EPN-TAP: the VO standard to share and access Solar System data.

S. Erard[1], B. Cecconi[1], P. Le Sidaner[2], M. Demleitner[3], M. Taylor[4], [1]LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Univ. Paris Cité, 5 place Jules Janssen, 92195 Meudon, France [2]DIO-VO/UMS2201 Observatoire de Paris/CNRS, Fr, [3]Heidelberg University, Ge [4]University of Bristol, UK

Abstract

A new protocol has been adopted in the Virtual Observatory to share and access data related to the Solar System in the broadest sense. This relies on the Table Access Protocol (TAP) standard and a specific vocabulary called EPNCore used to describe quantities of interest for Planetary Science, heliophysics, exoplanets, and related laboratory measurements. It is described here, together with possible usages beyond the basic access to public databases.

Introduction: In the last two H2020 Europlanet programs, the VESPA (Virtual European Solar and Planetary Access) activity consists in setting up an infrastructure to both distribute and access data related to the Solar System. This encompasses observations in Planetary Science, heliophysics, exoplanets, as well as related simulations and laboratory measurements (Erard et al., 2018 & 2022a). The data infrastructure relies on FAIR principles, and is light enough so that small research teams can make their data accessible easily, therefore enforcing the Open Science policy for research data.

The Virtual Observatory (VO) was readily identified as the most promising framework, as it fulfills these functions for astronomy and has developed many useful standards and tools to display and analyze similar data. Several elements were however missing in the VO, in particular support for moving celestial objects, many body-related coordinate systems, measurements in reflected light and in-situ, particles, electro-magnetic fields, physical samples, etc.

The central missing part was therefore a vocabulary of metadata to describe regular quantities for Solar system and Planetary Science data. The VESPA team has developed such a vocabulary, called EPNCore, based on 60+ data services encompassing the many fields of Solar System and exoplanet studies. EPNCore associated to TAP forms the EPN-TAP protocol. TAP (the Table Access Protocol, Dowler et al 2019), is a standardized interface allowing clients to query the metadata and content of remote databases using an SQL-like language. EPN-TAP is now a Recommendation of the International Virtual Observatory Alliance (IVOA), and has therefore become the international standard to publish Planetary Science data in the Virtual Observatory (Erard et al., 2022b).

Construction of data services: EPN-TAP data services are responsive to the IVOA TAP standard, with additional rules: each service consists in a single metadata table. Each table row describes a single data element ("granule") which may consist in a data file, a set of scalar parameters, or a call to a web service. Each row is described by a set of columns, or parameters; such parameters are standardized in terms of name and unit, and have an associated UCD (Unified Content Descriptor, specifying a physical quantity in the VO) (Preite Martinez et al., 2023). The most common parameters are mandatory, i.e. the columns must be present in the table even though their content may be blanck, so that services remain interoperable. In addition, standardized optional parameters can be used to describe quantities of importance in specialized fields. This system allows the user to send basic queries to many services at once (so as to discover the content of published data services), then to refine the search for specific configurations in selected data services pertaining to the same field.

EPN-TAP data services are declared in the IVOA registry with indication of the access protocol installed. Altogether, EPN-TAP is similar to the ObsCore standard used to share measurements of celestial objects, and in practice several parameters are identical to, or inspired by ObsCore ones.

EPNCore parameters: However, ObsCore parameters are not sufficient to describe typical situations encountered with Solar System data. EPNCore therefore also uses concepts adapted from data archive consortium standards in this field: PDS for space mission data (originating from NASA and maintained by the International Planetary Data Alliance, IPDA), SPASE (Space Physics Archive Search and Extract), IHDEA (International Heliophysics Data Environment Alliance) – in addition to more general standards from the IAU (International Astronomical Union) and RDA (Research Data Alliance). Other standards and vocabularies from the IVOA are used wherever relevant.

There are currently ~180 defined EPNCore parameters which can be grouped in 4 broad classes: provenance; content; coverages; data access.

1) Provenance parameters include usual quantities such as publisher, bibliographic references, creation

and modification dates, etc. *instrument_host_name* identifies a telescope, space mission, or laboratory facility, while *instrument_name* refers to the instrument itself (the camera, spectrometer, etc in use). A set of parameters is dedicated to identifying contributions to collaborative works and coordinated observations.

2) Content parameters include a unique ID in the service (*granule_uid*) and two other IDs providing independent groupings: related to the same observation (*obs_id*) or to a similar type of product (*granule_gid*); *dataproduct_type* describes the data organization depending on the data axes and their nature (e.g. image, spectrum, etc); *target_name* and *target_class* refer to the object(s) of interest; *measurement_type* introduces a UCD describing the physical quantity(ies) accessible; *processing_level* refers to the data calibration. The pair of values (*granule_uid*, *service_title*) provides a unique ID to the data element in the VO, while *modification_date* provides a sort of version number. These parameters are intended for cross-service references and call-back processing, they can also be used for reproducibility purpose and to handle updates in evolving databases.

3) Coverage parameters specify the ranges along 8 main axes: time, spectral, spatial (3 coordinates), and viewing geometry (3 angles). Each range is defined by two parameters providing min and max values; resolutions in time, space and frequency are introduced by other pairs of parameters. Coverage parameters also define the reference frames in use, e.g., time and spatial origins, time scale and measurement position, etc. Spatial references are intrinsically complicated: the "flavor" of coordinate system is given by *spatial_frame_type* which defines the natures of the 3 spatial coordinates (celestial, body-fixed, Cartesian, spherical, etc) while the actual system in use (e.g., IAU_Mars, ICRS, etc) is provided though *spatial_coordinate_description*. Spatial frames follow some rules: e.g., body-fixed coordinates are always planetocentric, and units are associated to each axis depending on frame type. Similarly, spectral ranges are always provided on a frequency axis in Hz.

It is important to realize that axis requirements refer to the ranges provided in the EPNCore table describing the data, to allow for uniform queries on all services. But the data themselves can be provided on different scales, and do not need to be converted (e.g., spectral axis can be provided in wavelength, body-fixed coordinates as planetographic, etc). The complexity of dealing with the data is left to the tools and applications which actually handle them.

Since spatial coordinates limits only define a large bounding box, more accurate spatial footprints can also be provided using existing VO standards, which support powerful 2D searches for intersections, inclusions, etc. First, the *s_region* parameter defines a contour of interest, usually a polygon. Second, the *coverage* parameter introduces a MOC (Multi-Order Coverage) which is a list of healpix cells at various orders. MOC prove to be more reliable in this context, especially for orbital images which may spread a large area on the planet and can describe disconnected regions. *Coverage* can be extended to ST-MOC, which provide a time range together with a spatial footprint; this is particularly handy, e.g., to follow the operation plan of a space mission, and can be crossed with secondary time axes (local time and season).

Other axes are indeed available to describe local time, solar longitude (season), distances from the observer and the Sun, coordinates of the disk center and subsolar point, etc. The spectral axis can be completed with indication of the frequency band (IVOA standard *messenger*) and the (broadband) *filter* parameter with reference to the VO Filter Profile Service (Rodrigo et al., 2020).

4) Data access parameters are used when data are not contained in the metadata table itself. This latter solution is however recommended if it consists in a set of scalar values (e.g., orbital parameters for a catalogue of small bodies), since including the data in the table makes it possible to search on their values directly. The most common case is however to provide a link to a file through the *access_url* parameter – format and size can be specified with other parameters, like in ObsCore. Since *access_url* provides a full URL, the data files can be located anywhere on the internet, not necessarily on the same server. *thumbnail_url* can be used to associate a small preview in query clients, and *file_name* is available for situations where the name itself bears information on the data. In EPN-TAP, *access_url* can also provide a query to a web service, typically using a protocol from outside the VO; examples include OGC protocols (from the Earth Observation community) often used for planetary surfaces, and the das2 protocol (from the radio community) used for time series with adaptive resolution (Cecconi et al., 2020).

Although only one file can be associated to a table row, secondary information can be associated through the IVOA Datalink standard; this provides a table of links which can be selected for download in a workflow. Datalink can point to a documentation related to the entire dataset or to progenitor data products, and is often used to send a query to an ephemeris server with parameters from the current row.

EPNCore extensions: The main bulk of EPNCore consists in 56 mandatory parameters. All services are therefore responsive to queries on these parameters, which are sufficient to identify the target, type of observation, and the main coverages. In addition, ~120 extra parameters are predefined to provide more detailed information, e.g. on surface region, features, physical characteristic, observational and instrumental conditions, etc. Although EPN-TAP supports the use of service-specific parameters, quantities used by several services should be uniformly handled and normalized to preserve interoperability. We therefore expect new parameters to be defined on a regular basis as the content of EPN-TAP services enlarges.

Normalization of new parameters is designed to be flexible and rapid, in contrast to the complexity and time scale of IVOA review and adoption processes. Extensions pertaining to a consistent field of study can be defined by agreement between data providers and standard designers, based on a minimum of two data services. Extensions encompass the usual quantities required to describe a field which is not already covered by the standard. Current extensions include orbital/rotational properties of Solar System bodies, maps, particle spectroscopy, events, experimental spectroscopy, etc. They are published on a web page with a Permanent ID:
https://hdl.handle.net/21.15110/epn_tap_extensions

Extensions are available for use but are not formally included in the EPN-TAP standard until they are finalized. Future versions of the standard will include recent extensions for review by a larger community.

EPNCore parameter values: some EPNCore parameters only accept values from predefined lists, or follow standard construction rules. For instance, *dataproduct_type* or *target_class* must be selected from a closed list; other parameters must be selected from the IVOA rendition of the Unified Astronomy Thesaurus, e.g. *target_region*, or existing IVOA vocabularies, e.g., *messenger*. Difficulties start with *target_name*: values must be selected from the IAU reference lists, but those include many synonyms and changing designations, especially for small bodies; although the *alt_target_name* parameter can accommodate a list of alternative designations, the best way to handle this is to build queries using a name resolver on the client side – the quaero resolver supports all Solar System bodies (Berthier et al., 2023). A similar case where no exhaustive reference list is currently available relates to *instrument_host_name* and *instrument_name*; in this case, usual acronyms are recommended for the time being, and an action has been started in the IVOA to finalize such lists and resolvers. The list of coordinate systems in use is another example of such a situation.

Some string parameters can be multivalued, with values introduced as hash-separated lists. In addition to *alt_target_name* already mentioned, this is required, e.g., to describe composite observations from several instruments. In special cases such as *target_class*, multiple values can also be used to circumvent fuzzy identifications, in particular for transitional objects (*comet#asteroid*) or evolving classifications (*dwarf_planet#asteroid*). The point is of course not to enforce a scientific interpretation, but to make the data findable to potential users.

Service validators: the EPNCore vocabulary is supported by several tools in the IVOA. First, the DaCHS framework from the University of Heidelberg includes an EPN-TAP mixin predefining all EPNCore parameters with unit and UCD – mandatory parameter definitions are automatically included and optional ones when declared. Second, the taplint validator from the University of Bristol includes a special mode to verify EPNCore parameters from public or private data services (not only name, unit, and UCD, but also values and ranges). Finally, the service validation portal at Paris Observatory uses taplint to monitor all services declared in the IVOA registry (https://voparis-validation-reports.obspm.fr/). A special mode has been implemented to also check services included in the local registry of the VESPA portal, so as to monitor its consistency.

EPN-TAP clients: since EPN-TAP is a restriction of TAP, all TAP clients can be used to query EPN-TAP services — this includes taphandle, tapsh, python VO libraries (pyvo, astroquery, etc), and VO tools such as TOPCAT, CASSIS and Aladin. Such clients query the services one by one, and do not really take advantage of the uniformity of service descriptions.

The VESPA portal (https://vespa.obspm.fr/) is a dedicated client conceived as a discovery tool: by default, it builds a query from text fields and sends it to all available EPN-TAP services (Fig. 1). It uses the quaero name resolver for disambiguation and completion of Solar System target names. A local registry is used to maintain a selection of services which have been reviewed by the VESPA/EPN-TAP team, but the portal can also access any on-line service, given its URL.

The VESPA portal can display the answer of individual services (the rows of the metadata table responding to the query), or send it to TOPCAT which has the ability to cross-correlate tables from various services. When global queries are sent to all services, the portal also gathers together results from all responding services in a single metadata table, again for further use in TOPCAT or VO tools.

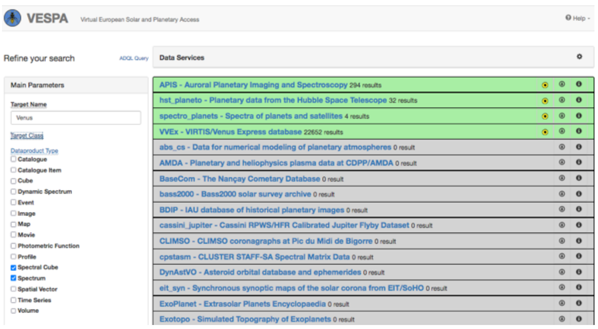

Fig. 1: The VESPA portal main query page: https://vespa.obspm.fr

An ElasticSearch database collecting the metadata of EPN-TAP services is also maintained in parallel to assess further cross-service capacities. It is already used internally to check the content of all services, public and in development. If flexible enough, it will become a standard access mode when several hundred services are published (Le Sidaner et al., 2023). Another application is to provide quick access to georeferenced data based on MOC footprints – including the extraction of regions of interest from HiPS (see below) and selection via converted ArcGIS shapefiles.

Finally, data products of interest selected in EPNCore tables can be forwarded to adequate VO tools for analysis and visualization, as identified by the *dataproduct_type* parameter. Functions dedicated to Solar System data have been included in TOPCAT, Aladin, CASSIS, AMDA, and 3DView in the recent years, e.g., to support radiance or reflectance spectra, planetary surfaces, atmospheres and magnetospheres, etc. TOPCAT now has the ability to include full spectra directly in the metadata table together with their footprints, and 60+ planetary HiPS (multiresolution maps) are available from Aladin (Fig. 2). More specific or higher-level processing can be installed on workflows platforms or Jupyter notebooks. Examples and tutorials are available here: https://github.com/epn-vespa/tutorials.

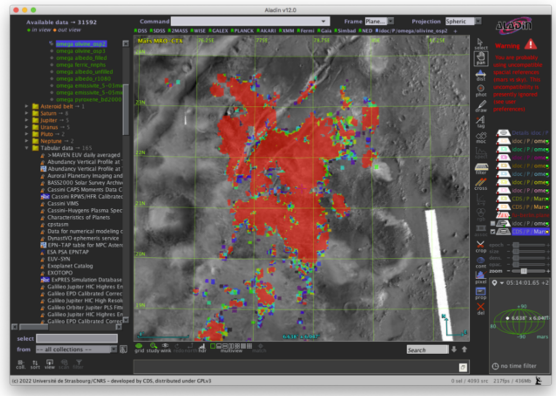
Fig. 2: olivine map from Mars-Express / OMEGA superimposed on MRO CTX HiPS in N Syrtis Major, as viewed in Aladin (Jezero crater is at the bottom)

Uses of EPNCore: The main use of the EPNCore vocabulary is of course to issue queries to EPN-TAP data services declared in the registry. Those include, at the time of writing, 64 services provided by Europlanet contributors and partner institutes. The largest ones include the Planetary Science Archive (ESA's repository of space mission data), while NASA's PDS Planetary Plasma Interactions (PPI) node is preparing their collections as individual EPN-TAP services (150+ of them, in review).

The objective is to get uniform access to archived data (raw and calibrated), but also to derived datasets, typically published by research teams in relation with a science article. EPN-TAP services are also a convenient way to fulfill the open access to the data required by programs publicly funded at national or European level. The use of EPNCore is therefore a good practice to organize any science work involving many data and observing configurations. Access to many astronomical data catalogues from published papers is granted by CDS (VizieR service), but planetary data are documented there with lesser accuracy; the VizieR_planets service describes those in more details using EPNCore.

Beyond access to individual repositories, EPNCore provides a common frame to integrate elements from various data systems, which is for instance the scope of NASA's Planetary Data Ecosystem (PDE) project (Milazzo, 2022). For instance, EPNCore provides a consistent way to expose Solar System-related content in telescope archives, which often already share astronomical data through TAP. Solar System content is usually difficult to identify from ObsCore or free-form TAP tables, mainly because it consists in moving objects. An EPN-TAP service has been set up for HST observations of planets and satellites to illustrate the usefulness of this approach. The same stands for experimental work: the SSHADE database of solid spectroscopy has an EPN-TAP interface in addition to its primary web interface (Schmitt et al., 2020). A coming EPNCore extension for bandlists will help identify observed spectral signatures from comparison with laboratory data.

An important requirement during the design phase was to address the particular situation of active instruments on board space missions. Those require a secure data management system which is often developed on the fly. EPNCore provides a standard frame to describe data catalogues consistently between ground calibrations and operations, and to track acquisition files and housekeeping parameters. EPN-TAP services can be installed behind an Authentication and Authorization Infrastructure (AAI) and the data files themselves can be protected during a proprietary period. Such systems would facilitate the production of public datasets expected by space agencies, and would provide an efficient off-the-shelf data management system to small missions (e.g., nanosats projects managed by Universities) or experimental facilities. The flexibility of EPNCore allows the use of PDS4 or FITS keywords for undefined quantities. A PDS4/EPNCore dictionary has been drafted as an IPDA action, and an extension of the FITS format and WCS (Word Coordinate System) to document observations of planetary surfaces has been published (Marmo et al., 2018).

EPNCore is also useful to support contributive work, in particular involving amateurs or citizen science. Such services are being designed in Europlanet 2024 to gather and share observations by large networks: the large PVOL database of amateur observations of the planets (Hueso et al., 2022); fireball networks; various pro-am collaborations supported by ESA; the PARSEC project to share data from the Europlanet Telescope Network (17 telescopes in the 0.4-2 m class across Europe, open to applications).

Service backup and maintenance: as part of the Europlanet VESPA activity, a preservation system has been set up for EPN-TAP services. It consists in a dedicated Gitlab at Paris Observatory, where server configurations and service definition files are stored. This area is protected by AAI, and is maintained by several teams. The Gitlab is used to review and comment the data services through issues, and to redeploy them if needed – for instance after a server crash. This greatly facilitates the maintenance of data services, e.g., to track non-conformities issues in older services when the final version of EPN-TAP was published.

EPN-TAP service publication: EPN-TAP is of course a standard available to the community, and any institute can publish their data independently: the infrastructure is designed so that the data remain in the research teams, close to the scientific expertise, and only need to be declared in the IVOA registry – the procedures have been documented to make it as light as possible. At least during the lifetime of the Europlanet program however, the VESPA team is happy to support new data providers and to help them make their services compliant, using the aforementioned Gitlab. In rare cases where a data server cannot be installed on site (e.g., because of local IT policies), a data service can be hosted by one of the VESPA contributors.

Prospects: The VESPA portal currently gives access to 64 data services. This number is expected to more than double in the coming year, with the publication of the current projects. This significant increase will of course impact the design of the VESPA portal, which is currently under study.

The EPNCore vocabulary will also be enlarged as new types of data need to be described. Possibly the most significant progress will relate to various lists of values in use: *instrument_host_name*, *instrument_name*, and *spatial_coordinate_frame*, which need both vocabularies and resolvers. Planetary Coordinate Reference Systems (CRS) in particular need to be made compliant with the OGC standards, so that the tools developed by the Earth Observation community can be used for planetary surfaces. A solution to encode IAU's planetary CRS has matured in the past years (Hare et al., 2018; Hare & Malapert, 2021), and is slowly progressing in OGC testbed reports. Planetary CRS definitions are available on this server: http://voparis-vespa-crs.obspm.fr:8080/ and will be included in astropy.

A next important step is to develop more workflows for mass processing of datasets, including machine learning algorithms, as started for radio observations (Louis et al., 2023). Such workflows can apply to various fields in the domain, in particular spectroscopy of surfaces and atmospheres.

Acknowledgements: The Europlanet-2024 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.
Support from Paris Astronomical Data Centre (PADC) is acknowledged.
VESPA web site: http://www.europlanet-vespa.eu

References:

Berthier J., Carry B., Mahlke M., Normand J. (2023) SsODNet: Solar system Open Database Network. *A&A 671, A151*. doi: 10.1051/0004-6361/202244878

Cecconi B, Loh, A, Le Sidaner, P, Savalle, R, Bonnin, X, Nguyen, QN, Lion, S, Shih, A, Aicardi, S, Zarka, P, Louis, C, Coffre, A, Lamy, L, Denis, L, Grießmeier, J-M, Faden, J, Piker, C, André, N, Génot, V, Erard, S, Mafi, JN, King, TA, Sky, J and Demleitner, M. (2020) MASER: A Science Ready Toolbox for Low Frequency Radio Astronomy. *Data Science Journal, 19(1), p. 12*. doi:10.5334/dsj-2020-012.

Dowler P., Rixon G., Tody D., Demleitner M. (2019) Table Access Protocol Version 1.1. *IVOA Recommendation, 27 Sept. 2019*. bibcode: 2019ivoa.spec.0927D. https://ivoa.net/documents/TAP/

Erard S, Cecconi B, Le Sidaner P, Rossi AP, Capria MT, Schmitt B, Génot V, André N, Vandaele AC, Scherf M, Hueso R, Määttänen A, Thuillot W, Carry B, Achilleos N, Marmo C, Santolik O, Benson K, Fernique P, Beigbeder L, Millour E, Rousseau B, Andrieu F, Chauvin C, Minin M, Ivanoski S, Longobardo A, Bollard P, Albert D, Gangloff M, Jourdane N, Bouchemit M, Glorian J, Trompet L, Al-Ubaidi T, Juaristi J, Desmars J, Guio P, Delaa O, Lagain A, Soucek J, Pisa D 2018 VESPA: A community-driven Virtual Observatory in Planetary Science. *Planet. Space Sc.*150, 65-85. ArXiv 1705.09727

Erard, S., Cecconi, B., Sidaner, P.L., Rossi, A.P., Tomasik, L., Ivanovski, S., Schmitt, B., André, N., Trompet, L., Scherf, M., Hueso, R., Demleitner, M., Manaud, N., Taylor, M., Alexeev, I., Määttänen, A., Millour, E., Waldmann, I., Fernique, P., D'Amore, M., Brandt, C., Rothkaeh, H., Molinaro, M., Génot, V. (2022a) Publishing Planetary Data In The Virtual Observatory: VESPA (Virtual European Solar & Planetary Access). *PSIDA conference 2022*.

Erard, S., Cecconi, B., Le Sidaner, P., Demleitner, M., Taylor, M. (2022b) EPN-TAP: Publishing Solar System Data to the Virtual Observatory, Version 2.0. *IVOA Recommendation, 22 August 2022*. bibcode: 2022ivoa.spec.0822E.
https://ivoa.net/documents/EPNTAP

Hueso R., J. Legarreta-Exegibel, A. Sánchez-Lavega, I. Ordoñez-Etxeberria, S. Erard, and I. Garate-Lopez (2022). Status of the PVOL image database and future prospects. *EPSC 2022, Granada, Spain*.
doi: 10.5194/epsc2022-854

Le Sidaner P., Azria C., Cecconi B., Chauvin C., Erard S. (2023) VESPA portal for access to heterogeneous data. PV2023, *this conference.*

Louis, CK, CM Jackman, SW Mangham, KD Smith, E O'Dwyer, A Empey, B Cecconi, P Zarka and S Maloney (2022) The "SPectrogram Analysis and Cataloguing Environment" (SPACE) Labelling Tool. *Frontiers in astronomy and space sciences, 9, 2022*. doi:10.3389/fspas.2022.1001166

Marmo C., T. M. Hare, S. Erard, M. Minin, F.-X. Pineau, A. Zinzi, B. Cecconi, A. P. Rossi (2018) FITS format for planetary surfaces: definitions, applications and best practices. *Earth and Space Science* **5**, 640-651. doi: 10.1029/2018EA000388.

Milazzo M. (2022) The Planetary Date Ecosystem: A Brief Summary. *2022 Meeting of Planetary Geologic Mappers, Flagstaff*. LPI Contribution No. 2684, id.7013

Hare T. M., Rossi A. P., Frigeri A., Marmo C. (2018) Interoperability in planetary research for geospatial data analysis. *Planet. Space Sc.* 150, 36–42. doi : 10.1016/j.pss.2017.04.004

Hare and Malapert (2021) Standards proposal for 2021 to support planetary Coordinate Reference Systems for open geospatial web services. 5th PSIDA conference. LPI Contrib. No. 2549

Preite Martinez A., Louys M., Cecconi B., Derrière S., Ochsenbein F., Erard S., Demleitner M. (2023) UCD1+ controlled vocabulary - Updated List of Terms, Version 1.5. *IVOA endorsed note 2023-01-25*. https://ivoa.net/documents/UCD1+/20230125

Rodrigo, C., Solano, E. (2020) The SVO Filter Profile Service.
https://ui.adsabs.harvard.edu/abs/2020sea..confE.182R/abstract

Schmitt B., Bollard P., Albert D., Bonal L., Poch O. (2020) SSHADE, its solid spectra databases and its future band list of molecular solids. *EPSC 2020*. doi: 10.5194/epsc2020-644.