

Connecting EPN-TAP and PDS4

S. Erard and the VESPA / Europlanet team

Observatoire de Paris-PSL

IVOA Interop, Malta Nov 14-16, 2024



Europlanet 2024 RI has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149

Space missions archives

- Stored / preserved by space agencies
- Common archive format = PDS4 (new missions \geq 2015)
 - Data products are grouped in datasets (\approx collections)
 - Associated documentation included
 - (most archives are still in PDS3, migrating)
- Variations in level of accessibility, in particular granularity
 - ESA at file level (TAP & EPN-TAP interfaces)
 - NASA PDS at collection (dataset) level
 - currently no PDS registry, and no complete catalogue in a dataset

=> how can we search for a particular target / configuration in PDS?

Space missions archives

- PDS4:
 - Data file + separated xml label
 - Labels include keywords from dictionaries
 - Many (≈ 60) dictionaries related to science field / PDS node, etc
 - Data file may be a PDS object or a standard format (including fits)

Q: can we provide an interface to these collections
at datafile level via an EPN-TAP table?

i.e.: can we build an EPN-TAP table from PDS4 information?
(of course, ESA is doing that somehow...)

Would allow detailed searches for configurations, and cross-searches

Assessment study

- Naive approach, based on 2 examples at NASA PDS Small Bodies Node
 - Two limited collections of asteroid spectra, with slightly different formatting:
ascii tables + xml labels + global tables with labels
 - Does not cover all possible situations

=> feasible, but not straightforward

<https://vespa.obspm.fr> (or via astropy)

services spectro_m_ast

spectro_trojans

service files available at:

https://voparis-gitlab.obspm.fr/vespa/dachs/services/padc/voparis-tap-planeto/spectro_m_ast
(via EduTEAMS or ORCID)

Assessment study

1) Read/parse individual labels

Not using PDS4_tools (xml.etree instead)

Only some (~10) global EPNCore parameters can be mapped from labels

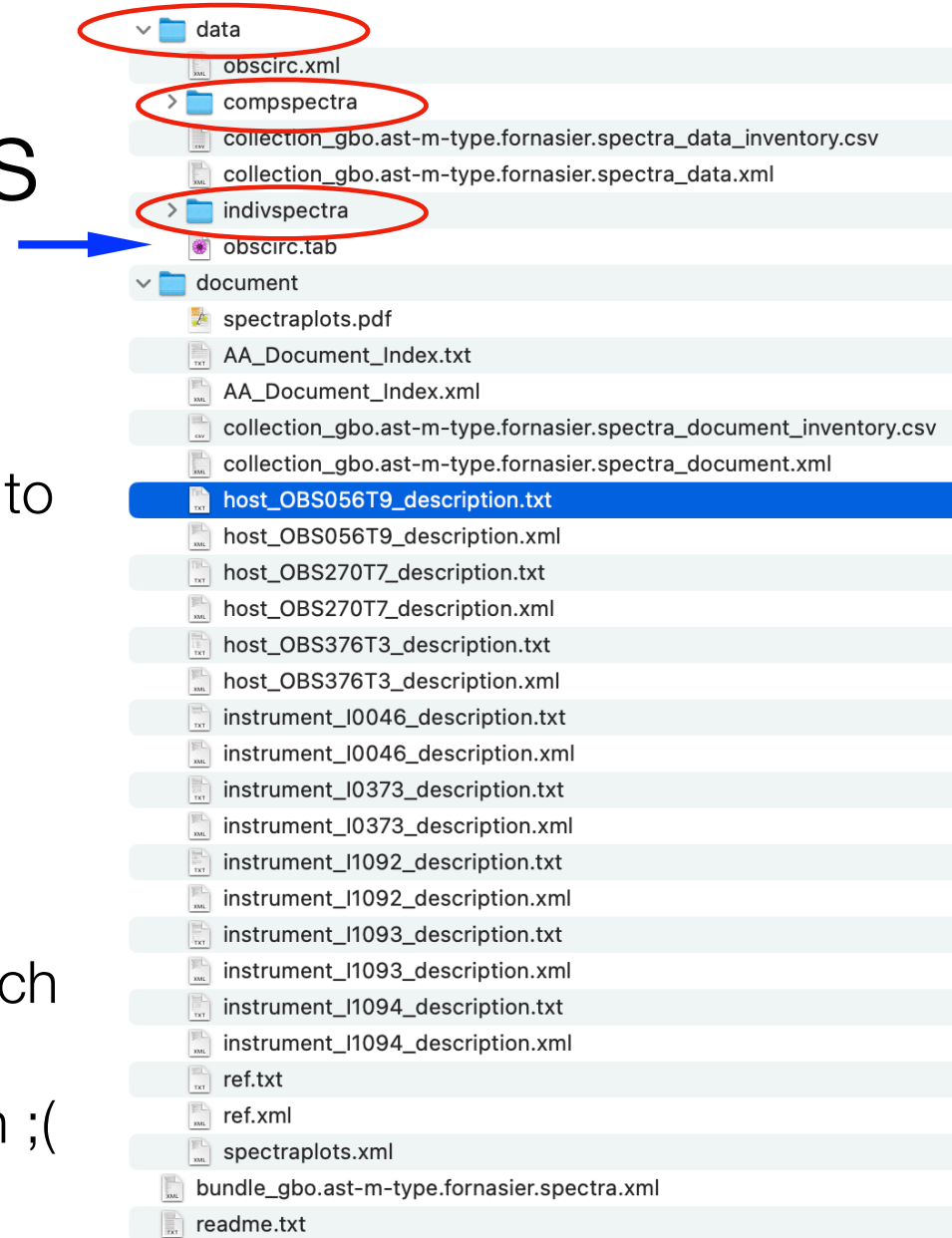
2) Add extra metadata from additional catalogues

```
... </Facet/Linked / Facet/...
... </Science_Facets>
... </Primary_Result_Summary>
... <Investigation_Area>
...   <name>None</name>
...   <type>Individual Investigation</type>
...   <Internal_Reference>
...     <lid_reference>urn:nasa:pds:context:investigation:individual:none</lid_reference>
...     <reference_type>data_to_investigation</reference_type>
...   </Internal_Reference>
... </Investigation_Area>
... <Observing_System>
...   <Observing_System_Component>
...     <name>DOLORES (Device Optimized for LOW RESolution)</name>
...     <type>Instrument</type>
...     <description>The description for DOLORES (Device Optimized for LOW RESolution) can be
document collection for this bundle.</description>
...     <Internal_Reference>
...       <lid_reference>urn:nasa:pds:context:instrument:orm.galileo_zeiss3m5.dolores</lid_re
...       <reference_type>is_instrument</reference_type>
...     </Internal_Reference>
...   </Observing_System_Component>
...   <Observing_System_Component>
...     <name>Roque de los Muchachos Observatory</name>
...     <type>Observatory</type>
...     <description>The description for Roque de los Muchachos Observatory can be found in t
collection for this bundle.</description>
...     <Internal_Reference>
...       <lid_reference>urn:nasa:pds:context:facility:observatory:orm</lid_reference>
...       <reference_type>is_facility</reference_type>
...     </Internal_Reference>
...   </Observing_System_Component>
...   <Observing_System_Component>
...     <name>3.5-m Galileo Zeiss Ritchey-Chretien altazimuth reflector</name>
...     <type>Telescope</type>
...     <description>The description for 3.5-m Galileo Zeiss Ritchey-Chretien altazimuth ref
found in the document collection for this bundle.</description>
...     <Internal_Reference>
...       <lid_reference>urn:nasa:pds:context:telescope:orm.galileo_zeiss3m5</lid_reference>
...       <reference_type>is_telescope</reference_type>
...     </Internal_Reference>
...   </Observing_System_Component>
... </Observing_System>
... <Target_Identification>
...   <name>(16) Psyche</name>
...   <type>Asteroid</type>
... </Target_Identification>
... </Observation_Area>
... <Reference_List>
...   <Internal_Reference>
...     <lid_reference>urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:document:bundle_descriptio
...     <reference_type>data_to_document</reference_type>
...   </Internal_Reference>
```

First try: M-type asteroids

Initial difficulties:

- Summary table provides little information, need to parse individual file labels - *is that standard ?*
- Dataset structure doesn't seem to be defined
=> *parsing the /data dir*
Need to handle specific subdirectories
- labels do not contain all observational / instrumental parameters (as opposed to fits), which can be provided in a separated table
- other ascii files may contain relevant information ;(



First try: M-type asteroids

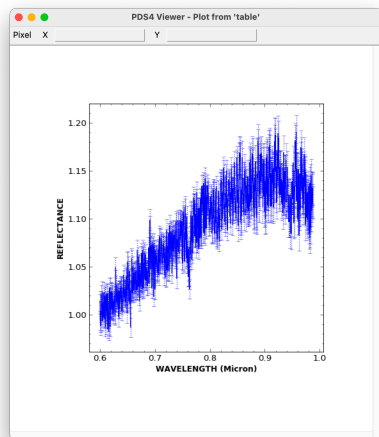
Formal difficulties:

- **target names** - apparently encoded as "(##) name" — but formatting is implicit
 - **dates**: ~ ISO strings — but incomplete / variable formatting
 - **string values**: case is variable (lower in EPNCore, except names)
 - **instrument & instrument host names**: not standardized, variable (as usual)
 - **coverages**:
 - Only time is included in the labels => need to parse files to get spectral range
 - **No phase angle, no Earth or Sun distance** ;(
 - => call IMCCE [ephemcc](#) to complement the EPNCore table
 - In image labels: mid-values are found for angles, coordinates, distances — not ranges
 - **Only ~10 global EPNCore parameters can be mapped from labels**
- Trial and error process: dictionary may grow with enlarged experience

First try: M-type asteroids

access_url: to xml label at SBN — PDS4 tables are supported by TOPCAT

Will be different for other datatypes: add SAMP to PDS4_viewer?



PDS4_viewer

Copy/Paste URL

urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_250bettina_nics_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_250bettina_dolores_mrb_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_250bettina_dolores_lrr_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_22kalliope_emmi_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_224oceana_dolores_mrb_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_224oceana_dolores_lrr_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_216kleopatra_nics_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_201penelope_dolores_lrr_tab	spectrum
urn:nasa:pds:gbo.ast-m-type.fornasier.spectra:data:indivspectra_201penelope_dolores_lrb_tab	spectrum

Download
Send Tables
Send Spectra
Send Images
Send CDF
Send VIRTIS PDS cubes
Send Das2stream
Send PDS4 table

Data Selection Metadata Selection All Data All Metadata

```
SELECT * FROM spectro_m_ast.epn_core
```

VESPA portal

SAMP

A screenshot of the TOPCAT software interface. The window title is "TOPCAT - Plane Plot (37)". It shows a table list on the left, current table properties in the center, and a plot of Reflectance (y-axis, 1.00 to 1.15) versus Wavelength (Micron) (x-axis, 0.6 to 0.9) on the right. The plot displays a noisy red line representing the reflectance data. The interface includes various toolbars and panels for table management and visualization.

TOPCAT

Second try: Trojan asteroids

Difficulties:

- Data subdirectory tree is different
- [need to grab data from additional data tables](#) (photometry)
 - May include non-standard names, to be handled manually
- [need to cross match tables](#):
 - With other observations*: the only match is (observation time + target)
Can be different from spectra (by several h if they are different measurements: spectra and broad filters)
 - With other documents*: instrument/host names etc, are different

Conclusion

Very feasible - and done

- Data structure is not normalized in PDS4
- Parameter names and values are not entirely standardized
- Only some parameters currently mapped, need to check with new trials
- Coverages are not easily found, except time
- Information is often spread on several files, difficult to merge. Reference to granules not always clear (may be through observing time, with tolerance)

=> Each dataset is a new project, requires hand tuning

- Need to check with images — Should be easier for embedded fits files