IHDEA Registry Working Group Report

Brian Thomas, Baptiste Cecconi

Outline

1. Purpose
2. Actions Past Year
3. Forward Actions

Purpose

To discuss the need and standards for sharing registries (services, data) for heliophysics data.

Past Year Actions

1. Series of virtual meetings (~1 / 2 months)
2. Discussion of
    a. Terminology
    b. Existing services (HDP, VESPA), relevant standards (SPASE, EPN-TAP, IVOA registries)
    c. Need / Use Cases
        i. Better enable machine use
        ii. Avoid single point of failure (HDP)
        iii. Spread operational effort (shared codebase/standards?)

Review of what is a Registry (IVOA context)

The IVOA Registry in short:

- A **Registry** is a collection of record (using the **VOResource** schema, or extensions of it), and gathered into a database.
- IVOA has a replicated **Searchable Registries**, and many **Publishing Registries** (local providers), which are harvested regularly by the searchable registries. All the registries are listed in the **registry of registries** (RoR)
- Publishing registries are managed by datacenter/providers, and **cover the resources they curate**.
- Several IVOA interfaces to the registries :
  a) The IVOA Registry Interface API,
  b) RegTAP (TAP interface to the registry),
  c) a harvesting interface based on OAI-PMH.
- IVOA registry resources are mostly: data services, data collections, catalogs..

How to find data products, then ?

- IVOA protocols allow to discover data of interest of various types (spectra, images, etc)
- IVOA services such as **ObsTAP** (astronomy) and **EPN-TAP** (solar system) are the **data product level registries**.
- ObsTAP and EPN-TAP services are findable from the Registry.

⇒ data discovery of Solar System products is a 2 steps process:
a) retrieve all EPN-TAP services access points from the registry
b) send the same query to all EPN-TAP services.

⇒ VESPA is doing this.

Review of what is a Registry (IHDEA context)

What we have in IHDEA:

- The SPASE registry is composed of SPASE XML trees. Now managed on git repositories (mostly HPDE github).
- The individual SPASE registries are organized by funding agencies.
- https://hpde.io is a view on the SPASE registries.
- Several searchable web interfaces over the SPASE registry (e.g., HDO and HDP), and a python module (hdpws). Limited to the SPASE registry content.

What we may need:

- Facilitating development of search interfaces over the SPASE registry
  - A registry of registries: a place to discover the location of the registry repositories
- A data product registry?
- A wider scope search interface, which allows to find SPASE resources in a wider context?
  - VESPA would do both.
- Explore how we could benefit from the IVOA registry?

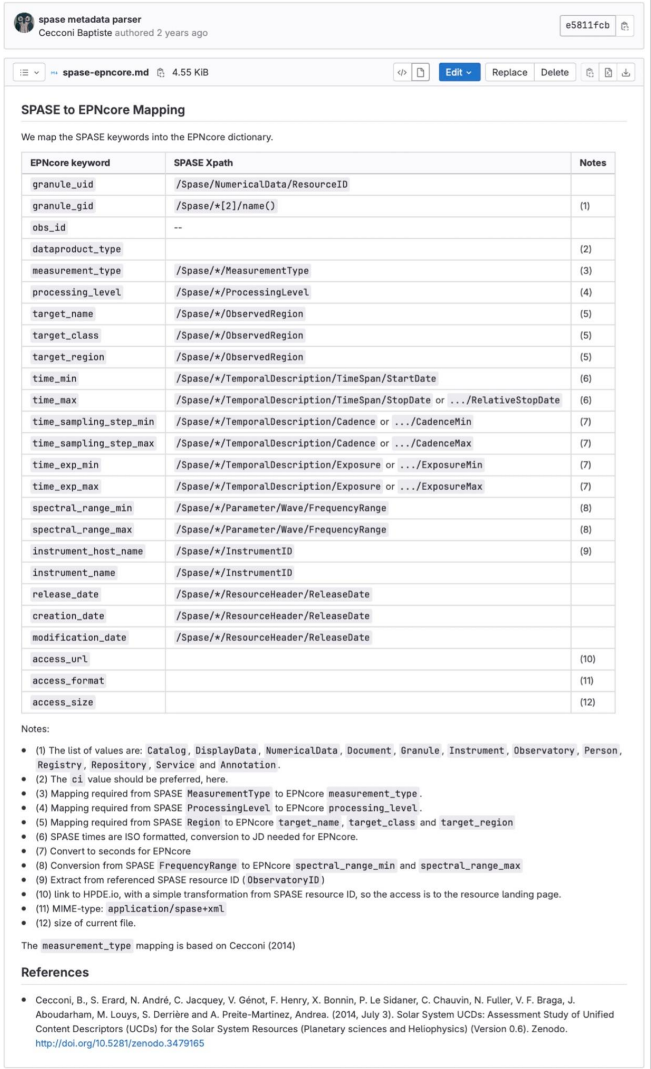
VESPA prototype

Goal: find SPASE records together with other resources available in VESPA

- Ingesting SPASE XML (from list of git repos)
- Updated nightly.
- Mapping SPASE keywords into EPNcore (still working on it)
- Accessible on sandbox TAP@ObsParis
- Access URL to https://hpde.io
- Plans to map SPASE:AccessInformation to IVOA:datalink

short demo

/!\ sandbox server: may not work in the future /!



spase metadata parser
Cecconi Baptiste authored 2 years ago
e5811fcb

spase-epncore.md 4.55 KiB
Edit Replace Delete

SPASE to EPNcore Mapping

We map the SPASE keywords into the EPNcore dictionary.

EPNcore keyword	SPASE Xpath	Notes
granule_uid	/Spase/NumericalData/ResourceID	
granule_gid	/Spase/*[2]/name()	(1)
obs_id	--	
dataproduct_type		(2)
measurement_type	/Spase/*/MeasurementType	(3)
processing_level	/Spase/*/ProcessingLevel	(4)
target_name	/Spase/*/ObservedRegion	(5)
target_class	/Spase/*/ObservedRegion	(5)
target_region	/Spase/*/ObservedRegion	(5)
time_min	/Spase/*/TemporalDescription/TimeSpan/StartDate	(6)
time_max	/Spase/*/TemporalDescription/TimeSpan/StopDate or .../RelativeStopDate	(6)
time_sampling_step_min	/Spase/*/TemporalDescription/Cadence or .../CadenceMin	(7)
time_sampling_step_max	/Spase/*/TemporalDescription/Cadence or .../CadenceMax	(7)
time_exp_min	/Spase/*/TemporalDescription/Exposure or .../ExposureMin	(7)
time_exp_max	/Spase/*/TemporalDescription/Exposure or .../ExposureMax	(7)
spectral_range_min	/Spase/*/Parameter/Wave/FrequencyRange	(8)
spectral_range_max	/Spase/*/Parameter/Wave/FrequencyRange	(8)
instrument_host_name	/Spase/*/InstrumentID	(9)
instrument_name	/Spase/*/InstrumentID	
release_date	/Spase/*/ResourceHeader/ReleaseDate	
creation_date	/Spase/*/ResourceHeader/ReleaseDate	
modification_date	/Spase/*/ResourceHeader/ReleaseDate	
access_url		(10)
access_format		(11)
access_size		(12)

Notes:

- (1) The list of values are: Catalog, DisplayData, NumericalData, Document, Granule, Instrument, Observatory, Person, Registry, Repository, Service and Annotation.
- (2) The ci value should be preferred, here.
- (3) Mapping required from SPASE MeasurementType to EPNcore measurement_type.
- (4) Mapping required from SPASE ProcessingLevel to EPNcore processing_level.
- (5) Mapping required from SPASE Region to EPNcore target_name, target_class and target_region
- (6) SPASE times are ISO formatted, conversion to JD needed for EPNcore.
- (7) Convert to seconds for EPNcore
- (8) Conversion from SPASE FrequencyRange to EPNcore spectral_range_min and spectral_range_max
- (9) Extract from referenced SPASE resource ID (ObservatoryID)
- (10) link to HPDE.io, with a simple transformation from SPASE resource ID, so the access is to the resource landing page.
- (11) MIME-type: application/spase+xml
- (12) size of current file.

The measurement_type mapping is based on Cecconi (2014)

References

- Cecconi, B., S. Erard, N. André, C. Jacquey, V. Génot, F. Henry, X. Bonnin, P. Le Sidaner, C. Chauvin, N. Fuller, V. F. Braga, J. Aboudarham, M. Louys, S. Derrière and A. Preite-Martinez, Andrea. (2014, July 3). Solar System UCDs: Assessment Study of Unified Content Descriptors (UCDs) for the Solar System Resources (Planetary sciences and Heliophysics) (Version 0.6). Zenodo. http://doi.org/10.5281/zenodo.3479165

Forward Actions

Honestly, TBD.

Thoughts (Brian):

- Solidify Use Cases
- Consider IVOA registry standard;
  - What do we get out of using it : Is this a good fit for us?
  - Impacts on SPASE
- Hold a workshop ??
- Be better about holding monthly meetings
- Expand membership (NO members outside of N. America, Europe)