

External Matching to the Meert Spectroscopic catalog

Alan

13 May 2013

The catalog of fitted SDSS spectroscopic galaxies described in Meert et al. (2013a) and Meert et al. (2013b) and used in several papers (e.g. Shankar et al. 2013; Huertas-Company et al. 2012; Bernardi et al. 2013b,a) can be easily matched to external catalogs using a simple python script. In this short paper, I will describe the structure of the MySQL database, and the method used to extract a preferred catalog.

1 THE DATABASE

The Meert catalog contains several cross-matched data sets:

- Selected SDSS DR7 photometric and spectroscopic information (Strauss et al. 2002; Abazajian et al. 2009)
 - Calculated values of angular diameter distance, distance modulus, k-corrections using Blanton & Roweis (2007)...
 - cross-matches to the BAC morphology from Huertas-Company et al. (2011)
 - the original data set used in Bernardi (2009) and Bernardi et al. (2007)
 - the JHU catalog of stellar masses and star formation rates¹
 - the Yang Group catalog (Yang et al. 2005)
 - the Simard sample (Simard et al. 2011)
 - the Lackner and Gunn disk galaxy sample (Lackner & Gunn 2012)
 - the NYU VAGC (Blanton et al. 2005)
 - as well as cross-matches to GALEX (Martin et al. 2005)
- UV data and UKIDSS (Hambly et al. 2008) IR data
- and other data ...

The heart of the database is a unique identifier referred to as **galcount**. This unique ID number serves a purpose similar to the **objid** parameter in SDSS catalogs. It allows quick matching across data sets and convenient reference to specific targets (references require only this 6-digit number to uniquely identify all other data associated with the galaxy). When matching internal data, using **galcount** is essential to allowing a quick, efficient search of the database. Additional unique identifiers for individual data tables are retained and are internally consistent for separate tables from the same data source. For example, a unique identifier provided in the Yang group catalog is **IGAL**. This identifier can still be used in the Meert database for queries specific to

tables originally part of the Yang catalog, but has no meaning beyond this scope. Efficient searching of the databases relies on use of a unique identifier. It is still possible to match data based on a non-unique column, but this has the potential to take prohibitively long and produce a meaningless search result, so caution is advised when attempting this task.

1.1 The Databases and Tables

The data is organized in a database/table structure. The databases are:

- catalog** Contains the bulk of the data including CasJobs, the Meert fits, the DR6 sample from Bernardi, and ancillary data
- yang** Contains the Yang group catalog
- simulations** Contains the simulation input data and output fits.
- JHU** Contains the JHU stellar masses and SFRs
- simard** Contains the raw values from Simard et al. (2011) a version of these tables is also available in the catalog database
- lackner** Contains the fits from Lackner & Gunn (2012)
- additional dbas** more databases exist. I will exhaust this list at a future date

Within each database is a set of tables. The tables are large in number (~ 300) and each table contains columns of data. Since the numbers are so large, there is little benefit to enumerating them at this time, but in the future, they will be outlined. In lieu of enumerating the tables here, I provide an example file that queries data later in the paper (Section 4).

2 THE MATCHING ALGORITHM

The method used to match these data is a simple nearest neighbor cross-match on ra/dec, requiring the two matched galaxies to be nearest neighbors in both directions. For example if galaxy *a* from catalog *A* and galaxy *b* from catalog *B* are matched this implies that *b* is the closest object to the ra/dec of *a* in the entire catalog *B* and galaxy *a* is the closest object to the ra/dec of *b* in the entire catalog *A*. Initial matching was carried out using Topcat², a Java-based utility used to manipulate large catalogs. The exter-

¹ see <http://www.mpa-garching.mpg.de/SDSS/>

² see <http://www.star.bris.ac.uk/~mbt/topcat/> for the Topcat program description

nal matching utility is powered by a Healpix mapping of the Meert catalog. Using the python module `healpy`, external catalogs can be quickly cross-matched. Operation of this basic cross-matching routine is available in Section 3. More sophisticated cross-matching is available by downloading the catalog and using your own matching code.

3 EXTERNAL CROSS-MATCHING

Given a catalog provided by the user, the program `crossmatch.py` will provide a matched catalog, retaining the same row order as the original catalog. This original catalog should contain at a minimum two columns (ra and dec). Additional columns are allowed, but they will be ignored by the program. `crossmatch.py` can be run at the command line. Here is the “help” string associated with the program:

Usage: `crossmatch.py` OPTIONS

This program will crossmatch an input catalog of ra/dec coordinates to our catalog and output the results. In general, you need to specify an input table, column file, output table, and a table name.

Options:

```
-h, --help      show this help message and exit
-i INCAT, --input-cat=INCAT
                input catalog to be cross-matched
-c IN_FILE, --columns=IN_FILE
                columns to be in final catalog
-o OUTFILE, --output-file=OUTFILE
                output file for cross-matched catalog
-t TBL_FILE, --table-file=TBL_FILE
                table file containing mysql table list
-n TABLENAME, --table-name=TABLENAME
                name of temporary table for
                cross-matched catalog
```

Note: this database is hosted on `chitou.physics.upenn.edu` for the time being. Any filepath you specify must be writable by both you and the mysql user (generally this restricts the output path to be either `/tmp/` or `/scratch/`. Be sure to choose an appropriate filepath. The column file `IN_FILE` is described in Section 4 because it lists the columns desired in the final output.

4 THE QUERY PROGRAM

The query program writes the desired data table to a text file and is executed at the command line. The help string for the program is:

Usage: `generate_table.py` OPTIONS

This program, when run alone, will generate an ascii txt table of data from all of our catalogs. In general, you need to specify at least a column file, output file name, and a table name.

Options:

```
-h, --help      show this help message and exit
-c INFILE, --columns=INFILE
                columns to be in final catalog
-o OUTFILE, --output-file=OUTFILE
                output file for cross-matched catalog
-t TBL_FILE, --table-file=TBL_FILE
                table file containing mysql table list
-l LEAD_TABLE, --lead-cat=LEAD_TABLE
                the table by which the selection
                will be made
```

It is also invoked directly by `crossmatch.py`. Column list is a file containing the desired output columns. These columns can be complex functions of actual data columns. There are two important rules for the column file. The first rule is that each value to be output into the final table should occupy its own line in the column list. Multiple outputs on the same line will break the program. The second rule is that the ordering of the final output is always based on the first entry in the column list. Therefore, the first column should utilize either the galcount or the rowcount of a table. Objects are sorted in ascending order. Most tables with an order (e.g. SSDR6) contain a rowcount variable that stores the original ordering of that catalog. Using `catalog.SSDR6.rowcount` as the first column would output a catalog that matches the SSDR6 data in row order.

The `LEAD_TABLE` sets the primary table of the matching program. The primary data table is joined on each table from which a data column is requested. The table identified by the `LEAD_TABLE` is the source of all the objects in the final table. If `crossmatch.py` is the calling program, then the temporary catalog fed to the program will be the leading table. This table will contain a `rowcount` column. Therefore, the column list provided to the crossmatch program should have

```
temp.crossmatch.rowcount
```

as the first row of the table.

5 AN EXAMPLE QUERY

We have already produced many tables based on the data of the SDSS DR4 sample, updated with DR6 values. I refer to this table as the SSDR6 throughout the paper. The column list for this common output is shown in Figure 5.

Placing the file from Figure 5 into a file called `SSDR6.columns`, we can create our standard table with row-number, ra, dec, z, sersic and ser+exp fit values, morphological probabilities, Vmax, absmagcorr, and SexMag by typing:

```
./generate_table.py -o /tmp/SSDR6_out.txt
-l catalog.SSDR6 -c SSDR6.columns
```

Note that I have omitted the table file. This causes the program to use a default table file with all the mysql tables in it. In general, this is the preferred method. The output is stored in `/tmp/SSDR6_out.txt` which I can then copy to my preferred location.

```

catalog.SSDR6.rowcount
catalog.CAST.ra_gal
catalog.CAST.dec_gal
catalog.CAST.z
catalog.r_rerun_ser.m_tot-catalog.CAST.extinction_r-catalog.DERT.dismod-catalog.DERT.kcorr_r
catalog.r_rerun_ser.n_bulge
log10(catalog.r_rerun_ser.Hrad_corr/sqrt(catalog.r_rerun_ser.ba_tot_corr)*catalog.DERT.kpc_per_arcsec)
catalog.r_rerun_ser.ba_tot_corr
catalog.r_rerun_serexp.m_tot-catalog.CAST.extinction_r-catalog.DERT.dismod-catalog.DERT.kcorr_r
catalog.r_rerun_serexp.n_bulge
log10(catalog.r_rerun_serexp.Hrad_corr/sqrt(catalog.r_rerun_serexp.ba_tot_corr)*catalog.DERT.kpc_per_arcsec)
catalog.r_rerun_serexp.ba_tot_corr
catalog.r_rerun_serexp.BT
catalog.M2010.probaE
catalog.M2010.probaE11
catalog.M2010.probaS0
catalog.M2010.probaSab
catalog.M2010.probaScd
catalog.DERT.Vmax
catalog.DERT.dismod+catalog.DERT.kcorr_r
catalog.r_rerun_fit.SexMag-catalog.CAST.extinction_r

```

Figure 1. The example column list used to produce a catalog similar to those commonly used in the past.

REFERENCES

- Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., Bailer-Jones C. A. L., Barentine J. C., Bassett B. A., Becker A. C., Beers T. C., Bell E. F., Belokurov V., Berlind A. A., Berman E. F., Bernardi M., Bickerton S. J., Bizyaev D., Blakeslee J. P., Blanton M. R., Bochanski J. J., Boroski W. N., Brewington H. J., Brinchmann J., Brinkmann J., Brunner R. J., Budavári T., Carey L. N., Carliles S., Carr M. A., Castander F. J., Cinabro D., Connolly A. J., Csabai I., Cunha C. E., Czarapata P. C., Davenport J. R. A., de Haas E., Dilday B., Doi M., Eisenstein D. J., Evans M. L., Evans N. W., Fan X., Friedman S. D., Frieman J. A., Fukugita M., Gänsicke B. T., Gates E., Gillespie B., Gilmore G., Gonzalez B., Gonzalez C. F., Grebel E. K., Gunn J. E., Györy Z., Hall P. B., Harding P., Harris F. H., Harvanek M., Hawley S. L., Hayes J. J. E., Heckman T. M., Hendry J. S., Hennessy G. S., Hindsley R. B., Hoblitt J., Hogan C. J., Hogg D. W., Holtzman J. A., Hyde J. B., Ichikawa S.-i., Ichikawa T., Im M., Ivezić Ž., Jester S., Jiang L., Johnson J. A., Jorgensen A. M., Jurić M., Kent S. M., Kessler R., Kleinman S. J., Knapp G. R., Konishi K., Kron R. G., Krzesinski J., Kuropatkin N., Lampeitl H., Lebedeva S., Lee M. G., Lee Y. S., French Leger R., Lépine S., Li N., Lima M., Lin H., Long D. C., Loomis C. P., Loveday J., Lupton R. H., Magnier E., Malanushenko O., Malanushenko V., Mandelbaum R., Margon B., Marriner J. P., Martínez-Delgado D., Matsubara T., McGehee P. M., McKay T. A., Meiksin A., Morrison H. L., Mullally F., Munn J. A., Murphy T., Nash T., Nebot A., Neilsen Jr. E. H., Newberg H. J., Newman P. R., Nichol R. C., Nicinski T., Nieto-Santisteban M., Nitta A., Okamura S., Oravetz D. J., Ostriker J. P., Owen R., Padmanabhan N., Pan K., Park C., Pauls G., Peoples Jr. J., Percival W. J., Pier J. R., Pope A. C., Pourbaix D., Price P. A., Purger N., Quinn T., Raddick M. J., Re Fiorentin P., Richards G. T., Richmond M. W., Riess A. G., Rix H.-W., Rockosi C. M., Sako M., Schlegel D. J., Schneider D. P., Scholz R.-D., Schreiber M. R., Schwobe A. D., Seljak U., Sesar B., Sheldon E., Shimasaku K., Sibley V. C., Simmons A. E., Sivarani T., Allyn Smith J., Smith M. C., Smolčić V., Snedden S. A., Stebbins A., Steinmetz M., Stoughton C., Strauss M. A., SubbaRao M., Suto Y., Szalay A. S., Szapudi I., Szkody P., Tanaka M., Tegmark M., Teodoro L. F. A., Thakar A. R., Tremonti C. A., Tucker D. L., Uomoto A., Vanden Berk D. E., Vandenberg J., Vidrih S., Vogeley M. S., Voges W., Vogt N. P., Wadadekar Y., Watters S., Weinberg D. H., West A. A., White S. D. M., Wilhite B. C., Wonders A. C., Yanny B., Yocum D. R., York D. G., Zehavi I., Zibetti S., Zucker D. B., 2009, *ApJS*, 182, 543
- Bernardi M., 2009, *MNRAS*, 395, 1491
- Bernardi M., Hyde J. B., Sheth R. K., Miller C. J., Nichol R. C., 2007, *AJ*, 133, 1741
- Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013a, *ArXiv e-prints*
- Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2013b, *MNRAS* submitted,
- Blanton M. R., Roweis S., 2007, *AJ*, 133, 734
- Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., Ivezić Ž., Knapp G. R., Lupton R. H., Munn J. A., Schneider D. P., Tegmark M., Zehavi I., 2005, *AJ*, 129, 2562
- Hambly N. C., Collins R. S., Cross N. J. G., Mann R. G., Read M. A., Sutorius E. T. W., Bond I., Bryant J., Emerson J. P., Lawrence A., Rimoldini L., Stewart J. M., Williams P. M., Adamson A., Hirst P., Dye S., Warren S. J., 2008, *MNRAS*, 384, 637
- Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, *A&A*, 525, A157+
- Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2012, *ApJ* submitted,
- Lackner C. N., Gunn J. E., 2012, *MNRAS*, 421, 2277
- Martin D. C., Fanson J., Schiminovich D., Morrissey P.,

- Friedman P. G., Barlow T. A., Conrow T., Grange R., Jelinsky P. N., Milliard B., Siegmund O. H. W., Bianchi L., Byun Y.-I., Donas J., Forster K., Heckman T. M., Lee Y.-W., Madore B. F., Malina R. F., Neff S. G., Rich R. M., Small T., Surber F., Szalay A. S., Welsh B., Wyder T. K., 2005, *ApJ*, 619, L1
- Meert A., Vikram V., Bernardi M., 2013a, In Prep,
- , 2013b, *MNRAS* submitted,
- Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, *MNRAS*, 428, 109
- Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnell A. W., 2011, *ApJS*, 196, 11
- Strauss M. A., Weinberg D. H., Lupton R. H., Narayanan V. K., Annis J., Bernardi M., Blanton M., Burles S., Connolly A. J., Dalcanton J., Doi M., Eisenstein D., Frieman J. A., Fukugita M., Gunn J. E., Ivezić Ž., Kent S., Kim R. S. J., Knapp G. R., Kron R. G., Munn J. A., Newberg H. J., Nichol R. C., Okamura S., Quinn T. R., Richmond M. W., Schlegel D. J., Shimasaku K., SubbaRao M., Szalay A. S., Vanden Berk D., Vogeley M. S., Yanny B., Yasuda N., York D. G., Zehavi I., 2002, *AJ*, 124, 1810
- Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, *MNRAS*, 356, 1293