The Planetary Archive

Paulo Penteado[1]*, David Trilling[1], Alexander Szalay[2], Tamás Budavári[2], César Fuentes[3]

[1]Northern Arizona University, [2]The Johns Hopkins University, [3]Universidad de Chile

*pp.penteado@gmail.com, http://www.ppenteado.net

Abstract

We are building the first system that will allow efficient data mining in the astronomical archives for observations of Solar System Bodies. While the Virtual Observatory has enabled data-intensive research making use of large collections of observations across multiple archives, Planetary Science has largely been denied this opportunity: most astronomical data services are built based on sky positions, and moving objects are often filtered out.

To identify serendipitous observations of Solar System objects, we ingest the archive metadata. The coverage of each image in an archive is a volume in a 3D space (RA,Dec,time), which we can represent efficiently through a hierarchical triangular mesh (HTM) for the spatial dimensions, plus a contiguous time interval. In this space, an asteroid occupies a curve, which we determine integrating its orbit into the past. Thus when an asteroid trajectory intercepts the volume of an archived image, we have a possible observation of that body. Our pipeline then looks in the archive's catalog for a source with the corresponding coordinates, to retrieve its photometry. All these matches are stored into a database, which can be queried by object identifier.

This database consists of archived observations of known Solar System objects. This means that it grows not only from the ingestion of new images, but also from the growth in the number of known objects. As new bodies are discovered, our pipeline can find archived observations where they could have been recorded, providing colors for these newly-found objects. This growth becomes more relevant with the new generation of wide-field surveys, particularly LSST.

We also present one use case of our prototype archive: after ingesting the metadata for SDSS, 2MASS and GALEX, we were able to identify serendipitous observations of Solar System bodies in these 3 archives. Cross-matching these occurrences provided us with colors from the UV to the IR, a much wider spectral range than that commonly used for asteroid taxonomy. We are retrieving the archived observations to derive spectrophotometry from searching for 640 thousand asteroids, from 0.1 to 2 μm. In the future we will expand to other archives, including HST, Spitzer, WISE and Pan-STARRS.

Moving body observations

Several data archives, particularly those of wide-field, wide-area surveys, contain serendipitous observations of Solar System bodies. But archive queries are usually on the RA/Dec domain. To allow easy collection of data from large numbers of Solar System objects, the archives would need to have databases with the occurrences of Solar System objects within its images, so that queries could be easily done for all moving bodies, or large lists of them.

There are two complementary approaches to populate such databases:

1) Known bodies: For each image in an archive, calculate if there is any known body which should fall on that image. These occurrences can be tagged, and the archive's source catalogs, when available, used to retrieve the corresponding photometry for the object.

2) New body identification: When there are sequences of images, or long exposure images, use them to identify moving bodies. If these do not correspond to any known body, tag them into a moving body database, which can be used for orbit determination, if there are enough observations, or in precovery searches for objects detected in any other images.

In this poster, we present the preliminary results of the identification and cross-match of known body observations, for a small sample of bodies and a few archives, to demonstrate the functionality we are building.

Identifying known body observations

We integrate into the past the orbits of the known bodies, to obtain their positions as a function of time. For this proof of concept, to reduce the computation time, we used only the first 25 000 objects. In the future, we will expand the database to include all the known bodies (currently, ~ 640 thousand).

The positions were calculated with the SPICE toolkit (Acton 1996), from the Navigation and Ancillary Information Facility (NAIF), from kernel files generated by the JPL HORIZONS system (Giorgini et al. 1996). Since accurate positions have to be calculated for the time and location of each archive exposure, and archives can have several million exposures, the positions are calculated in two steps:

1) A coarse grid is calculated with the position of the bodies into the past, at 1-day time intervals. We map both this grid and the observation footprints in a Hierarchical Triangular Mesh (HTM; Budavári et al. 2010) for the sky location, which allows efficient matching for large numbers of footprints and trajectories.

2) The candidates found in (1) have their positions precisely calculated for the time and location of each frame in the archive, to determine if they actually fall into the field of view.

The resulting matches are stored in a database, indexing all observations of each body. The bodies database contains multiple types of identifiers for the objects, and orbital and physical properties, which can be readily used in queries.

References

Acton Jr, C. H. (1996). Ancillary data services of NASA's Navigation and Ancillary Information Facility. Planetary and Space Science, 44(1), 65-70. http://dx.doi.org/10.1016/0032-0633(95)00107-7

Budavári, T.; Szalay, A.; Fekete, G. (2010). Searchable Sky Coverage of Astronomical Observations: Footprints and Exposures. PASP, 122 (897),1375-1388. http://dx.doi.org/10.1086/657302

Bus, S. J., & Binzel, R. P. (2002). Phase II of the small main-belt asteroid spectroscopic survey: A feature-based taxonomy. Icarus, 158(1), 146-177. http://dx.doi.org/10.1006/icar.2002.6856

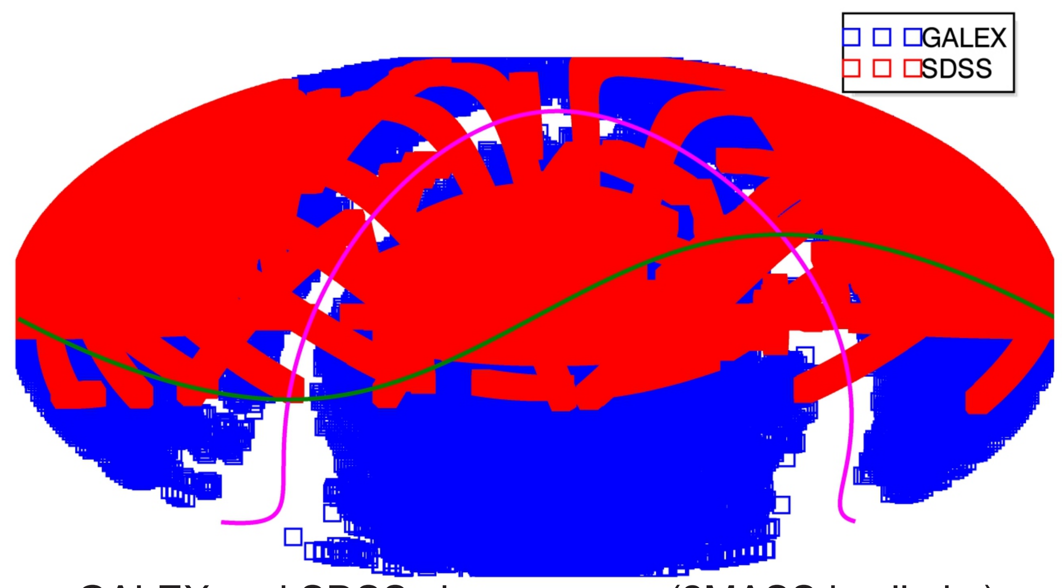
Giorgini, J. D., Yeomans, D. K., Chamberlin, A. B., Chodas, P. W., Jacobson, R. A., Keesey, M. S., ... & Wimberly, R. N. (1996, September). JPL's on-line solar system data service. In Bulletin of the American Astronomical Society (Vol. 28, p. 1158). http://adsabs.harvard.edu/abs/1996DPS....28.2504G

Tholen, D. J. (1984). Asteroid taxonomy from cluster analysis of photometry. Ph. D. Thesis, University of Arizona. http://adsabs.harvard.edu/abs/1985DiAbI..45.2201T

Observations found

For this initial test, we searched for the first 25 000 asteroids, in 3 wide-coverage surveys:

1) GALEX: 2 filters (FUV: 0.15μm, NUV: 0.23μm), 1.2° field of view, 4"/pix <~21 mag
2) SDSS DR10: 5 filters (u: 0.36μm, g: 0.47μm, r: 0.62μm, i: 0.75μm, z: 0.89μm), 9' field of view, 0.4"/pix, < ~22 mag
3) 2MASS: 3 filters (J: 1.25μm, H: 1.65μm, K: 2.17μm), 9' field of view, 1"/pix, < ~ 16 mag, allsky



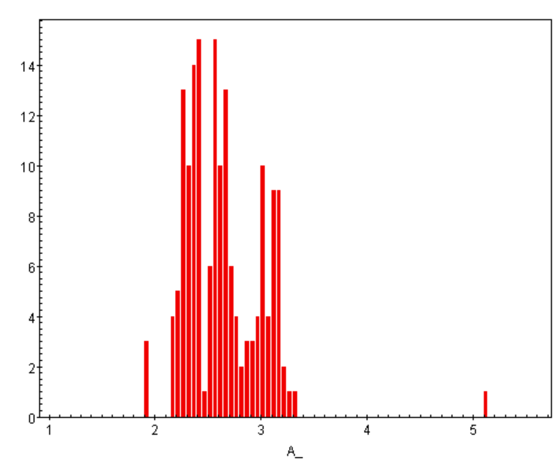
GALEX and SDSS sky coverage (2MASS is all-sky)
Green line: Ecliptic; Red line: Galactic Plane

When cross-matching the bodies between different data sources, the number of observations we find are

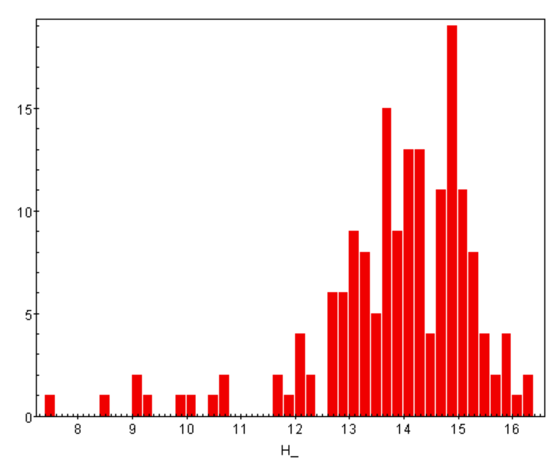
SDSS DR10: 1006 observations, 451unique objects.
2MASS: 720 observations, 493 unique objects.
GALEX: ~10^4 observations (estimate; match computation not yet completed)
SDSSx2MASS: 169 objects observed by both.

Extrapolating for the ~640 thousand known objects, this indicates ~4300 objects observed in 8 colors, from 0.36μm to 2.17μm.

Distribution of objects found in SDSS and 2MASS

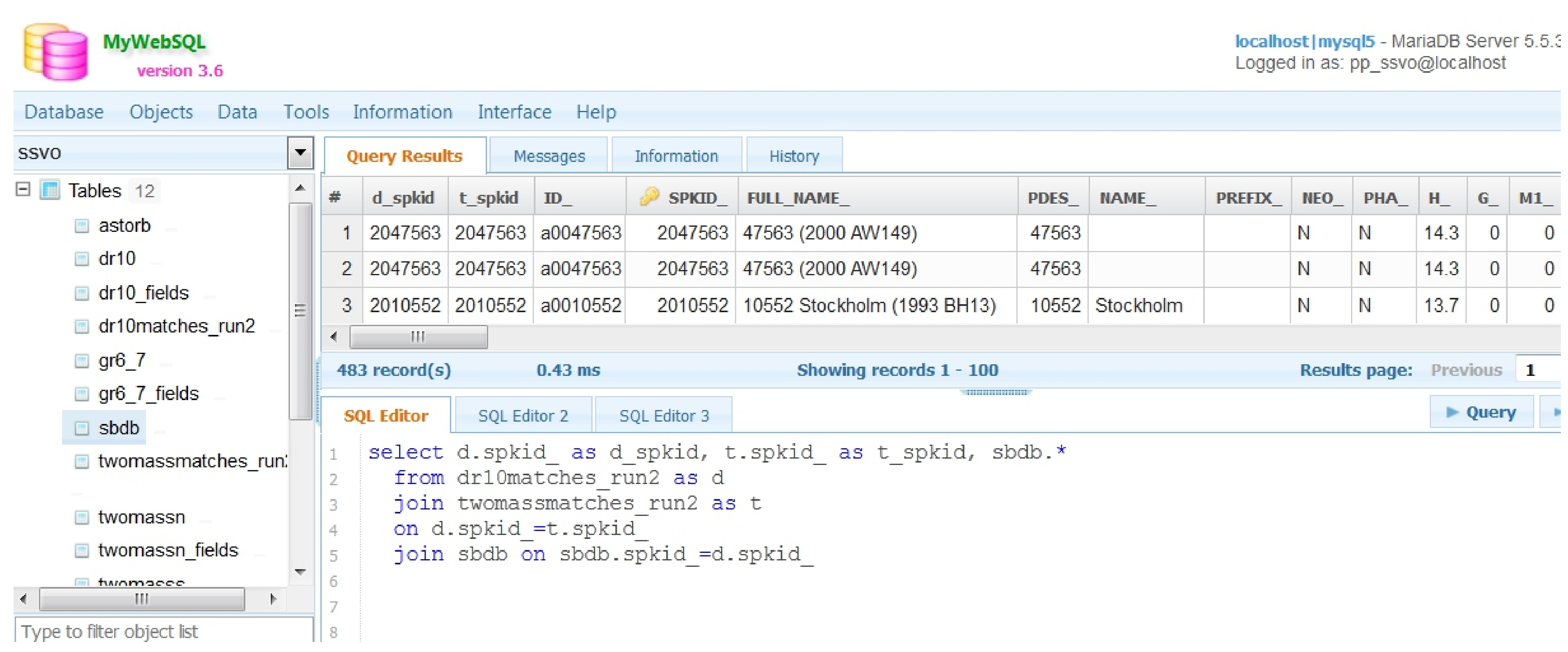


Semi-major axis (excluding one object with a=42 au)



Absolute magnitude

Prototype online service



A prototype interface for online access is being built, and is available at http://www.ppenteado.net/pp_ssvo

It already contains tables of object data, which can be used as query parameters, and is being populated with the matches from the catalogs we are processing.

Conclusions / Future work

We are creating the first database of archived observations of Solar System bodies, from matching the orbits of known objects with footprints of archived observations and catalog magnitudes.

These databases enable systematic analyses of large numbers of objects, in a wide wavelength range. In the LSST-era, they will encompass observations of millions of objects.

The final product will be accessible through a CasJobs interface and through a VO API, which will enable complex non-interactive queries.

We will create a name resolver function, which will eliminate the need for object identifiers in exact formats.

We will analyze the object distribution across multiple colors and its relation with orbital and physical parameters.

Support for this work comes from the Northern Arizona University Office of the Vice President for Research.

Poster available at http://www.ppenteado.net/ast/dps2014_pfp.pdf