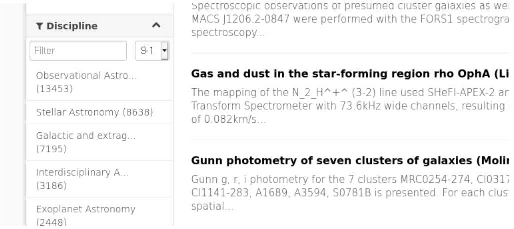
Fig. 1



Observational Astro... (13453)

Stellar Astronomy (8638)

Galactic and extrag... (7195)

Interdisciplinary A... (3186)

Exoplanet Astronomy (2448)

Spectroscopic observations of presumed cluster galaxies as wel MACS J1206.2-0847 were performed with the FORS1 spectrogra spectroscopy...

Gas and dust in the star-forming region rho OphA (Li

The mapping of the N_2_H^+^ (3-2) line used SHeFI-APEX-2 ar Transform Spectrometer with 73.6kHz wide channels, resulting of 0.082km/s...

Gunn photometry of seven clusters of galaxies (Molir

Gunn g, r, i photometry for the 7 clusters MRC0254-274, Cl0317 Cl1141-283, A1689, A3594, S0781B is presented. For each clus spatial...

Fig. 2

1. Bridging Semantic Gaps

Markus Demleitner
msdemlei@ari.uni-heidelberg.de

– or rather –
Improving the metadata for VO Resources in B2FIND using Semantics

- B2FIND and the Registry
- The UAT and the Registry
- Bridging the semantic gap

(cf. Fig. 1)

2. B2FIND

B2FIND[1] is a EUDAT-operated cross-disciplinary data search engine.

We feed it a subset of the VO registry metadata though the oai_datacite format.

It is actually rather straightforward to map large parts of VOResource to the Datacite metadata kernel; see a piece of XSLT[2] that is also behind the VOiDOI[3] DOI minter.

(cf. Fig. 2)

[1] `http://b2find.eudat.eu`

[2] `https://volute.g-vo.org/svn/trunk/projects/registry/dois`

[3] `http://dc.g-vo.org/voidoi/q/ui/custom`

3. Semantic Gap: B2FIND side

In the age of cheap full-text search, subject keywords only make sense if they're controlled.

B2FIND has its own keyword schema. For astronomy, they have the top-level concepts of the UAT[4]:

- Astrophysical processes
- Cosmology
- Exoplanet astronomy
- Galactic and extragalactic astronomy
- High energy astrophysics

- Interdisciplinary astronomy
- Interstellar medium
- Observational astronomy
- Solar physics
- Solar system astronomy
- Stellar astronomy

4. Semantic Gap: VO side

In VOResource's subjects, there's all kind of mess. Here's a sampling:

- Galaxies
- Proper_Motions
- Galaxies:Markarian
- EXTRAGALACTIC RADIO SOURCES
- TAP
- DK154
- ???
- Digitised photographic Schmidt field test photometry astrometry
- UKIDSS DR8, SIAP, Images

Of course, it's not supposed to be that way. In particular, the last example, where multiple keywords are joined into a single subject, is plain wrong: never squeeze more than one subject keyword into a `subject` element. You can have multiple subjects per content element, of course.

[4] `https://astrothesaurus.org/`

5. Fixing the VO Side

Since 2018, VO subjects should come from the UAT as per VOResource 1.1[5].

Trouble: we never said how to do that. Only with Vocabularies 2.0, http://www.ivoa.net/rdf/uat says what to do.

As part of Voc 2.0 proving, I've mapped all usable subjects in the registry to the UAT to build SemBaReBro[6].

I give you that SemBaReBro's utility at this point is probably questionable. I do consider fiddling something quite like it into WIRR[7], though.

Side effect: `rr.subject_uat`.

6. UAT Keywords in RegTAP

```
select top 15 * from rr.subject_uat
where ivoid like ’%g%’
```

ivoid	**uat_concept**
ivo://astron.nl/lofartier1/q_img/imgs	catalogs
ivo://astron.nl/lofartier1/q_img/cutout	catalogs
ivo://astron.nl/apertif_dr1/q/apertif_dr1_continuum_images	radio-astronomy
ivo://org.gavo.dc/bgds/l/meanphot	surveys
ivo://org.gavo.dc/bgds/l/meanphot	galaxy-planes
ivo://org.gavo.dc/bgds/l/meanphot	milky-way-galaxy
ivo://org.gavo.dc/bgds/l/meanphot	variable-stars
ivo://org.gavo.dc/bgds/l/meanphot	broad-band-photometry
ivo://wfau.roe.ac.uk/glimpse-dsa	infrared-astronomy
ivo://wfau.roe.ac.uk/glimpse-dsa/ceaapplication	infrared-astronomy
ivo://wfau.roe.ac.uk/galexgr6-dsa	ultraviolet-astronomy
ivo://wfau.roe.ac.uk/galexgr6-dsa/ceaapplication	ultraviolet-astronomy
ivo://bsdc.icranet.org/whsp/q/cone	catalogs
ivo://bsdc.icranet.org/whsp/q/cone	active-galaxies
ivo://bsdc.icranet.org/whsp/q/cone	bl-lacertae-objects

7. Synthesis

All UAT terms have one or more root terms.

And it's simple to find them from a desise object.

Hence, when we generate a oai_datacte record:

1. Retrieve UAT subjects for its IVOID from `rr.subject_uat`
2. Use IVOA UAT to figure out the top-level terms for these subjects
3. Add them to the subjects of the datacite record

[5] `https://ivoa.net/documents/VOResource/20180625/`

[6] `http://dc.zah.uni-heidelberg.de/sembarebro/q/ui/info`

[7] `https://dc.g-vo.org/WIRR`

8. Conclusion

B2FIND in the future gets subject keywords it can actually deal with for VO resources.

Ingredients:

- Formal semantics on VO resource records
- The UAT Vocabulary
- A dash of python code

There's a post on this including source code on
https://blog.g-vo.org.