INTEROPERABLE MARS ATMOSPHERE DATA SERVICES

S. Erard, Z. Yin, **B. Cecconi**, *LESIA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, Univ. Paris Diderot, Sorbonne Paris Cité (stephane.erard@obspm.fr)*, **E. Millour** *LMD, IPSL, CNRS,* **A. Määttänen** *LATMOS/IPSL, UVSQ Université Paris-Saclay, UPMC Univ. Paris 06, CNRS,* **P. Le Sidaner**, *DIO, Observatoire de Paris / PSL Research University*

Introduction:

Modern space borne instruments produce huge datasets, especially on long-lived missions such as Mars-Express. This calls for new ways to handle the data, not only to perform mass processing, but also more basically to access them easily and efficiently. Virtual Observatory (VO) techniques developed in Astronomy during the past 15 years can be adapted to address this problem, provided they are enlarged to include specificities of Solar System studies such as coordinate systems, target-related time scales (local time and season), or measured quantities. An effort to adapt VO techniques to Solar System studies has been started in the frame of the Europlanet program, first in FP7 (2009-2012) as a demonstrator, and now in Horizon2020 (2015-2019) in a more extensive way. In the current program Europlanet 2020, the VESPA activity deals with the infrastructure and implements new data services. VESPA stands for Virtual European Solar and Planetary Access and supports all aspects of Solar System science. Its developments are essentially based on the standards of the IVOA (International Virtual Observatory Alliance).

As part of this enlargement, several data services related to the Martian atmosphere were recently designed and released, including a set of SPICAM derived data and a demonstrator of VO access to the Mars Climate Database. In both cases, vertical profiles are distributed with the basic VESPA architecture. In this abstract, we present the two services and assess this type of access to a simulation service.

Architecture:

The VESPA data access architecture [1] is based on a new data access protocol, a specific user interface to query the available services, and intensive usage of tools and standards developed for the Astronomy VO [2]. The Europlanet data access protocol, EPN-TAP, relies on the general TAP (Table Access Protocol) mechanism associated to a set of parameters that describe the contents of a data service [1] [3]. These parameters are defined to enable queries on quantities relevant to the scientific user, including observational and instrumental conditions, and are defined to handle the specific diversity and complexity of Planetary Science. Services are required to return answers formatted as VOtables, which are handled by all standard VO tools.

Data services are installed at their respective providers institutes and are declared in the standard IVOA registries, so that they are always visible and reachable from query interfaces. Although accessible in many ways, EPN-TAP data services are best queried from the optimized VESPA user interface: http://vespa.obspm.fr

In the frame of TAP, all data services present a list of granules (usually data files) described by a series of parameters. The Europlanet data access protocol, EPN-TAP, defines a set of mandatory parameters introducing metadata that describe all granules; this is similar to the ObsTAP protocol from IVOA, which describes observational datasets in Astronomy. EPN-TAP metadata introduce both observational and instrumental conditions: ranges on several axes (spatial, temporal, spectral, photometric), measurement type, origin of data, and various references. Location is provided in the most appropriate coordinate system (e.g., sky or planetary coordinates); target-related time (local time and season, through Ls) can be provided when relevant. The VESPA interface uses the mandatory parameters to search for individual granules in all registered data services at once, allowing for discovery of data content unknown to the user. In addition, specific parameters may also be used to describe services in more details, and can be used to identify granules more precisely when querying a single service.

All granules provide a link to a data file, or include the data itself in the table when possible (e.g., for scalar quantities). Data description parameters are used to identify adequate VO tools to access, plot and handle the data correctly. They not only provide a description of the file format, but also specify the dimensions, units, and physical quantities, relying on IVOA data models extended for VESPA. For instance, spectra and images are handled in different tools, and spectra measured in radiance or in reflectance are handled differently by the spectral tools.

Mars data services content:

Currently, two such data services are implemented for the Mars atmosphere. Profiles measured by SPICAM during the first Martian year of Mars-Express operations [4] [5] are distributed through EPN-TAP in a standard way. The SPICAM service contains 616 temperature & CO_2 profiles plus 433 O_3 profiles, with arbitrary spatial distribution. It will be completed with aerosols profiles in the next future

[6]. Simulated vertical profiles from the Mars Climate Database (MCD) [7] [8] are accessible in two different ways. First, as a standard EPN-TAP service providing situations sampled on a regular grid of coordinates, local time and season, for some of the available scenarios (a total of 62,244 profiles in this demonstration step). While the MCD supports quasi-continuous variations in input parameters, sampling is required to implement data access via the TAP mechanism, which is originally designed to provide access to catalogues. Although the resolution of the service cannot be maintained in these conditions, a trade-off has been searched between the physical size of the resulting data service and the accuracy of the results. A second VO access to the MCD is discussed below.

The profiles themselves are provided as VOtables, a handy format for this type of data that is directly recognized by VO tools. Parameters included in MCD profiles are temperature, pressure, densities of H_2O, O_3, and CO_2 computed in 50 layers up to 250 km altitude, plus column densities of water vapor and water ice, and dust opacity. SPICAM profiles include 3 different sets of temperature and CO_2 profiles in the 60-140 km range, derived under 3 different assumptions on the temperature at the top of the atmosphere [4]. SPICAM measurements of O_3 are available separately in the 20-70 km range [5]. The vertical scale is provided as radial distance and as altitude above both the areoid and the surface (i.e., including MOLA topography) for the two services.

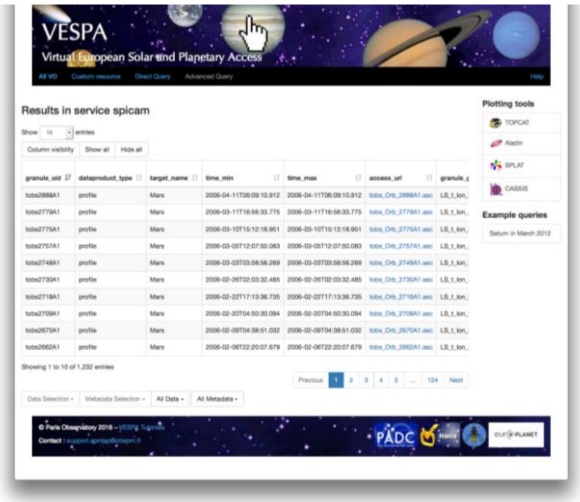

Fig 1: Result page for a query on SPICAM service

Search interface and basic tools:

The first function of the VESPA interface is of course to provide search capacities in the datasets; in this case, the obvious search parameters are location, season, local time, plus the scenario adopted for simulations. Since all VESPA services are queried simultaneously, this readily identifies corresponding situations in the observed and simulated datasets.

The distribution of answers to such queries can be visualized rapidly by plotting the metadata. Those can be sent from a service result page, either for all the answers or for a manual selection in this list; they are smoothly transferred from VESPA to VO tools according to the IVOA SAMP protocol, with no user intervention. MIZAR and Aladin immediately display the footprint of SPICAM profiles on a 3D view of Mars; TOPCAT can overplot other metadata such as local time or season on a 3D sphere.

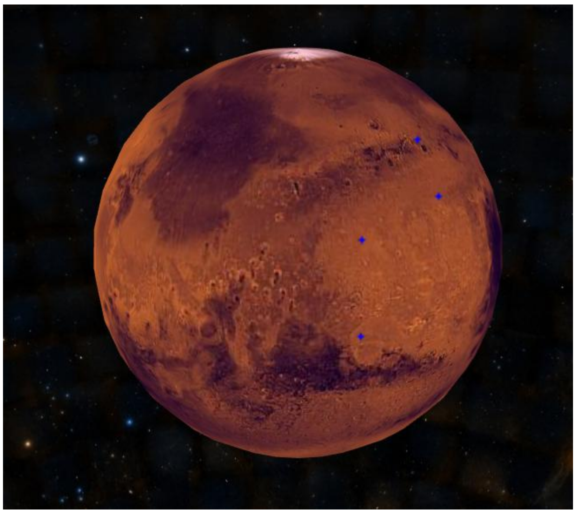

Fig 2: Footprints of selected profiles (punctual in this case) are directly plotted by Mizar in a browser

The data themselves can be transferred in a similar way. This allows the user to overplot observations and simulations with minimum manipulations in TOPCAT, the displaying tool of choice for profiles.

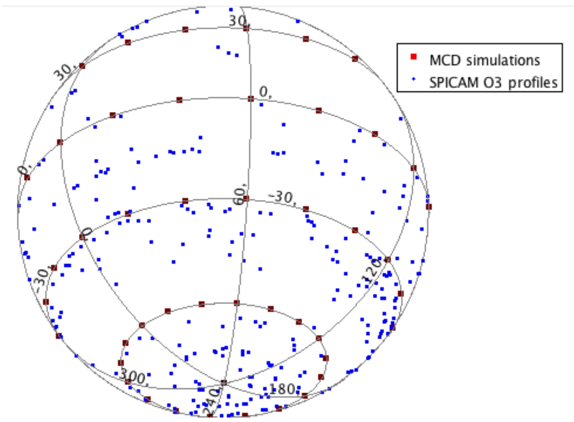

Fig 3: All footprints of SPICAM O_3 profiles plotted together with the current sampling grid of MCD in TOPCAT

Comparisons:

As seen on Fig. 3, the observations are actually grouped in a small part of the surface, which is only scarcely sampled in the current MCD service. The observations are even more closely grouped on the other two parameters, Ls and local time, so that finding a simulation really comparable to the measurement is difficult with the current sampling steps. For this reason, we're using here profiles extracted from the MCD with exact parameters, through the other mechanism described later on.

We're mimicking here a study of SPICAM and the MCD published in [4] and we're interested in comparing the two plots, regardless of the goodness of fit (Fig. 4). Local variations have been smoothed out in [4] by computing monthly averages in extended areas, which yields a reasonable match. Conversely, we're comparing single profiles as extracted from the VO services using the averages of the parameter ranges – this is a matter of minutes. Although this comparison validates the content of the services and query system, we notice that our fits are significantly poorer and probably less informative. This stresses the need to develop a procedure to average the retrieved profiles, which is currently difficult in TOPCAT.

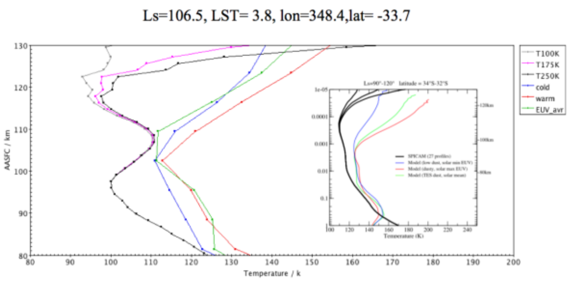

Fig 4: SPICAM profiles compared with MCD simulations: the large plot uses individual profiles retrieved from the VO services, the inlay from [4] uses averages.

Discussion:

Beyond the usual added value of VO access to observational data, this use case demonstrates the potential of comparing observations and simulations entirely in a VO environment. However, access to complex simulations via a strict TAP procedure appears to be restrictive; this requires sampling the parameters space and precomputing files corresponding to these situations. If the simulation uses many parameters, this is rapidly becoming inadequate: storing the files requires a large disk space but most of all, the details of the simulation are lost in the sampling procedure. This is clear in the current example: although the MCD sampling step is much too coarse to preserve the accuracy of the simulation (Fig. 5), it already results in a significant data volume (~ 100 GB of precomputed files to accommodate all scenarios at the current resolution). Although a moderately sampled EPN-TAP service such as the current one is valuable for public outreach and teaching purposes, increasing the resolution of the service is required to meet the needs of scientific users, say by a factor of 10 on each parameter, but this would lead to impracticable data volumes.

For this reason, a second VO-like access to the MCD has been implemented. This service launches a script that triggers on-line computation with the parameters provided in the query, similarly to the standard on-line interface of the MCD. The same script formats the resulting profile as a VOtable, which is directly ingestible in TOPCAT and other VO tools. TOPCAT provides an acceptable interface to this service, since it can call the script on the server and download the result. The current EPN-TAP service also makes use of this script system, which eliminates the need to store precomputed files and therefore allows for much smaller sampling steps (at the expense of more computation on the server side). Although the script can be called "manually" from VO tools, the EPN-TAP layer takes advantage of the querying syntax of the VO and TAP: it allows the user to search for various configurations of input parameters, and presents the results in a table identical to that of observational data.

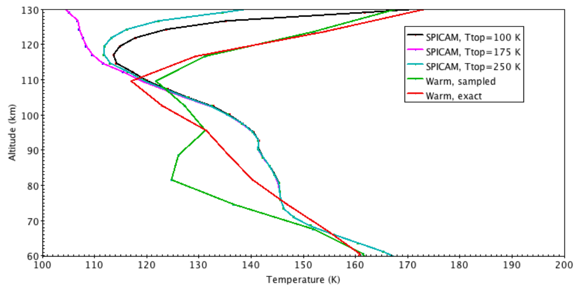

Fig 5: SPICAM profiles compared with MCD simulations: with exact parameters, or from closest sample. The sampled configuration is much further from the observations (Ls=13.2, LocalTime= 21.7, lon=247.8, lat= 44.4)

However, the available resolution of the EPN-TAP service is still somewhat limited by the table size (currently 500 MB, and could amount to some TB to be handled in postgreSQL). In order to preserve the flexibility of TAP access while maintaining the full resolution of the simulation, a different calling system is being studied. The next step will be to convert the EPN-TAP query issued by the VESPA interface into a direct call to the computing service. This would accept continuous values of the input parameters as the usual MCD interface does, or it could perform averages if ranges of parameters are provided. VESPA will experiment with scientific workflow engines already used by the IVOA, and study a possible implementation of this scheme. Other MCD outputs will be distributed in a similar way, e.g. local diurnal temperature curves.

APPENDIX - Other atmospheric services in VESPA

VESPA currently connects 27 data services in all areas of Planetary Science, Heliophysics, and Exoplanets, and about as much are being designed. In the area of planetary atmospheres, open services include SOIR profiles and VIRTIS calibrated data on

Venus, CIRS/Cassini profiles on Titan, and absorption cross-sections of various molecular species in the UV-visible range. Besides, a workflow engine will be implemented to allow for processing on demand. In addition to the MCD case described here, possible usages for planetary atmospheres include radiative transfer codes or inversion procedures applied to entire data services.

APPENDIX - VESPA prospects and development plan

VESPA compliant data services are expected to multiply in the coming years, to include data archives on one hand, and derived data on the other hand. Agreements or projects exist with ESA to make the whole PSA content available in this system; with PDS small bodies node to implement a bridge toward (recent) PDS4 archives, which could be extended to the other PDS nodes; with other space agencies which are also considering VESPA as an alternative way to make their data available. This will provide the science users with a handy way to search large datasets and easily identify data of interest, based on observational or operational conditions.

Distributing derived data is of course even more profitable to the community. Although this can be coordinated at the level of a whole mission, typically by the IDS (e.g., on Mars-Express or Cassini), research teams or individual researchers can also contribute and take advantage of the system directly by setting up a small data service to distribute results of analyses, e.g., from a published paper. All the VESPA documentation is freely available (although still partly in progress at this point). But most of all, VESPA opens a yearly call for new services; selected teams are invited to a 4-days implementation workshop where they will acquire the know-how, install their own data server, and design their service.

These calls are advertised in particular on the VESPA web site: http://www.europlanet-vespa.eu

Acknowledgements:
The Europlanet 2020 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654208.

Additional funding was provided in France by the Action Spécifique Observatoire Virtuel and Programme National de Planétologie / INSU.

References:

[1] Erard et al (submitted) VESPA, a Virtual Observatory infrastructure for Planetary Science and Solar System studies. Submitted to PSS special issue*, Nov 2016*.
[2] Erard et al (2014) Planetary Science Virtual Observatory architecture. *A&C* **7-8**, 71-80 http://arxiv.org/abs/1407.4886
[3] Erard et al (2014) The EPN-TAP protocol for the Planetary Science Virtual Observatory. *A&C* **7-8**, 52-61. http://arxiv.org/abs/1407.5738
[4] Forget et al (2009) Density and temperatures of the upper Martian atmosphere measured by stellar occultations with Mars-Express SPICAM, *JGR **114***, E1004, doi: 10.1029/2008JE003086, 2009
[5] Lebonnois et al (2006) Vertical distribution of ozone on Mars as measured by SPICAM/Mars-Express using stellar occultations, *JGR* **111**, E09S05, doi:10.1029/2005JE002643.
[6] Määttänen et al. (2013) : A complete climatology of the aerosol vertical distribution on Mars from MEx/ SPICAM UV solar occultations. *Icarus* **223**, 892-941, doi: 10.1016/j.icarus.2012.12.001.
[7] Navarro et al (2014). Global climate modeling of the Martian water cycle with improved microphysics and radiatively active water ice clouds. *JGR* **119**:1479-1495.
[8] Millour et al, this conference.