ESA's Planetary Science Archive (PSA):
Maximising the Long-Term Usability of Planetary Data

S. Martinez [1], David J. Heather [1], M. Barthelemy [1], N. Manaud [1], M. Szumlas [1], J. L. Vazquez [1], and the PSA Development Team [1]

[1] *ESA-ESAC*
PoBox 78, 28691 Villanueva de la Cañada, Madrid, Spain
Email: Santa.Martinez@sciops.esa.int

ABSTRACT

The European Space Agency's Planetary Science Archive (PSA) is responsible for the long-term preservation of all the scientific and engineering data returned by ESA's planetary missions, as well as for the provision of services to increase the accessibility and usability of the archived PSA data. This paper will describe the concept for a highly automated system that will streamline the current archiving process from end-to-end, providing support for data producers in the design, preparation and delivery phases, tracking data internally after delivery and through the standard validation, ingestion and release procedures, and delivering additional support tools for end-users wishing to visualize, analyse and manipulate data from the PSA. Such a system will increase the level and quality of the data and supporting information ingested into the PSA and will maximise the usability of the planetary data both now and in the long-term. By standardising and automating the archiving process, a better guarantee of conformance to more restricted standards can be ensured, also allowing for the development of improved data visualization, analysis and manipulation tools for the end-users.

Keywords: planetary, ESA, PSA, data, archive, preservation, long-term, validation

INTRODUCTION

Within the European Space Agency, planetary data are archived and distributed to the worldwide scientific community by the Planetary Science Archive (PSA). The primary focus of the PSA is to maximise the usability of the data both now and in the long term. In order to achieve this goal, PSA aims to provide all necessary services and resources to ensure a high-quality and scientifically useful archive in addition to ancillary data, software tools and technical expertise to fully assist the scientific community in the use and analysis of the data.

All PSA data are compliant with NASA's Planetary Data System (PDS) standards. At the time of the archive definition, PDS standards were widely accepted and understood within the planetary science community, having been used on data returned by the majority of previous planetary missions. Adhering to the PDS standards was therefore a first key decision to guaranteeing usability and long-term preservation.

Since the beginning, PSA efforts have concentrated on the design and preparation of the archive by establishing archiving standards, policies and procedures, and by providing expert consultancy and support to the data providers throughout the entire archiving process. With most data pipelines and routine deliveries now in place, the focus has shifted to providing services that will make the archived data more accessible to the end-users. In support of this, PSA organises workshops for the scientific community at large led by the instrument experts.

Archiving Process Overview

ESA's planetary missions are Principal-Investigator-led, and therefore the design, preparation and delivery of the data is typically the responsibility of the individual instrument teams. During all archiving phases, PSA provides consultancy to all of the instrument teams for the correct usage of the standards to describe their observational data, the calibration algorithms and any relevant supplementary information that could contribute to the usability and preservation of the observations and the knowledge of the instrument. This distributed setup imposes a significant effort in coordination to ensure there is consistency at mission level and to guarantee compatibility for potential cross-missions and cross-instrument data analysis. In addition to this challenge, instrument funding is in some cases limited for data archiving and analysis of the science data, requiring additional support from the PSA and usually leading to delays in data deliveries.

The overall process for archiving is captured in an Archive Plan including the definition of guidelines, procedures and agreed delivery schedules, while instrument specific information is documented in the Experiment to Archive Interface Control Document (EAICD), which provides a full description of the data sets being archived by a given instrument, including a summary of the data pipeline, the structure of the data products, data sets and any calibration or geometry information provided.

Once the data are delivered to the PSA, a set of rigorous and well-defined archiving procedures is followed to ensure not only compliance with the PDS standards but also completeness and usability of the data. These procedures stem from the requirements defined in the top-level archive documentation, and in the mission and instrument specifications. Validation includes syntactical and qualitative checks, in most cases demanding intensive manual interactions. Although several software tools exist to aid PSA archive scientists in the different validation steps, the lack of a unique system that would guarantee a minimum set of validation criteria may lead to inconsistencies in the data.

In addition to the internal quality assessment, and prior to ingestion into the archive, each phase of the archive process is controlled by a corresponding peer review, during which external experts are asked to validate the data and documentation against scientific quality criteria.

The interfaces and interactions during the archiving process are clearly defined. Nevertheless, only basic implementations for these interfaces are in place to follow up data set discrepancies as well as data deliveries, validation and ingestion status, such as emails, spreadsheets or feedback documents. This stresses the need for a more standardised software infrastructure to support uniform and consistent interfaces that would be able to notify the relevant parties when required. Besides, a centralised information management system would help to keep track of the data set status and the history of all reported issues.

STREAMLINING THE ARCHIVING PROCESS FROM END-TO-END

The procedures and policies that define the current archiving process (delivery, validation, review and ingestion) are based on the PDS3 standards and Data Providers Handbook recommendations. These procedures have evolved over the years as new PSA requirements have been established and mission-specific and instrument-specific requirements have been defined. Lessons learned and recommended practices from previous missions have also helped to refine all archiving procedures.

A review of these archiving procedures has been recently conducted by the PSA in order to streamline the process from delivery to ingestion, and to identify those areas where automated tools or other supporting applications could improve the efficiency of the process, pointing to the need for a more centralised and automated system. The proposed approach is a centralised software architecture supporting the overall archiving process with different degrees of automation and customised user interfaces depending on the user type. At the core of the archiving system, a well managed centralised repository to acquire information from and provide information to the different components of the system, keeping track of all information relevant to the science data archive from design to distribution.

As a direct effect of minimising the manual work, the turnaround times and the productivity will improve and the operational mistakes or misunderstandings will be reduced. Standardising procedures and interfaces will increase efficiency and improve the service to the end-users.

Figure 1 shows the three main parties involved in the archiving process and the corresponding responsibilities. The list of responsibilities is by no means complete but outlines the scope of the evaluated activities for this purpose.
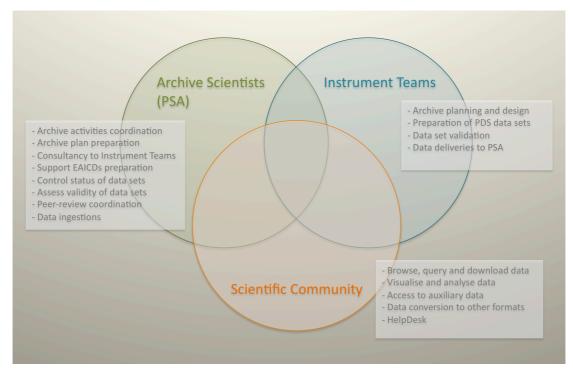

Figure 1: Main archiving process stakeholders and responsibilities.

The following is a set of recommended components foreseen for the enhanced archiving system derived from the streamlining process and aiming to highlight the areas of improvement. These recommendations will help PSA to focus efforts and to plan, design and implement a system that will satisfy the requirements, needs and expectations of all interested parties. Since the needs and requirements of the different user types can overlap, the system components have been grouped into two functional areas: archiving components and end-user components.

Archiving Components

Effective management of the archiving process is very important and depends upon common understanding of the requirements and effective monitoring of the activities involved. Several components are proposed to cover this area.

- A common layer supporting basic **PDS/PSA read, write and manipulation functionalities** is needed. Although the motivation for this component might not stem directly from the described activities for the system, providing a common component for this purpose will ensure consistent manipulation of data as well as avoid duplicating efforts in different components.
- With the purpose of standardising and automating the generation of PDS/PSA data products and data sets, a generic **data preparation and processing facility** is proposed to enhance the services for the data providers (i.e. mainly the instrument teams). This facility would provide interactive creation and editing of data sets content, catalogs and labels, as well as automatic options with configurable parameters. The data processing component would allow for the automatic conversion of telemetry data into PDS/PSA data products based on a generic processing core and with configurable options

depending on the instrument specifics. Allowing for the addition of external modules (plugins) to enhance the capabilities of the data processing component could also be an interesting area to explore.

- A robust and integrated *data validation* component should be present in the system providing means to execute and support a standard validation. This standard validation shall guarantee the data being checked is acceptable for ingestion and complies with all necessary standards for a given instrument and mission. This component would be used at all stages of the data development and should, therefore, cover all validation aspects at different levels. A minimum set of validation steps should be checked by default, offering also the possibility to define additional criteria. A robust set of configurable algorithms would be integrated into this component and available to its users. An improved interface allowing for an efficient tracing of the source of errors would be very convenient.
- Regarding data handling and management, the recommendation is to implement an *archive process controller* component that would support archive scientists in managing the status and lifecycle of all data, from the first data delivery to the distribution to the end users.
- In order to be able to offer a standard and uniform interface, a component supporting the *interactions of PI - Reviewers - Archive Scientists* is recommended. Tracking all exchanged information (mainly discrepancies, status and additional documentation) during the entire archiving process, and in particular during the peer review, is key to reducing both the turnaround time and the complexity of the process. Handling of schedules, notifications and effective tracking of discrepancies and responses should be the main focus.

End-Users Components

These components may not seem directly related to the archiving process. However, in addition to enhancing the services to the scientific community, end-user components would support and improve the services for the preparation and validation phases of the archiving process.

- Providing a *data visualisation* component that would provide means to quickly browse the content of a data set and visualise data products would not only be a service for end-users but could also be used to support the validation process to ensure readability and usability criteria are met.
- The motivation for a *data conversion* component is to provide the scientific community with means to convert data to more common scientific formats. This would allow end-users to import the data in more widely used tools and applications.
- A *data analysis* component could be available to perform basic image manipulations, extract derived information from the data e.g. identify geographical features in an image. Allowing this component to interface with already existing external applications for data analysis and visualization (e.g. ISIS, IDL or Matlab) should be also considered.
- PSA feels it is very important to monitor the need of the scientific community and make sure the services and resources available to them are actively developed to incorporate new requirements. Gathering feedback is key in this process and the proposed component is an improved *HelpDesk* that would provide external users with a centralised point of contact, possibly a web interface that they could contact to solve problems and obtain information about PSA data and software/services.

PROTOTYPING

Prototyping is a fundamental part of our archiving system development approach. By simulating the main functionalities of some of the key components of the proposed system, we will be able to better evaluate if the required capabilities are fulfilled and complete, and can satisfy the expectations from the defined user groups.

One of the prototyping efforts recently completed by the PSA that will contribute to enhance the services for the scientific community is a Planetary GIS Prototype. At present, the PSA lacks support for OGC web services. These well-known and widely used standards could enable easy access to map-projected data for GIS client applications. The development of a Planetary GIS Prototype has allowed PSA to explore the benefits, feasibility and limitations of an OGC compliant interface for the PSA and to support the definition of requirements for the future PSA
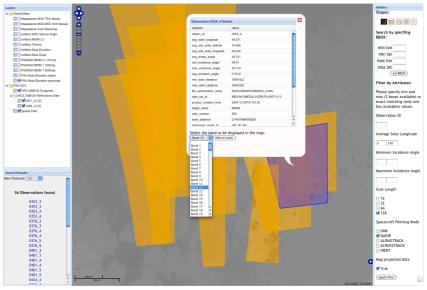

Figure 2: PSA Planetary GIS Prototype

and its geospatial data handling. This interface has been designed to accommodate future needs as additional planetary surface data becomes available or additional functionalities are required.

Another prototype implementation currently being evaluated is the data validation component. This prototype is being used to help PSA define not only the validation requirements but also other supporting functionalities that could be useful during the validation process.

CONCLUSION

Having centralised procedures and software to support the overall workflow of the data has many advantages:

- It will help to ease the load of the PSA staff (primarily the Archive Scientists) and will allow them to concentrate on enhancing both data and services provided by PSA.
- It will allow PSA to provide better and more standard support and consultancy to the instrument teams and end-users.
- It will ensure the availability of mechanisms to preserve knowledge and information that would allow PSA to maintain data and software after the end of the mission.

Replacing the current manual interactions with a centralised information management system and well-defined automated processes would not only increase the efficiency of the archiving process but also the level and quality of the data and supporting information ingested into the PSA and would maximise the usability of the planetary data both now and in the long-term. By standardizing and automating the archiving process from end-to-end, a better guarantee of conformance to more restricted standards can be ensured, also allowing PSA to concentrate efforts in providing better services and software tools to the end-users.

REFERENCES

[1] – Caltech, JPL, Planetary Data Standards Reference Version 3.8 JPL D-7669, Part 2, February 27, 2009