

MARS OBSERVER DATA PRODUCTION, TRANSFER, AND ARCHIVAL:  
THE DATA PRODUCTION ASSEMBLY LINE

David B. Childs

N 9 4 - 2 3 9 5 3

Jet Propulsion Laboratory  
Pasadena, California, U.S.A.

## ABSTRACT

This paper describes the data production, transfer, and archival process designed for the Mars Observer Flight Project. It addresses the developmental and operational aspects of the archive collection production process. The developmental aspects cover the design and packaging of data products for archival and distribution to the planetary community. Also discussed is the design and development of a data transfer and volume production process capable of handling the large throughput and complexity of the Mars Observer data products. The operational aspects cover the main functions of the process: creating data and engineering products, collecting the data products and ancillary products in a central repository, producing archive volumes, validating volumes, archiving, and distributing the data to the planetary community.

## 1. INTRODUCTION

### 1.1 Purpose

The purpose of this paper is to describe a proposed archive collection production process for the Mars Observer Mission. The Mars Observer archive collections are logical assemblages of science data products and their associated meta data, ancillary data, documentation, catalog information, and software that were identified for permanent archive and distribution to the general planetary science community. The contents of the archive collections were agreed upon by

Mars Observer and the Planetary Data System [Ref. 1-2]. The Planetary Data system acts as the contact between Mars Observer and the National Space Science Data System which serves as the permanent archive for planetary science data in NASA. Mars Observer is the first mission to be fully integrated with an operational Planetary Data System.

The archive collection production process includes the extraction of raw and reduced data products and their associated ancillary products from a centralized Project Database, repackaging the data products in some cases, organizing the products into archive volumes, premastering CD-ROMs, validating archive volumes, and coordinating the mastering and release of CD-ROM volumes. The proposed functional design of the archive collection production process was prepared by the Science Integrated Data System Task at the Jet Propulsion Laboratory (JPL). This team is a multiorganizational design team with representatives from the Mars Observer Project, the JPL Multi-mission Operations Systems Office, the Planetary Data System, and the JPL Standards Group. The functional design will be presented to the four organizations in December, 1992. Prototyping and detailed design began in October, 1992 by the Mars Observer Data Administration Team and the Science Data Validation Team.

### 1.2 Summary of the Mars Observer Mission

The Mars Observer Spacecraft was launched

in September of 1992. It will be placed into orbit around Mars in August, 1993. After a four month orbit insertion phase, the instruments on the spacecraft will begin mapping the atmosphere, surface, and interior for at least one martian year (687 Earth days). The spacecraft has seven instruments: the Gamma Ray Spectrometer, the Magnetometer/Electron Reflectometer, the Mars Observer Camera, the Mars Observer Laser Altimeter, the Pressure Modulator Infrared Radiometer, the Thermal Emission Spectrometer, and Radio Science. The objective is to produce a suite of data products that depict atmospheric, surface, and subsurface characteristics as a function of latitude, longitude, altitude, and season.

The Mars Observer ground data system includes a centralized mission operations component at JPL and a distributed component located at the home institutions of the Principal Investigators, Team Leaders, Interdisciplinary Scientists, Team Members, Co-Investigators, and Participating Scientists. Science Operations Planning Computer (SOPC) workstations will be electronically connected via NASCOM links to a centralized Project Database Personnel at the central site and those with remote workstations will use the Project Database to coordinate observation planning, deposition of raw science and engineering data, acquisition of data, deposition of reduced data and documentation, and extraction of reduced data for analysis.

The collective science teams and mission operations teams on Mars Observer will produce approximately 80 unique data products totalling to about 370 GB of data. The data and ancillary products will be organized into 15 archive collections: one for each of the six instruments and the Radio Science Experiment, one for each of the six Interdisciplinary Scientists, one for the

SPICE data (geometry data), and one for the engineering data. The archive collections will be stored on approximately 800 CD-ROM volumes.

The overall archive collection flow production is shown in Figure 1. Mars Observer is responsible for the production and validation of the archive collections and archive volumes (CD-ROMs). After a generation/validation period of six months (since the creation of the data product) the CD-ROMs are released to the Planetary Data System for distribution to the NASA sponsored planetary community and permanent archival. The National Space Science Data Center serves as the permanent archive and also distributes to the general science community. The archive collection production process includes only the tail end of downlink portion of the Mars Observer (MO) mission operations flow. The included functions are shown in Figure 2.

## 2. WHAT IS AN ARCHIVE COLLECTION

An archive collection is a logical assemblage of related data products and all the information necessary to access, understand, and use the data (fifteen archive collections for MO). Archive collections are organized into data sets (approximately 80 for MO) and data sets are groups of data product instances. Each data product contains the actual values for the measured parameters for one observation (for example, an image) and its associated meta data. The meta data identifies the observation, defines its properties, and defines the structure and format of the data values. Data sets may include unique versions of geometry data or calibration data. A data set may also include documentation catalog information, and software.

Documentation may include flight project documents, instrument papers, science

articles, or any other textual material deemed necessary to understand and use the data sets or software. Catalog information consists of completed, predefined catalog loading templates or outlines for describing missions, spacecraft, instruments, data sets, and personnel. Software includes executables, source code, installation procedures, user guides, and design documents. SPICE Kernels (or geometry data) consists of software and data files that together characterize the critical features of the Spacecraft ephemeris, Planet ephemeris or other target, Instrument characteristics, C-matrix/scan platform orientation, and Events associated with data sets. Calibration data is any data used in the calibration of the instrument which is necessary to understand or use the data sets or software. Finally, in some cases, product specifications such as Science Data Product Software Interface Specifications (SISs) or Archive Collection SISs may be included.

The data sets are packaged onto physical volumes such as CD-ROM, CD-WO (write once), or magnetic tape for archival and distribution. Most data sets require several volumes. Volumes are grouped into volume sets. Each volume and volume set has a unique identifier. Each volume contains one of several standardized directory structures, required files such as AAREADME.TXT and one containing a volume description, and a tabular index of the data products on the volume. Multivolume sets contain cumulative indices. The organization of data sets and volumes is standardized and described in the PDS Data Preparation Workbook, Volumes I and II [Ref. 3-4].

The organization and expression of the meta data for the data objects is standardized. The meta data identifies the observation, describes its characteristics, associates it with other objects, and gives the internal

organization and format of the data values. The meta data is contained in a catalog object. It is expressed in a "keyword equals value" notation (for example, TARGET = MARS) called the Object Definition Language [PDS 92]. The keywords and standard values are defined in the Planetary Science Data Dictionary [Ref. 5]. The keywords describing the internal organization of data objects are aggregated into structure objects (such as IMAGE, TABLE, QUBE). A wide variety of structure objects are standardized for use in the meta data for primary and secondary data objects.

### 3. OBJECTIVES OF THE MO ARCHIVE COLLECTION PRODUCTION PROCESS

The Mars Observer 'archive collection production system has the following objectives.

- (1.) Produce archive volumes containing quality data with the meta data and ancillary products necessary to access, understand, and use data. The volumes must be on a permanent, high density, and distributable media. The products and volumes must adhere to Planetary Data System standards.
- (2.) Produce all the archive volumes before the end of the mission (or the accounts close). Restoration is very expensive.
- (3.) Be cost effective, automated, and reliable. Development and operations must be accomplished within a fixed budget. The process must be automated where possible to accommodate the high throughput for Assemble Volumes with some products and validate Volumes for all volumes.

#### 4. MARS OBSERVER ARCHIVE COLLECTION PRODUCTION PROCESS

The high level functions of the MO archive collection production process are shown in Figure 3. These functions span from the creation of the bested telemetry packets through to use of the archive volumes by the general planetary community. Several of the functions are existing subsystems in the MO Ground Data System and are provided by the Multimission Operations Systems Office (MOSO). The archive production process was designed to minimize the impact on these existing functions. The archival and distribution function is performed by the operational Planetary Data System. The use of archive volumes is performed by the general planetary community. While these functions are not a part of the proposed functional design they are included in the description of the archive process because they are the interfaces to the proposed functions, the proposed functions require minor modifications to the existing functions, and they are required to understand the overall archive process flow.

The remaining functions, the package, assemble, validate, and coordinate release functions were designed recently for completion of the archive collection production process and to fully integrate the mission data output with the archive system input. The following sections describe the high-level archive collection production process (See Figure 3).

##### 4.1 Collect Telemetry and Create Level 0 Engineering and Science Products

Telemetry packets from the Deep Space Network arrive at the Collect Telemetry and Create Level 0 Engineering and Science Products function. This function is an abstraction of the existing front end of the

downlink of the ground data system. This function performs Reed/Solomon decoding, frame synchronization, extracting packets from frames, prepares channelized engineering data, performs besting of the data, and extracts anomaly records. The telemetry packets are stored on the centralized Project Database. The output of this function is the final or bested science and engineering telemetry packets (Level 0 products). This function is performed by the Mission Operations System Teams.

##### 4.2 Create Science Data and Ancillary Products

The telemetry packets are extracted from the Project database by the Principal Investigator Teams and transferred to their remote sites in the Create Science Data and Ancillary Products function. Then the data is decompressed, reconstructed, and reformatted. Various error corrections, calibrations, and science processing algorithms are applied to produce reduced science data products (Level 1+). This function is an abstraction of the science investigations performed by the Principal Investigator and Interdisciplinary Science Teams. The associated meta data is packaged with the data products and it is wrapped with Standard Formatted Data Unit (SFDU) labels. The data product is then ready to be transferred to the Project Database.

##### 4.3 Package Level 0 Engineering and Science Data Products

The Package Level 0 Engineering and Science Data Products systematically extracts the final level 0 telemetry (science and engineering) from the Project Database as the database nears its capacity and packages it for off-line storage and archival. Since the telemetry packets are stored in a DBMS, they must be extracted with queries based on a

predetermined extraction strategy (e.g. one orbit's worth or one day's worth). Each group of packets is written to a file, the associated meta data for the group is included, and the product is packaged into an SFDU. Other ancillary products and archive volume components such as documentation and volume indices for the Level 0 archive volumes are prepared here. The output of this function is a logical archive volume containing level 0 engineering and science data products ready for premastering onto CD-WO. This function is one of the new, proposed functions and would be performed by the MO Data Administration Team.

#### 4.4 Collect Data and Ancillary Products

The Collect Data and Ancillary Products function receives completed, reduced, science data products from the science teams, performs a cursory check of the packaging and meta data, catalogs the files based on values in the meta data, and stores them in a file management system (i.e. loads them into the Project Database). This function also accepts queries for the retrieval of data products. The Project Database is managed and operated by the MO Data Administration Team.

#### 4.5 Assemble Volumes

The Assemble Volumes function creates a manifest for each archive volume based on the volume specifications in the Archive Collection Software Interface Specifications (SIS). The manifests are expressed in keyword equals value notation and are human and machine readable. Using the manifest, the specified data products and ancillary products are extracted from the Project Database. The extraction step is verified and the manifest updated (some desired products may not be ready). Several validation checks are performed to ensure

that the products conform to standards. Some products may need to be reformatted because of chronic errors or known design compromises. Reformatting is performed based on instructions in the manifest. Some ancillary products may need to be repackaged because they were stored on the Project Database as a single aggregate unit (e.g. software source code). Then, the volume specific components are completed such as the AAREADME.TXT file, the volume description, and the volume indices. Next, a logical volume is created and used to premaster a CD-ROM volume image. The volume image is written to CD-WO. The capability to write the image to 8mm tape is also provided. Finally, the media is checked to verify the premastering steps. The level 0 engineering and science products enter this function as a completed logical volume and only go through the premastering and verification steps. This is one of the new, proposed functions and would be performed by the MO Data Administration Team.

#### 4.6 Validate Volumes

All of the completed archive volumes are sent to the Validate Volumes function for final validation before release. In this function, the file and subdirectory contents of the volumes are compared with the manifest and the volume index. The meta data in the data products are compared to the design of the data products in the Science Data Product SIS and the Planetary Science Data Dictionary [Ref. 5]. The volume organizations are compared to the designs of the volumes in the Archive Collection SISs. This function does NOT systematically check the quality of the science data. However, statistical sampling may be used to verify the correlation of the meta data with the data products. Also, a final error analysis is performed using the cumulative error log which accompanies each volume. The error logs are expressed in

keyword equals value notation and are human and machine readable. The results of the error analysis determine if the volume is released or returned for rework. A validation report is prepared. The error analysis also determines where in the process to return the volume for rework. This function may also perform minor error corrections and mastering of volumes. This is one of the new, proposed functions and would be performed by the Science Data Validation Team.

#### 4.7 Coordinate Release of Volumes

The Coordinate Release of Volumes function receives all of the completed, validated archive volumes and their validation reports. In this function, the completed volume and the validation report are reviewed by the MO Data Archive Working Group to determine if the volume should be released, checking that the six month generation/validation period has passed (MO policy), and obtaining approval from the Principal Investigator Team representative. If the volume is not ready for release, it is held for rework or pending further negotiation. When the volume is ready for release it is either duplicated in the Assemble Volumes function or sent to a CD-ROM mastering vendor. If less than fifteen copies are required, it is more cost effective to produce CD-WO copies than to master and duplicate. If mastering is required, the CD-ROM artwork, packaging artwork, distribution lists, and the purchase order are prepared for the mastering vendor. A copy of the mastered volume may be requested for verification before duplication. Upon completion of the mastering and duplication, any copies received by MO are distributed internally. The Planetary Data System and the National Space Science Data Center receive fifty to several hundred copies directly from the mastering vendor. The CD-WO copies produced internally are packaged

manually and the appropriate number of copies are distributed to the Planetary Data System directly. This is one of the new, proposed functions would be coordinated by the Science Data Validation Team.

#### 4.8 Archive and Distribute Volume

The Archive and Distribute Volumes functions are performed by the Planetary Data System (PDS) and the National Space Science Data Center (NSSDC). The PDS and the NSSDC receive copies of the final archive volumes either from the mastering vendor for large duplication orders or from MO directly for limited copies of CD-WO volumes. The PDS operates interactive catalogs of the archive volumes and the contained data sets. In some cases, the PDS has catalogs which inventory the archive volumes down to the individual data product level. The PDS accepts orders and distributes to the NASA funded planetary science community. Copies of the volumes are archived in the PDS local archives at its Discipline Nodes. The PDS scientists representing the six planetary science disciplines and the navigation area provide expert help in finding and using archive collections. Special processing of data sets is provided when requested. The PDS maintains standards for the contents and organization of archive collections [PDS 92] and provides (archive) data engineering support to missions. The NSSDC serves as the permanent archive for the planetary science archive collections. It also accepts orders and distributes to the remainder of the NASA funded scientists and the general public.

#### 4.9 Use Archive Volumes

The general planetary science community requests data sets, data products, expert help, and special processing to support scientific

investigations. The community also provides feedback on the accessibility, usability, and quality of the archive collections. Over time, the feedback is formalized into refinements to standards for preparing archive collections.

## 5. CONCLUSIONS

The development of the MO archive collection production process is currently in the detailed design phase for the package level 0 products, assemble volumes, validate volumes, and coordinate release functions. The remaining functions in the described archive process already exist. One of the key development objectives is to minimize the impact on the existing functions. The complete functional specification of the four proposed functions is contained in the "Mars Observer Archive Collection Production Specification" document [Ref. 6] prepared by the Science Integrated Data System Team. The functional specification will be reviewed by Mars Observer, the Multimission Operations Systems Office, and the Planetary Data System in December 1992. The Principal Investigator Teams and the Data Archive Working Group are in the process of designing the data products and CD-ROM volumes. Prototypes of the level 0 engineering archive volumes (CD-WO) were produced recently using the Package Level 0 Engineering and Science Data Products function and portions of the Assemble Volumes function.

The development process has provided a testbed for several archive production "technologies". These are:

- (1.) The approach for writing Archive Collection SISs which contain the contents and organization specifications for over 800 unique CD-ROM or CD-WO volumes.

- (2.) The means of expressing the volume specifications in machine and human readable form as manifests. The manifests will support automated extraction of products from the Project Database, automated volume assembly, and automated volume validation.

- (3.) The approach for describing and tracking errors in a machine and human readable form throughout the production process.

- (4.) The approach for scheduling and tracking volume assembly, volume validation, and release throughout the production process.

- (5.) An approach for integrating several disjoint data production) and validation tool sets into an efficient, automated assembly and validation process.

The Science Integrated Data System Team proved to be very successful as a multiorganizational system engineering and design team. Engineers representing Mars Observer, the Multimission Operations Systems Office, the Planetary Data System, and the JPL Standards Group worked together effectively to define the archive collection production process and bridge the gap between the MO ground data system and the Planetary Data System. Implementing the proposed archive production functions will ensure that the Mars Observer science and engineering products will be archived and distributed to the general planetary community in a timely manner. In addition, the enormous costs of data restoration can be avoided for the 370 GB of MO data to be archived.

## 6. REFERENCES

1. Mars Observer Project Data Management Plan, 642-33, October 1991, JPL D-6615.
3. Planetary Data System Data Preparation Workbook, Volume 1, Procedures, November 1992, JPL D-7669.
2. Mars Observer Archive Policy and Data Transfer Plan, 642-447, March 26, 1992.
4. Planetary Data System Data Preparation Workbook, Volume 2, Standards, November 1992, JPL 3-7669.
5. Planetary Science Data Dictionary Document, November 1992, PDS Version 2.0, Rev. B, JPL D-7116.
6. Mars Observer Archive Collection Production Specification, December 1992.

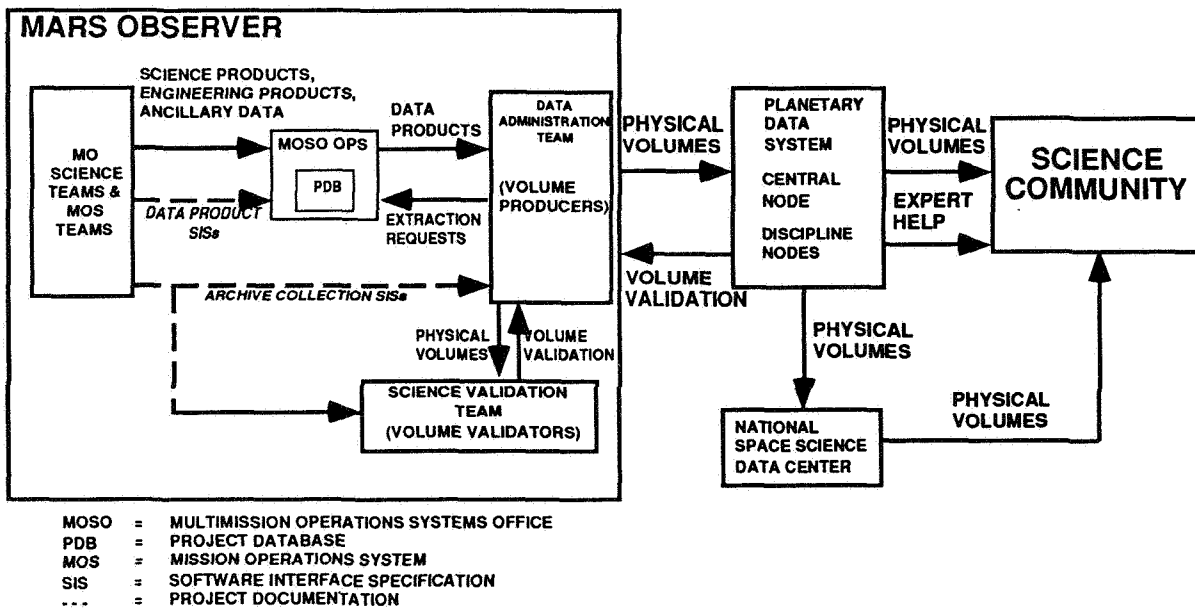
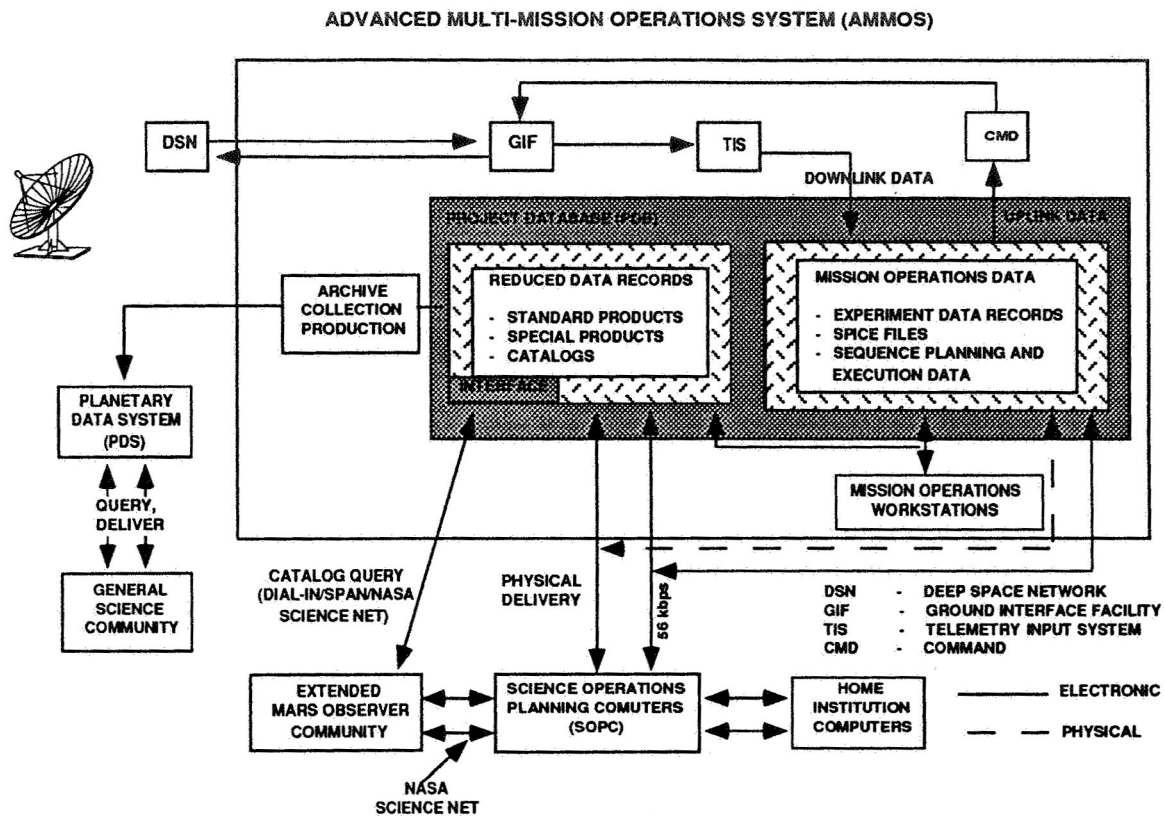
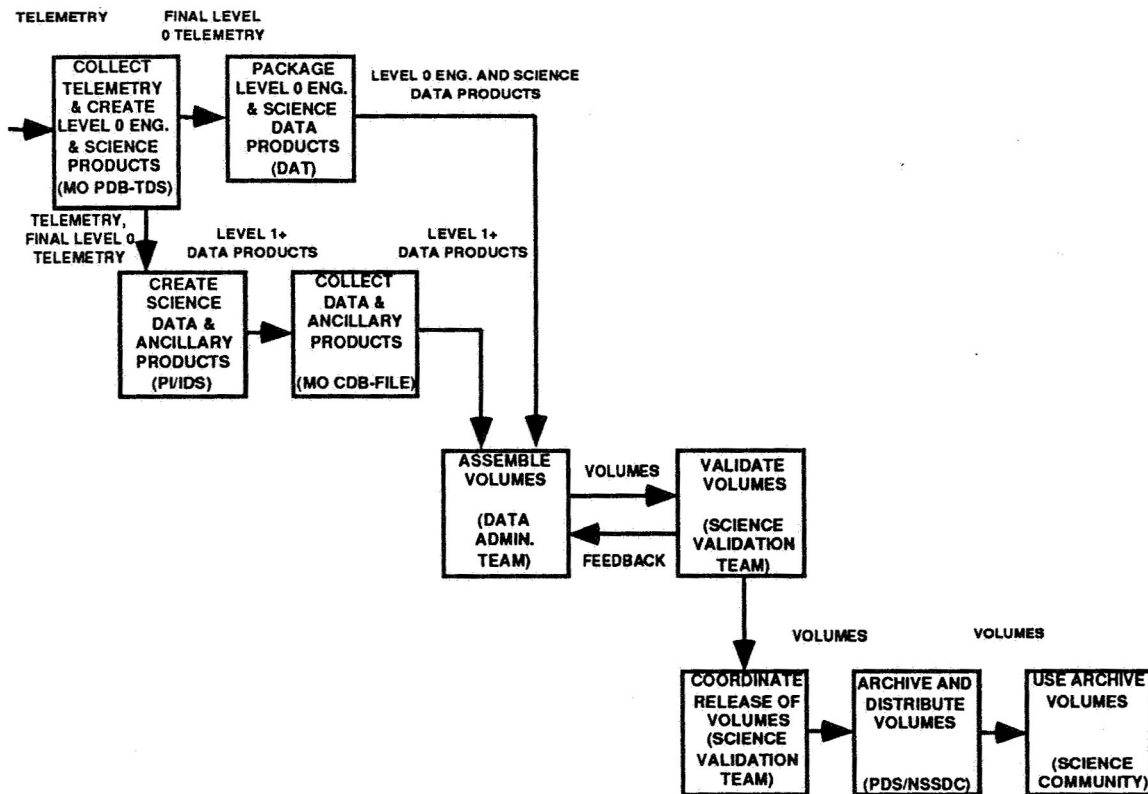


Figure 1. MO Archive Collection and Volume Flow





*Figure 2. Mars Observer Ground Data System*



*Figure 3. MO Archive Collection Production Process*