A Planetary Science VO prototype

S. Erard [(1)], P. Le Sidaner [(2)], J. Berthier [(3)], B. Cecconi [(1)],

F. Henry [(1)], L. Lamy[(1)], C. Chauvin [(1,2)], R. Savalle [(2)], I. Zolotukhin [(2)]

[(1)] LESIA, Observatoire de Paris/CNRS/UPMC/Univ. Paris-Diderot
5 pl. J. Janssen 92195 Meudon, France. EMail: stephane.erard@obspm..fr

[(2)] DIO-VO, UMS2201 CNRS, Observatoire de Paris
61 av. de l'Observatoire, 75014 Paris, France

[(3)] IMCCE, Observatoire de Paris/CNRS
61 av. de l'Observatoire, 75014 Paris, France

ABSTRACT

In the framework of the Europlanet-RI program, a prototype of Virtual Observatory dedicated to Planetary Science was defined. Most of the activity was dedicated to the elaboration of standards to retrieve and visualize data in this field, and to provide light procedures to teams who wish to contribute with on-line data services. The architecture of this VO system and selected solutions are presented here, together with existing demonstrators.

Keywords: Virtual Observatory, Planetary Science, Solar System, Data services

INTRODUCTION

VO-Paris Data Centre is a structure of the Observatory of Paris involved in Virtual Observatory (VO) activities. As a partner in the Europlanet/IDIS activity, VO-Paris has contributed to setup a prototype Virtual Observatory dedicated to Planetary Science. VO-Paris was a partner in the Joint Research Activity (JRA4) designing the VO infrastructure, and the co-leader of the task "Added Value Services". Other groups involved in this matter included CDPP (Toulouse), IPAG (Grenoble), and INAF/IAPS (Rome). VO-Paris is also a historical contributor to the IVOA (International Virtual Observatory Alliance) in Astronomy, and a partner in the VAMDC, HELIO, and CASSIS European programs [1], all focused on on-line data distribution and processing related to Solar System studies.

The aim of the activity was to facilitate searches in big archives and sparse databases, to make on-line data access and visualization possible, and to allow small data providers to share their data in an interoperable environment with minimum effort. This system makes intensive use of previous studies and developments led in Astronomy (IVOA) and by space-borne data archive services (IPDA, International Planetary Data Alliance), as well as in Solar Physics (HELIO program) and spectroscopy (VAMDC program). In particular, it remains consistent with extensions of IVOA standards.

The global architecture involves existing data services accessible through IVOA protocols (Cone Search, TAP…) or the IPDA protocol (PDAP) whenever relevant. However, a more general standard has been devised to handle the specific complexity of Planetary Science, e.g. in terms of measurement types and coordinate frames. This protocol, named EPN-TAP, is based on Table Access Protocol (TAP) and includes precise requirements to describe the contents of a data service. The data services are declared in the extended IVOA registry based at VO-Paris, which is queried by a client handling both EPN-TAP and PDAP protocols.

The client itself and some demonstrators are available from the VO-Paris IDIS node at: http://voparis-europlanet.obspm.fr/.

PROPOSED ARCHITECTURE

A general scheme is proposed in Fig. 1, which illustrates the sequence of steps in a typical working session. The user is working at his computer, sends queries to databases to identify data of interest, and gets answers. The data can then be loaded in memory, plotted in various forms (images, spectra…), and are possibly sent to more elaborated tools performing specialized functions or processing. The various steps are commented below.

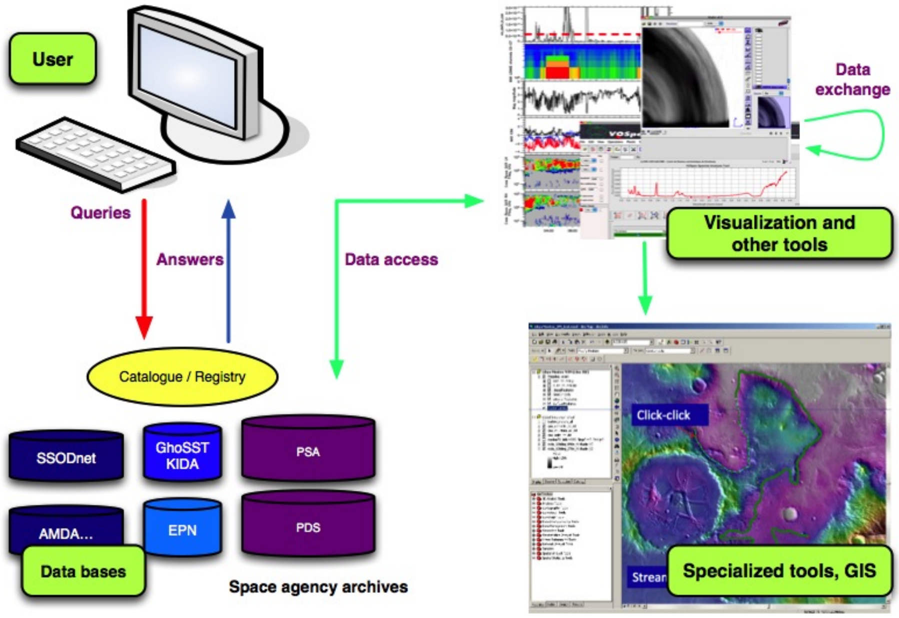

Figure 1: Overall scheme of the Planetary Science VO

DATA SCOPE

The perimeter of data to be accessed by the Planetary Science VO derives from the objectives stated in the Europlanet program proposal. It includes (Fig. 2):

- Data bases produced by various work packages during the Europlanet program (JRA4/task4).
- A selection of spaceborne data from planetary missions. This includes data from European space missions, i.e. access to ESA's Planetary Science Archive (PSA) [2].
- Specialized databases and tools related to participants of the EPN/IDIS activity are also linked to the system, e.g. GhoSST at IPAG/Grenoble, AMDA at CDPP/Toulouse or SSODnet at VO-Paris.
- Big data repositories including Planetary Science data and predating Europlanet are other natural targets to expand this system, e.g. the ESO and HST archives.
- Data sets directly published in a compliant form by data providers, typically as end product of a research activity, after scientific publication.

The Planetary Science VO is expected to be sufficiently open to allow external data providers to include their databases in the system with minimum efforts. This includes observational data derived from space missions or ground-based telescopes, but also reference data acquired in the laboratory and simulations.

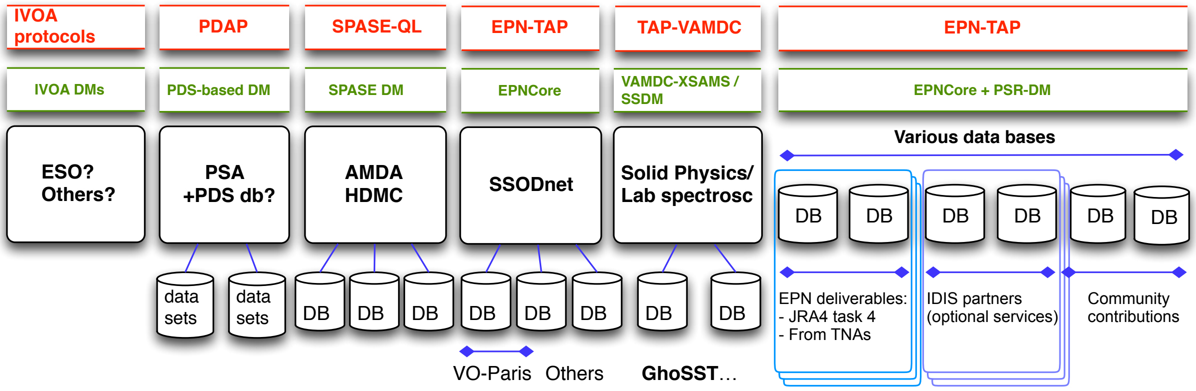

Figure 2: Data services expected to be included in the EPN VO,

with related access protocols (red) and data models (green)

DATA ACCESS PROTOCOLS / DATA MODEL

The user is expected to write queries describing his search. Such queries must be translated in a standard form and used to search a catalogue of available data. This procedure is called a Data Access Protocol in the IVOA framework.

Queries can be sent from a web form, from specialized software, or directly from visualization tools, as it is currently done in the IVOA system (e. g., from Aladin or TOPCAT). However, this requires the protocols to be implemented in the visualization tools, and therefore involves heavy developments. Several protocols are used in different contexts, in particular:

- IVOA protocols, which allow searching the data according to various criteria (i.e., Cone Search to locate objects near a position in the sky, TAP for tabular data…).
- The PDAP protocol, currently being defined in the IPDA framework. A typical use of PDAP is to address the entire contents of the PSA at ESA by querying their database globally, as a single data service [3].
- The Spase-QL protocol used in plasma physics, e.g., to access the AMDA service at CDPP.
- LineList is a VAMDC protocol to query spectroscopy databases.

These solutions were studied to figure out whether they could answer the needs of the Planetary Science VO. Most IVOA protocols appear to be related to objects located on the celestial sphere, and no solution currently exists to address data located e.g. on a planetary surface or in planetary atmospheres. Usability assessments of PDAP were performed at VO-Paris and CDPP [4]. The protocol proved able to address the data of interest, provided topical extensions and perhaps some adjustments; however, PDAP is closely related to an implicit Data Model associated to the PDS3 format, therefore adapted to observational spaceborne data only and providing little support for telescopic observations or laboratory data. Another drawback is that PDAP is not implemented in the IVOA visualization tools, as IVOA protocols are.

The outcome of this study therefore identified the need to develop a simple protocol based on existing ones, but including specific parameters permitting to handle Planetary Science data in general. The adopted solution was to define a restriction of the TAP protocol, called EPN-TAP. The TAP mechanism is here associated to a simple, specific Data Model (EPNcore). This Data Model is in practice a set of mandatory parameters describing a data set, which is implemented in all data services and used as query parameters by the protocol [5]. Most of these parameters are related to the description of the data axes (time, coordinates, spectral), target, measurement type, and origin of the data. The protocol also states that quantitative parameters have to be provided in standard units in the catalogue, and that non-quantitative parameters (such as instrument names) are associated with identified reference lists. Most of these reference lists originate from the IAU or other independent sources such as NSSDC. This standardization makes it possible to send uniform queries to all EPN-TAP services.

A more general Data Model has also been defined to describe the data of interest in a uniform way (Planetary Science Resource Data Model, or PSR-DM [6]). This Data Model could be used in the future with a more general protocol. Such applications should use common metadata defined in the data model and the associated dictionary.

Specific domains may not be included in a single data model though, and can be handled with different systems. In particular, other data models are being defined in relation with laboratory databases of solid spectroscopy (SSDM for the GhoSST data service) and atomic and molecular spectroscopy of gases (XSAMS), both in collaboration with the VAMDC consortium, which provides access to Atomic and Molecular Data.

As in Astronomy, a potential difficulty to write valid queries is related to the many names of potential targets/objects. But this problem is much worse for solar System objects, which cannot be identified by their position in right ascension / declination. A name resolver is therefore required to process user's queries correctly. Such a name resolver has been devised at VO-Paris as part of the SSODnet project, and is available as a web service.

CATALOGUE / REGISTRY

Queries are sent to a global catalogue containing a description of all accessible data services and their capabilities. This description allows the system to make a first order selection of services matching the user's query. In the IVOA, this is done using a system of mirrored registries where data providers can register their services. Declaring a new data service in the registry system is the normal way to make it available in the VO, and to publicize it. The PDAP 1.0 document mentions the possible use of a registry system, but the level of detail involved is still unclear. IVOA-like registries exist both at VO-Paris and ESAC. The VO-Paris registry is currently used to access EPN-TAP resources, and is the core of the Planetary Science VO system.

We call EPN Service Data Model the data model used to describe the content of the EPN registry itself, i.e. all data services accessible through the EPN registry. The solution adopted for the Planetary Science VO follows the IVOA scheme, which is easy to maintain.

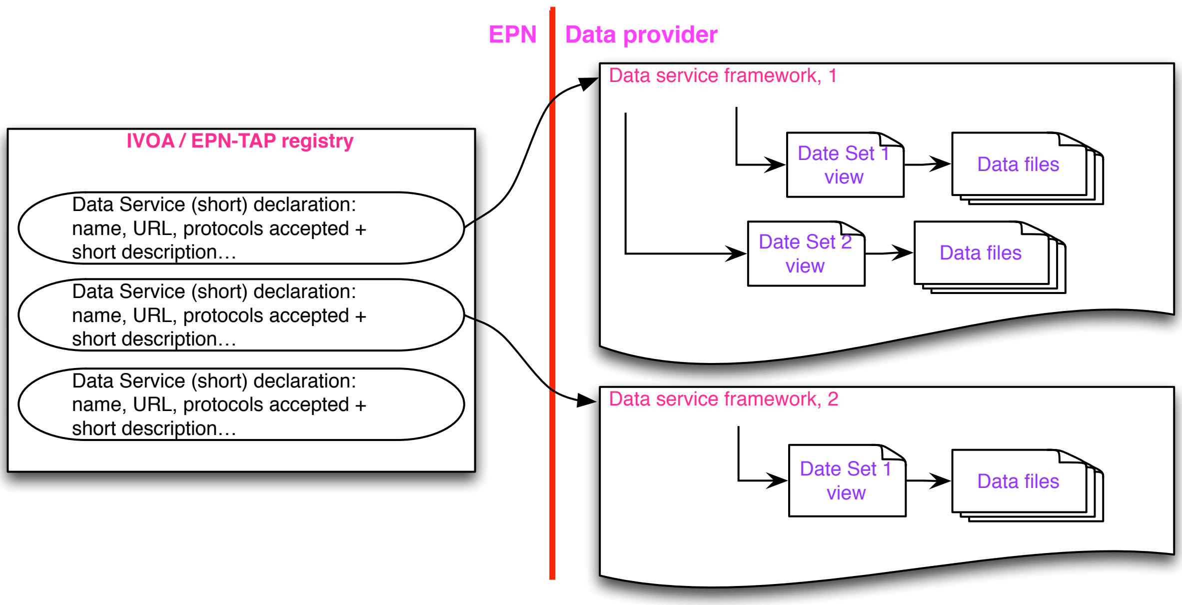

Figure 3: EPN registry and data services

The EPN registry only contains a short description of the available data services, including their address, and the protocols they support. The detailed description of the services is stored and maintained locally by the data providers. The queries are therefore processed at two successive levels (Fig.3): services of interest are identified at the registry level, the query is passed to them according to the protocol indicated, and is processed locally. In this context, a data service is a series of data sets located in a given place, accessible through the same protocol with content described in a unique, local, catalogue.

The registry system will need to be maintained on long time scales. Compatibility issues with other information systems are best anticipated by adopting the OAI-PMH standard when replicating the registry in a different institute. OAI-PMH is the widely used standard for metadata exchange, used in particular by the IVOA (this is a strong recommendation from the astronomy community). This standard includes the use of the so-called "Dublin core" metadata.

DATA ACCESS

The answer to a query is sent back by the service to the user. It typically provides a description, a link to the data, and information about data interface, but is in general limited to metadata and does not includes the data themselves. To be directly usable, this answer must also be formatted according to a known standard. This formatting can either be related to a data model (this is one of the solutions currently adopted by the VAMDC consortium) or remain at a general level (e.g. included in a VOTable). The EPN-TAP protocol states that the answer is returned in a VOTable, like PDAP and most IVOA protocols. In this case, a URL to the selected data is usually included in the VOTable. In some cases when the data are directly integrated in the database view, they can be embedded in the result VOtable and have to be parsed.

Reading the data can be practically difficult, given the wide variety of formats used in Planetary Science. A special problem arises with PDS3 data (used by all current space data archives), for which no versatile reader is available. Although PDS4 is expected to greatly simplify this issue in the future [7], availability of "historical" space data in PDS4 is an open question. PDS data of many types can be read on-line, and then sent to IVOA tools using a mechanism set up at VO-Paris (LecturePDS library ran in an IDL/GDL shell, then transmitted through a JavaScript interface, see Demonstrators below). IVOA tools handle some current data formats such as FITS or CSV, and TOPCAT now supports CDF.

TOOLS / VISUALIZATION

Many visualization tools for basic data types were developed in the IVOA framework. The baseline for Planetary Science is to use existing tools, which are maintained by external teams; possible adaptations are then reduced to interface layers, which may be discussed with the original software developers. The most flexible tools in this context appear to be Aladin and SAOImage/DS9 (for images and cubes), TOPCAT (for tables), SPLAT, VOPlot, SpecView, and VOSpec (for vectors and spectra). These IVOA visualization tools implement a data exchange protocol called SAMP (based on XML-RPC, or preferably HTTP). Once the data are loaded in one of those tools, they readily become available to other tools through this protocol.

Specialized data types however may require specific visualization and measurement tools, which may be part of the data service user's interface. For instance, spectral laboratory data can be handled with precision in the GhoSST interface provided by IPAG together with the database [8]. Similarly, AMDA includes a specialized environment for plasma and solar Physics [9].

Many planetary data need to be projected in a particular coordinate system. Different situations may occur in Planetary Science:

- Sky coordinates (used e.g. for telescopic images of a target on background sky). This situation is routinely handled by Aladin and similar tools [9]. The IMCCE Skybot service is accessible directly from Aladin to identify moving targets in a Field of View at a given time.
- Planetary coordinates (e.g. for orbital images of planetary surfaces). This is similar to geographical coordinates on Earth, although the coordinate systems are body-related. Apart from the geometric computation (which is expected to lie on the provider's side), plotting in such frames may also be an issue. High resolution imaging in particular requires a detailed description of planetary coordinates

frames, including control point networks, in the Data Model, which is currently not available. Converters between coordinate systems may also need to be developed in the long run.

Among the functions of interest for Planetary Science, accurate registration of imaging data has a special importance. A large fraction of our community works with Geographic Information Systems (GIS) to handle orthorectified data, either bitmap or vectorial. The GIS community has developed tools using standards elaborated in the framework of the Open Geospatial Consortium (OGC) which can be used for Planetary Science [10] [11] [12]. A demonstrator based on the Smart-1/AMIE images of the Moon is developed at VO-Paris.

Specialized tools for 3D plotting of solar System bodies and spacecraft trajectories are also available, e.g. 3DView (developed by CNES, used by the IMPEX program [14]) or MATISSE (developed by ASI [15]). They make use of Spice kernels to handle solar System body dimensions and motions.

LINK TO COMPUTING ENVIRONMENTS

Services on data may also include standard computation/inversion algorithms. The question arises of the environment used to perform such operations, and their interface to the Planetary Science VO. Several possibilities already exist:

- The Aladin Java plug-in system allows developers to implement basic operations on images (such as computing the average spectrum in a region of interest).
- IDL can exchange data with Aladin in some environments via SAMP, and VOTables are supported to a certain extent via user libraries. Conversely, the only “complete” support for PDS3 data is within IDL/GDL. Compatibility with IDL/GDL would give access to a very wide range of applications in public IDL libraries. This would also provide a link with ENVI, which is widely used for surface studies and imaging spectroscopy.
- The latest versions of IRAF include a VO interface (support for VOTable, registry, SAMP…).
- Some powerful libraries of image processing algorithms are available, such as the Orfeo Tool Box from CNES (developed for its Earth observation program). The French project Vahiné aims at providing a graphical interface to this library for remote sensing in Planetary Science [13].
- The IVOA has developed several systems to handle automated workflows, e.g. UWS [17].

EUROPLANET CLIENT

A global EPN-TAP client has been developed at VO-Paris, chaining all the functions described above [17]. It includes a user interface, queries/answers handling, and transmission to standard visualization tools through SAMP. Figure 4 illustrates these functions. The client first provides the user with an interface to send queries to the EPN-TAP registry. The latter returns a list of services possibly containing data of interest, with a description of their interface. The query is then transmitted to the various EPN-TAP data services; it is also translated as a PDAP request passed to preselected PDAP services. For each service, the answer is a VOtable providing a list of selected data files, which can be used either to restrain the query further, or to select data for quick-look. In this case, the data are transmitted to visualization tools via the common SAMP hub. The data files can also be passed to external, specialized environments for further processing. A current development is to convert the data from PDS to FITS whenever required. The SSODnet name resolver is used as a name completion service, both for solar System objects and exoplanets. Thanks to the VOSI mechanism, the client may query all the parameters describing a TAP resource, beyond the mandatory parameters of EPNCore.

The client currently accesses the Planetary Science services declared in the VO-Paris registry, plus the PDAP services at ESA/PSA and JAXA/DARTS (which are currently test services, and do not necessarily implement the entire PDAP protocol). The VO-Paris client is currently the only EPN-TAP client available, and is therefore the main entry point to the Planetary Science VO (it is also apparently the only public PDAP client). It is sufficient for users to search and retrieve data in the available databases.

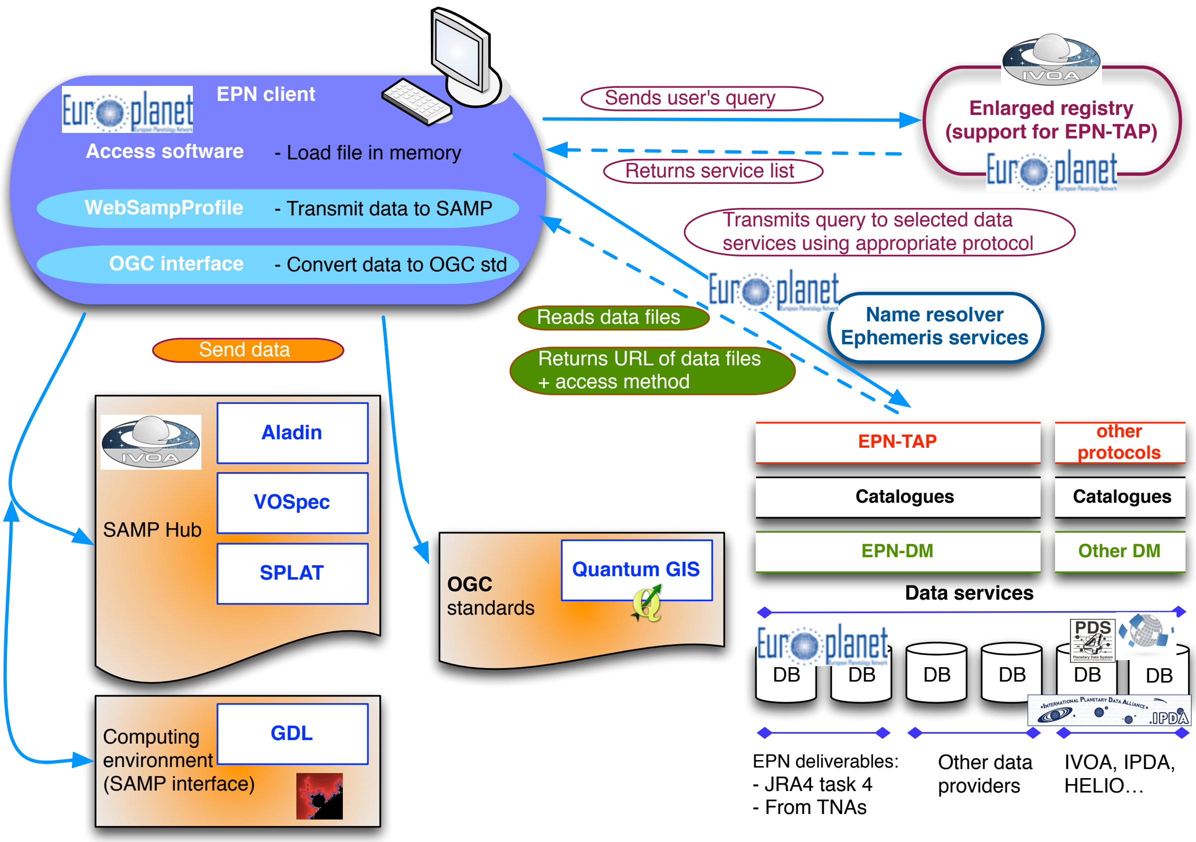

Figure 4: Overall data access through the EPN client. Logos indicate the origin of the various standards in use.

DATA SERVICES

Several EPN-TAP services are already publicly available: Auroral Planetary Images and Spectroscopy (APIS, from HST data), Vertical Profiles in Titan Atmosphere (from Cassini), BDIP (historical telescopic images), and the Extrasolar Planets Encyclopedia (comprehensive compilation of published data) at VO-Paris; the NASA dust catalogue at INAF/IAPS in Rome; AMDA uses the EPN-TAP protocol to retrieve data from APIS. Other projects are being finalized or studied.

New resources can be shared in a VO-interoperable manner by any data provider. The process is the following:

- Set up an SQL database + install a framework supporting the TAP protocol (DaCHS and VO-Dance have been used successfully). A tutorial to install such a system is available on VO-Paris EPN site.
- Ingest your data and create a view named epn_core describing all the “granules” (~individual data files). Pre-existing databases do not need to be changed, but this view must contain all the EPNCore mandatory parameters in the standard units. The descriptive parameters must be chosen from specific lists.
- The data themselves can be either linked through the “access_url” parameter, or included in the epn_core view. Data formats are preferably FITS, VOtable, CDF, or ascii/csv for convenient handling by plotting tools.
- Set up the framework to provide answers as a VOTable with compliant content.
- Fill up an XML file describing the data service and put it in the VO-Paris registry (see demonstrators below). Your service will be available from any existing EPN client, in particular the one at VO-Paris. It will also be accessible from TOPCAT directly (see doc on the VO-Paris EPN site).

A single DaCHS installation can accommodate several services; therefore grouping services may be a timesaving solution for small teams. Support can be provided at all these steps, e.g. efficient tools are available to write the XML descriptors (e.g., the CDPP demonstrator below), to check service validity, and to register new data services.

DEMONSTRATORS

Demonstrators of the above mechanisms are available from the VO-Paris IDIS node:

http://voparis-europlanet.obspm.fr

- The EPN client at VO-Paris is currently the main entry point to the system. It sends queries to all EPN-TAP services declared in the VO-Paris registry, and to PDAP services.

  http://voparis-europlanet-new.obspm.fr

  The “Custom resource” tab makes it possible to query a resource not yet declared, by providing the URL to its VO schema (this should also allow to query any TAP service). The “Advanced query form” link in the service description of the result page allows the user to query all parameters from this service.

- Access to PDS3 data has been demonstrated using VIRTIS/VEx spectral cubes. A cube is read in an IDL (or GDL) session on the server and stored in FITS, the reference is sent to Aladin and plotted in X/Y coordinates. In Aladin, spectra are extracted on mouse click and forwarded to VOSpec or (preferably) SPLAT for plotting.

  http://voplus.obspm.fr/samp/SAMPWebProfile+FITS/demo.php

- A service to test and validate VOTables and data services is also proposed at VO-Paris. It can be used as a support by external contributors to the Planetary Science VO who wish to share data services.

  http://voparis-validator.obspm.fr/

- The SSODnet name resolver can be used to search a solar System body by its different names, or by its coordinates at a given moment. The standard values are based on the IAU Minor Planet Center list.

  https://vo.imcce.fr/webservices/ssodnet/?forms

  The API is described here: http://api.voparis-tmp.obspm.fr/ssodnet/

Other demonstrators are available at CDDP:

- Access to planetary plasma data is implemented in the AMDA tool developed by the CDPP. Queries to distant databases (VEx/MAG in collaboration with IWF Graz, Cassini-MAPSKP…) are currently done using a SPASE-based connector:

  http://cdpp-amda.cesr.fr

- Access to external databases in AMDA has also been implemented using webservices (e.g., CDAweb). Current studies concern possible connections of PDS webservices with AMDA.

- A service to support writing of XML data file descriptors has been developed by CDPP. An interactive mode produces the XML files from a user-friendly web interface (which can be bypassed for pipe-line processing). This will soon be open for public use.

PROSPECTS

At the end of the Europlanet-RI program, the basic standards of the Planetary Science VO are defined and practical implementation has started. A registry system and a client have been developed to explore the data archives, and external VO tools are routinely used to provide on-line data visualization. Several test services are accessible and more are expected. A light framework and a procedure have been identified to allow small research teams to install data services, and two hands-on sessions have been organized in 2012 in Vienna and Paris. The next data services will focus on support to the on-going space missions, in particular Rosetta.

Some technical issues are still being worked out however, for instance to improve the interface with the client on the one hand, the SSODnet name resolver and ephemeris systems on the other hand. The use of UCDs and Utypes associated with measurement parameters is also still in discussion. More integrated PDS3 support is also being developed, as well as minor adaptations of existing VO tools to planetary data (e.g., plotting reflectance spectra may currently be difficult).

Several more fundamental points are also still under study to facilitate data mining, in particular concerning the referencing of coordinate systems in use for the solar System, and the identification of data sources (e.g., radio-telescopes and some orbital telescopes are not referenced by the IAU). Such points should be handled by a high-level authority such as IAU, or alternatively by consortia such as the IVOA or the IPDA. Planetary Science is more and more represented in IVOA activities, but the form this implication should take in the future is not yet very clear – either a liaison group with another entity (the IPDA or a « post-Europlanet » consortium), an « interest group » for Planetary Science, or direct participation of planetary scientists in many IVOA thematic working groups.

A VO-oriented sequel to the Europlanet program is currently considered in the Horizon 2020 framework. If this actually happens, the participants will focus on providing additional data contents and setting public outreach activities taking advantage of the VO system. New participants will certainly be welcome to contribute. Data providers, including small research teams, are also invited to consider sharing their data in this system – the more data are available, the more attractive the Planetary Science VO is expected to become to science users.

The EuroPlaNet-RI project was funded by the European Commission under the 7th Framework Program, grant 228319 "Capacities Specific Programme".

REFERENCES

[1] - Le Sidaner P. et al: Presentation of VO-Paris Data Centre and feedback on data archive. PV2013 Conference, ESRIN (2013)

[2] - D. Heather, M. Barthelemy, N. Manaud, S. Martinez, M. Szumlas, J. Vazquez, P. Osuna and the PSA development team: ESA's Planetary Science Archive (PSA): Status, activities and plans. 44th LPSC, LPI Contribution No. 1719, p.1930 (2013)

[3] - Planetary data access protocol (PDAP). IPDA draft 1.0 (16/4/2013):

https://planetarydata.org/projects/active-projects-2012-2013/PDAP%20Core%20Specification%20

[4] - M. Gangloff, S. Erard, B. Cecconi, P. Le Sidaner, C. Jacquey, J. Berthier, N. Bourrel, N. André, E. Pallier, J. Aboudarham, M.T. Capria, M. Khodachenko, N. Manaud, W. Schmidt, B. Schmitt, F. Topf, F. Trautan, A. Sarkissian, M.C. De Sanctis. An Assessment of the IPDA/PDAP protocol to access Planetary Data. EPSC-DPS Joint Meeting 2011, Nantes.

[5] - EPN-TAP documentation: http://voparis-europlanet.obspm.fr/xml/TAPCore/

[6] - EPN data model version 1.18a :

http://www.europlanet-idis.fi/documents/public_documents/Data_Model_v1.18a.pdf

[7] - L. Neakrase, L. Huber, S. Rees, M. Roybal, R. Beebe, D. Crichton, J. S. Hughes, M. Gordon, J. Mafi: The PDS4 archive: new structure, new possibilities. 3rd International dunes workshop, LPI Contribution No. 1659, id.2557 (2012)

[8] - B. Schmitt, D. Albert, P. Bollard, L. Bonal, M. Gorbacheva, P. Beck, E. Quirico and the SSDM expert working group: Last developments of GhoSST and its Solid Spectroscopy Data Model - The SSHADE project. 8th EPSC Meeting, London (2013).

[9] - Génot V.: CDPP, the French Plasma Physics Data Centre : available data, online analysis tools and future projects. PV2013 Conference, ESRIN (2013)

[10] - Marmo, C.: Development of a planetary Web GIS at the Photothèque Planétaire in Orsay. ADASS XXI Conference 2011, Paris.

[11] - Nass A., Van Gasselt S.: A platform for archiving and disseminating planetary geologic maps. 8th EPSC Meeting, London (2013).

[12] - Manaud N., Frigeri, A., Ivanov A.B: GeoMEx: Geographic Information System (GIS) Prototype for Mars Express Data. 8th EPSC Meeting, London (2013).

[13] - S. Douté, S. Girard; Vahiné web page, http://mistis.inrialpes.fr/vahine/dokuwiki

[14] - Génot V. et al: IMPEx : enabling model/observational data comparison in planetary plasma sciences. 8th EPSC Meeting, London (2013).

[15] - A. Zinzi, M. T. Capria, and L. A. Antonelli: MATISSE: Multi-purpose Advanced Tool for Instruments for the Solar System Exploration. 8th EPSC Meeting, London (2013).

[16] - UWS pattern: http://www.ivoa.net/documents/UWS

[17] - P. Le Sidaner, C. Chauvin, S. Erard, B. Cecconi, R. Savalle: EPN-TAP and EPN client: a good combination to access Planetary data. 8th EPSC Meeting, London (2013).