



Fig. 1

1. As If You Were There

Remote Queries in the Virtual Observatory

Markus Demleitner (*msdemlei@ari.uni-heidelberg.de*)

Agenda

- Locating data
- Sending queries
- Orchestrating servers

2. The Problem

(vgl. Fig. 1)

Most data currently acquired and much of the data previously obtained is “on the net” in some form or other. But how does one access it and use it, preferably rather ad-hoc, without large preparations? And how can one combine various data sets, preferably without having to pipe them through the local machine first?

```

tapsh> select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80
tapsh> start
tapsh> job
Job nicked zowize (remote id o8JfWP)
phase EXECUTING, destruction due 2010-09-05T09:52:02, time limit 3600
-----
select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80
tapsh> job
Job nicked zowize (remote id o8JfWP)
phase EXECUTING, destruction due 2010-09-05T09:52:02, time limit 3600
-----
select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80
tapsh> job
Job nicked zowize (remote id o8JfWP)
phase COMPLETED, destruction due 2010-09-05T09:52:02, time limit 3600
-----
select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80
tapsh>

```

Fig. 2

3. The VO's solution: TAP

The **T**able **A**ccess **P**rotocol transmits (ADQL) queries and ancillary data to servers, and lets you manage jobs and results, and provides for some service introspection.

The standard describing TAP was passed by the IVOA authorities in 2010. So, using TAP right now is a bit like having used the WWW in 1993: Useful, but wires may still be showing.

By the way, don't read the standard unless you're planning to write a TAP server or library – it's rather long and builds on numerous other standards you'd generally need to know as well to make sense of it all. Instead, peruse user-facing documentation, e.g., tapsh's manual.

Oh, and "introspection" just means you can find out which tables can be queried on a service and what columns they contain.

TAP is for running potentially large jobs, running for hours or even days. It would be unwise to require network connections staying up for such time spans (in particular for HTTPS). Thus TAP is designed to support "asynchronous" operation, i.e., you manipulate your job on the server, go away, and come back later to check how it has done.

TAP's typical pattern:

1. Define a job (set the query, add uploads, change limits)
2. Start it
3. Check progress (or not, if you're not curious)
4. When you see it's done, retrieve or use result
5. Clean up (if you're nice)

4. Async? Show me.

In case there's no net, a screenshot:

(vgl. Fig. 2)

The good news: Much of this can be hidden by software for "easy" cases. What you should take with you, though: The *creation* of a query doesn't mean it is *started*. On the contrary, the time between the creation of a job and its start is important in TAP, since this is when you manipulate jobs – upload files, ask for more time, etc.

Reference URL >>

Try ADQL to query our data.

Please report errors and problems to [GAVO staff](#). Thanks.
[Privacy](#) | [Disclaimer](#)
[Log in](#)

Vmag	V or other (see n_Vmag)	phot.mag;em.opt.V mag 0			verongsos_data	org.gavo.../tap/run
mv	Visual magnitude	phot.mag;em.opt.V mag 0			dmubin_main	org.gavo.../tap/run
vmag	Magnitude in Johnson V	phot.mag;em.opt.V mag 0			hipparcos_main	org.gavo.../tap/run
Hpmag	Median magnitude in Hipparcos system	phot.mag;em.opt.V mag 0			hipparcos_main	org.gavo.../tap/run
Hpscatt	Scatter on median magnitude	phot.mag;em.opt.V mag 0			hipparcos_main	org.gavo.../tap/run
Hpmax	Hpmag at maximum (5th percentile)	phot.mag;em.opt.V mag 0			hipparcos_main	org.gavo.../tap/run
Hpmin	Hpmag at minimum (95th percentile)	phot.mag;em.opt.V mag 0			hipparcos_main	org.gavo.../tap/run
Vmag	Optical V magnitude	phot.mag;em.opt.V mag 0			lspm_main	org.gavo.../tap/run
Vmag_e	Estimated V mag	phot.mag;em.opt.V mag 0			lspm_main	org.gavo.../tap/run

UCD UCD of column (many services still have 1.0 UCIDs).

Table words Terms in table description (case does not matter, patterns not allowed)

Column words Terms in column description (case does not matter, patterns not allowed)

Table Sort by Limit to items.

Output format

Fig. 3

You can also see that the server imposes restrictions on how long your job may run (time limit) and tells you when it will discard your job (destruction due). You can usually ask for less strict limits here (in tapsh, that's the limit and keepfor commands).

5. What's out there?

To find out what is already on TAP, you could ask the registry, use VODesktop, or query each server's TAP_SCHEMA yourself. Or you could just use GloTS¹.

GloTS is for "Global TAP Schema". The TAP_SCHEMA itself is a database schema each (compliant) TAP server has, containing tables named tables, columns, and keys (try `SELECT * from TAP_SCHEMA.columns`). At GloTS, we try to bring all this info together.

(vgl. Fig. 3)

¹ <http://dc.zah.uni-heidelberg.de/glots/q/plain/form>

Parameters

- Column words: redshift
- Table words: quasar

Result

Matched: 4

Column	Description	UCD	Unit	Index?	Table description	Table name	id	Src
z	Redshift	N/A	N/A	N/A	The catalog of all confirmed quasars in SDSS Data Release 3. <i>NOTE: The DRxQuasarCatalog a derived science table that may not be loaded for every release or	unknown.DR3QuasarCatalog wfau.roe.ac.uk/sdssdr7-dsa		http://wfau/sdssdr7-d

Fig. 4

6. Running Queries

Though many services will let your web browser operate a TAP services more or less crudely, you definitely want specialized TAP client software:

- VODesktop²
- TAP shell³
- tapquery⁴ – a python library that's part of GAVO's python votable package
- more soon, or so we hope...

Here, I'm running the TAP shell.

7. A quick example

Let's look for really red objects close to quasars in the redshift range between 0.1 and 0.12. First, check GloTS for tables about quasars having a redshift column:

(vgl. Fig. 4)

² <http://www.astrogrid.org/wiki/Install/Downloads>

³ <http://vo.ari.uni-heidelberg.de/soft>

⁴ <http://vo.ari.uni-heidelberg.de/soft>

	specObjId	SDSSName	RA	dec	z	p:
1	121820031053987840	073623.12+392617.8	114,09637	39,43828	0,1181	16
2	143214198736814080	110538.99+020257.3	166,41246	2,04927	0,1055	16
3	170234563912531968	134617.54+622045.4	206,57312	62,34596	0,1164	17
4	156723388675522560	093943.74+560230.4	144,93226	56,0418	0,1169	17
5	164886317182746624	142748.28+050222.0	216,9512	5,03946	0,1061	16

Fig. 5

8. Query WFAU

That's almost enough to let us query the service. In tapsh, using tab-completion, say:

```
server ivo://wfauroe.ac.uk/sdssdr7-dsa
select * from DR5QuasarCatalog as t where t.z between 0.1 and 0.12
run
send to topcat
```

Unfortunately, WFAU runs software expecting a rather ancient version of ADQL. That's why we need the weird "as t" nonsense. Sorry about that; as I said, the wires sometime show. But the service operators within the VO are nice and decent people, and if something doesn't work as expected, you are most welcome to ask.

Anyway, after this, topcat loads the result table. Display it to see:

(vgl. Fig. 5)

9. ...and search on

Now find a table with the UCD mag.phot;em.IR.K. Let's use PPMXL. We need to upload the last result; therefore, we give it a fixed nick.

```
nick quas
server ivo://org.gavo.dc/__system__/tap/run
select * from TAP_UPLOAD.uploaded as u join ppmxl.main as p on (1=contains(point('ICRS',
p.raj2000, p.dej2000), circle('ICRS', ra, dec, 0.5))) where kmag-jmag>0.5
upload result quas as uploaded
run
send to topcat
```

10. What next?

- Clean up: `echo "purge all" | tapsh`
- Download tapsh⁵ (and, of course, topcat⁶); see <http://www.g-vo.org>
- Complain to whoever made your favourite data set if it's not yet on TAP; tell them to talk to us if they don't know how to get it there.
- Report bugs, problems, and ideas.
- Be almost there!

⁵ <http://vo.ari.uni-heidelberg.de/soft/tapsh>

⁶ <http://www.star.bris.ac.uk/~mbt/topcat/>