Writing a Resource Record People Will Find

Markus Demleitner

IVOA Southern Spring Interop 2024, Malta Nov 14-17
Registry WG
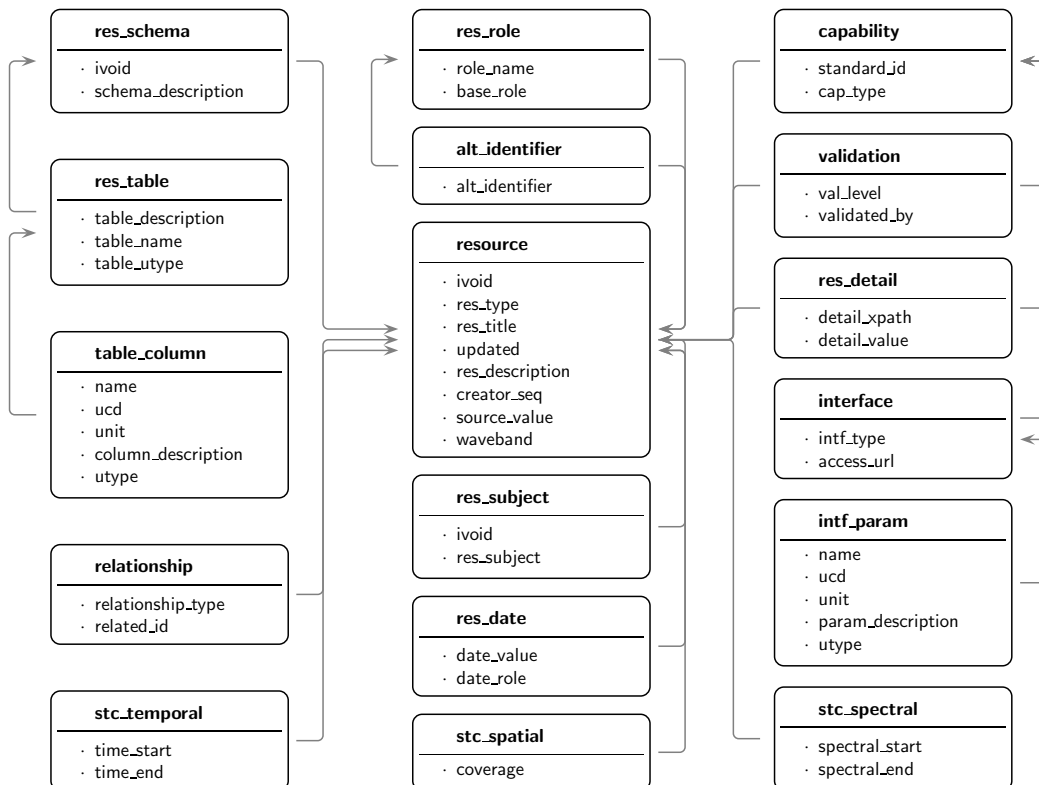
The Problem

In July 2024, I wanted to locate the exoplanet merged catalog ExoMerCat from all of TOPCAT's SCS and TAP windows, and with reasonable pyVO and WIRR queries.

- TOPCAT found exomercat, but WIRR and pyVO didn't

- "exoplanet merged catalogue" did work in pyVO but not in TOPCAT's TAP window

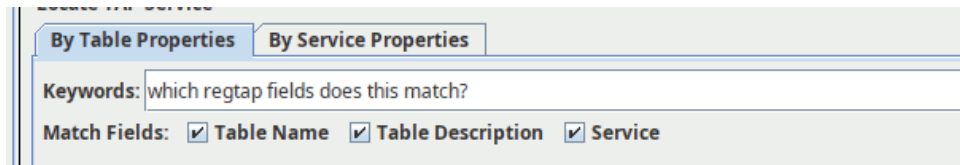- "exoplanet merged catalog" did not work anywhere

How come?

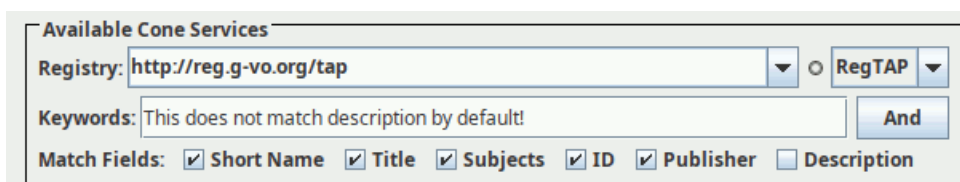**Background: The RegTAP Data Model**

RegTAP consists of many tables, which allows users to be very precise on what their constaraints mean and what they want to learn about the resources discovered. But it means that in seemingly simple cases like "full text search" you need to be aware what each piece of software does.

Who Matches What?

TOPCAT's default TAP search:

TOPCAT's SCS search:

pyVO and WIRR:

```
base_queries = [
    " 1=ivo_hasword(res_description, {{{parname}}})",
    " 1=ivo_hasword(res_title, {{{parname}}})",
    " rr.res_subject.res_subject ILIKE {{{parpatname}}}"]
```

Actually, it is even worse than that: TOPCAT's TAP search is (still) not even based on RegTAP. But that is yet another story. That we will hopefully close one of these days.

What Gives?

- Make sure that both your acronym and its expansion is present in both the *resource* title and, if it's a TAP table, in the *table* description.

- Whatever you write in your description will be found in pyVO (and WIRR) but not by default in TOPCAT.

- Authors are never considered by TOPCAT, and are not included in WIRR and pyVO full text searches (but there are special constraints for them).

- As a user, be aware that TOPCAT's RegTAP searches generally do not search description unless you switch it on.

Even Better: Shun Full Text Searches

The root evil is of course the full text search as such.

It would be so much better to search by physics: UCDs, UAT terms, coverage:

"Give me a catalogue-like resource with full-sky coverage which has a column with phys.mass UCD and subject of exoplanet-astronomy or a bit narrower."

I will give you that at this point, this yields fewer resources. But better curated ones, too.

https://dc.g-vo.org/WIRR can be used as a registry component supporting these for TOPCAT (except for TAP, regrettably).

Homework: a Registry Synonyms

That searches for catalog and catalogue yield different results is positively lame.

There is a way around this at least for postgres-based RegTAPs: synonym dictionaries.

Who would be in for maintaining such a thing?

I already have a few pairs.

When pulling this, make sure your browser does not wantonly "upgrade" to https, as the thing will then ask for your credentials.

If you are running RegTAP and have postgres as the backend database, install this file to `/usr/share/postgresql/<server_version>/tsearch_data/` on the database server, and then run

```
CREATE TEXT SEARCH DICTIONARY regtap_syns (
  TEMPLATE = synonym,
  SYNONYMS = regtap
);
ALTER TEXT SEARCH CONFIGURATION english
  ALTER MAPPING FOR asciiword
  WITH regtap_syns, english_stem;
```

Regrettably, this has to be repeated every time you do a major version step. Thus, when you do this, make sure there is a regression test in place alerting you when you forget it. In Heidelberg, I am asserting that

```
select top 1
  ivo_hasword('catalog', 'catalogue') as has_synonyms
  from tap_schema.tables
```

returns a 1.