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 ≳ 2015)
  - Data products are grouped in datasets (≃ 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

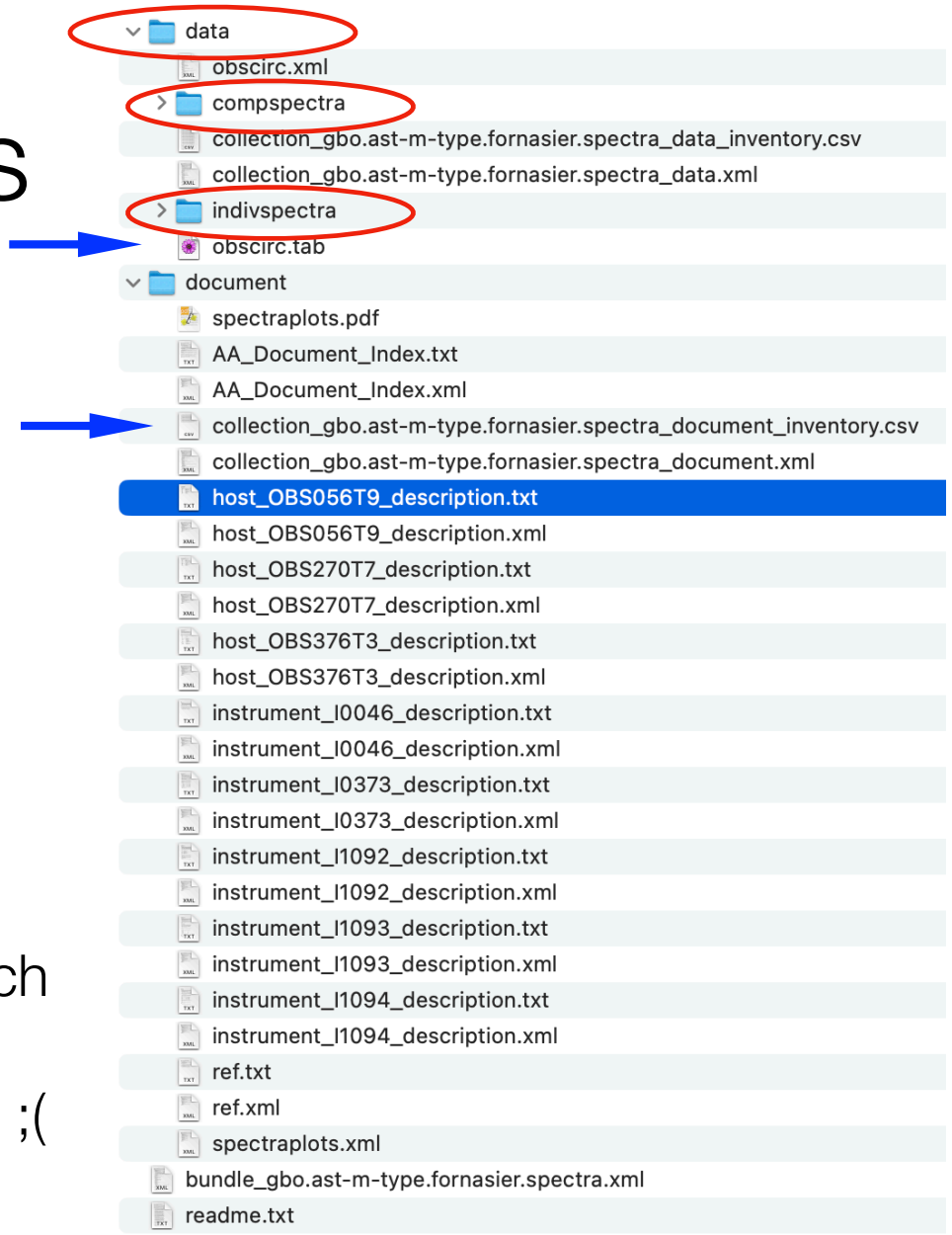
  https://vespa.obspm.fr (or via astropy)
  services spectro_m_ast
  spectro_trojans
  service files available at:
https://voparis-gitlab.obspm.fr/vespa/dachs/services/padc/voparis-tap-planeto/spectro_m_ast
(via EduTEAMS or ORCID)

First try: M-type asteroids

Initial difficulties:

- The inventory table provides little information
  => need to parse individual file labels
- 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 ;(



2-step process

1) Read/parse individual labels

Not using PDS4_tools (xml.etree instead)

2) Add extra metadata from additional catalogues - which are collection-specific

First try: M-type asteroids

Formal difficulties when parsing labels:

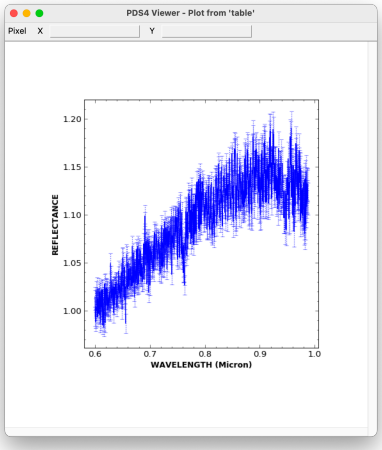
- Only ~10 global EPNCore parameters can be mapped from labels
Trial and error process: dictionary may grow with enlarged experience

- **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
- **Important quantities such as phase angle, Earth or Sun distances not found** ;(
  => call IMCCE ephemcc to complement the EPNCore table

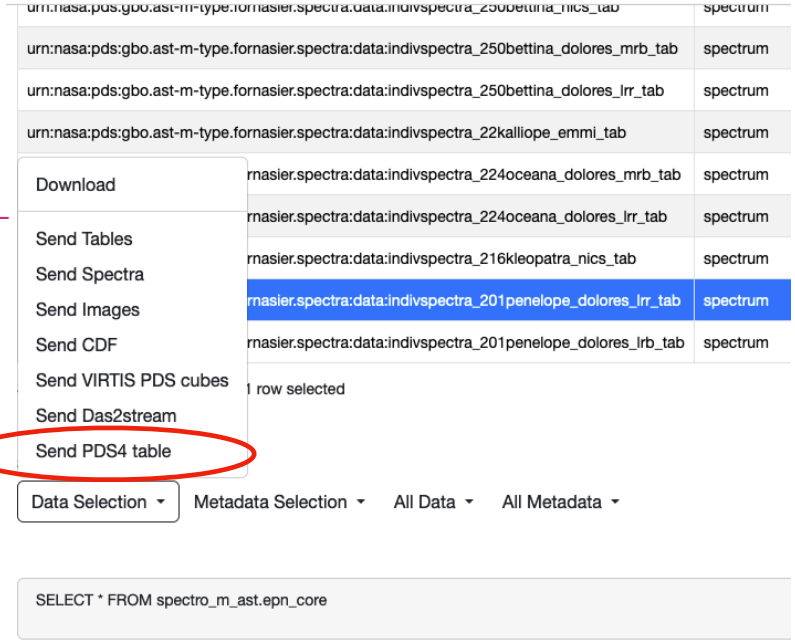
First try: M-type asteroids

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

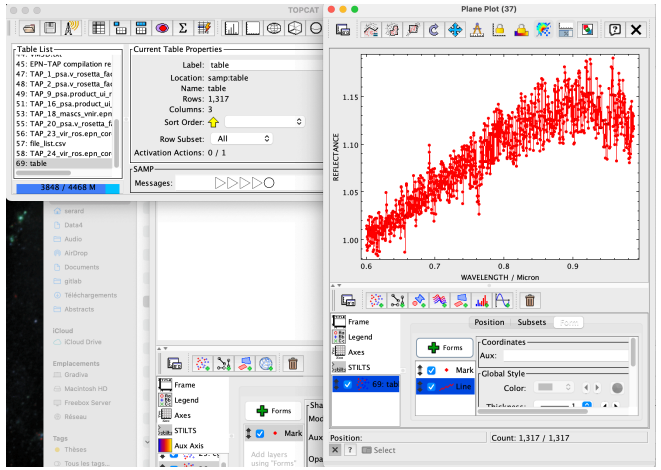
Will be different for other datatypes: add SAMP to PDS4_viewer?



Copy/Paste URL



SAMP



PDS4_viewer

VESPA portal

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

- Third try with an image collection: mid-values are found for angles, coordinates, distances in labels — not ranges

Conclusion

Conversion is 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 further checks with images — Should be easier with embedded fits files