Library and Information Services in Astronomy V
ASP Conference Series, Vol. 377, 2007
S. Ricketts, C. Birdie and E. Isaksson (eds.)

The Content of the CDS Services

F. Woelfel, F. Genova, F. Bonnarel, M. Brouty, C. Bruneau, C. Brunet, L. Cambrésy, P. Dubois, A. Eisele, P. Fernique, S. Laloë, S. Lesteven, C. Loup, A. Oberto, F. Ochsenbein, B. Vollmer, P. Vannier, P. Vonflie, M.-J. Wagner, M. Wenger, and S. Yi

CDS, Observatoire Astronomique de Strasbourg, UMR CNRS/ULP 7550, 11 rue de l'Université, 67000 Strasbourg, France

S. Borde

GEPI, Observatoire de Paris, 61 avenue de l'Observatoire, 75014 Paris, France

G. Chassagnard

Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France

E. Davoust

LAT, Observatoire Midi-Pyrénées, 14 rue Edouard Belin, 31400 Toulouse, France

G. Jasniewicz

Groupe de Recherche en Astronomie et Astrophysique du Languedoc, CC72, Université de Montpellier II, 34095 Montpellier Cedex 05, France

Abstract. The CDS value-added bibliographic services, SIMBAD and VizieR, are updated daily. Most of the information comes from the astronomical literature and the update mechanism is different for different types of information. The semi-automated SIMBAD data flow is described and the synergy between astronomers, specialized librarians and computer engineers is discussed.

1. Introduction

The objective of the Centre de Données astronomiques de Strasbourg (CDS – Genova et al. 2000) is "to collect, homogenize, distribute and preserve astronomical information for the usage of the whole community". Its aim is to build value-added databases and tools to help astronomers in their daily science work. This means a huge data curation work – *selection*, *validation*, *homogenization* of data – which is performed by an integrated team of specialized librarians, astronomers and computer engineers.

Section 2 describes the content of the CDS databases, and Section 3 illustrates the building of this content by describing in more detail the semi-automated entry of data from published tables in SIMBAD.

2. The CDS Content

The two main CDS bibliographic services, SIMBAD (Wenger et al. 2000) and VizieR (Ochsenbein, Bauer & Marcout 2000), are updated daily. SIMBAD provides basic data, cross-identifications, bibliography, and measurements for astronomical objects outside the solar system. In June 2006, it contained 3,700,000 astronomical objects, 170,000 bibliographic references, and 8,000 acronyms. In parallel to SIMBAD, the CDS collects and distributes catalogs and tables published in journals, in collaboration with the academic journals and the other data centers. In June 2006, the VizieR catalog browser contained 5,400 catalogs, published tables and observation logs. The *Dictionary of Nomenclature of Celestial Objects outside the solar system* (Lortet, Borde & Ochsenbein 1994) is another product of the scanning of the literature performed for SIMBAD and of the gathering of catalogs for VizieR. It keeps track of the object acronyms detected in the literature and in catalogs, with additional information about the original object list, the list object type, the instrument used to perform the observations when applicable, etc. In addition to the bibliographic services, more than 5.5 Tb of reference images are also distributed by CDS through the Aladin image server. The Aladin visualizer also gives access to many distributed image and catalog databases, in particular to large ground- and space-based observatory archives (Bonnarel et al. 2000).

The content of the CDS databases is growing fast, with about 500 catalogs and 15,000 references added each year in VizieR and SIMBAD respectively. The number of new objects added in SIMBAD during one year is variable, of the order of 200,000 new objects and 360,000 new identifiers. It can be larger when a large catalog is included during a particular year. More than 700 new acronyms are detected and documented each year in the *Dictionary of Nomenclature*.

3. Data Integration in SIMBAD

Information about objects is gathered from the systematic reading of 90 academic journals. When an object name is found in the text of an article, the corresponding SIMBAD name is identified – one important difficulty is the huge variability of the astronomical object nomenclature, and the frequent carelessness of authors in this domain. If the object does not exist in SIMBAD, it is created, which may also require the creation of a new acronym. Then the paper reference and data found in the paper, if any, are attached to the object.

For objects found in "long" tables (above 50–100 lines), advantage is taken of the availability of tables in electronic form. Many tables are available in the CDS catalog service and in VizieR with a standardized description, by agreement with the journals (the pioneering work of *A&A* and CDS in these matters is described in Ochsenbein et al. 2003), and most of the others can be downloaded from the journal electronic version. When it is decided to enter in SIMBAD a table which is not yet available at CDS, a standardized description, the ReadMe file, is built and the table is first entered in VizieR. The ReadMe file contains the list of all the columns of a table, and for each column the physical format (e.g., that bytes 31 to 35 contain an integer coded on 5 characters), the unit, the column label, and an explanation (e.g. Right ascension (J2000.0), Major

B band optical diameter, ...). A semi-automated method using the ReadMe file is then performed to enter the data in SIMBAD. The "raccord" tool is used to compare the data from the table with the data contained in SIMBAD, with the identification of a proper object name, and comparison of information found in the table such as coordinates, magnitudes, object types, etc..., with the information existing in SIMBAD. For a given object of the table, SIMBAD is searched by object name or around the object position to look for possible cross-identification with existing objects. The program lists the differences. When no cross-identification is found a new object is created. The program assigns a score to each object found in SIMBAD and when the score is low or when several likely cross-identifications are possible the choice is validated by one or several experts. The program then generates updating commands.

4. Conclusions

The updating procedures of SIMBAD have evolved significantly since the advent of electronic journals, and now include semi-automated procedures (another example is the use of tables of contents provided in electronic form by the journals to create new bibliographic references in the database). The implementation of these semi-automated methods has helped significantly to cope with the ever increasing data flow, with a team which is far from increasing at the same rate. However, a very important lesson learnt from more than 10 years of daily updates using these methods is that even with carefully prepared automated procedures, human intervention is required when one cares for data quality, e.g., to decide on cross-identification criteria, and to identify the data which should be improved in SIMBAD. Many objects require a case-by-case decision, and the scientific expertise of astronomers is required on the most complex cases.

Another important lesson learnt in the long term is that the construction of the content of added-value services is a very interesting, high impact profile for librarians: many other added-value services, in addition to SIMBAD, VizieR or NED, are emerging in the context of the astronomical Virtual Observatory development, and the importance of the reference services is growing with the increase in the number and heterogeneity of available on-line resources. This difficult work requires a high expertise, continuously updated to take into account the scientific evolution of astronomy and technical evolution, hence the importance of teams integrating librarians, astronomers and software engineers.

References

Bonnarel, F., et al. 2000, A&AS, 143, 33
Genova, F., et al. 2000, A&AS, 143, 1
Lortet, M.-C., Borde, S., & Ochsenbein, F. 1994, A&AS, 107, 193
Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
Ochsenbein, F., Bertout, C., Lequeux, J., & Genova, F. 2003, in Proc. Library and Information Services in Astronomy IV, ed. B. G. Corbin, E. P. Bryson & M. Wolf (Washington, DC: US Naval Observatory), 257
Wenger, M., et al. 2000, A&AS, 143, 9