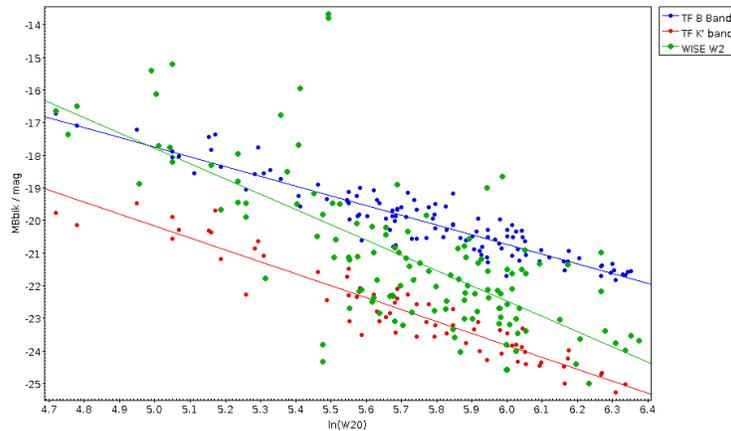


# Data Discovery Using the Virtual Observatory Registry



F. Rothmaier, S. Gabriel, M. Demleitner

March 14, 2023

## Abstract

The Virtual Observatory Registry (VO Registry) is a database for metadata on astronomical “resources” – essentially data services and the like. Its function is to help astronomers (or programs on their behalf) locate data relevant to their research. As an example, here we will look for resources giving V magnitudes and redshifts of quasars, and then for data pertinent to a specific Galaxy cluster.

There are many ways to access registry data – a fairly powerful one is the Web Interface to the Relational Registry (WIRR for short) at <http://dc.g-vo.org/WIRR>. To actually speak to the VO services discovered, clients speaking the service’s protocols are still required. In this tutorial, TOPCAT plays that role.

**Software:** a javascript-capable web browser of your choice, [TOPCAT](#)

## 1 Querying the Registry

- ▷ **1** *Prepare for the query with WIRR* – Open a web browser with <http://dc.g-vo.org/WIRR><sup>1</sup>. Try to get familiar with the search form by playing a little bit around. Change the search criteria (e.g., from the free text to a waveband search in the drop-down menu), and add (click + **Add constraint**) or remove one or several constraint lines (- button on the left-hand side) to the search form. The **Run Query** button sends your query to the database.
- ▷ **2** *Set constraints for our search* – Next, translate our problem “data with V magnitudes and redshifts of quasars” into WIRR’s criteria.
- The topical constraint “quasar” we cover with a full text match; set an **Any Text Field** constraint to **match quasar**. This will do a full-text search; an interesting alternative would be to use a constraint on subjects – see the **UAT Term** constraint. This will probably gain precision (i.e., fewer false positives returned) at the expense of recall (i.e., relevant resources whose publishers have neglected to properly give their subject keywords will be missing).
  - Add a new constraint line; let us interpret the “data” part of our problem statement as “catalog”. We are thus looking for catalog services. In WIRR, that means a **Service Type** constraint set to **Cone Search**. Note that depending on what you’re planning, it might actually be smarter to look for TAP services, but using these is a bit more involved.
  - For the physical constraints (existence of columns for magnitudes and redshifts), we employ *Unified Content Descriptors* or UCDs for short, which are a controlled vocabulary of concept labels covering much of what you will encounter in astronomical tables. Of course, you don’t have to know the UCDs for redshifts and V band magnitudes by heart. Instead, add a **Column UCD** like constraint and click on the **Pick one** button appearing to look for suitable UCDs. By typing **redshift** and hitting **Search**, we end up with a list of possible UCDs with a short description; in our case, choose **src.redshift**.
  - After adding another UCD constraint, type something like **V band magnitude** into the search field and pick **phot.mag;em.opt.V** from the options returned.

All in all, you should end up with something like Fig. 1.

---

<sup>1</sup>Make sure that JavaScript is enabled or at least grant the originating page execute privileges for JavaScript.



Figure 1: WIRR query form for redshifts and V magnitudes of quasars.

## 2 Using the Results

Click the **Run Query** button. Matching resources are displayed in the box below the search form. The results show up in a tabular layout with resource title, the capabilities (i.e., things you can talk to), and a contact info button. Click on the arrow in the beginning of each row to get more detailed information about a resource (Fig. 2).

- ▷ **3** *Connect WIRR to TOPCAT via SAMP and send the resource list* – Start TOPCAT.

At the bottom of the WIRR page, you will find a **Connect to SAMP** hub button – hit it; if this fails, make sure you actually have TOPCAT running.

TOPCAT will ask you whether the connection should be allowed – this is because in principle, whoever can access ports on your machine, can try to connect to your SAMP hub. Therefore, only allow hub access if you just initiated something that would cause a transfer to TOPCAT.

After WIRR has its SAMP connection, the text on the SAMP button changes to **Send to VO-Client(s): Cone search**. Additionally, the bottom box in TOPCAT's main window should look somewhat like Fig. 3.

The “GAVO” icon indicates that WIRR has successfully connected to TOPCAT.

Try to send the resource list to TOPCAT by clicking on **Cone search** in WIRR. A warning shows up saying that TOPCAT has received the list successfully but could not make use of it. This is because TOPCAT only accepts service lists if the dialog for running the respective services is open. Hence, select **VO** → **Cone Search** in TOPCAT's menu. In the window that then opens, make sure **Accept Resource Lists** is checked (it is by default).

Send the resource list again – it will now turn up in TOPCAT's Cone Search window.

- ▷ **4** *Run a few cone searches on different services* – In that Cone Search window,

Total matching: 66 [SQL](#)

- ▶ 1410MHz-spectra of quasars III. [SCS](#)  
- ▼ 5000 AGNs behind the Magellanic clouds [SCS](#)  

**Description**

We show that using mid-IR color selection to find active galactic nuclei (AGNs) is as effective in dense stellar fields such as the Magellanic Clouds as it is in extragalactic fields with low stellar densities using comparisons between the Spitzer Deep Wide Field Survey data for the NOAO Deep Wide Field Survey Bootes region and the SAGE Survey of the Large Magellanic Cloud. We use this to build high-purity catalogs of ~5000 AGN candidates behind the Magellanic Clouds. Once confirmed, these quasars will expand the available astrometric reference sources for the Clouds and the numbers of quasars with densely sampled, long-term (>decade) monitoring light curves by well over an order of magnitude and potentially identify sufficiently bright quasars for absorption line studies of the interstellar medium of the Clouds.

**VOA Identifier ("ivoid")**  
<ivo://cds.vizier/f/apj/701/508>

**Source**  
[2009ApJ...701..508K](#)

**Reference URL**  
<https://cdsarc.cds.unistra.fr/viz-bin/cat/J/ApJ/701/508>

- ▶ A 3.5mm polarimetric survey of radio-loud AGNs [SCS](#)  
- ▶ A survey for quasars near quasars - QNQ [SCS](#) [SCS](#) [SCS](#)  

Figure 2: WIRR results; we have clicked on the arrowhead to pop up the detailed information for one result.

you can now enter the name of your favorite quasar in the field **Object Name** (use 3C273 if you have no favourite quasar), hit **Resolve**, enter a small search radius and then query each service in turn by double-clicking it (hint: the pin icon in the top left corner – or **Window** → **Stay Open** – keeps the window open between queries). See Fig. 4 for how this should look like.

Of course, you can use the resulting services more usefully, too – one thing that doesn't work in this way is querying *all* of the matched services at once. For Cone Searches, that is somewhat difficult, as the different tables have vastly different structures – producing a union of the various results hence is everything but straightforward.

For images and spectra, joining results is somewhat easier, so many clients – like Aladin or Splat – support cross-server queries. TOPCAT, however, does



Figure 3: TOPCAT SAMP status after connecting WIRR.

not at this point. If you want to see how to do multi-service queries yourself, check out our little pyVO showcase at <http://docs.g-vo.org/pyvo.pdf>.

### 3 Using the Registry to Investigate a Galaxy cluster

As a variation of these techniques, let us find services giving data for the galaxy cluster Abell 242 and compare what they have to say.

- ▷ 1 *Remove unnecessary constraints* – Go back to the web browser with the WIRR page. We are still interested in redshifts, but we do not need V magnitudes and a restriction to quasars any more. Remove the corresponding constraints using the - button.
- ▷ 2 *Add a spatial constraint* – This time, we only want services that (may) have data for the object we are interested in, Abell 242. So, add the new constraint **Blind Discovery** → **ICRS footprint**<sup>2</sup> Type **Abell 242** (or any other name known to Simbad), and by clicking **Resolve name** the object’s coordinates are filled in.
- ▷ 3 *Obtain services and send them to TOPCAT* – Hit **Run Query** as before, and then again **Cone Search**. You only have to re-connect to SAMP if you reloaded the web page in between.
- ▷ 4 *Configure TOPCAT’s Cone Search* – Somewhat regrettably, TOPCAT does not know about the spatial constraint you set in WIRR. You will thus have to enter **Abell 242** again in TOPCAT’s cone search window and then hit its **Resolve** button. Also, set **Radius** to, perhaps, 6 degrees.
- ▷ 5 *Retrieve data from a few services* – Now, you can go through the list of services and see which of the services actually returns data useful to your task at hand – just try as many as you like. For every successful search, a new table pops up in TOPCAT’s main window, from where the table data can be further processed.

Hint: click on the pin icon in the left upper corner of the cone search window (or equivalently select **Window** → **Stay Open** in the menu). That way, you can double click a service entry to query the service but stay in the Cone Search window rather than have to navigate back to it from TOPCAT’s main window.

To reproduce what we did here, look for services with the short names **VII/249** (the 6dF-DR2 Galaxy survey from VizieR) and **aspic.6df SCS** (the ASPIC 6dF

<sup>2</sup>ICRS stands for the “International Celestial Reference System”, which basically means “equatorial coordinates”. Almost all VO protocols just deal with those, so if all you have is Galactic or Supergalactic coordinates, you will need to transform them; spatial coverage on, say, planetary surfaces cannot be queried through the VO Registry yet.

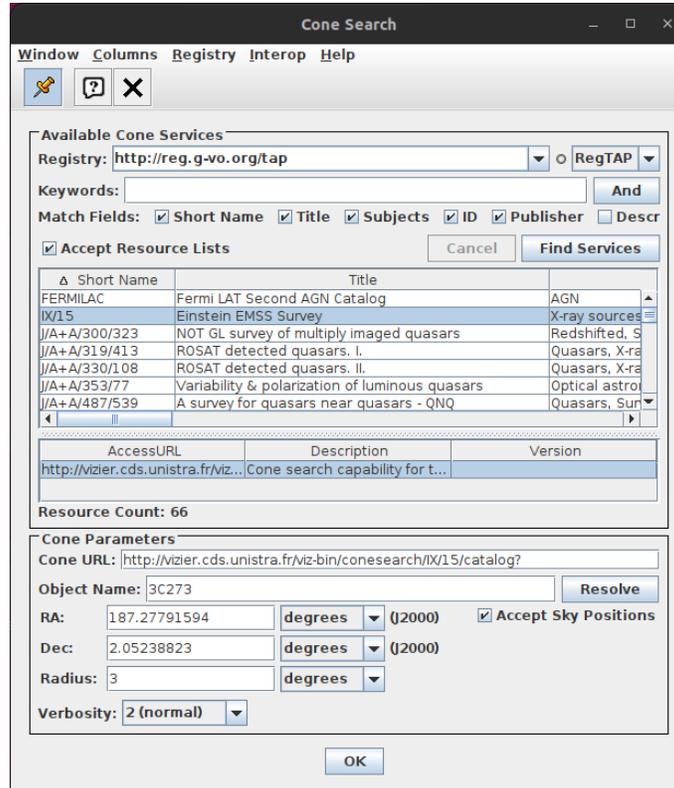


Figure 4: A TOPCAT Cone Search window populated with services coming from WIRR.

from CESAM) and double-click on each. This will retrieve the sources from each in the cone you defined above. But do yourself a favour and try a few other services, too.

- ▷ **6** *Crossmatch the two tables* – To combine the two data sets, click the  symbol in TOPCAT's main window (or, equivalently, choose **Joins** → **Pair match**). Select the tables you just retrieved in the drop-down menus and click **Go**, which creates a new table with the matched data. A new window (Fig. 5) should open, which informs you about the successful match; perhaps experiment a bit with the **Max Error** setting in the crossmatch window if TOPCAT misses matches you would have expected.
- ▷ **7** *Investigate the joined data* – For instance, you can see how well **cz** from 6dF-DR2 and **z\_spec** from ASPIC agree. See if you can figure out where the obvious outliers come from, or draw a color-magnitude diagram with infrared magnitudes

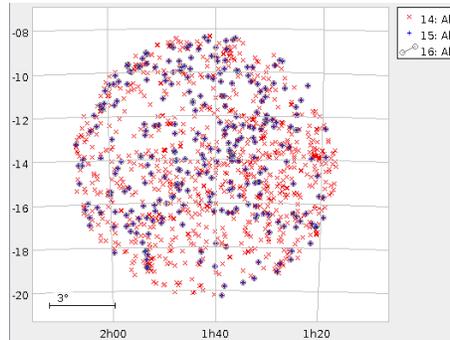


Figure 5: TOPCAT's rendering of a match between two resources.

from ASPIC and the `rFmag` from 6dF-DR2.

To check whether the matching of our objects actually worked, let's look at a plane plot (  in the main window) of the matched table. In the new opened window, choose `redshift` as X and `z` as Y in the drop-down menus to plot the redshifts of both datasets against each other (if you have chosen other services that we have, you have to look the variable name of the redshift up in the table browser ). You will probably see something like Fig. 6.

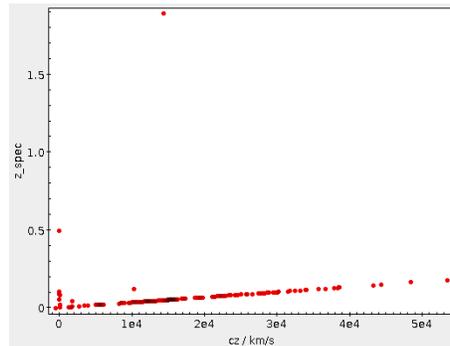


Figure 6: Correlation of the redshifts from our two resources, in a TOPCAT plane plot.

What have you been expecting and are those expectations met? If not, what could be the reasons for that?

## 4 The Tully-Fisher Relation in the Mid-Infrared

To round this exercise off, let us see whether we can find a correlation between properties we take from two different, independently discovered resources.

What we would like to reproduce is the Tully-Fisher (TF) relationship, which relates the maximum rotational velocity  $v_{\max}$  of a disk (spiral) galaxy to its luminosity like this:

$$L \propto v_{\max}^{\beta},$$

where  $\beta$  could be something like 3 or 4 depending on what part of the spectrum you measure  $L$  in. It stands to reason that brighter galaxies in the mean are more massive, and more mass means faster rotation. Still, let us see if we can find this relation in VO data.

As a proxy for the rotational velocity, it is common to use the width of the 21 cm line of neutral hydrogen; as a proxy for the (logarithm of) the luminosity, we will use some magnitude.

- ▷ **1** *Find 21 cm line widths with WIRR* – Go back to the WIRR page in your browser. Drop the constraints on the ICRS footprint and the service type; there are not that many 21-cm line width measurements to begin with. To find 21 cm data, create a UCD constraint and try “line width” in **Pick one**. Scrolling a bit, you can see there is a UCD `spect.line.width;em.line.hi`, exactly what we are looking for.

Browsing through the matches of such a request, you will find that **Tully and Pierce (2000)** have already done much of the work for us: There is a resource “Calibration of the Tully-Fischer relation”. Send over the resources to TOPCAT as before (after opening the cone search window if necessary).

- ▷ **2** *Grab the whole Tully-Fisher data and plot it* – Often, you want to avoid bulk downloading data. In this case, we have a relatively small catalogue, and we want to look at all of its objects anyway. So, let’s pull it whole; to do that from TOPCAT’s cone search window, set **RA** and **Dec** both to 0, and **Radius** do 180. For good measure, set **Verbosity** to 3, meaning you really want all the columns that are in the catalogue.

Send off that query. I will call the resulting table “TF table” in the following.

To see Tully-Fisher relationship, open a plane plot and plot `ln(W20)`<sup>3</sup> against `MBb1k`, the absolute magnitude in the blue band. You should see a rather well-defined straight line, as is to be expected for a power law in a log-log plot (note again that magnitudes are logarithms of luminosities). If you want, add a **LinearFit** in the Form tab to see the slope (which is the  $\beta$  in the formula above; it should be rather close to  $-3$  here).

- ▷ **3** *Overplot the TF relationship in the K band* – The  $\beta$  is somewhat dependent on the wave band. Let’s add a plot for the infrared K’ band (around 2.2 mi-

<sup>3</sup>Yes, you can enter reasonable expressions into the plot fields in TOPCAT; see [Help](#) → [Available Functions](#) to see what is available.

rometer). So, do [Layers](#) → [Add Position Control](#) in your plane plot window and configure that second plot in the **Table** box to show data from the TF table (at this point, that is probably the last entry in the list), and then set **X** to  $\log_{10}(W20)$  and **Y** to  $MK'_{\text{bik}}$  almost as before. This time, the beta should be closer to  $-3.7$ .

So... what would happen farther in the infrared?

- ▷ **4** *Find an all-sky survey in the mid-infrared* – Let’s look for resources in the mid-infrared in WIRR. You know the drill: Make or re-use a **Column UCD** constraint, hit **Pick one**, and this time try “Mid infrared”. The UCD `phot.mag;em.ir.4-8um` looks promising, and since that yields more than 1000 records and the data we would like to match is all sky, add an **Any Text Field** constraint with “All Sky” (yes, it would be good to have a better way to express such a property).

This still gives about 250 resources, but paging through them, the AllWISE Data Release shows up rather quickly; let’s use that ([Cutri and Wright et al., 2021](#)).

You could try to pull that catalogue in with a second cone search, but AllWISE is a dense catalog, and you certainly cannot (sensibly) download and cross-match it locally. So, the recipe we used above won’t (really) work<sup>4</sup>.

There is a much better protocol for this kind of thing (and many other kinds of things): TAP. With TAP, you can run the cross match on the server side. Regrettably, sending the TAP resources you have found to TOPCAT is not quite as straightforward as with the Cone Searches (yet).

- ▷ **5** *Prepare the TAP query* – Instead, open the TAP window in TOPCAT ([VO](#) → [Table Access Protocol](#)). Then, copy the link below the TAP link in WIRR’s line for the AllWISE Data Release to the **TAP URL:** field in TOPCAT’s TAP window; in most browsers, you can do that with a right click and “Copy link location” or such. Alternatively, you can probably also drag the “TAP” link from WIRR window into the TAP URL field, too.

Hit **Use Service** in the TOPCAT TAP window. This then retrieve the table metadata from the remote service, which is quite a bit. Wait for something to show up in the **Metadata** pane, and then navigate to the allwise table, e.g., by typing “allwise” into **Find**. The table you want is called `II/328/allwise`; select it. Yes, that is a bit cumbersome, and it’s worse on Vizier than on most TAP services. But on the other hand, they *do* have a lot of tables over there.

- ▷ **6** *Prepare the upload table* – What we will do in a moment is upload the TF table into the remote database so it can do the matching. To save on resources and to speed up things it is a very good idea to only upload the columns you actually need. Additionally, for the TF table some of the column names will not be accepted by the database engine, and if you try to upload the table without editing the names or hiding the columns, you will get an error like “Impossible

<sup>4</sup>And don’t even *think* of running a multicone search unless you are sure you have tried everything else: These are a drag for the service operators.

to create uploaded table”.

So, while you have the TF table selected in TOPCAT’s main window, go to [Views](#) → [Columns Info](#). In there, there are checkboxes in the **Visible** column, and you want to uncheck them all first (there’s a button for that). Then, check the columns you need in the later plot, that is: `_RAJ2000`, `_DEJ2000`, `mu` (the distance modulus, which we could use to make absolute magnitudes), `Vcmb` (the system velocity, which we’ll need as a stand-in for the redshift), and of course the line width, `W20`.

- ▷ **7** *Generate and execute the crossmatch query* – With this, you are ready to do the remote cross match. Make sure again you have your TF table selected in TOPCAT’s main window and `II/328/allwise` in the Metadata pane in the TAP window. If you have that, you can hit [Examples](#) → [Upload](#) → [Upload join](#) in the TAP window, and TOPCAT will write a query for you in a language called ADQL (that you don’t need to understand in any depth here).

In this case, you can just send it off, and a few moments later, you will have a new table that has the columns we uploaded as well as magnitudes in various mid-infrared bands. Let’s look at the W2 band (that’s the name of the bandpass, and it’s a bit unfortunate that it looks so much like our HI width W20 that has nothing to do with it whatsoever), right around 4.6 micrometer.

- ▷ **8** *Add the correlation of W2 and W20 into the plot* – Go back to your plot with the two TF relationships. Add another position control, and set its **Table** to your newly downloaded crossmatch. As **X**, use the usual `ln(W20)`. In **Y**, you need correct for the various distances of the objects. You can write `W2mag-mu` (the distance modulus is defined as the difference between apparent and absolute magnitude), but you will immediately notice that there are only very few objects left then. If you inspect [Views](#) → [Column Statistics](#) (in TOPCAT’s control window), you will see that `mu` is NULL quite often (for me, `nGood` is just 21 of 113).

But we can substitute for that with the velocity `Vcmb`. In the non-relativistic regime we are in here, the distance is  $d = v/H_0$  with the Hubble constant  $H_0$ . Then, given the absolute magnitude is  $M = 5 - 5 \log d + m$  (with  $d$  in parsec), with a bit of unit algebra you can work out that you want

$$W2mag + (5 - 5 * \log_{10}(Vcmb / 75 * 1e6))$$

in **Y**. With that, you end up with Fig. 7; yes, the (green) W2 data has quite a bit more scatter, and the blind regression is quite a bit steeper than the others (−4.5 for me) – but given our modest effort: isn’t this fairly impressive?

## 5 Other Modes of Using the Registry

The Registry data is queriable via an open protocol called RegTAP, and WIRR is by no means the only client using it. Actually, the **SQL** button on WIRR’s UI

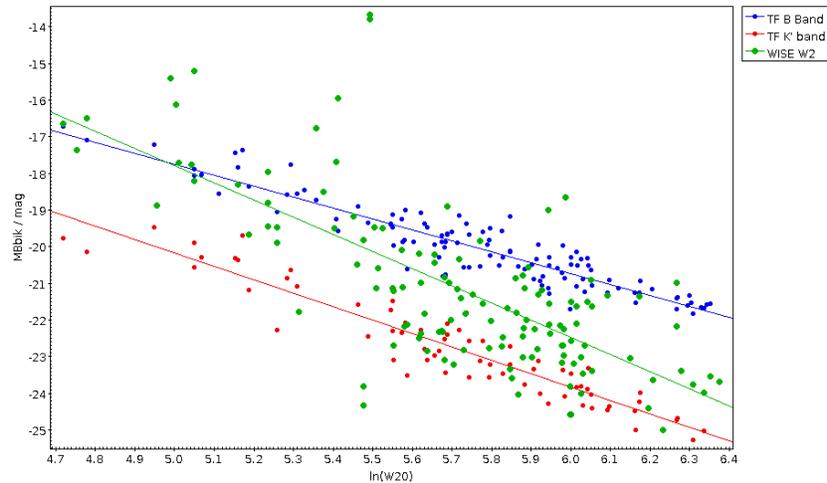


Figure 7: Our Tully-Fisher relationship in three bands. The *red* and *blue* data are from one VO resource, the *green* data were added from AllWise by us.

produces ADQL that you can execute on TAP services carrying the necessary tables (such as the GAVO DC TAP service you will find in TOPCAT’s TAP window).

Simple keyword searches are also possible directly from within TOPCAT. In [Aladin](#), the Registry feeds the discovery tree at the left side of the application.

You can also do sophisticated Registry queries from within pyVO (without directly having to figure out ADQL). A post on our blog<sup>5</sup> should get you started.

## References

Cutri, R. M., Wright, E. L., Conrow, T., Fowler, J. W., Eisenhardt, P. R. M., Grillmair, C., Kirkpatrick, J. D., Masci, F., McCallon, H. L., Wheelock, S. L., Fajardo-Acosta, S., Yan, L., Benford, D., Harbut, M., Jarrett, T., Lake, S., Leisawitz, D., Ressler, M. E., Stanford, S. A., Tsai, C. W., Liu, F., Helou, G., Mainzer, A., Gettnigs, D., Gonzalez, A., Hoffman, D., Marsh, K. A., Padgett, D., Skrutskie, M. F., Beck, R., Papin, M. and Wittman, M. (2021), ‘VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013)’, *VizieR Online Data Catalog* p. II/328.

<https://ui.adsabs.harvard.edu/abs/2014yCat..2328....0C>

<sup>5</sup><https://blog.g-vo.org/towards-data-discovery-in-pyvo.html>



Tully, R. B. and Pierce, M. J. (2000), ‘Distances to Galaxies from the Correlation between Luminosities and Line Widths. III. Cluster Template and Global Measurement of  $H_0$ ’, *ApJ* **533**(2), 744–780, arXiv:astro-ph/9911052.  
<https://ui.adsabs.harvard.edu/abs/2000ApJ...533..744T>