InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Overview

Case Study 08:
Mars Global Surveyor Data Records in the Planetary Data System

Peter Gagné, Université Laval

May 2006

The Creator Context / Activity

Creator: NASA (National Aeronautics and Space Administration)

Creator type: Scientific focus / Public sphere (agency)

Juridical context: Established by the National Aeronautics and Space Act of 1958 to carry out aeronautical and space activities of the United States. There also exist NASA regulations published under the United States Code of Federal Regulations, Title 14 – Aeronautics and Space, Chapter V, NAA, Parts 1200-1299.

Activity: As part of the NASA mandate to carry out the aeronautical and space activities of the United States, the creator must support the operation of the Mars Global Surveyor (MGS) mission and the maintenance of the Planetary Data System (PDS). The Mars Global Surveyor spacecraft was launched in 1996 to obtain global, synoptic views of the Martian surface and atmosphere in order to study meteorological, climatological and related surface changes.

The digital scientific data records originate from the MGS spacecraft in the form of instrument measurements (observations) of a target. The Command and Data Handling Subsystem records science and engineering data for later transmission or transmits this data in real-time through the Deep Space Network to Mission Ground Control where it is accumulated in a Project Data Base. The process of MGS archive generation, validation and transfer then occurs. In addition to the analysis of the data by the project's science team, the scientific data is incorporated into the Planetary Data System, which acts as a record-keeping system incorporating peer-reviewed scientific data. The PDS archive distributes scientific data from NASA planetary missions.

Specific activities include:

- Generating standard data products, documentation and index tables by the Mars Global Surveyor and Mars Orbital Camera team
- Generating archive volumes
- Validating archive volumes by the Science Data Validation Team and Planetary Data System

- Transferring archive volumes to the Planetary Data System
- Enabling access to the Planetary Data System by the planetary science community
- Transferring archive volumes to the NSSDC deep archives

Although the creator is a government agency, it is not doing (traditional) e-government activities. The activity is actually in the scientific, not governmental focus of InterPARES, since it is operating in the scientific field. So although the actor is governmental, the activity is scientific. In addition, the activity studied deals specifically with the experimentation of techniques to ensure that electronic records remain accessible by making them self-describing and independent of specific hardware and software – persistent object preservation.

Nature of Partnership

Although on the surface this case study does not seem to be a partnership, there are some aspects of partnership involved in it. First, the Mars Global Surveyor mission and the Planetary Data System are two distinct entities within NASA, and thus any sharing or collaboration between the two can be seen as a partnership between two separate divisions within the same agency.

Secondly, as the bureaucratic/organizational structure below shows, the project includes a "distributed component" among various institutions and the collaboration of principal investigators, co-investigators, "interdisciplinary" and "participating" scientists all contributing to the project and linked via a computer network. The final report uses the term "Mars Global Surveyor science community" (FR 23) and describes the following roles:

- Principal Investigators: responsible for the acquisition, reduction, and analysis of data from their instruments and for the generation of standard data products and archive volumes for their instruments.
- Interdisciplinary Scientists: enhance science analyses by participating in reducing data for specific instruments and/or by conducting analyses involving multi-instrument data sets to address specific science objectives.
- Co-Investigators: role not presented in the final report. It is assumed that they assist the principal investigators and may come from related disciplines.
- Participating Scientists: fulfill Co-Investigator, Team Member, or Interdisciplinary Scientist functions.
- Radio Science Team: utilizes both spacecraft and ground equipment to conduct experiments.

Lastly, the mission of the Planetary Data System is to distribute scientific data from NASA planetary missions to ensure the long-term usability of NASA data and to stimulate advanced research. As such, it must enter into collaborations, sharing or other sorts of partnerships with other scientific organizations and entities.

Bureaucratic/Organizational Structure

NASA has a large, bureaucratic administration similar to the governmental level to which it is related. The agency's headquarters is located in Washington, DC and falls under the leadership of a head Administrator as well as a Deputy Administrator who acts as the chief operating officer. NASA also has ten field agencies in addition to a number of installations conducting work in laboratories, on airfields, in wind tunnels and in control rooms. There is a large staff (e.g.,

officers, scientists, engineers etc.) working in a variety of offices under the divisions Mission and Mission Support. NASA is government funded.

The Mars Exploration Program is under the responsibility of the director of the Jet Propulsion Laboratory (JPL) and the Mars Exploration Program Office, one of two main offices of JPL. The Mars Global Surveyor ground data system includes a centralized mission operations component at JPL and a distributed component located at home institutions of the Principal Investigators, the Radio Science Team Leader, Interdisciplinary Scientists, Team Members, Co-Investigators, and Participating Scientists. Science Operations Planning Computer workstations located at Principal Investigator, Team Leader, and Interdisciplinary Scientist home institutions are electronically connected via NASCOM and Ethernet links to a Project Data Base at JPL.

For planning purposes, it is assumed that there will be five groups to cover science (atmospheres, polar processes, atmosphere-surface interactions, geodesy and geophysics) and one to cover data and archives.

The Planetary Data System is sponsored by NASA's Office of Space Science.

Digital Entities Studied

The digital entities being studied are the Mars Global Surveyor data records held within the Planetary Data System (as PDS data sets). Standard products form the core of the archives produced by MGS and transferred to the PDS for distribution to the science community. These products and associated raw data, SPICE (Spacecraft, Planet, Instrument, C-matrix, Events) files, and ancillary information are placed on archive volumes for validation and transfer to the PDS.

The various kinds of digital entities (types of digital objects) created are described on pages 13-15 of the Final Report. Although in the view of this case study they would be considered as components or elements of the digital entities studied, in PDS they are known as PDS structure objects.

Documentary Practices Observed

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

Records Creation and Maintenance

Although in the scientific focus there appears to be more emphasis on preservation issues than on records-making issues—and preservation is indeed the object of this project—in the Mars Global Surveyor mission and Planetary Data System, there are nonetheless many formal **procedures, rules and standards** for records creation. It is assumed that if these various norms and guidelines are followed, the digital objects in question will be ready and apt for preservation with a minimum of treatment necessary. So records creation seems to be structured with a view to preservation and may in fact be seen as the first step of preservation. What is more, these procedures, rules and standards are well documented (c.f. PDS Data Preparation Workbook, PDS structure objects & PDS Data Dictionary, below).

"Principal Investigators are responsible for generation of reduced science data records, documentation, algorithms/software to generate level 1 to 3 products, and archive volumes containing standard products and supporting information. The Radio Science Team Leader will not supply reduction software, but will instead **provide documentation** to explain how the processing is carried out. ... The author of each archive collection (and associated volumes) is responsible for publishing a Software Interface Specification document that delineates the format and content of the respective volumes." (FR 25)

The data sets have already been arranged and **described** before they are transferred to the Planetary Data System, in keeping with the Planetary Data System Data Preparation Workbook, a guide for the organization and preparation of data sets to be submitted to the Planetary Data System. The data sets in the Planetary Data System are **organized** according to mission, spacecraft instrument and target (sun, planet, moon, comet, etc.).

Digital entities are **aggregated** into data sets. "A dataset is an accumulation of data products, supplemental data, software, and documentation that will completely document and support the use of those data products." (FR 5) There are, in fact, several levels of aggregation in the project. "Data products, software and documents are aggregated into volumes, volumes aggregated into datasets, and datasets are associated with instruments and Targets, targets and instruments are associated with spacecraft and spacecraft with a mission." (FR 18)

In the Planetary Data System, each data product is assigned a permanent, **unique identifier**. "The PRODUCT_ID data element represents a permanent, unique identifier assigned to a data product by its producer. In the PDS, the value assigned to PRODUCT_ID must be unique within its dataset. The PDS Standards Reference also specifies the rules for dataset and volume names and ids....Each PDS dataset must have a unique DATA_SET_NAME and a unique DATA_SET_ID, both formed from up to seven components…Within datasets, there are unique volume IDs. Within volumes, the file names are unique." (FR 18)

What is known as "Planetary Data System structure objects" are **guidelines** that outline the format in which the science data appear in the Planetary Data System labels. This seems to be a way of both describing and organizing/aggregating the data. An explanation of each PDS structure object is included in a document known as the PDS Standards Reference. In that document, for each object there is text that describes the object, outlines its uses, and illustrates one or more examples. See pages 13-15 of the Final Report for a list and description of the various structure objects.

Other guidelines include a **controlled vocabulary** and **naming conventions**. "The Planetary Science Data Dictionary (PSDD) contains the set of definitions for all attribute names that are valid for use in resource descriptions across the PDS. The PDS data nomenclature standards define the rules for constructing Data Element and Data Object names." (FR 15) "The Planetary Science Data Dictionary is a NASA institutional standard for Planetary Science Metadata. The PDS procedures for assigning standardized names to data elements follows closely the NBS Guide on Data Entity Naming Conventions." (FR 33)

Description is done through what is known in the project as labels, which are used somewhat like **metadata**. "A label (product label) is a resource description stored in a file. If the label is in the same file as the resource, it is called an attached label. If it is in a separate file, it is called a detached label. The Object Description Language (ODL) is used to create labels (data descriptions) for data files and other objects such as software and documents. The PDS labels contain the key attributes of the digital objects…The label...describes the structure or format of the data." (FR 15, 16)

In addition, there appears to be integrated **metadata**-type information for the aggregated data sets. Each data set volume contains an index table that consists of a set of attributes that describes each data product. For each object, there exists text that describes the object, outlines its uses and illustrates one or more examples. Also, metadata is included for each data product aggregated into a data set. "The data product labels and ancillary product labels are included on data volumes. When data volumes are transferred to the PDS, they are automatically cataloged by mission, target name, target type, instrument, instrument type and dataset identifier. Standard terminology is maintained in the Planetary Science Data Dictionary...which is jointly maintained by the PDS and the multi-mission ground data system. The metadata values for new data products are compared with the PSDD and existing values used wherever possible." (FR 6)

The system tracks all **changes** or modifications (**transactions**) to the entities. "The Track function is the system function that reports on the status of a resource, starting from the time it is submitted to the PDS and continuing throughout its lifecycle." (FR 24) In turn, "the lifecycle of a resource is the set of distinct changes that a resource undergoes from its creation to its final place, if any, on the System and in the Archive…Any set of attributes can be defined…and thus be 'tracked' by System…the ARCHIVE_STATUS of a dataset tells which phase of the archive lifecycle the data set is in. The ARCHIVE_STATUS_NOTE element is available to describe the ARCHIVE_STATUS value in finer detail. The ARCHIVE_STATUS_DATE element provides the date that the archive status will in the future, or has in the past, changed." (FR 24) The PDS also logs (tracks) accesses to restricted areas of the system. User id, date, time and operation are logged.

Recordkeeping and Preservation

The Planetary Data System may be considered a long-term preservation strategy, in that it was designed to be an archive for the long-term preservation of planetary science data records. In other words, the Planetary Data System is a type of **persistent object preservation** (POP). It is a technique to ensure that electronic records remain accessible by making them self-describing in a way that is independent of specific hardware and software.

The Planetary Data System can be considered a **persistent archive** in that the data it holds remain accessible over time and through technological obsolescence. The PDS is a distributed data archive. There is a central node and seven other discipline nodes—Atmospheres, Geosciences, Imaging, Navigation Ancillary Information Facility, PPI, Rings and Small Bodies. The Final Report seems split on how to characterize the PDS as an archive. It states that "the PDS is referred to as an active archive, but it resembles a recordkeeping system, albeit of the records of scientific activities, rather than business activities," (FR 26) but then goes on to affirm that "the datasets have already been arranged and described at the time they are transferred to the

PDS. Therefore, the PDS could also be viewed as an archive for long-term preservation, rather than a recordkeeping system for active records." (FR 26)

Detailed **archiving** and **records management** plans are drawn up for all active and new projects and are heavily regulated and documented. "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. A Project Data Management Plan is required by NASA for all new projects. This plan provides a general description of the project data processing, cataloging, and communication plan. The Archive Policy and Data Transfer Plan (APDTP) provides a detailed description of the production and delivery plans for archive products for a project." (FR 10)

There are specific **roles** in the team for archiving and records management. "The MGS Interdisciplinary Scientist for Data and Archives...will advise the Project with regard to archiving and will work with MGS and the PDS to help ensure that detailed plans are in place for generation of Planetary Data System-compatible products and associated documentation, and that archive volumes are generated, validated, and transferred to the Planetary Data System." (FR 25) In addition, "MGS Data Administration will compile archive volumes…The Radio Science Team Leader will produce the Radio Science level 0 archive collection. Principal Investigators are responsible for generation of reduced science data records, documentation…and archive volumes containing standard products and supporting information." (FR 25)

The PDS regularly **captures** digital entities within the scope of its activity. "Ingest is the PDS function that receives, stores and catalogs resources. When volumes and datasets are transferred to the PDS, the catalog information for the dataset and the volume descriptions are used to automatically update the PDS catalog. Volumes are indexed by mission, target name, target type, instrument and instrument type and volume identifier attributes." (FR 20)

The peer review process that determines what records are incorporated into the Planetary Data System can be likened to an **appraisal** process. "The peer review process resembles the appraisal process for selection of records for long-term preservation." (FR 26)

Questions of **interoperability** have been addressed in the project. The data objects are displayed using the technology-neutral NASAView. "NASAView is the display tool for the PDS archive that runs on multiple platforms…The XVT Development Tool supplies the cross-platform GUI for NASAView…PDS labels have also been created to describe non-PDS data formats such as FITS and VICR labeled images. Using these "detached" PDS labels to describe non-PDS formatted data, standard [Object Access Library] function calls can be used to access the data." (FR 20)

In addition, "The PDS has been operational since 1989 and it has not been necessary to update (convert or migrate) any of the data products to other data formats. When hardware and/or operating system become obsolete, only PDS Label Library Light (L3), Object Access Library (OAL) and the NASAViewer need to be recompiled or translated to the new platform." (FR 31)

Accuracy, Authenticity and Reliability
Terms in *italics* were added to show links to other sub-sections in this section of the Overview.

Accuracy
All data incorporated into the PDS archives must undergo a peer review. The purpose of the review is to determine that the data is accurate, complete and *reliable*; that the data are suitable for archiving and that the PDS standards have been followed.

Accuracy of data is assured through a strict validation process. "The primary validation tool of the PDS is the Volume Verifier. The Central Node data engineers run this program on each product delivered from a project. It validates the format and content of all product labels, and validates the integrity of data files using checksums." (FR 22)

Authenticity
The term 'integrity' seems to be substituted for 'authenticity' in the project. "It is unlikely that the creators or maintainers of the records would customarily use that term to characterize their quality."(FR 23) "Project team members, PDS managers and engineers and other Planetary Scientists do not traditionally use the term "authentic" to characterize the data products that they create, maintain and use. They are concerned that the data records are complete, *reliable*, accurate and that the integrity of the data record is assured." (FR 23)

However, the scientists involved in the project do use measures to guarantee authenticity, even though they do not use that exact term. "One can conclude that, due to the emphasis on completeness and *reliability* of the planetary science data records, the peer review, role-based authentication of access to archived data products and data integrity checks, the scientific data records are maintained authentic." (FR 23)

To prevent tampering with and corruption of data, access to the Planetary Data System resources is controlled through user authentication. Access to restricted areas of the PDS is determined by the user's assigned role (i.e., role-based authentication).

Reliability
See accuracy and authenticity, above.