

Blind Discovery and VODataService 1.3

Markus Demleitner

Herbsttagung 2026 der Astronomischen Gesellschaft
Garching, 2026-09-08

Blind Discovery and VODataService 1.3

- What is Blind Discovery?
- What VODataservice 1.2 brought
- Column Statistics
- Data Types
- Data Sources

Blind Discovery

As of now, $\sim 90\%$ of registry queries are of the type “Where are SIAP services for XMM data?”: People *know* that there is some data and just want to find out where it is.

Blind discovery is when you would instead say: “I want X-ray data around 3C273 from the 1990ies going down to 20 mCrabs.”

And that’s been the promise of the VO all along: Finding data you didn’t know exists.

VODataService 1.2

“X-ray” and “around 3C273” is covered by VODataService 1.2’s coverage:

```
<coverage>
  <spatial>6/2623,2666-2667,2709,2752</spatial>
  <temporal>55819.5 55819.5</temporal>
  <spectral>2.95119e-19 4.08734e-19</spectral>
  <waveband>Optical</waveband>
</coverage>
```

(except that 31% of the resources still miss spatial, 77% spectral and 98% temporal coverage – help fix what’s fixable!)

Reaching 20 mCrabs

Sensitivity information is crucial (“...complete down to 23^m in G”) and hard. Exposure time, for instance, is an extremely lame proxy.

VODatService 1.3 proposes column statistics:

```
<column>
  <name>phot_g_mean_mag</name>
  <description>Mean magnitude in the G band. This is...
  <unit>mag</unit>
  <ucd>phot.mag;em.opt;stat.mean</ucd>
  <stats>
    <min>1.77328</min>
    <percentile03>15.5068</percentile03>
    <median>19.7229</median>
    <percentile97>21.0333</percentile97>
    <max>22.7507</max>
    <fillFactor>0.996994</fillFactor>
  </stats>
</column>
```

Why These Stats?

- min, max: UI generation, declaring queriable ranges
- percentile03, percentile97: the 2σ range for a Gaussian, more robust than min and max
- median: for estimating skews
- fillFactor (i.e., ratio of non-NULL values): don't bother when things are too incomplete

Product Types: What's Broken?

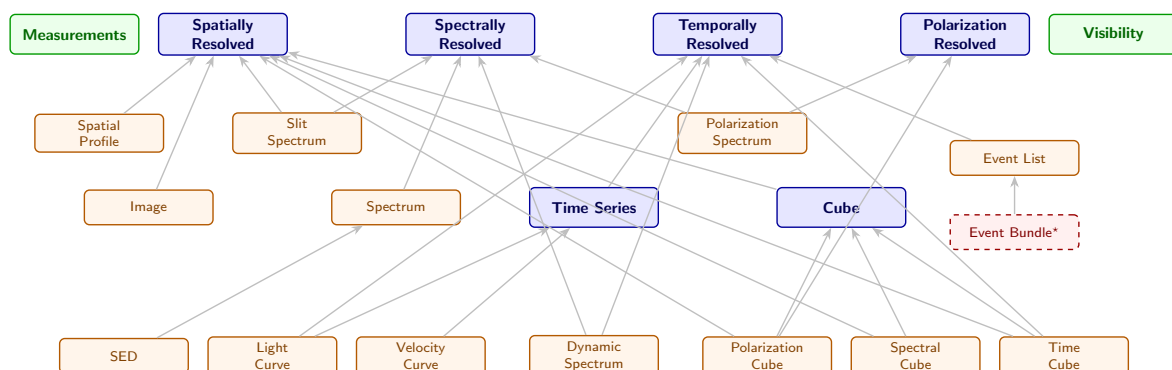
So far, when you look for spectra, you look for SSAP services.

That's not good:

- There are time series published through SSAP
- For images, there's now SIAv1, SIAv2, and Obscore, so this mode is becoming tedious
- Obscore can publish about anything and needs a way to say what it is anyway

Product Types: The Fix

For obscore, we have created the IVOA product-type vocabulary <http://www.ivoa.net/rdf/product-type>.



In your VODataService 1.3 Registry record, just give one or more of the corresponding identifiers:

```
<productTypeServed>velocity-curve</productTypeServed>
<productTypeServed>dynamic-spectrum</productTypeServed>
```

Data Sources

Outside of SSAP and SLAP, we have had no way to tell “theory” from observation. VODataService introduces a generic dataSource element admitting values from <http://ivoa.net/rdf/data-source>:

abstract-model The production of data artefacts from physical principles with the goal of modelling worlds without reference to real world objects or the present universe.

concrete-model The production of data artefacts with the goal of predicting observations of objects within our actual universe.

observation The use of instruments to directly measure or observe the actual world.

o/in-situ Doing measurements of, in or at actual astronomical objects (stars, galaxies, planets, etc).

o/laboratory Doing measurements using artificial samples in laboratory-like environments.

Can I Use This Now?

The Registry API is RegTAP, and that is not yet updated for the new facilities.

Also, almost nobody already produces registry records with this metadata.

But the registry at <https://dc.g-vo.org/tap> already exposes metadata from prototype services:

- 5400 records of numeric stats in rr.g_num_stat
- 12 records that have a detail xpath of /dataSource (assume observation by default, I guess)
- 15 records with a detail xpath of /productTypeServed (but these will probably be artificially backfilled by service types)

Pardon: What Did You Just Say?

Here’s a piece of ADQL that looks for resources that reliably reach $z = 4$:

```
SELECT ivoid
FROM rr.table_column
    NATURAL JOIN rr.g_num_stat
WHERE ucd LIKE 'src.redshift%'
AND percentile97>4
```

Yes, there’s almost everything missing still; but at least Vizier will be in there soon.

Comments? Contributions?

RFC is starting soon.

Join the fray! Subscribe to the Registry WG mailing list today at <https://mail.ivoa.net/mailman/listinfo/registry>.