



Fig. 1

1. Vocabularies: Lessons Learned

Markus Demleitner
msdemlei@ari.uni-heidelberg.de

- Vocabularies in the VO?
- The first three VEPs
- UAT: In scope?
- Next steps

(cf. Fig. 1)

2. What's it for?

- Standardise "datalink-like" vocabularies (also: reference positions, relationship types, time scales...)
- Terms can have hierarchy, can be deprecated.
- Make them easy to consume so clients figure out is-a relationships, descriptions, nice labels without having to know RDF.
- Clear, manageable rules for how to add terms as you need them, when you need them.

3. Adding Terms: VEPs

Vocabulary: <http://www.ivoa.net/rdf/datalink/core>
Author: francois.bonnarel@astro.unistra.fr
Date: 2019-10-15

New Term: documentation
Action: Addition
Label: Documentation
Description: Extra information on the item in human-readable text form, ranging from processing logs to weather reports to technical documents on instruments to related publications.
Rationale:
This term was proposed actually by Alberto Micol due to some ESO use cases. For example a text giving some details on observation conditions can be attached to a dataset. Could be also instrument configuration in free text, etc...

Discussion: Semantics took the liberty of modifying term, label and [...] there probably was and the single "documentation" was sufficient (<http://mail.ivoa.net/pipermail/dal/2019-October/008208.html>)

This is VEP-002 somewhat shortened to fit the slide. Basically, a VEP has some trivial metadata and then suggests a term, a good-looking label for it and an "official" description. Then, the submitter gives some rationale for why they think the term should be given.

Finally, the semantics WG provides a summary of the public discussion of the term, complete with links into the mailing list archive.

4. The First Three VEPs

All on Datalink (thanks, François!).

- VEP-001: *#associated-dataset*, *#associated-image*, *#associated-spectrum*, ... *#associated-timeseries-cube*. Lively discussion, although mostly on a datalink use case. Eventually abandoned for VEP-003.
- VEP-002: *#documentation*. Brief discussion, agreement reached (no TCG session since then).
- VEP-003: *#sibling*. Supercedes VEP-001. Brief discussion. Agreement?

5. Lessons Learned

- VEP evolution happens. To make that understandable later, we need to link VEPs with Superceded-by and Supercedes VEP fields.
- Having usage examples really helps. Hence: Used-in VEP fields.
- The TCG step hasn't been tried yet. Will someone look?
- What to do when the discussion somewhat fizzles out, as with VEP-003?

6. The UAT

The Unified Astronomy Thesaurus (UAT) is a general vocabulary of Astronomy.

Current IVOA use: VOResource 1.1 subject.

But: It's unclear how. UAT concepts look like <http://astrothesaurus.org/uat/1268> – we probably don't want to have that in VOResource.

1268? Hm. "Plutoids" would be optimal, but using the label would need a norm. And isn't stable.

And then, non-RDF-aware consumers of IVOA vocabularies couldn't do much with the files published by the UAT folks.

7. UAT with IVOA tooling?

We're studying if we can have a "UAT in IVOA format":

- Human-readable terms ("martian-satellites" rather than 1009)
- Stable mapping, published through `skos:exactMatch`
- Simple-to-use files from ivoa.net
- Automatic generation as UAT pushes new terms

Two URLs for the same thing isn't pretty, but then that's in the RDF spirit. But should we (the Exec?) give UAT carte blanche to modify an IVOA vocabulary without TCG participation? In the end, we already have, since VOResource recommends the UAT, but such things always look a bit different when the technical details need to be worked out.

8. Trisduvop

Part of the current draft is `revovo`, a Python module extracting from our RDF/XML:

- a mapping of terms to (label, description)
- a list of deprecated terms
- a list of preliminary terms
- a hierarchy mapping term to [wider terms]

What if we just dumped this into a JSON file and had that info with essentially zero tooling for clients that asked specifically for it?

The question might seem like a no-brainer (in particular because we'd still have our interoperable, standard representations), but of course throwing *yet another* format into the already large pool of RDF serialisations is a step we shouldn't lightly take. For reference, the UAT people let you download the UAT in RDF/XML, N-Triples, Turtle, Notation3, N-Quads, RDF/JSON, JSON-LD, TriX, TriG, and Sesame Binary RDF. For half of that bunch I have no idea what it is either.

If we want it, I'd say we should call the format `trisduvop` ("Trivial semantics for dumb VO programs") and use the `text/json;content=trisduvop` media type for it. That way, at least nobody will stumble into it by accident...

9. Where do we go from here?

- Let's see how VEP-002 fares at the next TCG.
- Next WD of `VocInVO 2` in March (with `Trisduvop`? With language on translated external "endorsed" vocabularies?)
- Further VEPs? Any takers?
- Please review use cases and requirements in the WD.