ESA's Planetary Science Archive: from data preservation to data exploitation

M. S. Bentley[1], R.Docasal[2], F. Felix[2], J. Oliveira[2], J. Osinde[3], F. Raga[3], J. Saiz[3], R. Valles[3], B. Merín[1], D. Coia[4], E. Grotheer[5], D. Heather[1], T. Lim[4].

[1]ESA/ESAC (Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain)
[2]Aurora Technology BV for ESA (Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain)
[3]Rhea for ESA (Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain)
[4]Telespazio UK Ltd for ESA (Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain)
[5]Serco for ESA (Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain)

Introduction: ESA's Planetary Science Archive (PSA) is the long-term home for all scientific data returned from ESA Solar System missions. Its remit to date has been primarily to preserve the data and knowledge needed to use it. Moving forward the hope is to expand this remit to provide environments, tools and guidance needed to exploit the data.

Data preservation: Given the low frequency of space missions and long lead times to receive new data, preservation is a critical part of the PSA. This means not only ensuring that we store the data, but that the documentation and meta-data clearly describe to future users how to calibrate and use those data. Adopting the NASA PDS format has been a big part of this, along with PSA-specific convention designed to help both data providers and users approach the data in a uniform way. In addition, the archive scientists working for each planetary mission, together with the instrument teams, are key to ensure that the data are usable for the scientific community. Finally, peer review of both data and documents ensures the scientific community are involved in this process.

Finding data: The PSA offers a variety of ways to search for planetary data – via table, image gallery and maps views in the user interface, as well as a simple file system browser, FTP access and, finally, two APIs (PDAP and EPN-TAP). These interfaces offer a multitude of ways to find data, but are limited by only being able to query a sub-set of the meta-data, and ultimately the end result is a download of data to the user's local computer. Even then, the road from downloading data to reading the documentation, finding or writing the right software and eventually doing science can be long and rocky.

In the coming years the hope is to improve this in several areas – by exposing more meta-data via new and updated APIs, by providing integration with the ESA DataLabs project and by providing tutorials and executable notebooks to help users get started.

Improved APIs: Currently the PSA supports two APIs. The first, PDAP, is built around the older version of the PDS data standard (PDS3) and offers limited query capabilities. To this end it doesn't offer the flexibility and performance needed by many users. The second, EPN-TAP, offers a planetary flavour of an interface commonly used throughout astronomy archives and beyond. The PSA uses this service to publish basic meta-data of individual data products, but cannot populate the rich meta-data possible due to the large variety of mission and instrument types. However, improvements are being made by, for example, adding geometry data to the tables to allow geometric search (e.g. by latitude and longitude).

Finally, the PDS Engineering Node has developed a PDS-wide registry and search capability which the PSA hopes to adopt. This provides the ability to search arbitrary meta-data and will eventually lead to global searches being made possible across the federation of PDS-compatible archives.

These steps will dramatically increase the flexibility of searches available to the end user, to narrow their selection before downloading or consuming the data.

DataLabs integration: Another key project for future integration is ESA DataLabs. This environment adopts the "bring the code to the data" philosophy and allows users to run their own code in web (e.g. Jupyter Notebooks) or desktop environments. Together with some additional API fields, this will allow users to run powerful searches and to rapidly process data in a secure environment, without having to download the products. This will open the possibility to run data-intensive workflows such as machine learning algorithms on many data products. Integration with the UI will eventually be possible to couple the visual browse and search capability with user code.

Data tutorials: Whilst data tutorials have always been a part of the PSA, modern development like Jupyter Labs (and eventually DataLabs) allow for a much more dynamic experience. Future tutorials will serve as templates for new users to copy, modify and explore as they learn the ins and outs of a new dataset.