NNH15ZDA012L. Preparation for the Development of a Community-Based Roadmap for Planetary Data Services

The proposal has been submitted to NASA:

- NOI Number: N6-PDSRFI15-0021
- Submission Date: Mon Jan 25 19:26:19 EST 2016

Documents submitted

Cecconi-Baptiste-RFI.docx

Cecconi-Baptiste-RFI.pdf

Call Announcement

Response Date: 25 Jan 2016

NASA is preparing to work with the planetary science community to develop a Roadmap for its Planetary Data Services - the data and sample management architecture that supports the robotic exploration of the Solar System - particularly, the Planetary Data System (PDS). This Roadmap will address actions for the years 2017-2026. NASA is seeking input from the planetary and data science communities in the form of a Request For Information (RFI) designated NNH15ZDA012L that has been posted on NSPIRES:

http://nspires.nasaprs.com/

Select Open Solicitations, then search on NNH15ZDA012L.

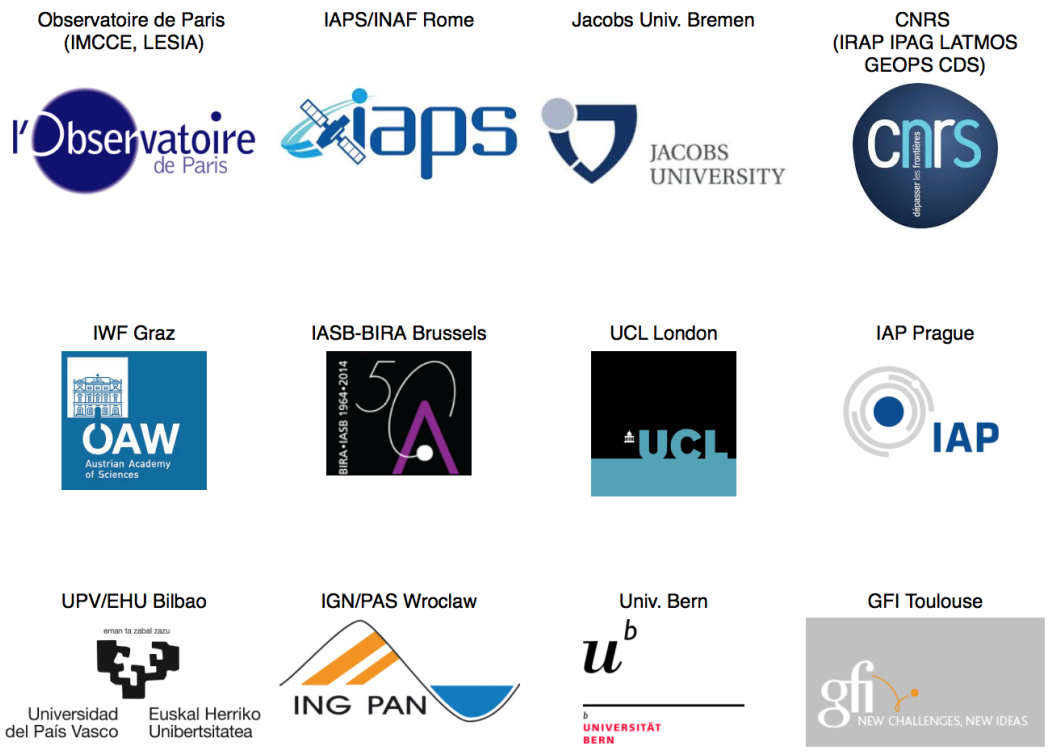
NASA especially seeks input on the following topics [edited]:

1. What tools, resources, workflows, tutorials, and interfaces
2. Making the archiving process seamless, less costly and more efficient
3. The role of PDS relative to other archiving alternatives (e.g., journals), in providing the public access to NASA-generated data
4. Integration of PDS data products and USGS, MPC, other products
5. The role should the PDS play in encouraging the development of higher-order data products
6. Identifiable improvements to the current search capabilities of the PDS

Introduction

VESPA (Virtual European Solar and Planetary Access) is an activity in the Europlanet 2020 Research Infrastructure programme funded under the European Commission's Horizon 2020 programme. It aims at building a Virtual Observatory for Planetary Science, connecting all sorts of data in the field, and providing modern tools to retrieve, cross-correlate, and display data and results of scientific analyses. The guiding principle of VESPA is to reuse existing technologies, so that minimal developments are required, and adoption of standards and new tools is made easier. VESPA is a common activity from 17 institutes in Europe, open to contributions from the community.

This document presents the VESPA project, and how these concepts can be used to enhance the science return of the NASA/PDS archive.



The Europlanet-H2020-RI/VESPA teams.

The VESPA Project

The Virtual Observatory (VO) is intended to make it easy to locate, retrieve, and analyse data from archives and catalogs worldwide, and it assumes that data is distributed rather than centralised. Thus, the Virtual Observatory is concerned with data discovery, data access, and data integration, the hallmarks of cyberinfrastructure projects.

The goal of VESPA (Virtual European Solar and Planetary Access) is to build a Virtual Observatory for Solar System Sciences.

The VO term covers two different meanings. It can be either “a virtual observatory”: a web-based portal providing access to remotely distributed data resources using online forms with scientific parameters; or “the virtual observatory”: a series of standards and interoperable tools that can share data transparently. In the first case, the user connects to a VO and search for data, while in the latter, the user is using tools to display data and the VO is the invisible machinery that allows him to work efficiently.

The VESPA project includes both aspects:

- It will provide the first multi-thematic Solar System and Planetary Science VO portal.
- It will also improve the connexions between existing VO tools taking into account the specificities of Solar System Science.

The Virtual Observatory in Planetary Science

This vision will tremendously increase the science return of the shared datasets. In the proposed infrastructure the shared datasets will be reachable either using the VESPA web-portal, or through existing visualisation tools in use in the science community. The scientists will be able to search for data using a simple interface and a series of limited scientific parameters. They will not have to worry about data location or data formatting. The system will provide equal access to all shared dataset. Hence, even small teams contributing to the VESPA effort will have the same visibility as large space agency databases in the system. The amateur community also has a place in this system as valuable data providers. All planetary and solar system science fields will be available through the same interface. It will thus allow a very efficient cross-fertilization between neighbouring fields. Several examples of use can be provided here:

- searching in a methan lake region of Titan, the user will be able to retrieve the atmospheric conditions depending on season or altitudinal density and composition profiles;
- looking for auroral images at Jupiter, the user will have also access to observations in the low frequency radio range, as well as thermospheric models, upstream solar wind or magnetospheric conditions;
- looking for spectral observations of comets, the user will also be able to get laboratory observations for various composition of ices.

The VESPA effort will also improve the overall VO efficiency by upgrading existing VO standards to adapt them to Solar and Planetary sciences. The main role of VESPA will be the addition of Planetary Science specific capabilities in existing data visualisation VO tools (TOPCAT, Aladin, CASSIS, AMDA, 3Dview…). VESPA will thus provide common data mining capacities, advanced visualisation, cross-comparison potential, and data analysis functions to all connected data services.

The VESPA project is connected to other data-access related efforts such as the astronomical VO (IVOA, International Virtual Observatory Alliance), the International Planetary Data Alliance (IPDA, which is a coordinated project to share planetary science archive from space agencies), or SPASE (Space Physics Archive Search and Extract). The VESPA core team members have been regularly interacting with these groups in the past 5 years. Hence all new developments will be done in coordination with existing standards, ensuring this way the project sustainability. The team also participated in many VO-related EU-funded FP7 programmes linked with solar system sciences such as Europlanet-RI, PlanetServer, HELIO, VAMDC and IMPEx in particular.

The goal of the VESPA project is to provide such an infrastructure, through a JRA (Joint Research Activity, i.e., developments) work package and a VA (Virtual Access Activity, i.e., implementation) work package, building on the prototype forged during Europlanet-RI (2009-2012). The JRA-VESPA will provide efficient visualisation and analysis tools to the Planetary Science VO, while the VA-VESPA will enlarge its content with new data services and by building a community of users and data providers, actually answering a growing demand in this field. The JRA-VESPA will also prepare new VO-compliant data services with particular interest and impact in several thematic fields of Solar System studies. The scientific community will participate to this project through workshops organised twice a year (at EPSC and EGU, the two major Planetary Science conferences in Europe), where VESPA will be showcased, tutorials will be proposed during hands-on sessions, feedback and needs from the user will be collected.

Evolutions

A first prototype was devised during the EU-funded programme Europlanet-RI (2009-2012), in the IDIS activity. VESPA has evolved and matured since the end of Europlanet-RI. Its development is now a part of the Europlanet 2020 programme. An increase by a factor of 5 of

the number of connected data services is expected during the programme, as well as the training of 15-20 teams in Europe.

Interoperability

Tools and interfaces

The main concept behind the VO is the interoperability. In Solar System Sciences, and more specifically in Planetary Sciences, there are already several interoperability standards used by the community. The 2 main ones are SPASE for Planetary Space Physics and GIS for Planetary Surfaces. VESPA is also reusing standards from the astronomical VO (IVOA), where they are not tied to sky coordinates. Being interoperable means being compatible with existing tools, which have been developed to fulfill the needs of a community. Hence databases and archives should be able to connect with these existing tools, either generic tools (e.g., TOPCAT) or discipline specific (e.g., AMDA, 3Dview...). The interoperability must be bidirectional:

- A visitor on a PDS node web portal should be able to send data from the web page to his preferred tools
- A tool user should be able to retrieve PDS data from the tool interface.

The first bullet is already being tested by the PDS-PPI node. On a data page, the user have access to a local menu allowing him to either view the data in a new page (the data is translated into a web page, using the label), view the data with a plotting library on a web page, view the data in Autoplot (a plotting tool developed for the Space Physics community), or view the data in TOPCAT (a plotting tool developed for the Astronomy community). The latter option is even more generic, as it implements the SAMP protocol (Simple Application Messaging Protocol). The data can then be sent to any SAMP-enabled tool (either web based to desktop tools).

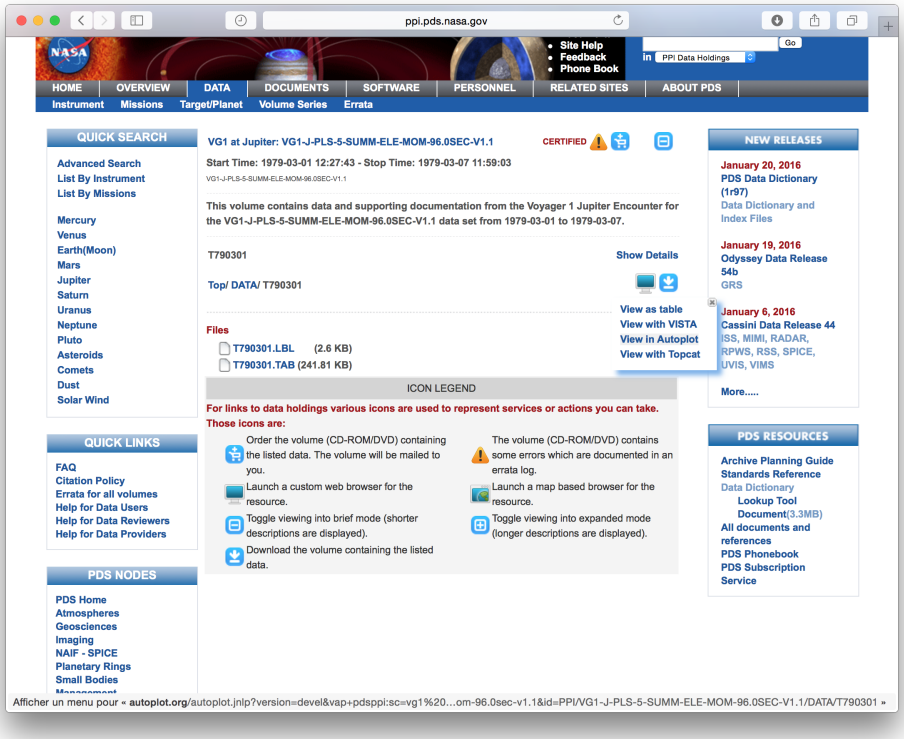


Figure 1. Screenshot of PDS-PPI node page, with the drop-down menu showing display options: view as table, view with VISTA, view in Autoplot and view with TOPCAT.

For the second bullet, the tools must act as clients to the PDS data server. This can most probably be done with PDS4 technology, but using existing interoperable standards (through RESTful interfaces or dedicated web services) would accelerate the adoption and connections by external developers. Developing of mappings between internal PDS4 discipline dictionaries and external data models (such as VESPA or SPASE) is really the key to efficiently implement interoperable search services.

As for VESPA, reuse of technology should be the guiding principle. But when there is nothing existing, new tools can be developed. Web GeoCalc is a good example for this. This online interface to the SPICE library is very helpful for data analysis, such as data selection based on viewing geometry. This tool would greatly benefit of little additions to include interoperable formats output and interfaces, such as:

- add VOTable output: VOTable is an XML-based table format, including formatted metadata, which is used in many disciplines (Astronomy, Space Physics, Heliophysics...)
- add SAMP interface: SAMP is a protocol to exchange messages between applications, such as data availability broadcasting (including data format signature, so that tools that understand the format, can download and display the data).

In Earth science, another community developed tool going in the same direction is OpenDAP. We have not rested this technology within VESPA yet, but it seems very promising. It provides a datacenter interface allowing to specify selection on data, operations on data and formatting output, in the RESTful query. This helps interoperability: the client sending the query can specify how the data should be formatted, organized and sampled, so that he can read them correctly. A similar development is in the PDS-PPI node project list, allowing to select the data format output at time of data download, with a conversion on the fly.

Workflows

Workflows are an other aspect of new capabilities to be developed in Planetary Sciences. Efficient workflow development and management requires interoperability. Having compatible input and output interfaces for workflow elements is really the key. Good workflow could nurture and/or ease cross-disciplinary studies, which is a key aspect in Planetary Sciences. It would also be very promising for data re-analyses with updated or recalibrated datasets. The main workflow engine developed in Europe, and used in several scientific domains (biology, astronomy, chemistry medicine, music, meteorology, social sciences...), is called Taverna. In the frame of the HELIO project, Taverna has been tested to Heliophysics studies with some success. In VESPA, we are going to study its application to Planetary Sciences. This workflow engine is connected a social-network-like web portal: myExperiment.org, where users are sharing their workflows to others. Such community-based helps a lot new-comers with easy to use already written and documented workflows.

Tutorials

Finally, last but not least the training of the community is essential. In the frame of VESPA, we are organizing splinter session during the two major Planetary Sciences conferences in Europe: EGU and EPSC. These splinter session are hand's-on session, where the participant are training on and discovering new tools. This is also the occasion to test the tools with new use cases, and gather the comments from the users. This workshop are helping adopting the new tools. Training workshops are held for user training as well as for data providers.

Dedicated workshops are also very useful. Last year, a workshop dedicated to planetary surfaces has been organized by PSA. This workshop was a great success. Many participants were already using GIS tools, but new comers really improved their understanding and experience on such tools. Discussions on how to interact with other communities (i.e., on being interoperable) was also very fruitful and some use cases were presented and discussed (e.g., mapping in-situ plasma measurements around Mars, Europa or Ganymede down to the surface and using GIS tools to plots the different maps).

In the US, to our knowledge, there are already similar workshops organized by USGS for planetary GIS tools, and CCMC for heliophysics modeling tools.

Better and easier archiving

The PDS4 infrastructure is going in the right direction. With the XML Schema and Schematron tools, scripting tools can automatically create or check product labels. These technologies are also making the archive itself more consistent, that the metadata are modeled more accurately. In order to facilitate even more the archiving and increase the quality of the archive, PDS should allow community-used formats. This kind of adoption must be preceded by an assessment of the archivability of these formats.

This has already be done for the CDF format, with restrictions on the CDF use, so that the files will be readable without a CDF library in the future. The use of CDF benefits to the user, the data provider, as well as to the archive quality process. The users already using CDF can go on using its fully developed libraries and tools that use CDF, and they don't have to write yet another reader every time a new dataset is archived. The data provider already using CDF can keep doing so and slightly adapt its processing pipeline to produce PDS4-compliant CDF files. Finally the benefit on the data quality is in the metadata of the CDF file. It is possible to add all relevant metadata in the CDF file itself, with predefined data modeling (as done for Space Sciences), so that the ingestion into the archive and the XML label generation can be automated.

Easier archiving could be proposed if the data provider is using a standard and acceptable data format. The archive specification may then concentrate on the file structure and metadata. If the data provider follows the specification, then the archive process can be rather straightforward, or even more, automated. In support of this proposition, we have been preparing two Cassini/RPWS higher order datasets, that have been delivered to PDS-PPI last summer. Those two datasets are using CDF as a file format, and a rather complete series of global attributes (metadata). The PDS4 labels of the data products have been generated automatically from each CDF file, using scripts developed by the PDS-PPI node team. The metadata included into those files are making them compliant with ISTP (Space Physics archive), PDS4 (Planetary Science Archive) and VESPA for easier distribution. In the case of VESPA, the metadata table used for data discovery use also generated automatically from the file global attributes.

Synergy with editors

The synergy between journal editors and data centers must be improved.

As an example, it is noticeable that in astronomy, the CDS (Centre de Données Astronomique de Strasbourg, France) is gathering all published data (bibliographic references, as well as data tables) for every astronomical object. This is a huge work initiated with Astronomy & Astrophysics. Now the authors submitting to A&A are encouraged and receive technical support to submit their data tables into a VO compliant format (including metadata).

It is probably more complex with planetary data (in astronomy, the RA and Dec coordinates are used to link observations), but this should be studied.

Higher-order data products and simulations runs

Interdisciplinary studies usually require higher-order data product, as they can be conducted by scientists from neighboring science domains (e.g., magnetospheric sciences and planetary sciences). These higher-order data product are derived from the PI teams, and their delivery to the community is a huge benefit to all. NASA/PDS should support as much as possible to sharing and archiving of these data products. The same way, NASA/PDS should also support as much as possible the participating scientists to share and archive the result of their project.

Together with higher-order data products, supporting data-products should also be considered, in particular, modeling run results. When the simulated data products is comparable to a observational data product, archiving should be rather straightforward. However, for propagation or transmission analysis, which are closer to a calibration than to an observation, archiving is probably more tricky. This should be assessed, and could end up with a "Planetary-CCMC" project. The European IMPEx project has studied the distribution of Simulation Runs, in the scope of Planetary Plasmas. The IMPEx model and infrastructure can be easily extended to any multidimensional modeling of a medium (atmospheres, interiors...). We plan to study this during the VESPA project.

NSPIRES Submission

The response will be submitted on the NSPIRES website, as a NOI (Notification of Intent). Baptiste Cecconi will submit for VESPA.

The submission material must be a pdf file (5 pages maximum), with a submission title. There can be several submissions from the same team, addressing various aspects/projects.