

Vocabularies in the Virtual Observatory

Markus Demleitner

NFDI4Earth Terminologies Workshop, 2024-05-22

Virtual Observatory?

The Virtual Observatory is a global infrastructure for finding, accessing, and investigating astronomical data.

It is built on a set of standards (<https://ivoa.net/documents/>): Formats, search protocols, job management and auth, a resource registry, semantics, IPC...

Vocabularies

Interoperable terminologies and their management are defined in <https://ivoa.net/documents/Vocabularies/>.

Basically: Our vocabularies are available in RDF, HTML, and a custom, trivial-to-use format ("Desise", Dead Simple Semantics) from <http://www.ivoa.net/rdf>.

Usable with a Browser

Our RDF URIs look like <https://www.ivoa.net/rdf/product-type#spectrum>. You can resolve them with a browser:

resolved-dataset	Resolved	intended for retrieval. To annotate datasets, use a narrower term.		Narrower
spectrum	Spectrum	Flux or magnitude as a function of a spectral coordinate.	#spectrally-resolved-dataset	Narrower
temporally-	Temporally	A dataset with a time axis. This term is mainly		Narrower

0% Javascript required here.

Non-Browser clients do content negotiation to retrieve Turtle, RDF/X, or Desise.

This last thing of course is not our invention. It follows W3C best practices as of 2008 ("Cool URIs").

Showcase

What are we doing with these vocabularies? Here are a few examples:

- Subject keywords and query expansion
- Mapping subjects for interdisciplinary search engines

- Machine-readable linking between datasets and associated artefacts
- Humanising machine responses
- Linking data resources and publishing services
- ... and all that with minimal code.

Subject Keywords

In our Registry (~ 30'000 resources), there are keywords from the Unified Astronomy Thesaurus <http://ivoa.net/rdf/uat>.

This gives you reliable labels in data discovery:

```
... WHERE res_subject='magnetic-fields'...
```

Yes, our data discovery queries are mostly formulated in SQL, and yes, many astronomers can write SQL. But sure, there are interfaces making this a bit less nerdy.

The relationships within the UAT let you do query expansion in case metadata has more specialised terms:

```
... WHERE 1=gavo_vocmatch('uat', 'magnetic-fields', res_subject) ...
```

In case you want to play around a bit with where such relationships might come in useful, there is the proof-of-concept concept chooser [SemBaReBro](#) that lets you navigate the UAT and shows where there are resources with the respective concepts.

Coarsifying Subjects

We are delivering our records to b2find. Non-astronomers may have no idea what virgo-cluster, say, might mean.

By following skos:broader relationships, we arrive at some “root” term suitable for a wider audience (galactic-and-extragalactic-astronomy, in this case):

.. _SMAKCED: <http://smakced.net> .. _resulting decompositions:
<http://smakced.net/data.html>

dwarf galaxies early type galaxies galactic and extrag...
infrared photometry observational astro...
surface photometry virgo cluster

Identifier

Roles in Datalink

Datalink is an IVOA standard to link all kinds of artefacts (e.g., calibration data, derived data products, preview...) with a dataset.

The nature of the artefact is described using the vocabulary <http://www.ivoa.net/rdf/datalink/core>, where, for instance, #preview-plot is narrower than #preview.

In the popular library pyVO, you can say:

```
match.bysemantics("#preview", include_narrower=True)
```

and you will get all artefacts that are some kind of preview.

Formatting Datalink Documents

The hierarchical vocabulary can also be used to format datalink documents and enrich them with explanations. Left is a tabular rendering, right a simple browser UI for a datalink document.

	access_url	semantics	description
44	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1985-01-11	#counterpart	1985-01-11 time series for this field.
45	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1985-01-12	#counterpart	1985-01-12 time series for this field.
46	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-18	#counterpart	1990-01-18 time series for this field.
47	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-19	#counterpart	1990-01-19 time series for this field.
48	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-20	#counterpart	1990-01-20 time series for this field.
49	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-21	#counterpart	1990-01-21 time series for this field.
50	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-22	#counterpart	1990-01-22 time series for this field.
51	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-24	#counterpart	1990-01-24 time series for this field.
52	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-24	#counterpart	1990-01-24 time series for this field.
53	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-29	#counterpart	1990-01-29 time series for this field.
54	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-30	#counterpart	1990-01-30 time series for this field.
55	http://dc.zah.uni-heidelberg.de/flare_survey/q/ts/dl/1990-01-30	#counterpart	1990-01-30 time series for this field.
56	http://dc.zah.uni-heidelberg.de/flare_survey/q/m/dl/st...	#preview-image	Mid-resolution plate photo.
57	http://dc.zah.uni-heidelberg.de/flare_survey/q/m/dl/st...	#documentation	Photo of the plate envelope.
58		#proc	This service produces datalinks for
59	http://dc.zah.uni-heidelberg.de/getproduct/flare_sur...	#this	The full dataset.
60	http://dc.zah.uni-heidelberg.de/getproduct/flare_sur...	#preview	A preview for the dataset.

▶ Preview
low fidelity but easily viewed representation of the data

▶ Processing
server-side data processing result

▶ Documentation
Structured or unstructured metadata helping to understand, interpret, or work with #this. Such information can range from processing logs to weather reports to technical documents on instruments to related publications.

▼ Counterpart
Data products sharing the target of the experiment or observation that led to #this but of unrelated provenance. This could be observations of the same object in different wavelengths or along different axes (time, spectrum), but spectra of dust of common origin but different laboratories would be #counterparts as well.

[1985-01-07 time series for this field.](#) (92 MiB)

[1985-01-10 time series for this field.](#) (92 MiB)

Linking Services to Data

In the Registry, there are records for tables (“A catalogue of high-redshift quasars”).

There are services serving many of them at a time. To link data to where it can be queried, a resource record defines a relationship with http://ivoa.net/rdf/voresource/relationship_type#IsServedBy:

```
<relationship>
  <relationshipType>IsServedBy</relationshipType>
  <relatedResource ivo-id="ivo://org.gavo.dc/tap">GAVO Data Center TAP service</relatedResource>
</relationship>
```

Note that, against the usual DataCite relationships like `IsSupplementedBy`, `IsServedBy` is actually operationally relevant.

...With Little Code

All this *can* be done without scary RDF libraries. See <https://wiki.ivoa.net/internal/IVOA/InterOpMay2021Semantics/voc-action.pdf> for sample code in Python and Javascript.