish the metadata, registration and classification schemes used in the recordkeeping system.
Controls: Record Forms, Record-making Metadata Schemes, Integrated Business and Documentary Procedures
Input: Design Requirements
Output: Recordkeeping Metadata Schemes, Registration Scheme, Classification Scheme[6]

Discussion: The Record-making Metadata Schema is for records. The Recordkeeping Metadata Schemes are for collections, datasets, and volume directories.

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 3, 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.

Data products are registered by entering a unique value for PRODUCT-ID in the label of a data product, which is also the filename of the data product within a volume. The filename is also entered in an index table in the index directory of the volume.

Registration of supplementary documents consists of assigning the filename to the keyword FILE_NAME in the PDS label of the document, placing the document in the DOCUMENT directory of a volume and including the filename in the Index file of the volume.

According to the COP model, a *Classification Scheme* is "a plan for the systematic identification and arrangement of business activities and records into categories according to logically structured conventions, methods and procedural rules." The primary classes in the Planetary Science Data Model are Mission, Spacecraft, Instrument and Target. Missions are associated with Spacecraft, and Spacecraft with Instruments and Targets. Datasets (which are aggregations of data records, documents and software) are associated with instruments and targets. Hence, this part of the data model makes up the classification scheme.

[5] In the IDEF0 Activity Diagram A1.2, *Recordkeeping Retrieval Requirements* occurs twice as an output of activity A1.2.2, *Recordkeeping Access Privileges* does not occur.
[6] In the version of the COP Model used in the walkthrough, the output *Classification Scheme* of activity A1.2.2.1 was mislabeled *Retention Schedule.*

```
DIRECTORY/FILE          CONTENTS
-------------------     ------------------------------------------
<root>
|-INDEX.HTM             Starting point for web-browsing CD contents.
|-AAREADME.TXT          The file you are reading (ASCII Text).
|-ERRATA.TXT            Description of known anomalies and errors
|                       present on the volume set (optional file).
|-VOLDESC.CAT           A description of the contents of this volume in
|                       in format readable by both humans and computers.
|-<CATALOG>             Catalog Directory
|  |-CATINFO.TXT        Describes Contents of the Catalog directory
|  |-DATASET.CAT        Dataset description.
|  |-DSMAP.CAT          Map Projection description.
|  |-INSTHOST.CAT       Spacecraft description.
|  |-MISSION.CAT        Mission description.
|  |-PERSON.CAT         Contributors to this dataset.
|  |-REFS.CAT           References
|  |-MOCINST.CAT        MOC instrument description.
|
|-<DOCUMENT>            Documentation Directory.  The files in this
|  |                    directory provide detailed information
|  |                    regarding the MOC DSDP archive.
|  |-DOCINFO.TXT        Description of files in the DOCUMENT directory
|  |-VOLINFO.TXT        Documentation regarding the contents of this
|  |                    CD Volume Set.
|  |-VOLINFO.LBL        PDS Label file describing the VOLINFO documents
|  |
|  |-MOCSIS.TXT         Mars Observing Camera Software Interface
|                       Specification document.
|-<INDEX>               Directory for the image index files.
|  |-INDXINFO.TXT       Description of files in <INDEX> directory.
|  |-IMGINDX.TAB        Image Index table.
|  |-CUMINDX.TAB        Cumulative Image Index table for all volumes.
|  |-IMGINDX.LBL        PDS label for IMGINDX.TAB.
|
|-<data directories>    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.
|-<BROWSE>              The files in this directory provide HTML
|                       browsing of the files contained on this volume.
   |-MAP.HTM            Map-based browser page.
   |-TAB.HTM            Table-based browser page.
```

Figure 3. The Directory and File Structure of a Volume

A1.2.2.2 Develop Retention Schedule

Definition: To determine and record the disposition of each series and/or class of records.
Controls: Integrated Business and Documentary Procedures,
Input: Design Requirements, Appraisal Decisions, Classification Scheme
Output: Retention Schedule

Discussion: King [1998] discusses the relationship between NASA's long-term preservation of scientific data for future scientific use and the National Archives and Records Administration's (NARA's) preservation for historical purposes of "seminal observations having extraordinary impact on the evolution of Space and Earth Sciences."

NASA's Records Retention Schedule [NASA 2003] specifies the kinds of observational data from space flights that are to be retained permanently by NARA.

> Schedule item 8/101
> If the records pertain to programs/projects relating to both manned and unmanned space flight, aerospace technology research, and basic or applied scientific research AND meeting one or more of the following criteria: are "first of a kind," establish precedents, produce major contributions to scientific or engineering knowledge, integrate proven technology into new products, or are/have been subject of widespread media attention or Congressional scrutiny.
>
> and consist of records essential for understanding the history of a program/project from inception to completion defined by the stages in program/project's life. Note 1 contains a list of eight stages and potential records that might be created in each, which are held at office of record, then the records are **permanent**. Cut off records at close of program/ project or in 3-year blocks for long term programs/projects. Transfer to records center storage. Transfer to National Archives 7 years after cutoff. Special media records will be transferred in accordance with 36 CFR § 1228.270 (digital records), 36 CFR § 1228.266 (audiovisual records), 36 CFR § 1228.268 (cartographic and architectural records), and/or current transfer instructions specific to individual formats. < N1-255-04-3>

Observational data is one of the phases listed in Note 1.

Observational data. Records containing physical sciences observational data created by both space- and earth-based platforms that are unique and cannot be extrapolated from other datasets or observations at a reasonable cost, including:
* Copies of definitive datasets
* Instrument operations
* Copies of processed data and metadata, its analysis, and proposals for analysis, and related finding aids
* Laboratory notebooks and logbooks
* Proceedings
* Studies and reports

A1.2.2.3 Establish Procedures for Maintaining Authentic Records

Definition: To develop procedures to ensure that records maintain their integrity as they are managed in the recordkeeping system.
Controls: Record-making Access Privileges
Input: Design Requirements
Output: Procedures for Maintaining Authentic Records, Recordkeeping Access Privileges[7]

Discussion: Project team members, PDS managers and engineers and 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.

Access to the 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 access authentication. The PDS logs accesses to restricted areas of the system. User ID, date, time and operations are logged. There is a checksum on the data objects to ensure data integrity. The PDS Central Node provides system backup.

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, that there are PDS procedures for maintaining authentic scientific data records.

A1.2.2.4 Design Recordkeeping Retrieval System

Definition: To develop a system to enable the retrieval of records or information about records upon request.
Controls: Recordkeeping Access Privileges
Input: Design Requirements
Output: Recordkeeping Retrieval System[8]

Discussion: Figure 4 shows the PDS Architecture. The PDS is a distributed active archive (recordkeeping system). The seven small, darkened circles are the Nodes of the PDS. Five of these nodes are called Discipline nodes. They correspond to the scientific disciplines that make up planetary science. They are:

- Geology and geophysics (Geo)
- Atmospheres (Atm)
- Plasma physics (PPI)
- Rings Rings)
- Comets and asteroids (SBN)

[7] Perhaps a better name for the output *Recordkeeping Access Privileges* would be *Recordkeeping Access Control Procedure* since access privileges are not assigned to users until after the recordkeeping system is implemented.
[8] The walkthrough team suggests that the name of the output *Recordkeeping Retrieval System* from activity A.1.2.2.4, Design Record Retrieval System, be changed to *Recordkeeping Retrieval System Design.*

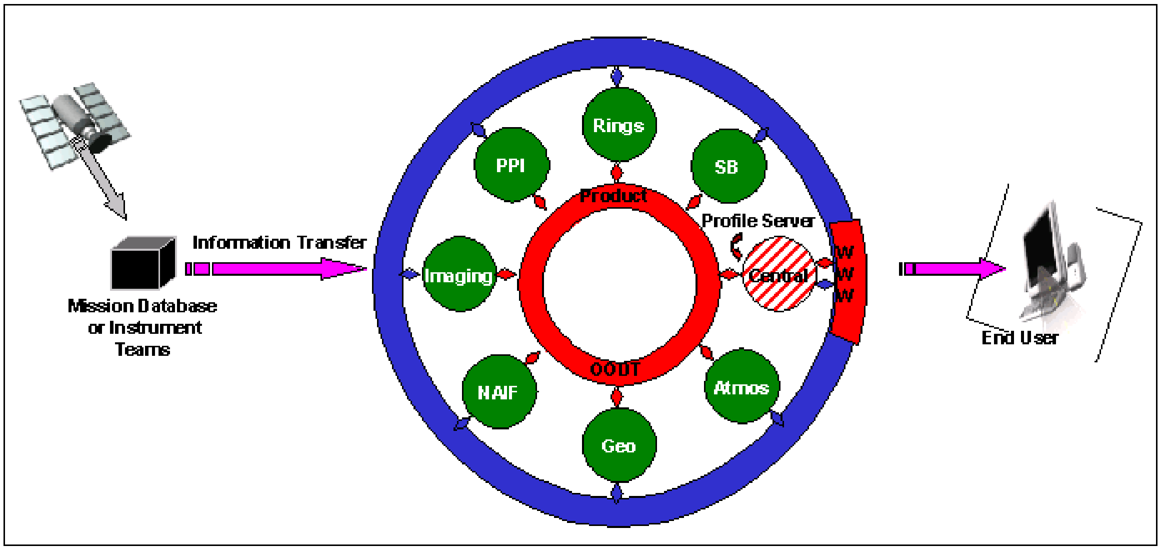


Figure 4. PDS Architecture

There are two support nodes that require expertise across disciplines.

- Imaging (Imaging)
- Observation geometry and events (NAIF)

Data volumes are stored at these nodes and can be accessed from these nodes. The Central Node, shown as a small, hashed circle, provides management of the PDS, validates data products for standards compliance, maintains and documents PDS standard, maintains the planetary science data catalog, and supports education and outreach activities.

The outer circle represents an HTML Web interface to the Nodes. The PDS supports online data distribution. Users can search the online catalog at the Central Node and order datasets. Data requests are directed to the Discipline Nodes where the orders are filled. The discipline nodes distribute data on CD-ROM and DVD-ROM. Users can also search science node data inventories and place orders directly with the appropriate node. Most nodes support online data delivery. Individual files can be ordered from most nodes.

There are two ways to find and access data products. As shown in Figure 5, the first method, PDS-D, provides an HTML Web interface to the catalog at the central node. Queries to the catalog are encapsulated in XML. Queries to the inventory of data products at the nodes are also encapsulated in XML as are the transfer of data products.

The second method of access is illustrated in Figure 6. An HTML interface is provided to the Distributed Inventory Tracking and Data Ordering System (DITDOS).

DITMOS enables the user to browse the file system and augments listings with metadata extracted from individual PDS label files.

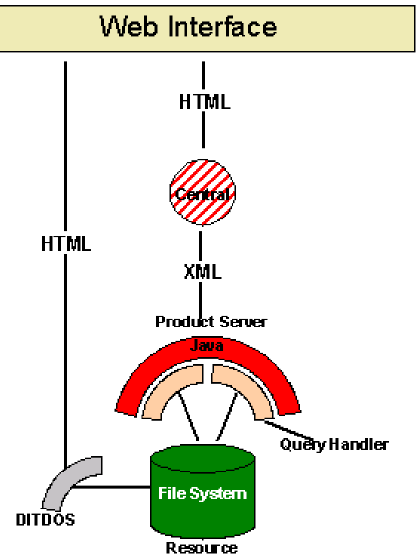


Figure 5. PDS Data Product Access Methods

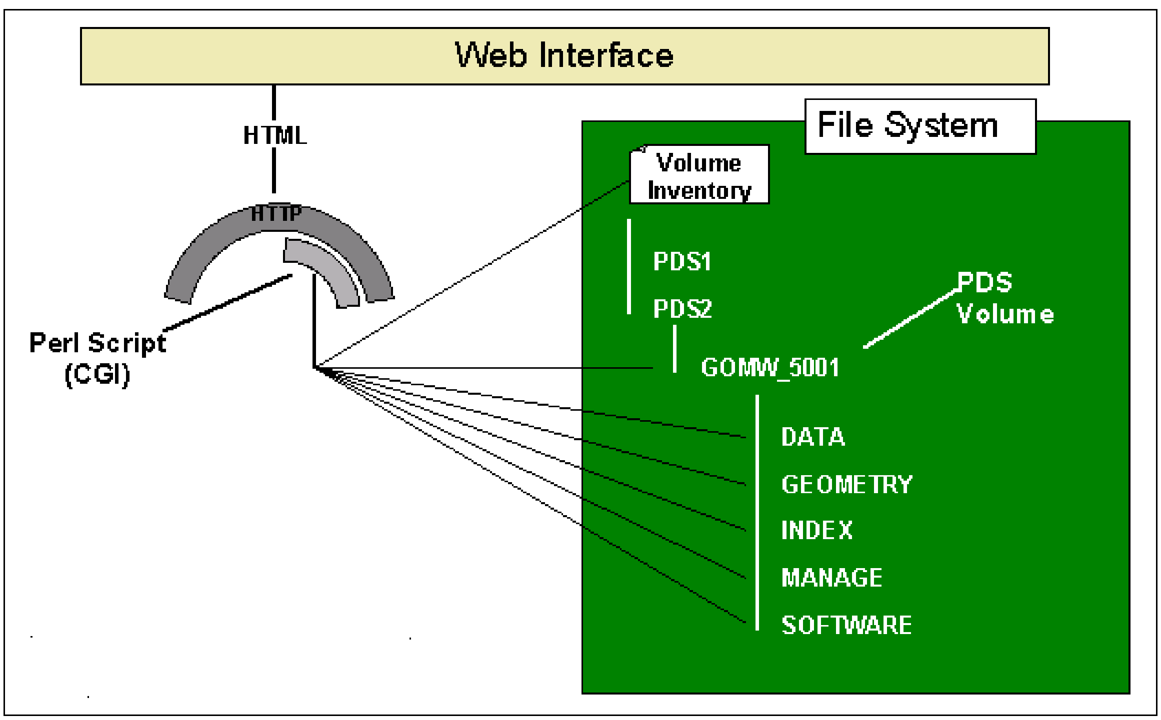


Figure 6. Web Interface to DITMOS

A1.2.2.5 Determine Recordkeeping Technological Requirements

Definition: To specify the hardware and software needed for the recordkeeping system.
Controls: Design Requirements, Record-making Technological Requirements, Recordkeeping Metadata Schemes, Registration Scheme, Classification Scheme, Retention Schedule, Procedures for Maintaining Authentic Records, Recordkeeping Access Privileges, Recordkeeping Retrieval System
Input: Information Concerning Available Technology
Output: Recordkeeping Technological Requirements

Discussion: The Central Node and Discipline Nodes need Web servers. The Central Node provides the middleware for communication among Nodes. The operating system should be Linux and Solaris.

A1.2.3 Design Permanent Preservation System

This is not a lowest-level activity in the COP model, but is included here to record the facts that the outputs of it lower-level activity are merged into a single output of this activity.

Output: Permanent Preservation System Design (Preservation System Design, Selection System Design, Descriptive System Design, Preservation Retrieval System Design)

A1.2.3.1 Design Selection System

Definition: To develop appraisal strategies, monitoring procedures, and disposition rules and procedures, and establish the technological requirements for the design of a selection system.
Input: Information Concerning Available Technology, Design Requirements
Output: Selection System Design (Appraisal Strategies, Monitoring Procedures, Disposition Rules and Procedures)

Discussion: All data incorporated into the PDS archives must undergo a peer review. [JPL 2003]. The purpose of the peer review is to determine that:

- The data is accurate, complete and reliable.
- The data are suitable for archiving.
- The PDS standards have been followed.

A1.2.3.2 Design Descriptive System

Definition: To develop descriptive rules and procedures, descriptive strategies, and to establish descriptive technological requirements for the design of a descriptive system.
Input: Information Concerning Available Technology, Design Requirements
Output: Descriptive System Design (Descriptive Rules and Procedures, Descriptive Strategies, Descriptive Technological Requirements)

Discussion: Catalog information is "Descriptive information about a dataset expressed in the Object Description Language (ODL) and suitable for loading into a catalog." The DATA_SET catalog object is used to submit information about a dataset in the PDS. The procedure for creating this object is contained in the PDS Standards Reference. [JPL 2003, Appendix B]. The Catalog subdirectory of a volume contains the Catalog object files (for mission, instrument, data, etc.). There is one required file in this subdirectory, CATINFO.TXT, which identifies and describes each file in the Catalog subdirectory.

A1.2.3.3 Design Preservation System

Definition: To develop preservation rules and procedures, preservation strategies, and to establish preservation technological requirements for the design of a preservation system.
Input: Information Concerning Available Technology, Design Requirements, Information about Record Creator's Technology
Output: Preservation System Design (Preservation Technological Requirements, Preservation Strategies, Preservation Rules and Procedures)

Discussion: During the 1980s and 90s, NASA's Office of Standards and Technology (NOST) developed a technology known as self-describing data files. 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, and thus are a type of preservation strategy. The designers of the PDS adopted self-describing data files as a preservation strategy. 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. There are methods for creating, reading and displaying objects in these classes.

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 fifty 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 fifteen 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]

While the PDS is a preservation system as well as a recordkeeping system, it is not the NSSDC deep archive. The NSSDC has a preservation system for other space science data sets than those transferred from the PDS active archive, but no case study information was collected on that design.

A1.2.3.4 Design Preservation Retrieval System

Definition: To develop retrieval rules and procedures, retrieval strategies, and to establish retrieval technological requirements and access privileges for the design of a preservation retrieval system to enable retrieval of records upon request.
Input: Information Concerning Available Technology, Design Requirements
Output: Preservation Retrieval System Requirements (Preservation Retrieval Technological Requirements, Preservation Retrieval Strategies, Preservation Retrieval Rules and Procedures, Permanent Preservation Access Privileges)

Discussion: No case study information was collected on the Design of the NSSDC Preservation Retrieval System. However, examples of the NSSDC Preservation Retrieval System are shown in activity A4.5.1, *Manage Retrieval Requests*.

A1.3 Maintain Framework[9]

Definition: To assess information about the performance of the record-making, recordkeeping, and permanent preservation systems and to make recommendations on the revision of the framework design.
Controls: Record-making System Design, Recordkeeping System Design, Permanent Preservation System Design
Input: Preservation Performance Information, Recordkeeping Performance Information, Record-making Performance Information
Output: Information about Records Creator's Technology, Recommended Revisions

Discussion: The PDS Central Node develops and maintains the PDS and the PDS Master Catalog. It also performs system assessment. Feedback from users of the Pilot Planetary Data System (discussed in the next activity) indicated that the PPDS was overly centralized and that more attention was needed to data quality. In response PDS standards were developed. The PDS evolved into a distributed archive, and Nodes developed online distribution systems.

[9] Suggest renumbering activities so that *Implement Framework* is A1.3 and *Maintain Framework* is A1.4.

A1.4 Implement Framework

Definition: To acquire, test and activate all the components of the record-making, recordkeeping, and permanent preservation systems, and issue information about implementation problems.
Input: Record-making System Design, Recordkeeping System Design, Permanent Preservation System Design
Output: Record-making System, Recordkeeping System, Permanent Preservation System

Discussion: A Pilot Planetary Data System (PPDS) was developed. The concept of data curation was developed. Documentation standards were developed for the planetary data archive. A data catalog was developed. In addition to distributable CD-ROMs, datasets are accessible on-line.

The PDS Toolbox [JPL 1995, Appendix B] and PDS Label Library Light [Hughes et al 1998] is the *Record-making System*. The PDS Toolbox consists of

- PDS label Verifier (lvtool)
- PDS Table Browser (tbtool)
- PDS Table to Label Generator (tab2lab)
- PDS File to Label Generator (file2lab)
- PDS Simple Label Editor (sled)
- PDS Utilities
    - PDS Label Formatter (labform)
    - PDS Label Expander (explab)
    - Label Dealiasing Tool (dealias)
    - Add Columns to Table Utility (addcols)
    - Make Data Dictionary Index Utility (make index)
- PDS Toolbox Data Dictionary, Version 3

PDS Label Library Light (lablib3) is an evolving set of routines written in C that can be used to read, process, and write PDS labels. The Label Library reads and writes labels in the Object Description Language (ODL). The Label Library may be linked with any program that must deal with PDS labels. It was developed in order to encourage others to use PDS labels, to make it easier to mass produce PDS labels, and provide a method for enforcing ODL standards.

A2.1.1 Identify Persons, Actions and Dates of Documents

Definition: To attach to the documents the following metadata: the names of the authors, the addressees, and the writers; the names of the action or matter; and the dates of the compilation.
Controls: Record-making Metadata Schemes
Input: Information about Documents Contexts, Documents
Output: Identified Documents, Reports on Operation of Record-making Activity

Discussion: In the case study, data from the MGS spacecraft instruments are stored in the Project Database. PDS Structure Objects outline the format in which the Science data appear in PDS labels. A PDS Structure object corresponds to the COP Control *Record-making Metadata Schemes.*

The image in Figure 7 was taken with the Mars Orbital Camera, January 1, 1998. It is of a system of ridges located near the south pole of Mars. The origin of the ridges is unknown. They may be wind-created features which have been cemented or frozen with looser material in the centers removed by wind. The dark spots are 20 to 100 m across. The frame is 20 x 14 km and is centered at 81.5 S, 64.6 W. North and is at 11:00.

The PDS label shown in Figure 8 was created for the data product shown in Figure 7. The author is the Producer ID = MGS-MOC_Team. The Subject or Matter is the Rationale_Description = Rectilinear Ridges in South Polar Region. The date of compilation is the Product_Creation_Time = 1999-01-15T 21:26:15. The writer is the same as the author, and there are no addressees. Other metadata in the PDS label for this data product will be described in subsequent record-making activities.

MOC products are produced by the *makepds* program from the format internally used at the MOC Mission Operations Facility (MOF). This program reads a raw image file (see the MOC Software User's Guide), extracts some information from its headers, formats and attaches the PDS labels, and appends the image data.

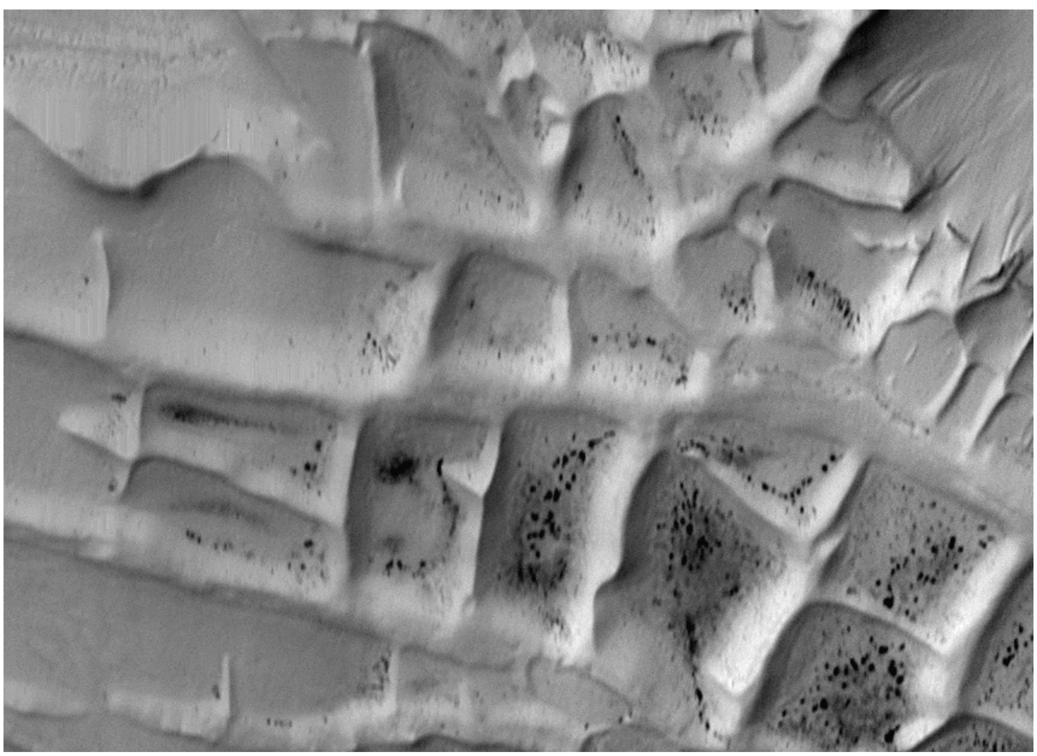


Figure 7. MOC Image of Rectilinear Ridges Near the Martian South Pole

The COP model does not indicate in which record-making activity the file format of the electronic record is identified. The PDS data preparers do more than indicate the format. They describe the format. For this data product the file format is described as 1409 records, each record of length 2048 bytes. There is one label record. The image begins at the second record.

The image is described as having 1408 rows (records), 2048 samples (pixels) per row. Each sample or pixel is represented by an unsigned integer. The checksum for the image is the integrity metadata. The last two lines of Figure 8 are the beginning bytes of the image data.

Supplementary documents are usually included with archive products to improve their utility. These documents provide assistance in understanding the data products [JPL 2003]. Typical archive documents include:

- Flight project documents
- Instrument papers
- Science articles
- Volume information
- Software Interface Specifications (SISs)
- Software user manuals

Science articles included in a volume of a dataset must also have labels, so this step would consist of adding values to the keywords:

AUTHOR_FULL_NAME
TITLE
JOURNAL_NAME
PUBLICATION_DATE
DOCUMENT_FORMAT

A2.1.2 Classify Documents

Definition: To intellectually set aside records by assigning classification codes from the classification scheme to documents and adding these codes to the identifying metadata.
Controls: Recordkeeping System (Classification System), Record-making Metadata Scheme
Input: Identified Documents
Output: Classified Records, Reports on Operation of Record-making Activity

Discussion: The Walkthrough Team has concluded that the PDS Record Makers are also performing the Recordkeeping function of records classification and registration. In this step, the following keywords (attributes) in the PDS Label of the data product are assigned values (classification codes). (See Fig. 6)

DATA_SET_ID = MGS-M-MOC-NA/WA-2-DSDP-LO-V1.0
SPACECRAFT_NAME = MARS_GLOBAL_SURVEYOR
INSTRUMENT_ID = MOC-NA
TARGET_NAME = MARS

DATA_SET_ID is similar to a record series id. Within the DATA_SET_ID, components are separated by hyphens. MGS is the instrument host; M (MARS) is the target, is MOC-NA/WA is Mars Orbital Camera - Narrow Angle/Wide Angle; LO is the description; V1.0 is the Version.

```
PDS_VERSION_ID = PDS3
FILE_NAME = "AB107908.IMG"
RECORD_TYPE = FIXED_LENGTH
RECORD_BYTES = 2048
FILE_RECORDS = 1409
LABEL_RECORDS = 1
^IMAGE = 2
SPACECRAFT_NAME = MARS_GLOBAL_SURVEYOR
MISSION_PHASE_NAME = "AB-1"
TARGET_NAME = MARS
INSTRUMENT_ID = "MOC-NA"
PRODUCER_ID = MGS_MOC_TEAM
DATA_SET_ID = "MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0"
PRODUCT_CREATION_TIME = 1999-01-15T21:26:15
SOFTWARE_NAME = "makepds 1.3"
UPLOAD_ID = "moc_p079_v1.sasf"
PRODUCT_ID = "AB-1-079/08"
START_TIME = 1998-01-01T01:04:09.36
IMAGE_TIME = 1998-01-01T01:04:09.36
STOP_TIME = 1998-01-01T01:04:12.61
SPACECRAFT_CLOCK_START_COUNT = "568083867:136"
SPACECRAFT_CLOCK_STOP_COUNT = "N/A"
FOCAL_PLANE_TEMPERATURE = 232.1 <K>
GAIN_MODE_ID = "AA"
OFFSET_MODE_ID = "72"
LINE_EXPOSURE_DURATION = 2.310000 <MILLISECONDS>
DOWNTRACK_SUMMING = 1
CROSSTRACK_SUMMING = 1
EDIT_MODE_ID = "0"
RATIONALE_DESC = "RECTILINEAR RIDGES IN SOUTH POLAR REGION"
OBJECT = IMAGE
LINES = 1408
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#216C6A13#
END_OBJECT = IMAGE
END

~}{}{~~~~~~~~~~~}|~~~|}~~|~|~|y~|~~~|}~~~~~}}}}~{|{~|~~{}}|}~z}
}s}{~{yq{yw}wv}x{x}{{~}~~}{~|z{~{ntm{z~xrq{y~~}v~~wzv|y~|~v}}^
```

Figure 8. PDS Label for Image ab107908

Classification of supplementary documents corresponds to assigning a value to the keyword DOCUMENT_TOPIC_TYPE. There are many such values, for example:

ASTEROID INFORMATION
CARTOGRAPHY
COMET HALLEY
COMETS
CURRENTS IN SATURN'S MAGNETOSPHERE
EXPERIMENT RESULTS
GEOLOGY
JUPITER ELECTRONS
MARS GRAVITY

PLANETARY ATMOSPHERES
PLANETARY MAPPING
PROJECT FINAL REPORT

A2.1.3 Register Records

Definition: To assign to records registration numbers based on the registration scheme, and to add these numbers to the identifying metadata.
Controls: Recordkeeping System (Registration Scheme), Record-making Metadata Scheme
Input: Classified Records
Output: Records to be Stored, Documents to be sent, Reports of Record-making Activity

Discussion: Registration of a data product consists of entering a unique identifier, PRODUCT_ID = AB-1-079/08 in the label of the data product and entering corresponding Filename in an index table in the index directory of a volume. This is also the FILE_NAME = AB107908.IMG.

Registration of supplementary documents consists of assigning the filename to the keyword FILE_NAME in the PDS label of the document, placing the document in the DOCUMENT directory of a volume and including the filename in the Index file of the volume.

A2.1.4 Send Out Documents

Definition: To transmit documents to external juridical and natural persons, add the dates of transmission (and if applicable, information on any attachments) to the identifying metadata, and make record copies of the sent documents to be stored in the recordkeeping system.
Controls: Record-making Metadata Scheme
Input: Documents to be Sent
Output: Outgoing Documents, Record Copies of Outgoing Documents, Reports on Operation of Record-making Activity

Discussion: In the case study, no evidence was found of documents in datasets that were made to be sent to other organizations. If the definition of this activity is expressed as an if-then rule, then it would be clearer that this activity only applies if there are records to be sent to other organizations, for example, "If documents are transmitted to external juridical or natural person, a record copy should be made of the sent document, and the date of transmission added to the record's metadata."

A2.2.1 Identify Persons, Actions and Dates of Incoming Documents

Definition: To attach to the incoming documents the following metadata: the names of the authors, the addressees, and the writers; the names of the action or matter; and the dates of compilation, transmission, and receipt.
Control: Registration Scheme
Input: Incoming Documents, Information about Incoming Documents' Context
Output: Identified Incoming Documents, Reports on Operation of Record-making Activity

Discussion: On the one hand, it is not clear that any of the documents that occur in a PDS volume are incoming documents from juridical or natural persons. On the other hand, all of the raw scientific data transmitted from instruments on spacecraft might be considered incoming scientific data. It is better to express the definition of this activity as an if-then rule. "If there are incoming documents, then attach to the metadata the name of the authors, the addressees, and the writer; the name of the action or matter; and the dates of compilation, transmission, and receipt."

A2.2.2 Classify Incoming Documents

Definition: To assign classification codes from the classification scheme to incoming documents and add these codes to the identifying metadata.
Control: Classification Scheme, Registration Scheme
Input: Identified Incoming Document
Output: Classified Received Records, Report on Operation of Record-making Activity

Discussion: Incoming documents were not discovered in the case study. Hence, it might be better to express the definition of this COP activity as "If there are incoming documents, then assign classification codes from the classification scheme to the documents and add these codes to identifying metadata."

A2.2.3 Register Received Records

Definition: To assign registration numbers to received records based on the registration scheme and add these numbers to the identifying metadata.
Controls: Registration Scheme, Record-making Metadata Scheme
Input: Classified Received Records
Output: Received Records, Reports on Operation of Record-making Activity

Discussion: In the case study, no evidence of incoming documents was found. Hence, it might be better to express the definition of this COP activity as "If there are incoming documents, then assign registration numbers to the received records based on the registration scheme and add these numbers to the identifying metadata."

A2.3 Monitor Performance of Record-making System

Definition: To assess the efficacy of the performance of the record-making system by analyzing reports on the operation of record-making activities and examining records, and issue information on the performance of the record-making system for use in continued maintenance of the chain of preservation framework.
Control: Record-making System
Input: Reports on Operation of Record-making Activity
Output: Record-making Performance Information

Discussion: Management at the PDS Central Node assesses system performance.

A3.1.1 Manage Information About Kept Records

Definition: To provide overall control and co-ordination of the various informational components of records in the recordkeeping system in response to information received via feasibility reports, documentation about records update activities, and, as necessary, through examination of the records themselves.
Input: Received Records, Feasibility Reports
Output: Records with Integrity Metadata, Information about Digital Components of Kept Records, Information about Kept Records, Reports on Operation of Recordkeeping Activity

Discussion: This activity corresponds to the operation of the PDS Archives and Catalog System. The PDS Project Manager at the Central Node coordinates the operation of the entire PDS, assisted by a PDS Management Council consisting of the Discipline Node Principal Investigators/Managers, Science Advisors, and representatives of Central Node Staff.

A3.1.2 Manage Storage of Kept Records

Definition: To place the digital components of records and their metadata into storage and maintain them.
Input: Records to be Stored, Record Copies of Outgoing Documents, Records with Integrity Metadata, Information about Digital Components of Kept Records, Information about Kept Records
Output: Kept Records, Kept Records that need updating, Reports on Operation of Recordkeeping Activity

Discussion: After Peer Review, the PDS Central Node data engineers validate dataset volumes using the Volume Verifier. It validates the format and content of all product labels, and validates the integrity of data files using checksums. Next, the dataset volumes are stored and cataloged. The catalog information for the volumes is used to automatically update the PDS catalog. Dataset volumes are indexed by mission, target name, target type, instrument and instrument type, and volume identifier attributes.

A3.1.3 Update Records

Definition: To carry out necessary conversions on the records to keep them accessible, legible, and intelligible over time.
Input: Kept Records that need Updating, Information about Digital Components of Kept Records, Information about Kept Records
Output: Updated Records, Information about Activities Done to Updated Records

Discussion: The preservation strategy for records in the PDS includes self-describing files. This strategy obviates the need to convert to other formats. However, the PDS Toolkit and Label Library Light provide the capability to convert many data formats to persistent form.

This is another case where it might be desirable to state the definition of the activity as an if-then rule. "If necessary to convert the records to keep them accessible, legible, and intelligible over time, then convert them via some method such as migration, standardization, or transformation to persistent form."

A3.2 Facilitate Access

Definition: To satisfy requests by retrieving and reproducing records and/or retrieving information about records in the recordkeeping system.
Input: Kept Records, Request for Kept Records and/or Information about Kept Records
Output: Reproduced Kept Records Issued to Users, Reproducible Kept Records Issued to Users, Information about Kept Records Issued to Users, Creators Certificates of Authenticity, Reports on Operation of Recordkeeping Activity

Discussion: When datasets are transferred to the PDS, they are automatically cataloged by mission, target name, target type, instrument, instrument type and dataset identifier. Scientists and the public can access the MGS and other Planetary Science datasets at http://pds.jpl.nasa.gov/data_services/. The Web page in Figure 9 will be displayed.

Figure 9. PDS Data Services

Examples of two of the services, Data Search and PDS Explorer, will be shown. If one selects Data Search, the Web interface in Figure 10 is shown.

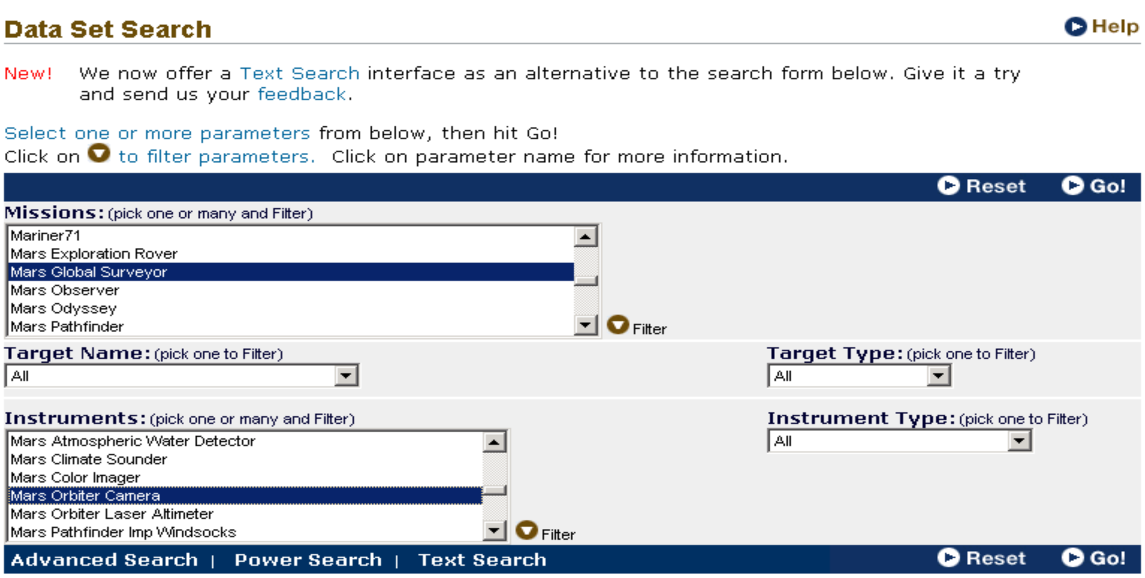


Figure 10. PDS Dataset Search

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 following page is displayed.

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 11. Datasets for the MGS Mars Orbital Camera

If one selects the dataset that is being used in this walkthrough, MGS-M-MOC-NA/WA-2-DSDP-l0-v1.0 (Mars Orbital Camera, Decompressed Standard Data Products), the page in Figure 12 is displayed.

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 12. Dataset Descriptions

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, which is shown in Appendix A.

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, as is shown in Figure 13.

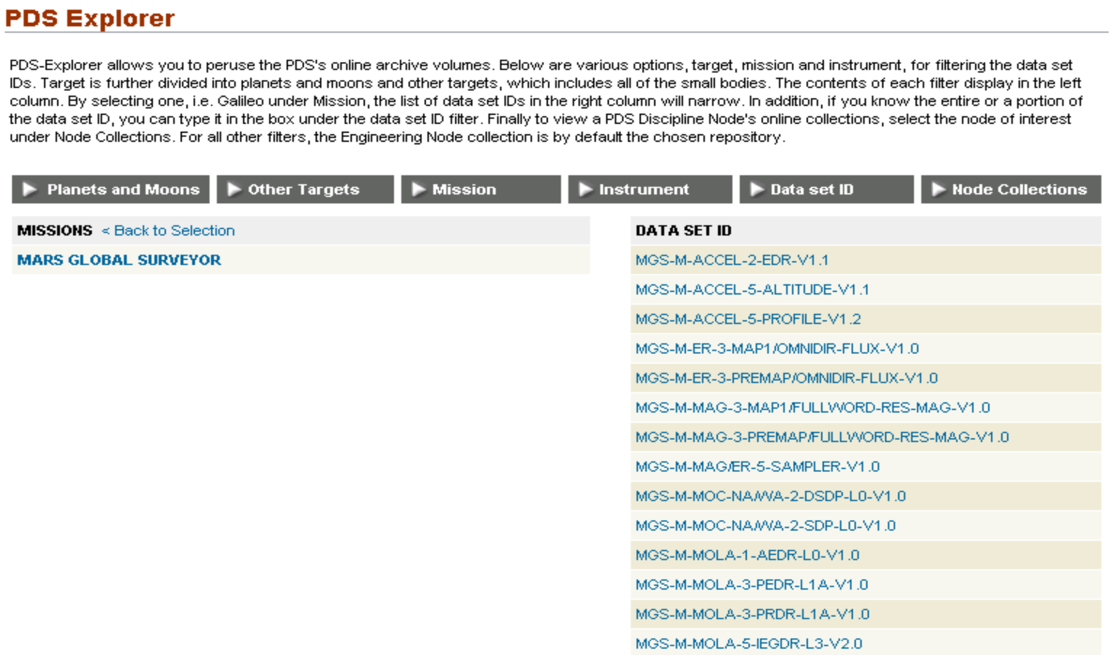


Figure 13. PDS Explorer for Browsing PDS Datasets

If one selects for the MGS Mission, the dataset ID used in this walkthrough, the following Web page is displayed.

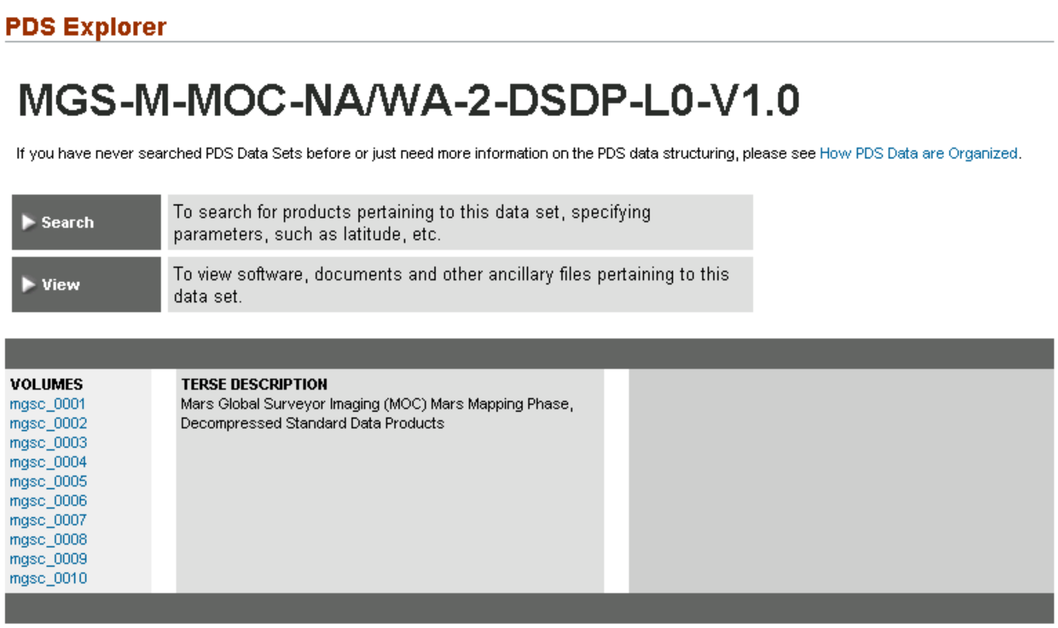


Figure 14. Volumes in the MGS-M-MOC-NA/WA-2-DSPS-L0-V1.0 Dataset

If one selects the mgsc_001 volume, the following Web page is displayed.

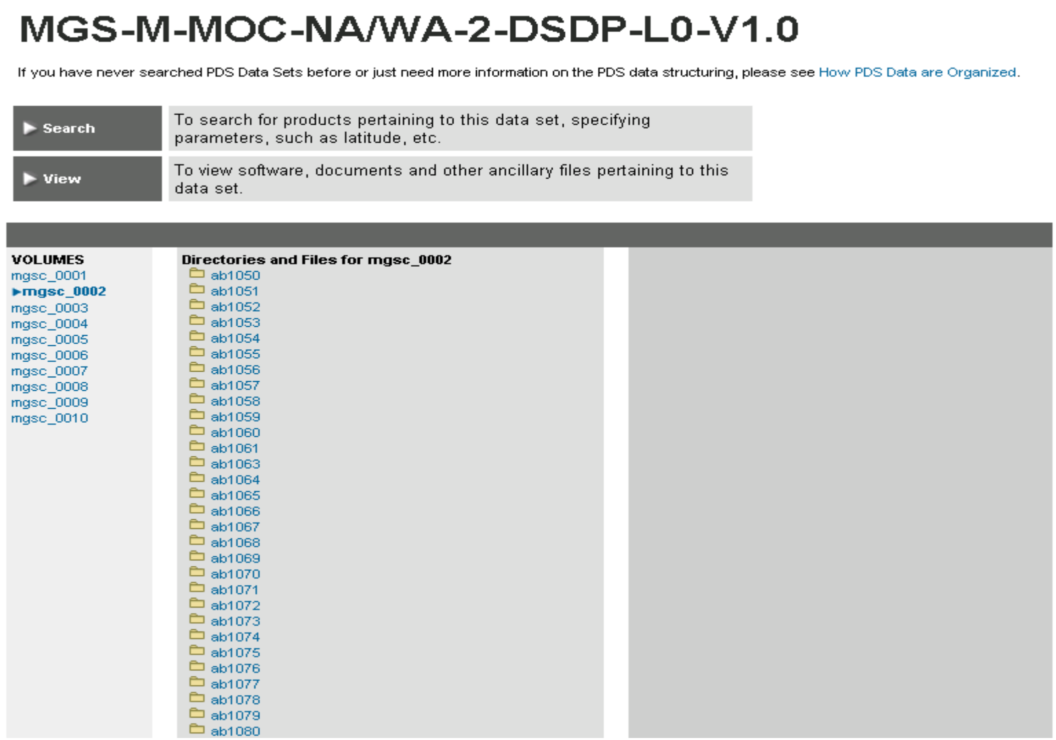


Figure 15. Files in the mgsc_0001 Volume

If one downloads the file ab107908.img and uses NASAView to view it, one sees the image shown in Figure 16.

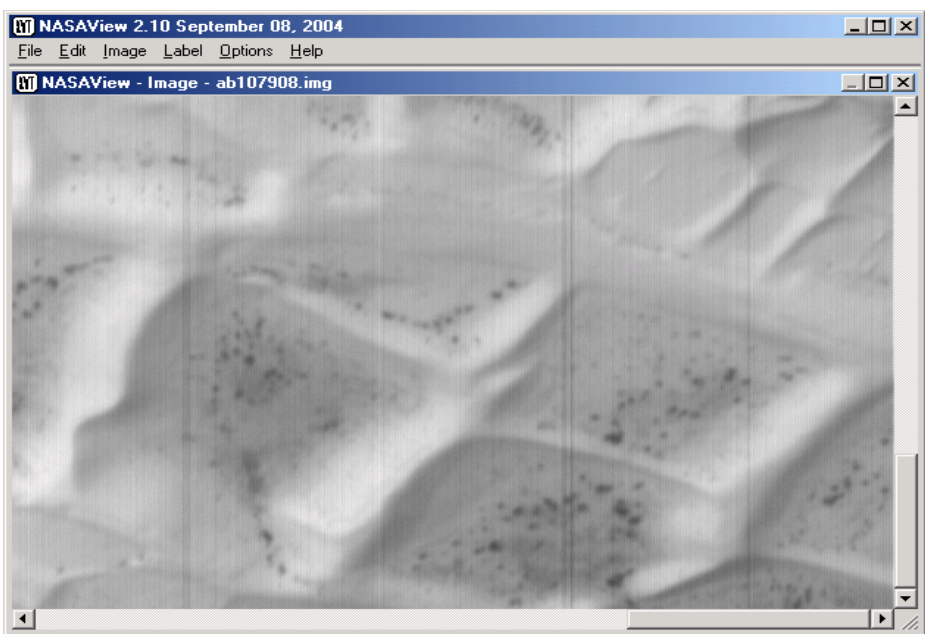
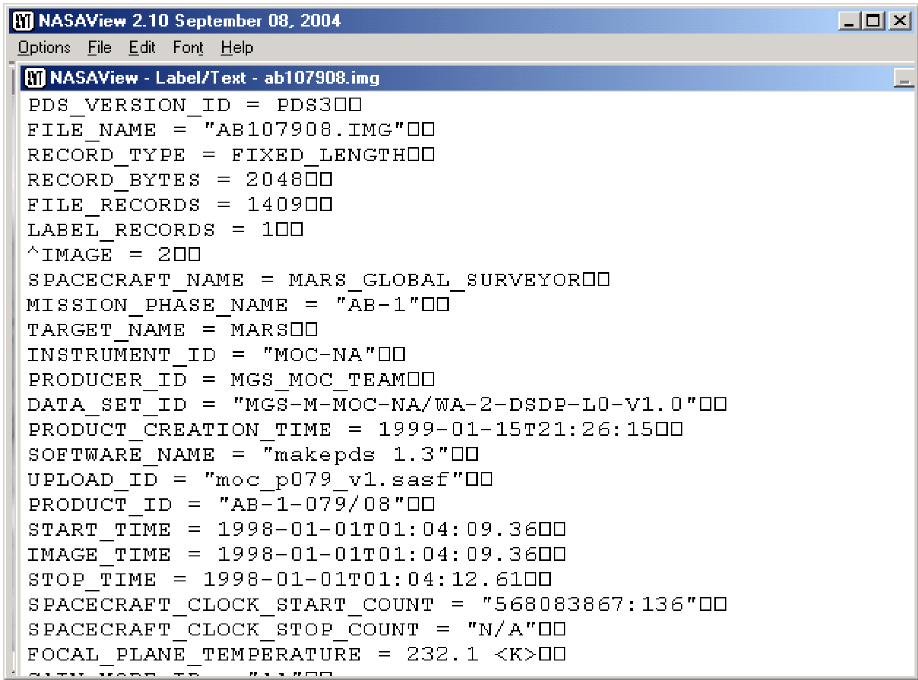


Figure 16. NASAView Display of Mars Orbital Camera Image of Mars Surface

If one selects Label from the Menu bar, one will see the label for the data product, which is shown in Figure 17.


NASAView 2.10 September 08, 2004

Options File Edit Font Help

NASAView - Label/Text - ab107908.img

```
PDS_VERSION_ID = PDS3□□
FILE_NAME = "AB107908.IMG"□□
RECORD_TYPE = FIXED_LENGTH□□
RECORD_BYTES = 2048□□
FILE_RECORDS = 1409□□
LABEL_RECORDS = 1□□
^IMAGE = 2□□
SPACECRAFT_NAME = MARS_GLOBAL_SURVEYOR□□
MISSION_PHASE_NAME = "AB-1"□□
TARGET_NAME = MARS□□
INSTRUMENT_ID = "MOC-NA"□□
PRODUCER_ID = MGS_MOC_TEAM□□
DATA_SET_ID = "MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0"□□
PRODUCT_CREATION_TIME = 1999-01-15T21:26:15□□
SOFTWARE_NAME = "makepds 1.3"□□
UPLOAD_ID = "moc_p079_v1.sasf"□□
PRODUCT_ID = "AB-1-079/08"□□
START_TIME = 1998-01-01T01:04:09.360□□
IMAGE_TIME = 1998-01-01T01:04:09.360□□
STOP_TIME = 1998-01-01T01:04:12.610□□
SPACECRAFT_CLOCK_START_COUNT = "568083867:136"□□
SPACECRAFT_CLOCK_STOP_COUNT = "N/A"□□
FOCAL_PLANE_TEMPERATURE = 232.1 <K>□□
```

Figure 17. Label for data product ab107908

A3.3.1 Manage Disposition Activities

Definition: To provide overall control and co-ordination of disposition activities through the issuing of disposition activity directives and orders formulated in response to ongoing feedback/information received via records transfer notifications and disposition activity reports.
Input: Notification of Receipt of Transfer, Notifications of Rejection of Transfer, Reports on Disposition Activity
Output: Order to Rectify Rejected Transfer, Disposition Activity Directives, Reports on Operation of Recordkeeping Activity

Discussion: For the PDS, this activity is the responsibility of the PDS Project Manager at the Central Node.

A3.3.2 Apply Retention Decision

Definition: To segregate records to be destroyed or transferred to the permanent preservation system.
Control: Retention Schedule, Disposition Activity Directives
Input: Kept Records, Order to Rectify Rejected Transfer
Output: Records to be Destroyed, Records to be Transferred, Reports on Disposition Activity

Discussion: In the PDS, this activity involves identifying dataset volumes to be passed on to the NSSDC for "deep archiving" along with appropriate catalog and ancillary information. It also involves transfer to NASA Records Management of copies of some project (mission) datasets identified by NARA as being of historical value.

A3.3.3 Destroy Records

Definition: To obliterate records identified for destruction, and provide documentation about the records destroyed.
Control: Disposition Rules and Procedures, Disposition Activity Directives
Input: Records to be Destroyed
Output: Documentation about Destroyed Records, Reports on Disposition Activity

Discussion: This activity does not apply to the collection of scientific data records in the PDS, cause they all have long-term value to space scientists. However, raw datasets including telemetry errors that have been used to create definitive, derived datasets should be destroyed six months after the acceptance of the derived data set into the PDS.

A3.3.4 Transfer Records

Definition: To prepare records for transfer in accordance with the terms and conditions of transfer, and send them to the office responsible for the permanent preservation function with the accompanying documentation necessary for permanent preservation.
Controls: Disposition Activity Directives, Disposition Rules and Procedures, Terms and Conditions of Transfer
Input: Records to be transferred, Order to Rectify Rejected Transfer
Output: Transfer Documentation, Records Transfer

Discussion: Datasets transferred from the PDS to the NSSDC are transferred as CD-ROM copies of volumes. Datasets to be transferred to NARA are transferred by NARA Records Management seven years after the close of a project. Information was not collected on the latter activity.

A3.4 Monitor Performance of Recordkeeping System

Definition: To assess the efficacy of the performance of the recordkeeping system by analyzing reports on the operation of recordkeeping activities and by examining kept records, and issue information on the performance of the recordkeeping system for use in continued maintenance of the chain of preservation framework.
Control: Recordkeeping System
Input: Reports on Operation of Recordkeeping Activity
Output: Recordkeeping Performance Information

Discussion: This activity is the responsibility of the PDS Project Manager at the Central Node.

A4.1 Manage Permanent Preservation System

Definition: To assess the efficacy of the performance of the permanent preservation system by examining reports on the operation of permanent preservation activities, issue permanent preservation activity directives as necessary, and issue information on the performance of the permanent preservation system for use in continued maintenance of the chain of preservation framework.
Input: Reports on Operation of Permanent Preservation Activity
Output: Preservation Performance Information, Permanent Preservation Activity Directives

Discussion: This activity is the Responsibility of the NSSDC Director and his staff. This activity was beyond the scope of the PDS case study, so no data was available for the walkthrough.

A4.2.1 Analyze Information About Records

Definition: To collect, organize, record, and assess relevant information from the kept records being appraised and about their juridical-administrative, provenancial, procedural, documentary and technological contexts.
Controls: Permanent Preservation System (Selection System), Permanent Preservation Activity Directives
Input: Information from Kept Records, Information about Context, Recommendations of Need for Change
Output: Information for Appraisal, Reports on Operation of Permanent Preservation Activity

Discussion: Each PDS volume in a dataset undergoes a review modeled after the peer review process used by scientific journals. The peer review process resembles the appraisal process for selection of records for long-term preservation.

The PDS lead node selects a review panel. The review panel consists of scientists who are experts in the use of similar datasets, members of the instrument team, and PDS standards experts. Outside science experts are asked to try to use the data and documentation provided. PDS experts review volumes for compliance with PDS standards. Since the PDS is the entry point into the NASA archival environment for planetary science data, this activity is performed before entry of volumes into the PDS.

A4.2.2.1 Assess Continuing Value of Records

Definition: To analyze and judge the capacity of records being appraised to serve the continuing interests of their creator and society.
Input: Information for Appraisal
Output: Technical Description of Records Proposed for Preservation, Valuation Information

Discussion: If the PDS review panel determines the datasets meet the PDS standards, then they serve the continuing interests of NASA and space scientists. There is also an appraisal by NARA appraisal archivists to determine the value of certain data sets as having value for historical purposes as "seminal observations having extraordinary impact on the evolution of Space and Earth Sciences." No data was collected on this NARA activity.

A4.2.2.2 Assess Authenticity of Records

Definition: To analyze and judge the grounds for presuming records being appraised to be authentic.
Input: Information for Appraisal
Output: Reports on Operation of Permanent Preservation Activity, Technical Description of Records Proposed for Preservation, Assessments of Authenticity

Discussion: This is not a lowest-level activity in the COP Model. Information was not found in the case study corresponding to the three activities in its decomposition. This is not surprising as the criteria and method of assessment of the authenticity of digital records was a new research contribution of InterPARES 1 [Duranti 2005, Appendix 2]. The criteria for a presumption of authenticity of a creator's digital records are shown in Figure 18.

The method of assessment is described as follows:

> A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created and maintained. The evidence that supports the presumption that the record creator created and maintained them authentic are enumerated in the Benchmark Requirements Supporting the Presumption of Authenticity of Digital records (Requirement Set A). A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. The requirements are, therefore, cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity. This is why these requirements are termed "benchmark" requirements.

A4.2.2.2.1 Compile Evidence Supporting the Presumption of Authenticity

Definition: To collect, organize, and record evidence of the identity and integrity of records being appraised and about the procedural controls applied to them, to support the presumption of authenticity of those records.
Input: Information for Appraisal
Output: Evidence for the Presumption of Authenticity, Reports on Operation of Permanent Preservation Activity

Discussion: This activity was not found in the PDS standards or procedures. However, the kinds of evidence with respect to the PDS that would need to be collected are indicated in the discussion of the previous activity.

A4.2.2.2.2 Measure Evidence Against Requirements For Authentic Records

Definition: To compare the evidence compiled about the identity, integrity, and procedural controls of the records being appraised with the requirements for authentic records.
Input: Evidence for the Presumption of Authenticity,
Output: Reports on Operation of Permanent Preservation Activity, Technical Description of the Records Proposed for Preservation, Assessments of Authenticity

To support a presumption of authenticity the preserver must obtain evidence that:	
REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record	the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records. ***A.1.a*** Identity of the record: ***A.1.a.i*** Names of the persons concurring in the formation of the record, that is: • name of author • name of writer (if different from the author) • name of originator (if different from name of author or writer) • name of addressee ***A.1.a.ii*** Name of action or matter ***A.1.a.iii*** Date(s) of creation and transmission, that is: • chronological date • received date • archival date • transmission date(s) ***A.1.a.iv*** Expression of archival bond (e.g., classification code, file identifier) ***A.1.a.v*** Indication of attachments ***A.1.b*** Integrity of the record: ***A.1.b.i*** Name of handling office ***A.1.b.ii*** Name of office of primary responsibility (if different from handling office) ***A.1.b.iii*** Indication of types of annotations added to the record ***A.1.b.iv*** Indication of technical modifications;
REQUIREMENT A.2: Access Privileges	the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;
REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records	the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;
REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;
REQUIREMENT A.5: Establishment of Documentary Forms	the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;
REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;
REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;
REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

Figure 18. Benchmark Requirements Supporting the Presumption of Authenticity of E-Records

Discussion: In InterPARES 2 general study 06 (Testing of the ATF's Method of Assessment with the Benchmark Requirements for a Presumption of Authenticity) [Underwood and Isbell 2007], a Bayesian Belief Network (BBN) is described that aids archivists is collecting evidence supporting the assessment of a presumption of authenticity of digital records in a recordkeeping system. Evidence from the Planetary Data System Case Study is used to demonstrate the use of this tool. The result of that demonstration is that one can have a high degree of belief that the PDS maintains authentic planetary science data records.

A4.2.2.2.3 Verify Authenticity

Definition: To establish alternative verification methods for presuming the authenticity of records being appraised in cases where there is insufficient evidence to meet the requirements for authentic records.
Input: Need for Verification
Output: Reports on Operation of Permanent Preservation Activity, Assessments of Authenticity

Discussion: "In any given case, there may be an insufficient basis for a presumption of authenticity, or the presumption may be extremely weak. In such cases, further analysis may be necessary to verify the authenticity of the records. A verification of authenticity is the act or process of establishing a correspondence between known facts about the record and the various contexts in which it has been created and maintained, and the proposed fact of the record's authenticity. In the verification process, the known facts about the record and its contexts provide the grounds for supporting or refuting the contention that the record is authentic. Unlike the presumption of authenticity, which is established on the basis of the benchmark requirements, this verification involves a detailed examination of the records themselves and reliable information available from other sources about the records and the various contexts in which they have been created and maintained. Methods of verification include, but are not limited to, a comparison of the records in question with copies that have been preserved elsewhere or with back-up tapes;" comparison of the records in question with entries in a register of incoming and outgoing records; textual analysis of the record's content; forensic analysis of the medium, script, and so on; a study of audit trails; and the testimony of a trusted third party. [Duranti 2005, Appendix B]

The fact that satellite images have been admitted into evidence in a few legal cases provides us with an alternative method for verifying the authenticity of scientific data [Hodge 1997]. While the legal cases in which satellite images were admitted involved Earth Science data and Environmental Law, they seem extensible to Planetary Science Data and future "Lunar and Planetary Law." "The admissibility of remote sensing information must be examined within the context of the general requirements for admission of scientific evidence and expert opinion." A litigator seeking the admission of remote sensing data as evidence must (1) qualify an expert, (2) authenticate and prove the contents of the data, and (3) establish that proper and accepted processing techniques were employed. The use of an archive history file accompanying the final satellite imagery exhibit provides the potential for objective, external authentication and establishes that appropriate techniques and methodologies were employed in the creation the exhibit. An *archive history file* is a document listing (1) all the data used in the creation of the final exhibit, (2) all the tools used in the creation of the final exhibit, and (3) all the processes and methods used to create an exhibit.

A PDS data set with reduced data would contain references to the raw data in other data sets and copies of the software used to produce the reduced data set. This seems very similar to an archive history file that along with expert testimony might be an alternative method of authenticating the data.

A4.2.2.3 Determine Value of Records

Definition: To establish the value of records being appraised based on assessments of their continuing value and their authenticity.
Input: Assessment of Continuing Value, Assessments of Authenticity
Output: Valuation Information,[10] Reports on Operation of Permanent Preservation System

Discussion: The PDS does not assess the authenticity of its datasets or data products, but the peer review panel does check for the conformance to PDS Standards. Conformance to the PDS Label standard and inclusion of checksums for data integrity are part of the requirements for authenticity, so this activity seems to be part of the PDS activities.

A4.2.3.1 Determine Record Elements to be Preserved

Definition: To identify the necessary documentary components (e.g., record profile, attachments, annotations, etc.) and elements of form (e.g., author, date, subject line, etc.) of records to be preserved to determine which record elements must be preserved to protect the authenticity of those records.
Input: Information for Feasibility, Valuation Information
Output: List of record Elements to be Preserved, Reports on Operation of Permanent Preservation Activity

Discussion: Because the PDS was designed to be a recordkeeping and preservation system for scientific data, volumes containing datasets that have been reviewed and accessioned into the PDS have the necessary documentary components and elements of form to be preserved. The components of a volume contain the record elements to be preserved.

A4.2.3.2 Identify Digital Components to be Preserved

Definition: To identify the digital components that manifest the record elements that need to be preserved to protect the authenticity of records earmarked for permanent preservation.
Input: List of Record Elements to be Preserved, Technical Description of Records Proposed for Preservation
Output: List of Digital Components to be Preserved, Reports on Operation of Permanent Preservation Activity

Discussion: The volume contents and PDS labels indicate the digital components to be preserved.

[10] Valuation information is an output of activity A4.2.2.3 and an input to activity A4.2.4, Make Appraisal Decision. Valuation Information is also an external input to the entire COP model and an external input to activity 4.2.3, Determine feasibility of Preservation. Shouldn't the Valuation Information input to A4.2.3 just come from activity A4.2.2?

A4.2.3.3 Reconcile Preservation Requirements with Preservation Capabilities

Definition: To determine whether the digital components manifesting the record elements that need to be preserved to protect the authenticity of records earmarked for permanent preservation can in fact be preserved given the preserver's current and anticipated preservation capabilities.
Control: Preservation System
Input: List of Digital Components to be Preserved, Information Concerning Available Technology
Output: Information about Digital Components to be Preserved, Feasibility Reports, Reports on Operation of Permanent Preservation System

Discussion: Since the PDS data products will not need to be converted to other formats, the NSSDC should have no difficulty in preserving them. However, the NSSDC preserves scientific datasets in several standard formats. It has been necessary for them to convert legacy scientific products to other formats. No case study data was collected on this NSSDC activity.

A4.2.4 Make Appraisal Decisions

Definition: To decide on and document the retention and disposition of records based on valuation and feasibility information, and to agree on and document the terms and conditions of transfer of the records to the preserver.
Controls: Permanent Preservation System, Permanent Preservation Activity Directives
Input: Feasibility Reports, Information about Digital Components to be Preserved, Valuation Information
Output: Information about Appraised Records, Appraisal Decisions, Terms and Conditions of Transfer, Information about Appraisal Decisions, Reports on Operation of Permanent Preservation Activity

Discussion: For PDS data sets, the peer review panel points out problems with the dataset and documentation and suggests improvements. Deficiencies are recorded as liens against the data set. When all outstanding liens against a data set are resolved, it is added to the PDS archive.

A4.2.5 Monitor Appraised Records

Definition: To keep track of changes to appraised records and/or their context that might make it necessary to adjust or redo an appraisal.
Controls: Permanent Preservation System, Permanent Preservation Activity Directives
Input: Information about Appraised Records, Appraisal Decisions, Information about Appraisal Decisions
Output: Recommendations of Need for Change, Reports on Operation of Permanent Preservation Activity

Discussion: With regard to PDS datasets, when the peer review panel approves a dataset for ingestion into the PDS archive, there is not need for a reappraisal. However, King [1998] points out that ensuring data integrity and usability requires periodic data renewal cycles. Some such cycles involve media refreshment. Other such cycles might involve data reorganization, reformatting or recreating related software. Particular data renewal cycles may require approval to release (destroy) the data set rather than renew it. This may require reappraisal of a data set's continuing value to determine whether the data renewal costs are justified.

A4.3.1 Register Transfers

Definition: To record registration information about received transfers and issue notifications of receipt to the persons transferring the records.
Input: Record Transfers, Transfer Documentation,
Output: Notification of Receipt of Transfer, Registered Transfers, Reports on Operation of Permanent Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.3.2 Verify Authorization for Transfers

Definition: To verify the authority for transfer of records selected for preservation, and, in cases of unauthorized transfers, issue notifications of rejection of transfer to the persons transferring the records.
Control: Terms and Conditions of Transfer
Input: Registered Transfers, Transfer Documentation
Output: Authorized Transfers, Notifications of Rejection of Transfer, Reports on Operation of Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.3.3 Verify Transfers

Definition: To determine whether transfers of records selected for preservation include all records and aggregates of records specified in the terms and conditions of the transfers, and, in unverified cases, issue notifications of rejection of transfer to the persons transferring the records.
Control: Terms and Conditions of Transfer
Input: Authorized Transfer, Transfer Documentation,
Output: Notification of Rejection of Transfer, Accepted Transfers, Reports on Operation of Permanent Preservation Activity

Discussion: The primary validation tool of the PDS is the Volume Verifier [JPL 2003]. 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.

This activity is performed after peer review and before data sets are stored in the PDS. It does not need to be re-performed upon transfer of PDS datasets to the NSSDC deep archive.

A4.3.4 Confirm Feasibility of Preservation

Definition: To confirm that the determinations of the feasibility of preservation made during the process of appraisal are still valid, and, in unconfirmed cases, issue notifications of rejection of transfer to the persons transferring the records.
Control: Permanent Preservation System (Preservation System)
Input: Accepted Transfers, Feasibility Reports
Output: Records to be Accessioned, Information for Preservation, Notification of Rejection of Transfer, Reports on Operation of Permanent Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.3.5 Accession Records

Definition: To formally accept records selected for permanent preservation into custody and document transfers in accessions documentation.
Control: Permanent Preservation System (Descriptive Rules and Procedures)
Input: Records to be Accessioned, Transfer Documentation
Output: Accessioned Records, Information about Accessioned Records, Reports on Operation of permanent Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.4.1.1 Compile Information for Preservation, Description and Output

Definition: To collect, organize and record relevant appraisal, acquisition, accession and preservation information about acquired records for their preservation, description, and output.
Input: Information for Preservation, Information about Accessioned Records, Documentation about Destroyed Records, Retrieval Requests, Accessioned Records, Updated Information for Preservation
Output: Information for Description, Arranged Records, Additional Information to Satisfy Requests, Reports on Operation of Permanent Preservation Activity, Information about Digital Components of Preserved Records, Information about Digital Components of Requested Records

Discussion: This is essentially the Database Management System, Metadata Schema, and data for the Preservation System. It also supports the function of the retrieval of information about datasets, volumes and data products including their location in archival storage. This is a NSSDC activity and system and was not part of the PDS case study.

A4.4.1.2 Describe Acquired Records

Definition: To record information about the nature and make-up of records acquired for permanent preservation and about their juridical-administrative, provenancial, procedural, documentary, and technological contexts, as well as information about any changes they have undergone since they were first created.
Input: Information for Description, Arranged Records, Information about Appraised Records

Output: Descriptive Instruments, Described Records, Reports on Operation of Permanent Preservation Activity

Discussion: The PDS Standards Reference prescribes the description of Planetary Science datasets at the time of creation (record-making) [JPL 2003]. Hence, the NSSDC, the deep archives, does not need to re-describe the already described datasets. At a minimum, the description of a PDS dataset includes:

DATA_SET_HOST
DATA_SET_INFORMATION
DATA_SET_REFERENCE_INFORMATION
DATA_SET_TARGET
DATA_SET_MISSION

There are templates for each of these indicating the Attributes that must be included in the description.

A4.4.1.3 Update Information on Preservation Actions

Definition: To record information about actions taken to update digital components of records acquired for permanent preservation or their storage.
Input: Information about Updated Digital Components, Information about Updated Storage
Output: Updated Information for Preservation, Reports on Operation of Permanent Preservation Activity

Discussion: Because the preservation strategy for PDS datasets is Persistent Object Preservation, the files are self-describing. Digital components of PDS data records should not need to be updated. Hover, the NSSDC does need to convert digital components of legacy data records to standard formats such as CDF, HDF, and FITS. However, information of these activities was not collected because it was beyond the scope of the PDS case study.

A4.4.1.4 Retrieve Information for Requests

Definition: To gather the information required, from descriptive instruments and other preservation information, to satisfy requests for records and/or information about records.
Input: Descriptive Instruments, Additional Information to Satisfy Requirements, Accounting for Unsatisfied Requests
Output: Retrieved Information about Required Digital Components, Retrieved Information about Requested Records, Reports on Operation of Permanent Preservation Activity

Discussion: This is a NSSDC activity and was beyond the scope of the PDS case study.

A4.4.2.1 Place Digital Components in Storage

Definition: To place digital components of records acquired for permanent preservation into one or more digital files for preservation purposes.

Input: Accessioned Records, Updated Digital Components, Digital Components on Refreshed Medium, Corrected Storage
Output: Stored Digital File, Reports on Operation of Permanent Preservation Activity, Information about Updated Storage

Discussion: PDS datasets and collections of datasets are stored in volumes, which contain directories, subdirectories and files. The NSSDC retains these on CD-ROM or copies them to their Digital Linear Tape System. However, information on this activity for non-PDS datasets was beyond the scope of the PDS case study.

A4.4.2.2 Monitor Storage

Definition: To monitor operation of the storage system, the media on which the digital components are stored, the digital components, and the facilities where the system and components are located.
Input: Stored digital file
Output: Decisions to Correct Storage, Decisions to Refresh Medium, Reports on Operation of Permanent Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.4.2.3 Correct Storage Problems

Definition: To take the actions prescribed by the preservation strategy to eliminate problems in storage.
Control: Permanent Preservation System (Preservation Strategy)
Input: Decisions to Correct Storage, Stored Digital File
Output: Corrected Storage, Information about Updated Storage, Reports on Operation of Permanent Preservation Activity

Discussion: Information not collected on this PDS or NSSDC activity.

A4.4.2.4 Refresh Storage of Digital Components

Definition: To convert storage of digital components from one medium to another or otherwise ensure that the storage medium remains sound.
Input: Decisions to Refresh Medium, Stored Digital File
Output: Digital Components on Refreshed Medium, Information about Updated Storage, Reports on Operation of Permanent Preservation Activity

Discussion: Information not collected on this NSSDC activity.

A4.4.2.5 Retrieve Digital Components from Storage

Definition: To output copies of retrieved digital components of records in the permanent preservation system in response to requests for those components, and, cases where digital components are encountered that need updating, redirect them to be updated.

Input: Stored Digital File, Retrieved Information about Required Digital Components
Output: Digital Components that Need Updating, Retrieved Digital Components, Reports on Operation of Permanent Preservation Activity

Discussion: The NSSDC will provide CDs of volumes containing specific PDS datasets. The NSSDC retains preservation copies of all scientific datasets in its "deep archive." Copies of these can be retrieved by scientists from FTPWeb[11] and from the CDAWeb.[12] For more information on CDAWeb, see http://cdaweb.gsfc.nasa.gov.

A4.4.3 Update Digital Components

Definition: To carry out necessary conversions on the digital components of records in the permanent preservation system to keep the records accessible, legible, and intelligible over time (such as by migration, standardization, or transformation to persistent form), and record information about updated digital components.
Input: Information about Digital Components of Preserved Records, Digital Components that Need Updating
Output: Updated Digital Components, Information about Updated Digital Components, Reports on Operation of Permanent Preservation Activity

Discussion:[13] NSSDC has migrated numerous datasets to different media and formats. For example, they have migrated legacy 9-track and 3480 cartridges to Digital Linear Tape for long-term archiving [Sawyer et al 2005]. However, these were not migrations of planetary science data, and so the PDS case study did not collect information on this activity at the NSSDC.

The datasets transferred from the PDS to the NSSDC contain self-describing files, and there are readers and viewers for the labels and data products, It is concluded that the NSSDC will not need to update digital components of the PDS datasets, that is, to migrate to current or standard formats or to transform to persistent form. If the platform for the viewer or Object Access Library changes, it will only be necessary to convert these two resources to the new platform.

A4.5.1 Manage Retrieval Requests

Definition: To register requests, translate them, define request specifications, generate retrieval requests, and account for any problems with requests.
Input: Request for Preserved Records and/or Information about Preserved Records, Reports on Operation of Permanent Preservation Activity
Output: Accounting for Unsatisfied Requests, Retrieval Requests, Reports on Operation of Permanent Preservation Activity, Request Specifications

Discussion: Requestors can query the NSSDC on-line Master Catalog by spacecraft (mission), by discipline category and/or by launch date.[14] Figure 19 shows a query on spacecraft name. It is then possible to retrieve information about experiments (instruments) on the spacecraft and/or datasets available, as shown in Figure 20.

---

[11] http://ftpbrowser.gsfc.nasa.gov.
[12] ftp://cdaweb.gsfc.nasa.gov/pub/istp.
[13] This Activity needs to have a component of the Permanent Preservation System as a control, namely the Preservation Strategy.
[14] http://nssdc.gsfc.nasa.gov/about/about_nmc.html.

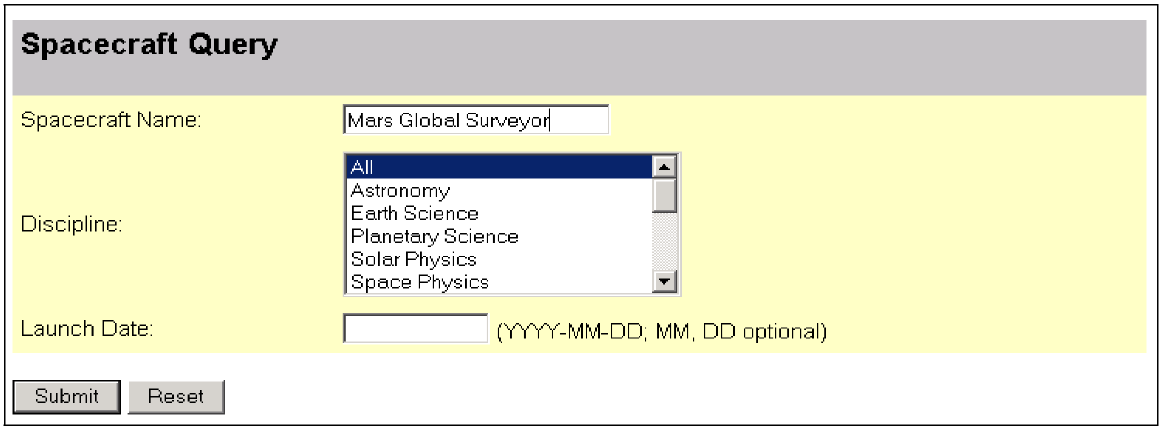


Figure 19. Query Interface to NSSDC Master Catalog

Dataset Name	Spacecraft,Experiment
MGS Pre-Mapping Phase Pilot DVD-ROM Archive	Mars Global Surveyor
Mars Orbiter Camera Standard Data Product Image Archive in Compressed Format	MGS,Mars Orbital Camera (MOC)
MGS Mars Orbiter Camera DSDP Archive (PDS)	MGS,Mars Orbital Camera (MOC)
Mars Global Surveyor Science Sampler CD-ROM (PDS)	MGS,Mars Orbital Camera (MOC) MGS,Thermal Emission Spect (TES) MGS,M.O. Laser Altimeter (MOLA) and 2 others
Mars Orbiter Camera Standard Data Product Image Archive in Compressed Format - Extended Mission	MGS,Mars Orbital Camera (MOC)
MGS Thermal Emission Spectrometer Data Archive (PDS)	MGS,Thermal Emission Spect (TES)
MGS Thermal Emission Spectrometer Data Archive 2 (PDS)	MGS,Thermal Emission Spect (TES)
MGS Mars Orbiter Laser Altimeter (MOLA) Data Archive on CD-ROM (PDS)	MGS,M.O. Laser Altimeter (MOLA)
MOLA Aggregated Experiment Data Record (AEDR)	MGS,M.O. Laser Altimeter (MOLA)
MOLA Experiment Gridded Data Record (MEGDR) Archive	MGS,M.O. Laser Altimeter (MOLA)
MOLA Precision Experiment Data Records (PEDR) Archive	MGS,M.O. Laser Altimeter (MOLA)
Mars Global Surveyor Radio Science Raw Data Archive	MGS,Radio Science (RS)
Mars Global Surveyor Mapping MAG/ER Level 1 Data	MGS,Magnetom/Elect Ref (MAG/ER)
Mars Global Surveyor Premapping MAG/ER Level 1 Data	MGS,Magnetom/Elect Ref (MAG/ER)

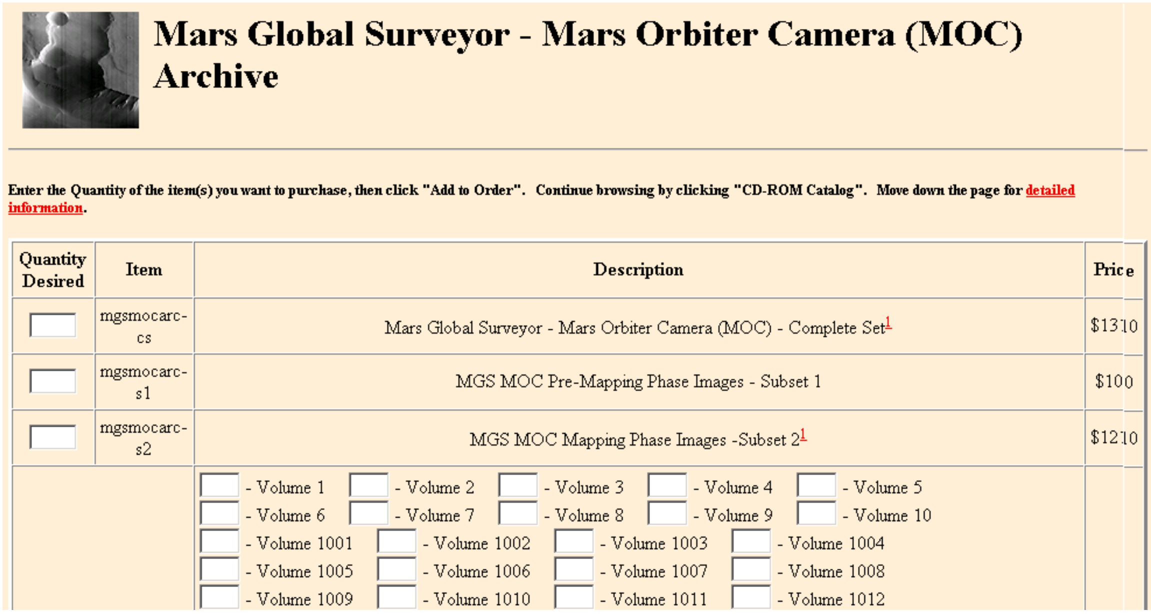
Figure 20. Relevant Datasets and Experiments

When datasets are selected, information is provided about the datasets and an order from is displayed, as shown in Figure 21.

A4.5.2 Review Retrieved Components and Information

Definition: To determine whether all components and information necessary to satisfy requests for preserved records and/or information about preserved records have been received and can be processed for output.
Control: Request Specifications
Input: Retrieved Digital Components, Information about Digital Components of Requested Records, Retrieved Information about Requested Records
Output: Requested Digital Components, Information about Preserved Records Issued to Users, Reports on Operation of Permanent Preservation Activity

Discussion: All the required components of PDS data products are in the requested dataset, so this activity is not required of PDS datasets. However, no information was collected on components and information for other datasets archived by NSSDC.



Mars Global Surveyor - Mars Orbiter Camera (MOC) Archive

Enter the Quantity of the item(s) you want to purchase, then click "Add to Order". Continue browsing by clicking "CD-ROM Catalog". Move down the page for detailed information.

Quantity Desired	Item	Description	Price
	mgsmocarc-cs	Mars Global Surveyor - Mars Orbiter Camera (MOC) - Complete Set[1]	$1310
	mgsmocarc-s1	MGS MOC Pre-Mapping Phase Images - Subset 1	$100
	mgsmocarc-s2	MGS MOC Mapping Phase Images -Subset 2[1]	$1210
		- Volume 1 - Volume 2 - Volume 3 - Volume 4 - Volume 5 - Volume 6 - Volume 7 - Volume 8 - Volume 9 - Volume 10 - Volume 1001 - Volume 1002 - Volume 1003 - Volume 1004 - Volume 1005 - Volume 1006 - Volume 1007 - Volume 1008 - Volume 1009 - Volume 1010 - Volume 1011 - Volume 1012	

Figure 21. Order Form for Dataset Volumes

A4.5.3 Reconstitute Records

Definition: To link or assemble all the components of requested records as necessary to reproduce and output records.
Control: Request Specifications
Input: Requested Digital Components,
Output: Requested Reconstituted Records, Reports on Operation of Permanent Preservation Activity
Discussion: The NSSDC does not provide online access to the data in Planetary Science datasets. However, the NSSDC does provide on-line anonymous FTP access to many scientific data sets. This activity was beyond the scope of the PDS case study.

A4.5.4 Present Records

Definition: To present the requested and reconstituted records with the appropriate extrinsic form, and, if requested, produce a Certificate of Authenticity for the records, or produce reports on problems.
Control: Request Specifications
Input: Requested Reconstituted Records, Retrieved Information about Requested Records
Output: Reproduced Preserved Records Issued to Users, Preserver's Certificates of Authenticity, Reports on Operation of Permanent Preservation Activity

Discussion: The NSSDC does not provide online access to the data in Planetary Science datasets. See discussion of activity 4.5.3.

A4.5.5 Package Outputs

Definition: To combine the digital components with information on how to reproduce the records, and, if the process is unsuccessful, produce a report on any problems.
Control: Request Specifications
Input: Retrieved Information about Requested Records, Requested Digital Components
Output: Reproducible Preserved Records Issued to Users, Reports on Operation of Permanent Preservation Activity

Discussion: The NSSDC provides copies of datasets from the PDS on CDs. NASAView can be used for viewing scientific datasets, PDS Label Library Light for accessing PDS labels and Object Access Library to read the data products in the dataset.

The NSSDC is beginning to store its datasets in Archival Information Packages [CCSDS 2002]. In particular, they have participated in the Consultative Committee on Space Data Systems (CCSDS) Workshops to define an XML Formatted Data Unit (XFDU) packaging standard for scientific data.

4. Conclusions

NASA refers to the Planetary Data System as an active archive. Copies of the scientific datasets are transferred to the National Space Science Data Center (NSSDC) for long-term preservation. The NSSDC is referred to as a deep archives. The PDS activities of data preparation and management of datasets in the PDS Archive are similar to the activities *Managing Records Creation* and *Manage Records in a Recordkeeping System*. The management at the NSSDC of scientific datasets from other space science disciplines appears to be similar to the activity *Select and Preserve Records*. However, the Planetary Data System has been used as an example of the Open Archival Information System (OAIS) Reference Model, which is supposedly a model of a data/record archive that is concerned with the long-term preservation of those data/records [CCSDS 2002].

One of the novel aspects of the Planetary Data System is the choice and implementation of a preservation strategy that obviates the need to convert data products to other file formats. That strategy is to prefix the scientific data records (and supplementary documentation) with labels that describe the context in which the scientific data was collected as well as a description of the format of the data. This label is described in a language called the Object Description Language. There are software tools to interpret the labels, including the description of the structure of the data, and other tools that use this description to read and/or display the data.

An interesting aspect of the walkthrough using MGS/PDS data is that the Description activity seems to take place during record-making rather than as in the COP model after transfer to the organization responsible for long-term preservation. Another interesting aspect of this case study is that parts of appraisal and validation activities take place in the PDS before recordkeeping activities; while in the COP model they take place after recordkeeping. A possible explanation for this is the PDS management decision to actively involve scientists in the archiving process. The scientists who create and use the data products are better able to appraise and describe them than archivists (or scientists) far removed from the mission and data creation. Because of the expense of space science missions, the investment dictates early description, appraisal and validation of the data sets.

Figure 22 summarizes the correspondence of COP activities and MGS/PDS activities as identified in the walkthrough. It also shows COP activities for which case study data was not collected.

Of the sixty-eight lowest-level activities in the COP model, data from the PDS and MGS case study was found to correspond to forty-six of those activities. There is no corresponding data in the PDS case study for seven of the COP activities. No data was collected for 15 of the NSSDC activities that would correspond to the COP long-term preservation activities. The walkthrough of the COP model using PDS and MGS case study data demonstrates that there is an interpretation of the record-making, recordkeeping and some of the preservation activities of the COP model in the domain of archives of scientific data records. That is to say, the COP model is *satisfiable* in this domain. A thorough validation of the model would require walkthrough of the

COP activities for which there are corresponding PDS or MGS activities	COP activities for which there are no corresponding PDS or MGS activities	COP activities for which case study data was not collected
A1.1.1		
A1.1.2		
A1.1.3		
A1.1.4		
A1.2.1.1		
A1.2.1.2		
A1.2.1.3		
A1.2.1.4		
A1.2.2.1		
A1.2.2.2		
A1.2.2.3		
A1.2.2.4		
A1.2.2.5		
A1.2.3.1		
A1.2.3.2		
A1.2.3.3		
		A1.2.3.4
A1.3		
A1.4		
A2.1.1		
A2.1.2		
A2.1.3		
	A2.1.4	
	A2.2.1	
	A2.2.2	
	A2.2.3	
A2.3		
A3.1.1		
A3.1.2		
A3.1.3		
A3.2		
A3.3.1		
A3.3.2		
A3.3.3		
A3.3.4		
A3.4		
		A4.1
A4.2.1		
A4.2.2.1		
	A4.2.2.2.1	
	A4.2.2.2.2	
	A4.2.2.2.3	
A4.2.2.3		
A4.2.3.1		
A4.2.3.2		
A4.2.3.3		
A4.2.4		
A4.2.5		
		A4.3.1
		A4.3.2
A4.3.3		
		A4.3.4
		A4.3.5
		A4.4.1.1
A4.4.1.2		
		A4.4.1.3

COP activities for which there are corresponding PDS or MGS activities	COP activities for which there are no corresponding PDS or MGS activities	COP activities for which case study data was not collected
		A4.4.1.4
A4.4.2.1		
		A4.4.2.2
		A4.4.2.3
		A4.4.2.4
A4.4.2.5		
A4.4.3		
A4.5.1		
		A4.5.2
		A4.5.3
		A4.5.4
A4.5.5		

Figure 22. Correspondences of COP and MGS/PDS Activities

COP preservation activities using case study data with regard to the NSSDC preservation activities and walkthroughs using case study data from other archival domains.

Information was not found for PDS activities corresponding to COP activities for classifying or registering incoming or outgoing scientific data records or other supporting documents. This is because the PDS does not keep records for all mission activities, but only scientific data sets. This does not invalidate the COP model, but just emphasizes that it is a more general model of recordkeeping activities than the activities of scientific data recordkeeping.

Information was not found in the case study corresponding to the three activities in the decomposition of COP activity A4.2.2.2, *Assess Authenticity of Records*. This is not surprising as the criteria and method of assessment of the authenticity of digital records was a new research contribution of InterPARES 1. However, it is demonstrated in another InterPARES 2 report [Underwood and Isbell 2007] that the assessment activity can be carried out with data from the MGS/PDS case study resulting in a high degree of belief in a presumption of authenticity of the records maintained in the PDS.

Technological change in the form of obsolete file formats is not the only threat to the accessibility and intelligibility of e-records. The computer hardware, operating system, database system and programming languages may become obsolete. One of the controls on the COP activity A1.3, *Maintain Framework*, should be something like "Technological change" or "Technology obsolescence." One should capture explicitly in this activity that it would be necessary to migrate the record-making, recordkeeping and preservation system to new hardware, operating systems and software to overcome technological obsolescence. The Preservation strategy of Persistent Object Preservation, which uses abstraction mechanisms for data management and storage systems, is one of the strategies that should be considered to economically and effectively achieve this goal.

It is recommended that the PADS case study be extended to include a study of the transfer of PDS datasets to the NSSDC and preservation activities of the NSSDC with respect to PDS datasets. This would support validation of some of the COP preservation activities for which case study data was not collected.

To more thoroughly validate the COP model, the list of lowest-level activities should be used to collect data from record-making, recordkeeping and preservation systems for other types of digital records. For example, another interesting study would be a walkthrough to the COP Model based on a case study of Mars Global Surveyor Mission/Project digital records other than the scientific data records.

References

[Arvidson 1986] R. Arvidson. Committee on Data Management and Computation, Space Science Board, Assembly of Mathematical and Physical Sciences, National Research Council, *Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences*, National Academy Press, Washington, DC, 1986.

[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

[Berstein 1982] R. Bernstein. Committee on Data Management and Computation, Space Science Board, Assembly of Mathematical and Physical Sciences, National Research Council, *Data Management and Computation - Volume 1: Issues and Recommendations*, National Academy Press, Washington, DC, 1982.

[CCSDS 2002] CCSDS 650.0-B-1. *Reference Model for an Open Archival Information System (OAIS).* Blue Book. Issue 1, January 2002. (IS) 14721:2002).
http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.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. Jet Propulsion Laboratory, *Specification for the Object Description Language*, NASA, Pasadena, CA, 1990. [also published by U. of Colorado Press, Boulder]

[Duranti 2005] L. Duranti (ed.) *The Long-term Preservation of Authentic Digital records: Findings of the InterPARES Project*. Archilab, San Miniato, Italy, 2005.
http://www.interpares.org/book/index.cfm

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

[GAO 1990] U. S. General Accounting Office. "Space Data: NASA's Future Data Volumes Create Formidable Challenge," GAO Report, November 1990.

[Hodge 1997] S. H. Hodge. "Satellite Data and Environmental Law: Technology Ripe for Litigation Application," *Pace Environmental Law Review* 14(1997): 691-718.

[Hughes 1998] J. S. Hughes, "The Planetary Data System A Case Study in the Development and Management of Meta-Data for a Scientific Digital Library," presented at Conference on Research and Advanced Technology for Digital Library. Heraklion, Crete, Greece, Sept 1998.

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

[Hughes et al 2002] J. S. Hughes, S. K. LaVoie, J. Wilf, R. Joyner and D. J. Chrichton. "PDS-D—The Planetary Data Distribution Subsystem," presented at Lunar and Planetary Science Conference XXXIV, 2002.

[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," presented at American Geophysical Union (Fall Meeting), San Francisco, California, 2003.

[InterPARES 2 2005] Modeling Cross-domain. Chain of Preservation - Activity Model, InterPARES 2, May 30 2005.

[InterPARES 2 2004] Modeling Cross-domain, Chain of Preservation - Entity Model, InterPARES 2, June 11, 2004.

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

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

[JPL 2003] Jet Propulsion Laboratory, *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/

[Kieffer 1984] *Planetary Data Workshop. Parts 1 and 2*, NASA Conference Publication 2343, Washington, D.C., 1984.

[King 1998] Joseph 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 (1996). "Overview of the Planetary Data System," *Planetary Space Science,* 44(1): 3-12.

[Moore et al 2000a] R. Moore, C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, W. Schroeder and A. Gupta (2000). "Collection-Based Persistent Digital Archives-Part 1," *D-Lib Magazine* 6(3).

[Moore et al 2000b] R Moore, C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, W. Schroeder and A. Gupta (2000). "Collection-Based Persistent Digital Archives - Part 2," *D-Lib Magazine* 6(4).

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

[NASA 2003] NASA Records Retention Schedules NPR.1D (with change 3 1/31/06) NRRS8 Program Management Records.
http://nodis.gsfc.nasa.gov/npg_img/N_PR_1441_001D_/N_PR_1441_001D_.pdf

[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

[Russell 1988] C. Russell. Committee on Data Management and Computation, Space Science Board, Assembly of Mathematical and Physical Sciences, National Research Council, Executive Summary, National Academy Press, Washington, DC, 1988.

[Sawyer et al 2005] D. Sawyer, H. K. Hills, P. McCaslin and J. Garrett. "Performing a Migration into the Framework of the OAIS Reference Model: NSSDC Case Study," presented at Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data (PV-2005], Edinburg, November 21-23, 2005.

[Underwood 2005] W. Underwood (2005). "InterPARES 2 Project - Case Study 08 Final Report: Mars Global Surveyor Data Records in the Planetary Data System."
http://www.interpares.org/display_file.cfm?doc=ip2_cs08_final_report.pdf

[Underwood and Isbell 2007] W. Underwood and S. Isbell. (2007). "InterPARES 2 Project - General Study 06 Final Report: Testing of the Appraisal Task Force's Method of Assessment with the Benchmark Requirements for a Presumption of Authenticity."
http://www.interpares.org/display_file.cfm?doc=ip2_gs06_final_report.pdf.

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

Appendix A: Dataset Description (dataset.cat) for MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0

```
PDS_VERSION_ID                    = PDS3
RECORD_TYPE                       = STREAM
SPACECRAFT_NAME                   = "MARS GLOBAL SURVEYOR"
TARGET_NAME                       = MARS
OBJECT                            = DATA_SET
  DATA_SET_ID                     = "MGS-M-MOC-NA/WA-2-DSDP-L0-V1.0"

  OBJECT                          = DATA_SET_INFORMATION
    DATA_SET_NAME                 = "MOC DSDP ARCHIVE"
    DATA_SET_COLLECTION_MEMBER_FLG = "N"
    START_TIME                    = 1997-09-15
    STOP_TIME                     = 1997-11-25
    DATA_SET_RELEASE_DATE         = 1999-01-15
    PRODUCER_FULL_NAME            = "MALIN SPACE SCIENCE SYSTEMS"
    DETAILED_CATALOG_FLAG         = "N"
    DATA_OBJECT_TYPE              = "IMAGE"
    DATA_SET_DESC                 = "

Dataset Overview
=================

This CD 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.

This CD contains also ancillary data files and browse images in a JPEG
format, HTML documents that support a web browser interface to the
CDs, an index file ('imgindx.tab') that tabulates the contents of the
CD, and documentation files.

For more information on the contents and organization of the CD volume
set refer to the 'CD CONTENTS, DIRECTORY, AND FILE NAMING CONVENTIONS
' section of the aareadme.txt file located in the root directory of
the data volumes.

Using a web browser, open the 'index.htm' file located in the 'root'
directory of the CD. The HTML document will direct you to other
informational documents and the image browser for rapidly viewing the
image collection.

Parameters
==========

Although this dataset has not been calibrated, and the algorithms for
calibration are still being developed, we here describe some of the
relevant calibration parameters.

The MOC uses programmable gain and offset states, commanded on the
ground prior to image acquisition, to condition the CCD output signal
prior to its digitization to 8 bits.  The very wide potential dynamic
```

range of MOC images has required a large number of gain states (16 for the NA and 20 for the WA) and offset states (256 possible) compared to, for example, the Viking cameras, which had only two gain and two offset states. This leads to the operational complexity of predicting the scene brightness in advance and selecting appropriate parameters.

The GAIN_MODE_ID and OFFSET_MODE_ID fields in the image headers describe the gain/offset selection. The GAIN_MODE_ID is a two-digit hexadecimal number which is the value of the MOC hardware register that selects the gain. The allowable flight values are

Narrow Angle

gain	hex	gain	hex
1	F2	7.968	EA
1.465	D2	11.673	CA
2.076	B2	16.542	AA
2.935	92	23.386	8A
4.150	72	33.067	6A
5.866	52	46.740	4A
8.292	32	66.071	2A
11.73	12	93.465	0A

Wide Angle

gain	hex	gain	hex
1.000	9A	16.030	96
1.412	8A	22.634	86
2.002	7A	32.092	76
2.832	6A	45.397	66
4.006	5A	64.216	56
5.666	4A	90.826	46
8.014	3A	128.464	36
11.34	2A	181.780	26
16.03	1A	256.961	16
22.67	0A	363.400	06

where the gain value given is the nomimal multiplicative factor from the lowest gain state.

The OFFSET_MODE_ID is the value of the MOC hardware register that selects the offset. Offsets are commanded in units of 5 (five) Data Numbers (DN), so an OFFSET_MODE_ID of '1' would correspond to a DN offset of 5. All offsets are positive.

The simplified MOC response equation (without pixel-to-pixel variation terms) is as follows:

```
dn = a*(r*ex+dc*ex+g)+(z-off)
```

where r is the average signal level being generated at the focal plane (in DN/msec at minimum gain), z is the fixed zero offset, off is the commanded variable offset in DN (note that the offset is subtracted), dc is the dark-current term (in DN/msec at minimum gain), g is the gain-dependent offset (in DN at minimum gain), a is the system gain (where minimum gain is 1 and all other gains are >1, as given in the above tables), and ex is the exposure time (given in the image headers as the LINE_EXPOSURE_DURATION.)

In-flight values for the fixed parameters in the above equation are still being derived from flight data. The values from ground testing at ambient conditions are

system	z	dc	g
NA prime	25.5767	-0.0529099	0.381963
NA spare	28.934	-0.0099495	0.371922
WA red	27.5633	0.0013369	0.196468
WA blue	27.9424	0.0008232	0.264303

The significance of the negative dark-current terms for the NA systems is suspected to be due to other system noise sources in ground testing; the NA systems should have negligible dark current, even at room temperature, because of the short exposure times.

The calibration algorithm will consist of two independent parts: removal of the pixel-to-pixel variation, which causes the visually apparent ‘streaking’ in the downtrack direction in MOC images, and conversion to either relative or absolute flux units (for purposes of mosaic construction, photometry, etc.) Work is ongoing to define these algorithms. Future volumes will include more information.

Processing
==========

Processing included packet decommutation, removal of the MOC communications protocol headers, and decompression. No additional geometric or radiometric processing was done.

For most of the pre-mapping phase of the MGS mission, data quality did not allow error-free transmission of the instrument data to Earth. The MOC protocols (in particular, the formats for compressed image data) were designed for the bit error rates expected in mapping. As a result, considerable data losses were incurred in the image data. The majority of processing for pre- mapping data was done to minimize the effects of this data loss. These efforts are ongoing; corrections for significant losses may appear on future volumes.

MOC image data are broken up into ‘packets’ of approximately 1000 bytes. A typical data loss is that of one or two packets, due to uncorrectable bit errors caused by noise in the space-to-Earth communications path, momentary loss of receiver lock caused by a transition between the one-way and two-way tracking modes, or loss in the Earth segment of the Deep Space Network.

For uncompressed images, a packet loss leads to loss of ‘line sync’ in the image. Since the amount of actual image data in a packet is variable and cannot be determined precisely without the packet, such errors must be corrected by hand. This has been done for as many images as practical. The majority of NA images were acquired using the lossless predictive compression mode of the MOC. However, when a packet is lost from this compressed data stream, the decompression algorithm cannot realign itself to the compressed pixel boundaries, and must skip ahead to the next sync marker, which occurs only every

128 lines in the image. The effect of decompressing the data between the site of packet loss and the next sync marker is unpredictable, but usually results in either semi-random variations in pixel brightness (with the general morphology of the original image still visible) or essentially random noise patterns.

A second type of loss is that of tens or hundreds of packets caused by bad weather, hardware failure, or operator error at the DSN stations, or miscommanding of the telemetry playback on the spacecraft. For these errors in a compressed data stream, over 128 lines of the image were lost, making it impossible to recover even the original downtrack size of the image. Such images are described as 'PARTIAL' in the NOTE field of each image header.

The browse images were subsampled via averaging and then auto-ends stretched to create visually acceptable contrast. No other processing was performed. Subsampling was intended to produce an image of an approximately fixed size, so the subsampling employed varied depending on the original image's dimensions.

Media/Format
============

The MOC DSDP archive is delivered to the Planetary Data System using CD media. Formats are based on standards for such products established by the Planetary Data System (PDS) [PDSSR1992]."

CONFIDENCE_LEVEL_NOTE = "

Geometric Accuracy

Latitude and longitude coordinates for the images given in the imgindx.tab file were computed using the best-available spacecraft position and orientation information, in the form of SPK and CK kernel files for the NAIF SPICELIB software. The versions used were recommended by the MGS Project and were retrieved from the NAIF FTP server (naif.jpl.nasa.gov):

mgs_ab1.bsp: Mars Global Surveyor Aerobraking-1 SPK file, MGSNAV Solution, Created by Boris Semenov, NAIF/JPL, October 2, 1998

mgs_spo.bsp: Created 1998-09-26/12:50:00.00.

mgs_spo2_gsfc.bsp: Mars Global Surveyor SPO-2 SPK file, GSFC Solution, Created by Boris Semenov, NAIF/JPL, October 2, 1998

mgs_sc_ab1.bc: Created by Boris Semenov, NAIF/JPL November 29, 1998

mgs_sc_spo1.bc: Created by Boris Semenov, NAIF/JPL November 29, 1998

mgs_sc_spo2.bc: Created by Boris Semenov, NAIF/JPL November 29, 1998

de403s.bsp: Dated 14-NOV-1995, Created 1995-06-01/12:14:42.00.

Latitude is given in areographic form using the IAU 1994 definition of

```
the Martian equatorial and polar radii (3397.0 and 3375.0 km,
respectively).  Coordinates are computed using the 1994 IAU spin
vector values.

Because of uncertainty in the MOC-to-S/C frame offset and limitations
of the processing software, the MOC offset ('I kernel') was not
applied; this should make a difference no more than 1/2 MOC NA FOV,
probably less.

It has been observed by MSSS that the USGS MDIM images were
constructed based upon a definition of Mars' orientation from the
Viking period.  It can be shown that this results in a systematic
shift between the 'old' and 'new' systems of 0.213 degrees in
longitude.  To place an image footprint onto the MDIM, one should
subtract 0.213 degrees from the longitudes tabulated on this data
volume.  Any residual error in the location of the image is caused by
further uncertainties in the MDIM and/or in the position and
orientation information of the MGS spacecraft.  Obviously, the best
available SPICE information should be used for geometric calculations.

In cases where only a portion of the lines of the image were actually
recovered on the ground due to the data loss described above, the
lat/lon coordinates given in the table are those of the center and
corners of the image as received, with the caveat that in rare
instances, lines may have been lost from the top of the image.  In
such cases, the start time of the image is that commanded, not the
actual line time of the first line of received data, and it is not
possible to determine what the true footprint of the image is, without
matching features seen in the image to preexisting image data.

In a few cases, spacecraft pointing information was not available for
an image.  In these cases, a nominal nadir pointing attitude has been
assumed.  This may lead to large errors in the footprint information,
which should be considered advisory only.

Map Projections of Images
-------------------------

High-precision map projections of the images may be generated using
the parameters given in the image header and/or the imgindx.tab file,
the appropriate SPICE kernels, and map-projection software capable of
processing line-scan imagery.

Lacking such software, however, a first-order map projection may be
produced by using the lat/lon coordinates of the image corners given
in the imgindx.tab file, transforming these four points from
rectangular image space to the essentially arbitrary quadrilateral in
map-projection space using the desired map-projection equations, and
then performing a four-point bilinear warp.  Such a warp can be done
in commercial packages such as Photoshop, as well as software
specifically for planetary image analysis (PICS, ISIS, VICAR, etc.)

Users wishing simply to correct for the effects of imaging flipping,
non-square pixel aspect ratio and image skew may also find the
USAGE_NOTE, PIXEL_ASPECT_RATIO and IMAGE_SKEW_ANGLE fields in the
imgindx.tab file useful.  The USAGE_NOTE indicates if the image should
be flipped left-for-right prior to additional processing.  If
```

```
IMAGE_SKEW_ANGLE is not too far from 90 degrees, the image can be
rectified to square-pixel form by expanding it in the vertical axis by
a factor of PIXEL_ASPECT_RATIO (noting that values less than 1 result
in shrinking rather than expansion.)  Skew angles far from 90 degrees
can be corrected by skewing the image from a rectangle to a rhomboid
with a base angle of the given skew angle."

END_OBJECT                       = DATA_SET_INFORMATION

OBJECT                           = DATA_SET_TARGET
  TARGET_NAME                    = MARS
END_OBJECT                       = DATA_SET_TARGET

OBJECT                           = DATA_SET_HOST
  INSTRUMENT_HOST_ID             = MGS
  INSTRUMENT_ID                  = MOC
END_OBJECT                       = DATA_SET_HOST

END_OBJECT                       = DATA_SET

END
```