Astronomical Data Analysis Software and Systems XVI
ASP Conference Series, Vol. 376, 2007
R. A. Shaw, F. Hill and D. J. Bell, eds.

ESA Planetary Science Archive

Christophe Arviset, John Dowson, Inaki Ortiz, Esther Parrilla, Jesus Salgado

European Space Astronomy Centre, Madrid, Spain

Joe Zender

ESA-European Space Research and Technology Centre, Noordwijk, The Netherlands

Abstract. The (ESA Planetary Science Archive[1] (PSA) hosts all the data from ESA's planetary missions into a single archive. It currently contains data from the *Giotto*, *Mars Express*, *Rosetta*, and *Huygens* spacecraft, some ground-based observations, and will host data from the *Smart-1*, *Venus Express*, and *BepiColombo* spacecraft in the future. Based on the NASA Planetary Data Systems (PDS) data dictionary, all datasets provided by the instrument teams are scientifically peer-reviewed and technically validated by software before being ingested into the Archive. Based on a modular and flexible architecture, the PSA offers a classical user-interface based on input fields, with powerful query and display possibilities. Data can be downloaded directly or through a more detailed shopping basket. Furthermore, a map-based interface is available to access *Mars Express* data without requiring any knowledge of the mission. Interoperability between the ESA PSA and the NASA PDS archives is also in progress, re-using concepts and experience gained from existing IVOA protocols. Prototypes are being developed to provide functionalities like *GoogleMars*, allowing access to both ESA PSA and NASA PDS data.

1. Introduction

A few years ago the European Space Agency (ESA) decided to build its Planetary Science Archive (PSA) to gather altogether all scientific data from its planetary missions (see Figure 1). The first version of the PSA was released in 2004 March and it now contains historical *Giotto* data as well as more recent *Mars Express*, *Rosetta*, and *Huygens* mission data. Soon data from the *Smart-1* and *Venus Express* missions will also be incorporated into the PSA, as will all future ESA planetary mission data (e.g., *BepiColombo*). PSA follows an active development cycle, with a few major releases per year; PSA v2.8 should be released by the time these proceedings are published.

Under the responsibility of the Solar System Science Operations Division in ESA's Research and Scientific Support Department, the PSA is a joint project between the PSA team located at ESTEC (Netherlands) and the PSA team at the European Space Astronomy Centre (ESAC) in Spain. The PSA team at

[1] http://www.rssd.esa.int/psa

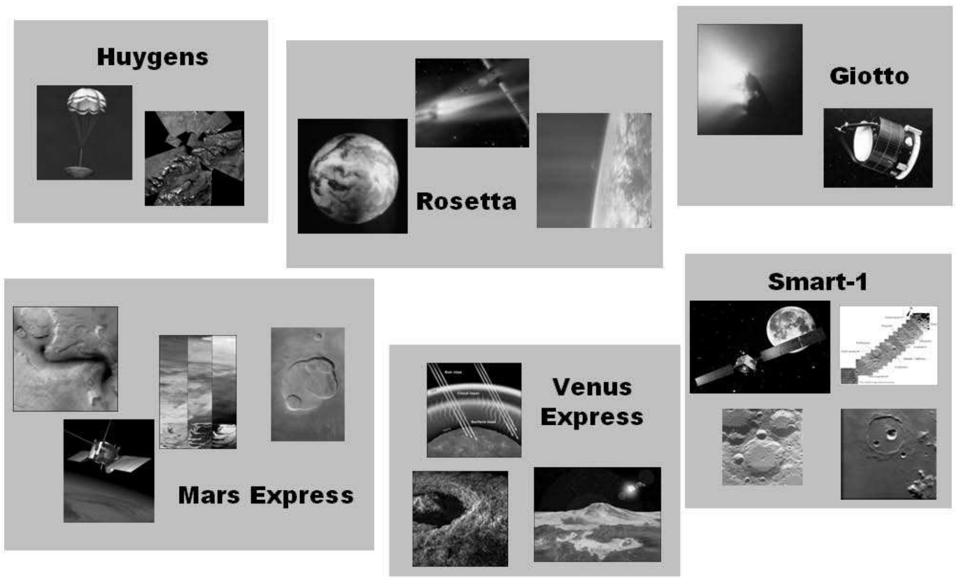

Figure 1. PSA: One archive, several missions.

ESTEC is responsible for all interfaces with Instrument Teams, the validation of the datasets received from the Instrument Teams before their ingestion into the PSA, and for setting up the user requirements for the Archive systems. The PSA Team at ESAC, within the Science Archives Team, is responsible for the design, development, and operations of all the PSA archives systems. The PSA is physically located at ESAC, located in Villafranca del Castillo, near Madrid, Spain. ESAC is becoming the center for all ESA Astronomy and planetary space missions. That will include the science operations center for such missions, as well as the scientific archives for data storage and data distribution to the scientific community. Hence, ESAC also aims at becoming the European Virtual Observatory (VO) node for space-based astronomy (see Arviset 2007).

2. *Venus Express* Data Flow

As is common for ESA projects, data flow for a mission can be quite complex. To illustrate this, we will develop the case of the *Venus Express* (VEX) mission. First, the VEX Science Operations Centre at ESTEC (Netherlands) prepares the planning of the mission which is sent to the ESA Satellite Operations Control Centre (ESOC) in Germany. Commands are sent to the spacecraft through the powerful new 35 m antenna located in Cebreros, near Avila, Spain. Raw data are then sent back by the spacecraft to the antenna and stored on the operational servers at ESOC in Germany (see Figure 2). Then corresponding raw data are disseminated from ESOC to each of the instrument team's institutes across Europe (Sweden, France, Germany, Italy, and in the case of *Venus Express*, Austria). Over a period of 6 to 12 months the instrument teams will process these raw data into scientific datasets, which will then be sent to ESTEC where

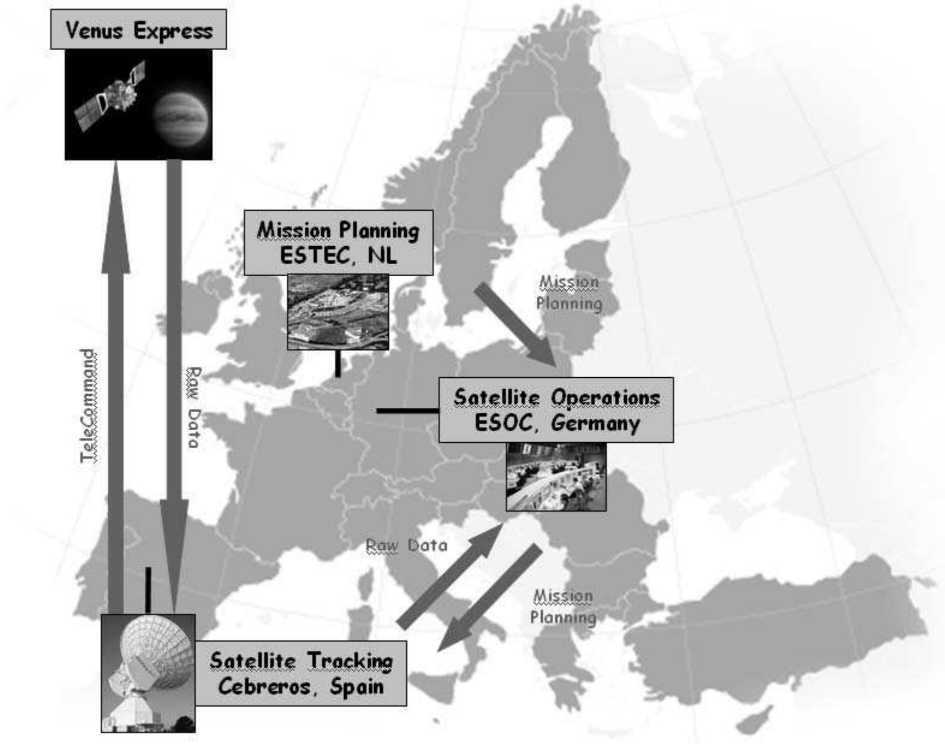

Figure 2. Uplink/downlink data flow.

they will be validated. Once validated, datasets can be ingested into the PSA and made accessible to the scientific community worldwide (see Figure 3).

3. PSA Architecture

The PSA 3-tier architecture (see Figure 4) is based on what had been developed already for other astronomical archives by the Science Archives Team at ESAC. The architecture is modular and flexible, and it aims it to separate the data (dataset storage and database) from their presentation (user interfaces). All subsystems described hereafter have been developed in Java, with extensive use of XML, to make these applications configurable and re-usable for other archive projects.

3.1. PVV: Data Validation

From the beginning of the PSA project it was decided to adopt the Planetary Data System (PDS) format as the standard for all datasets in the archive. As mentioned earlier, the instrument teams are responsible for generating and delivering the scientific projects to the PSA. As data was coming from various and different providers, we developed a system, the PSA Volume Verifier (PVV) to

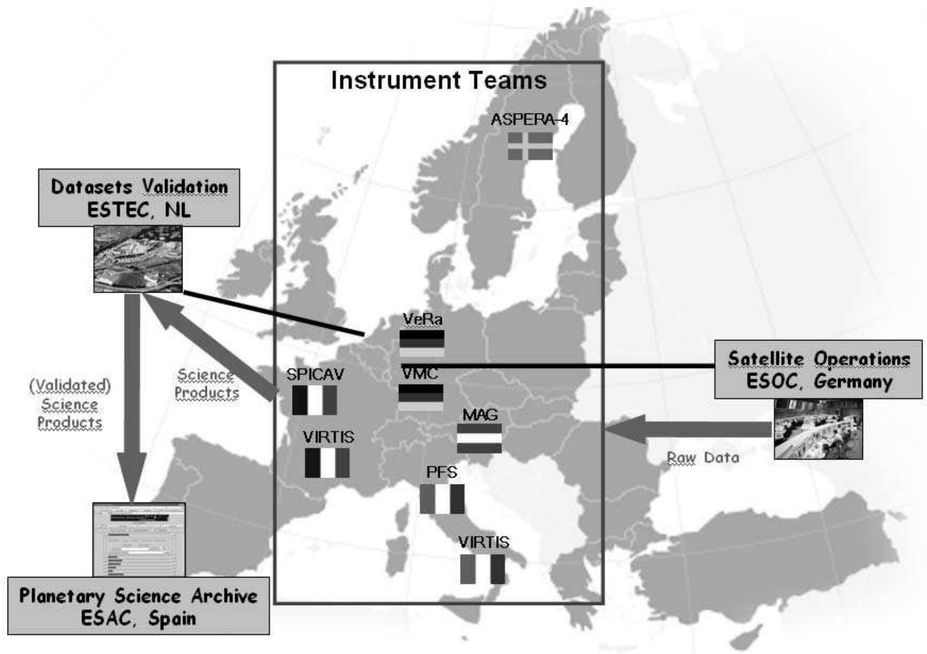

Figure 3. Data processing and distribution data-flow.

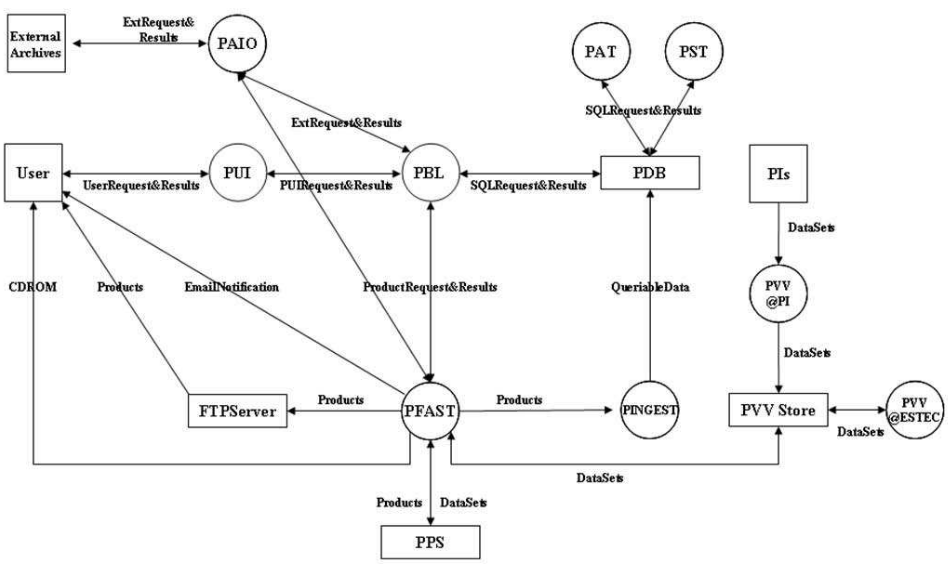

Figure 4. PSA Top Level System Design.

ensure that the delivered datasets are compliant with the PDS format[2]. The PVV was first released in 2003 November and is distributed to all the instrument teams to validate their datasets before being delivered to the PSA. The PVV is a Java command-line application that works on most operating systems (Windows, Mac, and most UNIX and Linux platforms). The PVV first scans the dataset (*PVV SCAN*) and builds an XML image of it, then it checks that the structure of the dataset is compliant with the PDS directory structure. Finally, the PVV will check the compliance of these data with the PDS dictionary and some extra PSA dictionary definitions (*PVV VERIFY*).

3.2. PSA Back-end Systems

The PSA back-end systems mainly deal with data and metadata storage. The PSA Product Store (*PPS*) is a disk array where all PSA datasets are stored. That represents around 4 TB of disk space and is regularly growing as new data are incorporated. The PSA File Access SysTem (*PFAST*) is the system that interfaces with the PSA Product Store. Its role is to copy the incoming datasets into the Product Store repository structure and to prepare and distribute the datasets which are requested by end users. The PSA Business Logic (*PBL*) is the system that interfaces with the front-end user interfaces and the back-end storage systems (both data repository PPS and database PDB). This middle-tier application can also be seen as an application server. It ensures modularity and will hide the complexity of the back-end storage systems to the end-users. The PSA DataBase (*PDB*) is a relational database where all the datasets metadata are stored so queries from end-users can be performed efficiently. It is based on SYBASE and consumes a small amount of space (around 1 GB) since only metadata are saved. The PDB is populated by the PSA ingest system (*PINGEST*) which extracts all metadata from the datasets that are necessary to ingest them into the database.

3.3. PSA User Interfaces

By default, users can access the archives through a user-friendly but powerful GUI[3]. Developed in Java and made configurable using XML, the PSA User Interface (*PUI*) main characteristics are the following (see Figure 5):

- Queries are organized by panels (missions) and tabs (instruments per mission)
- Configurable results display with navigation buttons through a hierarchical display (DataSets and DataProducts)
- Image preview and products headers display at one click
- Public data are accessible to everyone, proprietary data are accessible only by the owner or privileged users
- Quick data download via a single click or more complex and customizable data retrieval via a shopping basket

An even more intuitive, map-based user interface has been developed for visualizing *Mars Express* images (see Figure 6). Without the need to be a

[2] http://pds.jpl.nasa.gov/documents/sr/index.html

[3] http://www.rssd.esa.int/psa

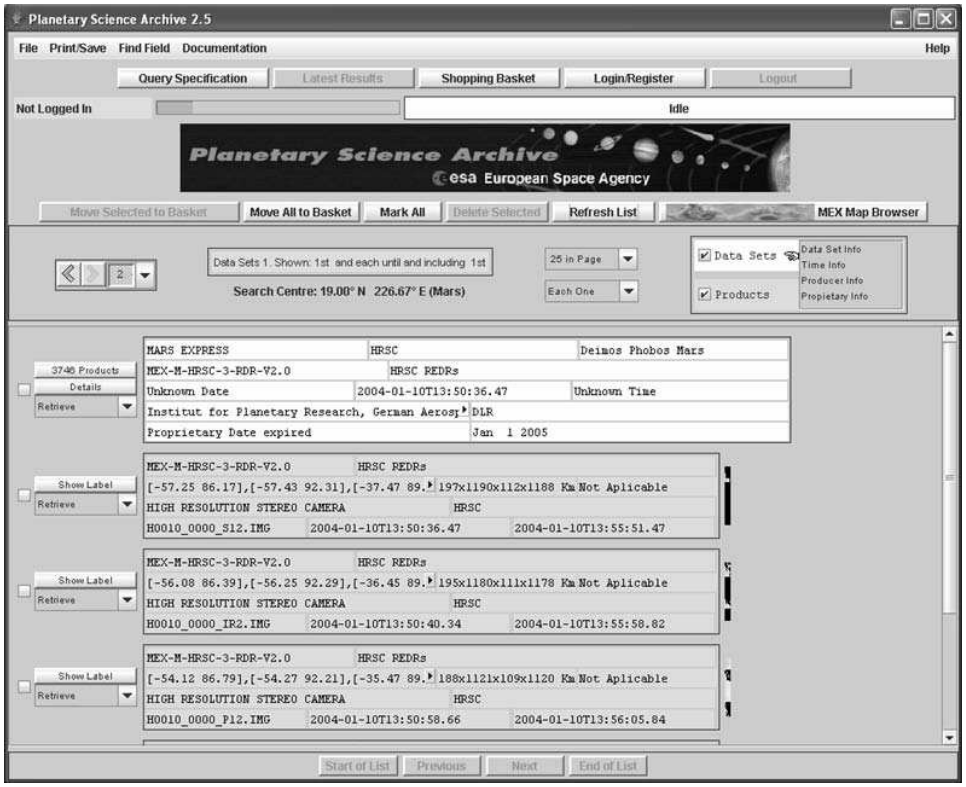

Figure 5. PSA Standard Query User Interface.

planetary expert, user can just select an area of interest with their mouse on the Mars map and have direct access to the corresponding images. Both the standard user interface and the map based interface are connected so the user can simply switch from one to the other.

The PSA *FTPServer* is a clever FTP server that has been developed in Java that creates links to the datasets, rather than copying the datasets from the PSA Product Store into an FTP area. That permits reducing the size of the FTP area to almost nothing, since the links do not occupy much disk space, while still ensuring proprietary access to the real datasets.

3.4. PSA Administration and Statistics Tools

Some extra tools have been developed mainly for internal use. The PSA Administration Tool (*PAT*) enables modifying user account details such as quotas, privileges to access data, notification of the release date for public datasets, as well as a dataset release-date setup. Furthermore, the PSA Statistics Tool (*PST*) enables the generation of PSA usage statistics like queries (see Figure 7), data requests, volume of data retrieved, user distribution by country, etc.

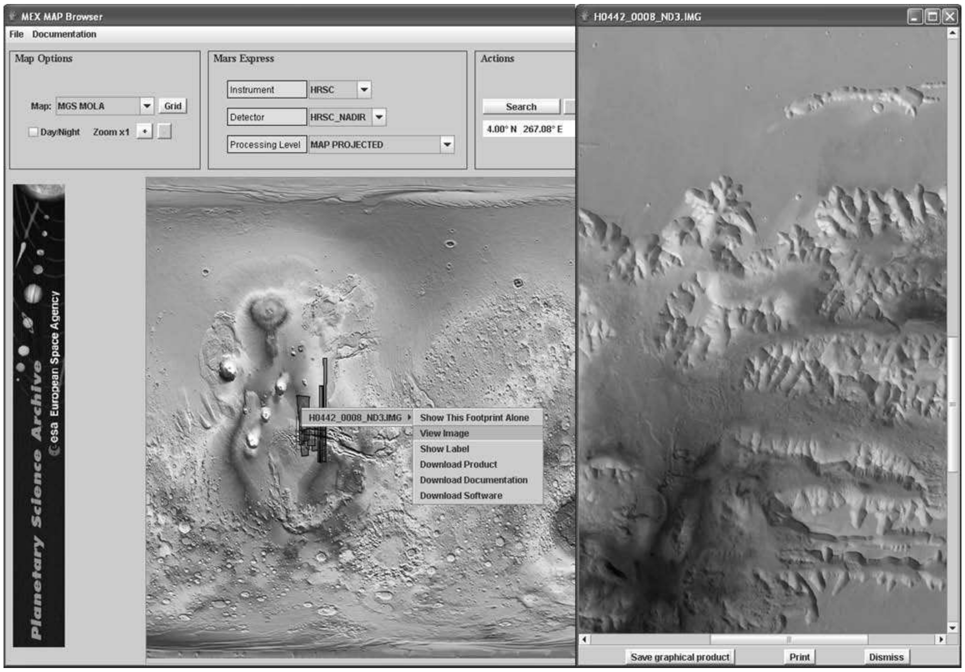

Figure 6. PSA Mars Express Map-based Interface.

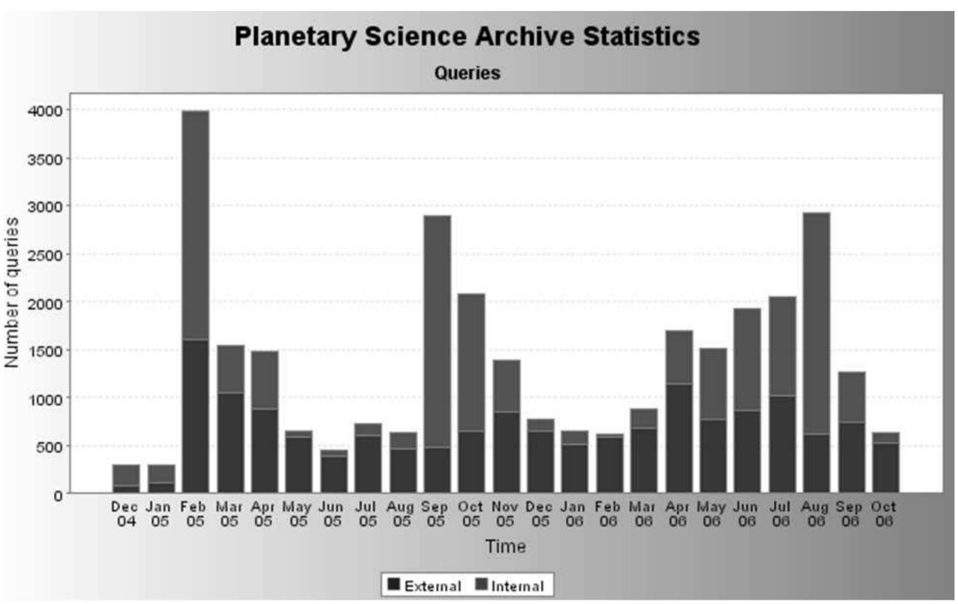

Figure 7. PSA Queries monthly statistics.

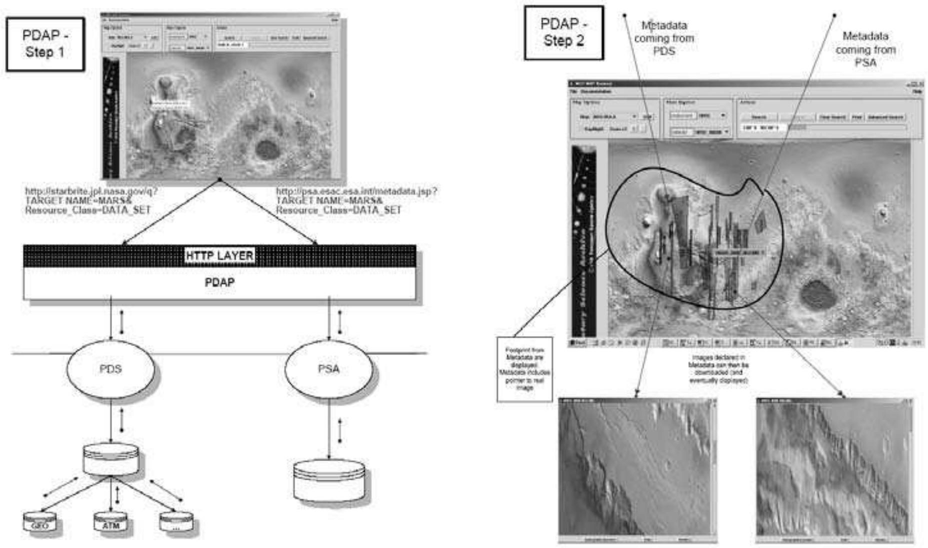

Figure 8. PDAP: Interoperability between planetary data archives.

4. Planetary Data Interoperability Between ESA and NASA

NASA and ESA have agreed to start facilitating interoperability between their planetary science archives by implementing a protocol for data access. This protocol has been called the Planetary Data Access Protocol (henceforth, PDAP). The PDAP defines the input and output parameters, plus the formatting syntax that shall be used to give seamless access to the different data pertaining to the two agencies. In the near future other agencies might want to join the initiative to allow even wider planetary data inter-operability.

The PDAP is basically a two-step protocol (see Figure 8). In the first step (*metadata access*), the archive is asked for the metadata pertaining to the data that obey a certain restriction in the input parameters. Among the compulsory output parameters of this metadata is a reference to the real file holding the data. This allows the second step of the protocol (*data access*) to be executed, as the data are readily made available through the declared link.

The PDAP has been designed re-using experience gained with conceptually similar protocols (Simple Image Access Protocol, SIAP) developed in the context of the International Virtual Observatory Alliance (IVOA[4]), making use of the VOTable format for metadata exchange in step 1. Since the PSA was based on the architectures of other astronomical archives, and has been developed by the same ESAC Science Archives and VO Team (Barbarisi 2007), the development of the PSA Archive Inter-Operability system (PAIO) which publishes PSA data holdings through the PDAP protocol, has been quite straightforward. Clients can be built on top of this protocol to demonstrate their usability. NASA and ESA have agreed to build their respective prototype clients independently to

[4]http://www.ivoa.net

demonstrate how we can make use of the available protocol to implement the inter-operability in real-world cases.

Acknowledgments. The authors would like to thank all the people who have worked or are still working on the PSA, in particular the ESAC Science Archive Team (I. Barbarisi, N. Fajersztejn, M. Freschi, and P. Osuna), the ESTEC PSA Team (D. Heather, N. Manaud, M. Barthelemy, and O. Witasse) and the people who participated in earlier phases of the project (J. Hernandez, G. San Miguel, A. Venet, J. Diaz, M. Castro, and J. Vasquez).

References

Arviset, C., et al. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco: ASP), 703

Barbarisi, I., et al. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco: ASP), 611