Implementing and mapping the VESPA metadata dictionary in the PDS4

Baptiste Cecconi[1], Stéphane Erard[1], Pierre Le Sidaner[2], and Cyril Chauvin[2]

[1]LESIA, Observatoire de Paris, CNRS, PSL
[2]DIO, Observatoire de Paris, CNRS, PSL

November 23, 2022

Abstract

In the frame of the Europlanet-RI-2020 program, the VESPA (Virtual European Solar and Planetary Access) team has developed a metadata dictionary dedicated to data discovery in solar system sciences. The dictionary is called EPNcore and includes metadata for data coverage (temporal, spatial, spectral...), data content (target, measured parameter...), data origin (instrument, publisher...) and data access (format, URL...). In order to interoperate with the NASA/PDS4, the main archive information model in planetary sciences, it is first necessary to implement the EPNcore dictionary into the PDS4 information model. This is done with setting up a Local Data Dictionary (LDD). The second step deals with the mapping between terms already defined in the PDS4 information model and the EPNcore dictionary. We present the EPNcore LDD and a first version of the term mapping, as well as a prototype interface to query PDS4 from the VESPA portal. The Europlanet H2020 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654208.

Implementing and mapping the VESPA metadata dictionary in the PDS4

B. Cecconi[1], S. Erard[1], P. Le Sidaner[2], C. Chauvin[2]

(1) LESIA, Observatoire de Paris, PSL, CNRS, Meudon, France;
(2) DIO, Observatoire de Paris, PSL, CNRS, Paris, France.

Abstract

In the frame of the Europlanet-RI-2020 program, the VESPA (Virtual European Solar and Planetary Access) team has developed a metadata dictionary dedicated to data discovery in solar system sciences. The dictionary is called EPNcore and includes metadata for data coverage (temporal, spatial, spectral...), data content (target, measured parameter...), data origin (instrument, publisher...) and data access (format, URL...). In order to interoperate with the NASA/PDS4, the main archive information model in planetary sciences, it is first necessary to implement the EPNcore dictionary into the PDS4 information model. This is done with setting up a Local Data Dictionary (LDD). The second step deals with the mapping between terms already defined in the PDS4 information model and the EPNcore dictionary. We present the EPNcore LDD and a first version of the term mapping, as well as a prototype interface to query PDS4 from the VESPA portal.

The Europlanet H2020 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654208.

VESPA: uses TAP + EPNcore for data discovery.
PDS: NASA planetary science archive

Aim: build a VESPA-EPNcore keyword dictionary to allow the use of EPNcore keywords in PDS4

EPNcore keyword dictionary

https://voparis-wiki.obspm.fr/display/VES/EPN-TAP+V2.0+parameters

Name,	type	Unit / Format / Range	Description
Mandatory parameters			
granule_uid	Text		Unique ID in data service.
granule_gid	Text		Common to granules of same type
obs_id	Text		Associates granules derived from the same data
dataproduct_type	Text		Organization of the data product, from enumerated list
target_name	Text		Standard IAU name of target (must match target_class), case sensitive.
target_class	Text		Type of target, from enumerated list
time_min	Double	d (date as JD)	Start time (in JD). UTC measured at time_origin location (default is observer's frame)
time_max	Double	d (date as JD)	Stop time (in JD). UTC measured at time_origin location (default is observer's frame)
time_sampling_step_min	Double	s	Min time sampling step
time_sampling_step_max	Double	s	Max time sampling step
time_exp_min	Double	s	Min integration time
time_exp_max	Double	s	Max integration time
spectral_range_min	Double	Hz	Min spectral range (as frequency)
spectral_range_max	Double	Hz	Max spectral range (as frequency)
spectral_sampling_step_min	Double	Hz	Min spectral sampling step
spectral_sampling_step_max	Double	Hz	Max spectral sampling step
spectral_resolution_min	Double		Min spectral resolution (resolving power)
spectral_resolution_max	Double		Max spectral resolution (resolving power)
c1min	Double	(1) Longitude from 0. to 360. RA from 0. to 24.	Min of first coordinate
c1max	Double	(1)	Max of first coordinate
c2min	Double	(1) Latitude and Dec from -90. to +90.	Min of second coordinate
c2max	Double	(1)	Max of second coordinate
c3min	Double	(1)	Min of third coordinate
c3max	Double	(1)	Max of third coordinate
s_region	spoly	PgSphere spoly	ObsCore-like footprint in 2D (celestial or body-related frames)
c1_resol_min	Double	(1)	Min resolution in first coordinate
c1_resol_max	Double	(1)	Max resolution in first coordinate
c2_resol_min	Double	(1)	Min resolution in second coordinate
c2_resol_max	Double	(1)	Max resolution in second coordinate
c3_resol_min	Double	(1)	Min resolution in third coordinate
c3_resol_max	Double	(1)	Max resolution in third coordinate
spatial_frame_type	Text	(1)	Flavor of coordinate system, defines the nature of coordinates. From enumerated list. Use "none" if undefined.
incidence_min	Double	deg	Min incidence angle (solar zenithal angle)
incidence_max	Double	deg	Max incidence angle (solar zenithal angle)
emergence_min	Double	deg	Min emergence angle
emergence_max	Double	deg	Max emergence angle
phase_min	Double	deg	Min phase angle
phase_max	Double	deg	Max phase angle
instrument_host_name	Text		Standard name of the observatory or spacecraft
instrument_name	Text		Standard name of instrument
measurement_type	Text		UCD(s) defining the data
processing_level	Integer		Dataset-related encoding, or simplified CODMAC calibration level
creation_date	Timestamp	(ISO-8601 String)	Date of first entry of this granule
modification_date	Timestamp	(ISO-8601 String)	Date of last modification
release_date	Timestamp	(ISO-8601 String)	Start of public access period (set to creation_date if no proprietary period)
service_title	Text		Title of resource = schema name
Common optional parameters			
access_url	Text		URL of the data file, case sensitive (additional files may be linked through datalink_url). Can point to a script. If present, next 2 parameters must also be present.
access_format	Text	(mime type in lowercase)	File format type
access_estsize	Integer	kbyte	Estimate file size in kbyte (with this spelling)
access_md5	Text		MD5 Hash for the file when available (real file, not script)
thumbnail_url	Text		URL of a thumbnail image with predefined size (png ~200 pix, for use in a client only)
file_name	Text		Name of the data file only, case sensitive
datalink_url	Text	(url)	Provides links to files or services on the server
species	Text		Identifies a chemical species, *case sensitive*
waveband	Text		Electro-magnetic band, from enumerated list
alt_target_name	Text	can be a hash list	Provides alternative target name(s)
target_region	Text		Type of region or feature of interest
feature_name	Text		Secondary name (e.g. standard name of a region of interest)
publisher	Text		Resource publisher
bib_reference	Text		Bibcode or doi preferred; can be a URL or anything else
internal_reference	Text		Related granule_uid(s) in the current service
external_link	Text	(url)	Web page providing more details on the *granule*.
spatial_coordinate_description	Text		ID of specific coordinate system and version / properties
spatial_origin	Text		Defines the frame origin
time_origin	Text		Defines *where* the time is measured (e. g., ground vs spacecraft). Default is observer's frame.
time_scale	Text		Always UTC in *data* services - from enumerated list
subsolar_longitude_min	Double	deg	Minimum sub-solar point longitude
subsolar_longitude_max	Double	deg	Maximum sub-solar point longitude
subsolar_latitude_min	Double	deg	Minimum sub-solar point latitude
subsolar_latitude_max	Double	deg	Maximum sub-solar point latitude
subobserver_longitude_min	Double	deg	Minimum sub-observer point longitude (sub-Earth for ground based observations)
subobserver_longitude_max	Double	deg	Maximum sub-observer point longitude (sub-Earth for ground based observations)
subobserver_latitude_min	Double	deg	Minimum sub-observer point latitude (sub-Earth for ground based observations)
subobserver_latitude_max	Double	deg	Maximum sub-observer point latitude (sub-Earth for ground based observations)
ra	Double	deg	Right ascension
dec	Double	deg	Declination
radial_distance_min	Double	km	Min distance from observed area to body center
radial_distance_max	Double	km	Max distance from observed area to body center
altitude_fromshape_min	Double	km	Min altitude of observed area above shape model / DTM
altitude_fromshape_max	Double	km	Max altitude of observed area above shape model / DTM
solar_longitude_min	Double	deg	Min Solar longitude Ls (location on orbit / season)
solar_longitude_max	Double	deg	Max Solar longitude Ls (location on orbit / season)
local_time_min	Double	h	Min local time at observed region
local_time_max	Double	h	Max local time at observed region
target_distance_min	Double	km	Min observer-target distance
target_distance_max	Double	km	Max observer-target distance
target_time_min	Timestamp	(ISO-8601 String)	Min observing time in target frame
target_time_max	Timestamp	(ISO-8601 String)	Max observing time in target frame
earth_distance_min	Double	au	Min Earth-target distance
earth_distance_max	Double	au	Max Earth-target distance
sun_distance_min	Double	au	Min Sun-target distance
sun_distance_max	Double	au	Max Sun-target distance

NB: Draft EPN-TAP v2.0 IVOA standard in preparation. It will be available in: https://www.github.com/ivoa-std

NASA/PDS4 implementation

- Local Data Dictionaries (LDD): extension of PDS4 IM. Possibility to add new concepts into information model.
- LDD = XML schema + Schematron rules. LDD is built with LDDtool: input = simple XML file
- See here: http://sbndev.astro.umd.edu/wiki/Installing_and_Configuring_LDDTool

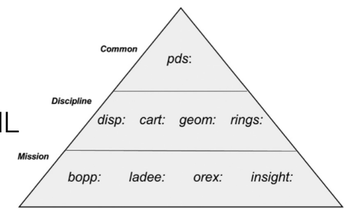
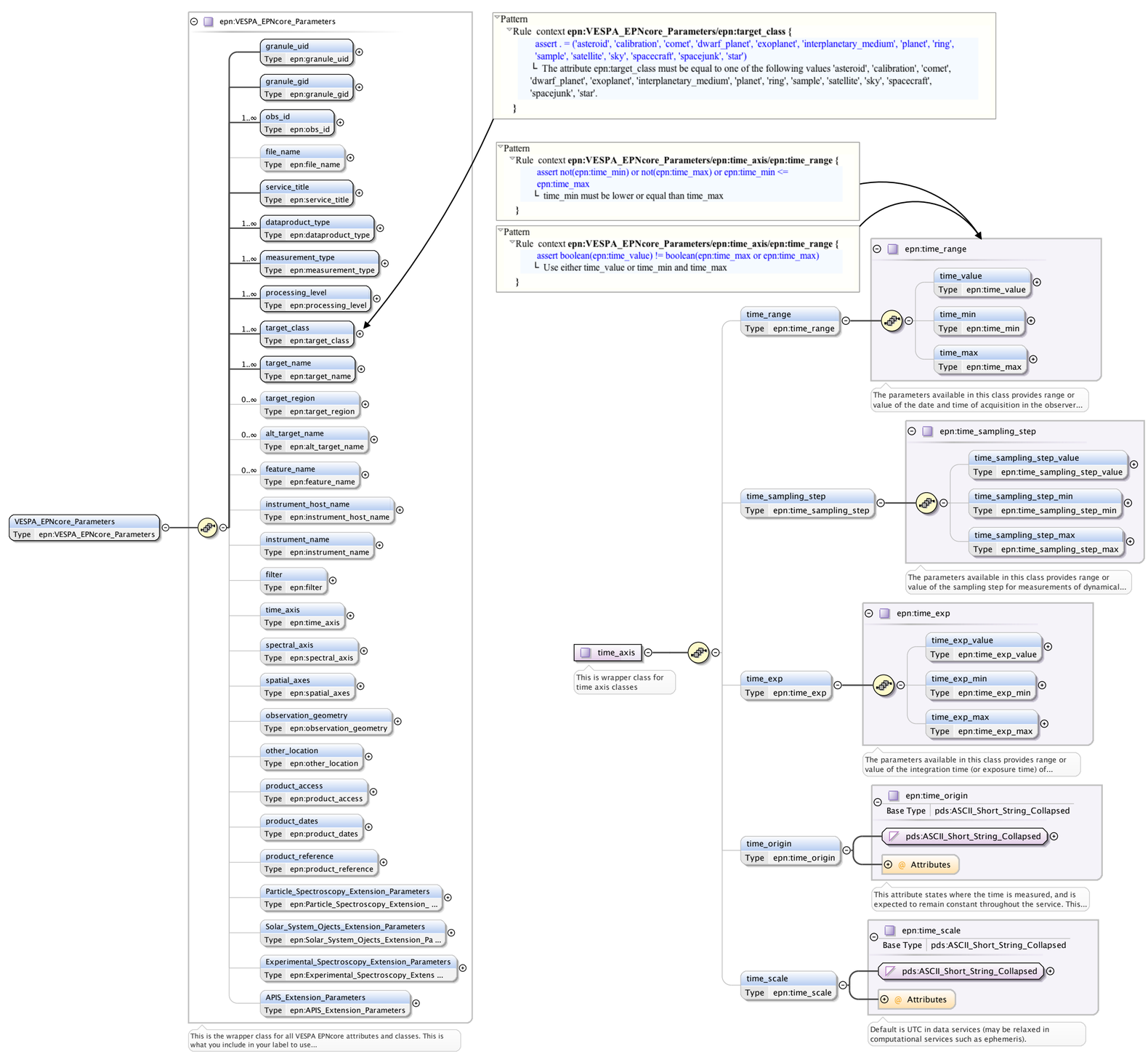


Fig. 1. - Multi-level Governance in the PDS4 Information Model.

The EPN VESPA LDD propose a nested architecture corresponding to EPN VESPA attributes classification in the EPN VESPA documentation.



Example: APIS 1D Spectrum from HST

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-model href="Ingest_LDD_VESPA_EPN_1x8000.sch" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<VESPA_EPNcore_Parameters xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns="http://pds.nasa.gov/pds4/epn/v1"
    xsi:schemaLocation="http://pds.nasa.gov/pds4/epn/v1 Ingest_LDD_VESPA_EPN_1x8000.xsd">
    <granule_uid>o4bda5kaq_1D_pdf</granule_uid>
    <granule_gid>1D_spectra</granule_gid>
    <obs_id>o4bda5kaq</obs_id>
    <service_title>APIS</service_title>
    <dataproduct_type>sp</dataproduct_type>
    <measurement_type>phot.flux.density;em.UV.100-200nm</measurement_type>
    <processing_level>5</processing_level>
    <target_class>planet</target_class>
    <target_name>Saturn</target_name>
    <target_region>magnetosphere</target_region>
    <feature_name>planetary aurorae</feature_name>
    <instrument_host_name>hst</instrument_host_name>
    <instrument_name>STIS</instrument_name>
    <filter>52X0.5</filter>
    <time_axis>
        <time_range>
            <time_min unit="julian day">2450787.7134216</time_min>
            <time_max unit="julian day">2450787.7249787</time_max>
        </time_range>
        <time_exp>
            <time_exp_value unit="s">1000.1899</time_exp_value>
        </time_exp>
    </time_axis>
```

Mapping: Terminology_Entry_SKOS

```
7
8     <!-- VESPA_EPNcore Attributes -->
9
10    <!-- [[ epn:granule_uid ]] -->
11
12    <!-- epn:granule_uid is a _exactMatch_ to pds:Product_Components/Identification_Area/logical_identifier -->
13    <Terminological_Entry_SKOS>
14        <identifier>epn.granule_uid</identifier>
15        <namespace_id>epn</namespace_id>
16        <steward_id>epn</steward_id>
17        <title>granule_uid</title>
18        <description>Unique granule ID in collection: logical_identifier of a Product</description>
19        <referenced_identifier>logical_identifier</referenced_identifier>
20        <skos_relation_name>exactMatch</skos_relation_name>
21        <model_object_id>0001_NASA_PDS_1.pds:Product_Components/Identification_Area/logical_identifier</model_object_id>
22        <model_object_type>attribute</model_object_type>
23    </Terminological_Entry_SKOS>
24
```

Next Steps

- Extended check on real cases, available on VESPA and planned for archive at NASA/PDS4
  E.g.: Cassini/RPWS/QTN Higher level data collection

  => goal: use VESPA_EPNcore_LDD on this data collection and prototype query of EPNcore keywords using PDS4 search interface.
- Build PDS4 “mapping” product between EPNcore and PDS IM (and other LDDs)
  => include SKOS into Ingest_LDD?
- Improve search VESPA interface, with NASA/PDS4 registry query using EPNcore keywords.
  NASA/PDS4 registry already available in VESPA page with a limited set of keywords, please check: http://vespa.obspm.fr
- Leads to improved specification of EPNcore model