



Fig. 1

## 1. Confessions of a Registry Janitor

Markus Demleitner  
 msdemlei@ari.uni-heidelberg.de

This is a collection of common defects in registry records I keep coming across – none of these will make your registry record invalid, so I can't come after you with the force of electronic law. But they *will* make your record and hence the whole Registry a good deal less useful.

- Coverage!
- Free-form person names
- Personal addresses as contacts
- Non-UAT subjects
- Short descriptions
- Missing Tablesets
- Column descriptions too short or too long
- Bad column UCDS

Funded by e-inf-astro, BMBF FKZ 05A20VH5



Distributed under CC0



## 2. Coverage!

Tell the world where in space, time, and spectrum your data is.

That's the number one thing needed for blind discovery. It also lets aladin colour matching resources:

(cf. Fig. 1)

## 3. Broken Author Names

If I'm looking for data created by Gerard de Vaucouleurs, it would be great if I could just write `role.name like 'de Vaucouleurs, G%'` and not have to worry about any of:

- van der Waarden, B; de Vaucouleurs, G.
- Gerard de Vaucouleurs
- G. de Vaucouleurs
- Vaucouleurs, G. de

So, please only put in names last name first, with a comma at the end of the last name. And no, computers cannot normalize names themselves, even for western names. Think of E.-M. Zeidler-K. T., an actual astronomer active at Kiel until the 2000s.

And don't get me started on "et al. MNRAS (in press)", "(NOT GIVEN)", or "under a cooperative agreement with the NSF on behalf of..." as creator names.

## 4. Fragile Contact Info

Curation/contact is an important item, both for janitors and for our users (for bug reports and/or help requests) – have you noticed the `get_contact` method on pyVO registry results?

Help make it more robust by avoiding personal addresses. Instead, use addresses you can hand on when you move on.

Try talking your IT department into giving you `vo-support@your.institute.edu`.

## 5. Free-form Subjects

Since last year, it's finally clear that subjects should *really* be the identifiers from `http://www.ivoa.net/rdf/uat` ("abell-clusters"). So, please no:

- "DR1, SIAP, Images"
- "???"
- "optical lines equivalent width catalog"
- "Galaxy:general"

There's much cool stuff you can do with machine-readable subjects – don't miss out!

## 6. Short Resource Descriptions

A resource description should be enough to give a non-specialist astronomer an idea if the data can be useful for what they're trying to do. And it should work for full-text searches.

So, please no:

- "Survey (LoLSS)."
- "USET Group form"
- "Convolution of normalized synthetic stellar spectra" (though that might work as a title)

Guideline: If you briefly address each of What, From where, What for, that's probably already a useful description.

## 7. Relationships

If a table underlying your data is queryable in a TAP service, don't forget to declare a ServedBy relationship there.

And please don't use related-to any more. It's deprecated for a reason.

Check out our relationship.type vocabulary (and have it amended if it's missing what you need).

## 8. Missing Tablesets

If your registry record has this:

```
</capability>  
</ri:Resource>
```

you're probably doing it wrong: Try to define a tableset, i.e., the table(s) you return or that can be queried.

For TAP services, I need it so I can one day shut down GloTS. In general, it's great for discovery ("Where can I get position angles of ellipsoids fitted to galaxies?").

## 9. Bad Column Descriptions

As for resource descriptions: Non-specialist astronomers should have a chance to figure out what it is. So, please no:

- ?, ??, or ???
- "RA" (at least tell users where it came from, a bit human-readable info on frame, epoch and such certainly is nice, too, at least as long as we don't have robust machine-readable annotation to that effect)
- "This is the excess [about 2k chars follow] Algorithms, and Software Implementation." (really long texts will lead to excessive false positives)
- NULL

On descriptions of excessive length: Let's work on column footnotes in VOTables!

## 10. Bad Column UCDs

It'd be great if I could reliably discover all position angle columns in the VO by constraining `ucd='pos.posAng'`.

But: there are 117 columns in the VO called `pa` without a UCD.

And: 30342 times `??`, 41 times `???`, and 70 times `????`.

Oh, and don't use `pre-1+` UCDs any more ("POS.EQ\_RA").

Trivia: the longest UCD in the VO has 74 characters.

## 11. Thank You For...

...improving the VO Registry!

The long version of this talk is at

<http://blog.g-vo.org/registry-a-janitor-speaks-out.html>