WHAT IS IN SIMBAD ?

Oberto Anaïs[1], Loup Cécile[1], Allen Mark [1], Lesteven Soizick[1] and CDS Team[1] : Boch T., Bot C., Bonnarel F., Brouty M., Brunet C., Buga M., Cambrésy L., Delacour T., DerriÃÍre S., Eisele A., Fernique P., Genova F., Guéhenneux S., Landais G., Louys M., Nebot A., Neuville M., Ocvirk P., Perret E., Pineau F.-X., Schaaff A., Siebert A., Son E., Vannier P., Vollmer B., Vonflie P., Woelfel F.

[1]*Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France*

Abstract. SIMBAD is a dynamic database of astronomical objects. It provides the bibliography, as well as basic information such as the nature of the object, its coordinates, magnitudes, proper motions and parallax, velocity/redshift, angular size, spectral or morphological type, and the multitude of names (identifiers) given in the literature. The information in SIMBAD is a compilation built from what is published in the literature with expert cross-identification performed at the CDS based on the compatibility of several parameters, in the limit of reasonably good astrometry. We show the data integration processes in SIMBAD.

1. Introduction

The SIMBAD astronomical database (Wenger et al. (2000)) is developed and maintained by the Strasbourg astronomical Data Center (CDS). It is the world reference database for identifications and bibliography of astronomical objects outside the Solar System. As its name 'Set of Identifications, Measurements and Bibliography of Astronomical Data' suggests, the SIMBAD database contains basic information such as the nature of object, the multitude of names (identifiers), magnitudes, velocity/redshift, size, spectral or morphological type, bibliography, and many measurements for astronomical objects given in the literature. For example, see the figure 1. At first, (in 1971 as the Catalog of Stellar Identifications), SIMBAD was focused on only stars in the Milky Way, but since about 1982 it has been extended to all objects outside the Solar System. One of the key tasks for SIMBAD is linking the astronomical objects to the multitude of articles in which they appear and adding various supplemental data when available.

Distinct work processes are involved for populating the SIMBAD database depending on the original data, and requiring the expertise of both documentalists and astronomers. Authors' data come to CDS through different manners and authors should follow some rules and principles for an efficient process entry (Cambrésy et al. (2011)).

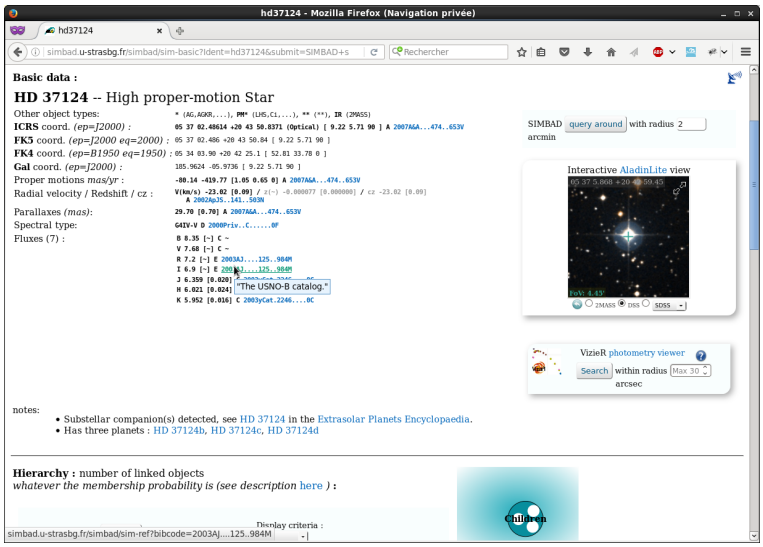

Figure 1. SIMBAD - Basic data

2. Data directly in the article

Some papers are written by focusing on one astronomical object with many details on all its aspects and can give information directly in the text (at least names). For these kind of articles, documentalists explore the whole text and attached data to look for all information. CDS has developed 'DJIN' (Detection in Journal of Identifiers and Names) (Lesteven et al. (2010)), a software that suggests to the documentalist a list of SIMBAD object automatically recognized in each article. In this way, since 2008, the references are linked to the SIMBAD astronomical objects rapidly after publication. The new objects and those that are not yet in SIMBAD are added manually in the database with the respective bibliographic codes and data. The documentalist team uses its expertise to validate the possible matches proposed by DJIN and corrects all errors or omissions by the software. Astronomical data (with enough accuracy) are potentially also added by the team to the objects identified in the title, abstract, figures, text, and small tables. The treatment of large tables is assigned to the second documentalists team.

3. Data from large tables

If the objects are cited in large tables (about more than one hundred), data will first be added into the VizieR Catalogue Service (Ochsenbein et al. (2000)) which regroups catalogues containing tables in a standardized form (original files either published in the articles or sent by the authors to CDS) and homogeneous metadata. According to many scientific criteria, astronomers and documentalists of CDS decide of the priority level for adding tables. This selection is required due to the growing number of publications but also to the team's limited number of documentalists (Loup (2017)). Documentalists use an internal program called 'COSIM' (Comparison of Objects for SIMBAD) to include large tables in SIMBAD. It checks scores of compatibility of input objects with

all SIMBAD objects around. It calculates probabilty of likelihood for comparable data, such as coordinates, velocity/redshift, magnitudes, and other data available for both the input object and the SIMBAD candidate.

4. Huge catalogues

For very large catalogues of objects, such as catalogues of infrared sources (2MASS, AKARI), stars (Tycho, Washington Double Star), planets or supernova, a selection is obviously necessary, and astronomers decide if it should be added in SIMBAD, partially or totally. The cross-identification of these large surveys is done either by the documentalists in close relation with the astronomers or, if specific knowledge is needed, by one of the astronomers at CDS. This work is done very carefully, and with the most accuracy possible to avoid an error of cross-identification that could have bad impact on the database. That is why, generally, such catalogues are not entered totally in SIMBAD.

5. Maintenance of data

As a general rule, all objects cited in a paper are stored in SIMBAD with a reference to this paper and relevant data if any. New identifiers are added if it enables the retrieval of the object in the paper or other papers. As mentioned earlier, SIMBAD does not contain 'all objects', but rather those that have been studied in the literature. SIMBAD has a system for users to make annotations, for example to point out a wrong cross-identification or a mis-print, and the corrections are made after a verification.

The SIMBAD team maintains the evolution of fundamental data on each object (basic data such as the nature of the object, its coordinates, magnitudes, spectral or morphological type, ...) as shown in the screenshot in the figure 1. Astronomical objects may have many identifiers. The general syntax of an identifier is the abbreviated catalogue name, or acronym, followed by one or several fields, which can be numbers, coordinates, strings, etc.

SIMBAD manages a dictionary for identifications in collaboration with the International Astronomical Union (IAU), who defines unambiguous astronomical nomenclature. Some simple names could be ambiguous because it could yield to several different possible objects (for "Perseus" example : a molecular cloud, a star association, a superbubble, a spiral arm, ...).

The SIMBAD team also maintains the bibliography (links for each new paper about an astronomical object), so that users may get the full set of papers and citations for a given object.

5.1. Added value

Cross-identification is the core of the work of the SIMBAD team, this means that it is necessary to organize as homogenously as possible a huge quantity of collected data coming from different tables and over the whole publication record. It is not the task of the CDS to claim that data are not coherent with other papers or one object has been misclassified. This is the work of the researchers and referees, SIMBAD is a record of what has been published. Thanks to this diffult work, a user can start from any of the known names of an object (could be among 125), to access the relevant information in SIMBAD.

The list of scientific papers focused on one astronomical object is provided and SIMBAD offers to the users few filters to find only papers with table data or to sort papers on a CDS score, based on many criteria such as location and occurency of the object citation, year of publication, etc ... (Oberto et al. (2015)). An example is shown on the screenshot in figure 2.

There are many links between the SIMBAD and VizieR databases (The CDS VizieR database provides a complementary system which aims to be as complete as possible for published catalogues). The output of a SIMBAD query includes links to the full data in VizieR (this is done as far as it is possible if there is a direct link from one SIMBAD object to one row in the table). In the opposite way, a new column is added virtually to the table in VizieR containing the link to the SIMBAD object, also, as far as it is possible.

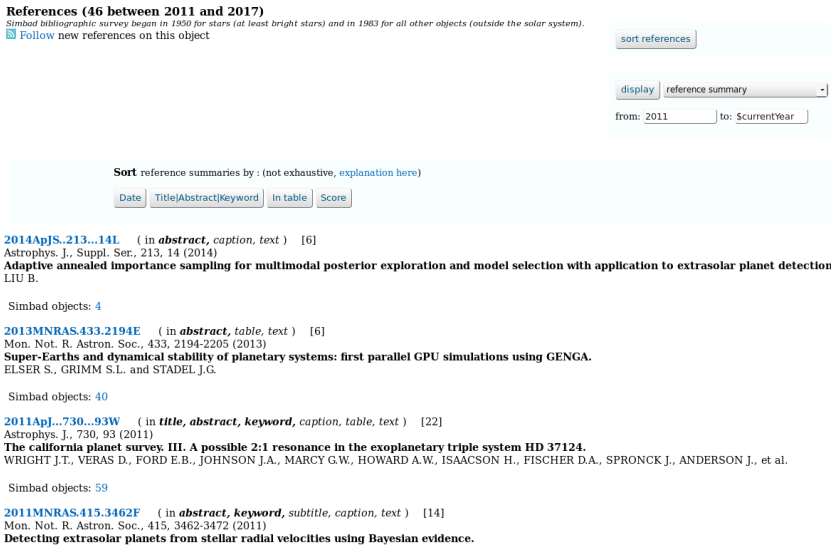

Figure 2. Ordered bibliography on the object in SIMBAD

References

Cambrésy, L., Genova, F., Wenger, M., & Loup, e. a., C. 2011, in EAS Publications Series, edited by C. Sterken, vol. 49 of EAS Publications Series, 135

Lesteven, S., Bonnin, C., Derriére, S., & Dubois, P. e. a. 2010, in Library and Information Services in Astronomy VI: 21st Century Astronomy Librarianship, From New Ideas to Action, edited by E. Isaksson, J. Lagerstrom, A. Holl, & N. Bawdekar, vol. 433 of Astronomical Society of the Pacific Conference Series, 317

Loup, C. 2017, in ADASS XXVI, edited by TBD (San Francisco: ASP), vol. TBD of ASP Conf. Ser., TBD

Oberto, A., Lesteven, S., Derriere, S., & Bonnin, C. e. a. 2015, in Open Science at the Frontiers of Librarianship, edited by A. Holl, S. Lesteven, D. Dietrich, & A. Gasperini, vol. 492 of Astronomical Society of the Pacific Conference Series, 284

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23. astro-ph/0002122

Wenger, M., Ochsenbein, F., Egret, D., & Dubois, P. e. a. 2000, A&AS, 143, 9. astro-ph/0002110