FROM YOUR PAPER TO VIZIER AND SIMBAD

L. Cambrésy[1], F. Genova[1], M. Wenger[1], C. Loup[1], F. Ochsenbein[1] and T. Boch[1]

Abstract. Scientific results strongly rely on previous studies, experiments, and observations. A huge quantity of data is produced by astronomers and needs to be available to the whole community. To support astronomers in their daily research work, the *Centre de Données astronomiques de Strasbourg* (Strasbourg astronomical data center – CDS) collects, verifies, homogenizes, and organizes information – in particular, those published in academic journals – in the most appropriate and comprehensible way. This CDS goal can reach the highest level of quality only through a close collaboration with journals, authors, and referees. This paper presents the succession of processes leading published data to the CDS databases, focusing on the strategy that maintains a high level of quality. Authors and referees are strongly encouraged to actively contribute to this endeavor. A few examples of how CDS databases are used through CDS services are also presented.

1 Introduction

The *Centre de Données astronomiques de Strasbourg* (Strasbourg astronomical data center – CDS) was created in 1972 by the French agency in charge of ground-based astronomy, which is now called INSU (*Institut National des Sciences de l'Univers*), as a joint venture with Strasbourg Louis Pasteur University. Louis Pasteur University merged on January 1, 2009, with the two other local universities to form the *Université de Strasbourg*. The CDS was created as the *Centre de Données Stellaires* (Stellar Data Center), but in 1983 it was decided to extend its scope and to consider all astronomical objects beyond the Solar System.

The high-level goal of CDS is to *"collect, homogenize, preserve and distribute astronomical information for the usage of the whole astronomy community"*. Its initial charter is still relevant:

- collect "useful" data on astronomical objects in electronic form

---

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

- improve them by critical evaluation and combination

- distribute the results to the international community

- conduct research using these data.

From the beginning, CDS has gathered two main types of information:

- astronomical catalogs, with a catalog service shared with several other data centers around the world, which was extended to include tables published in journals. These catalogs were distributed on demand on magnetic tapes and then through an ftp service. The VizieR service, available since 1996, has added a browsing capability (Ochsenbein, Bauer & Marcout 2000).

- object cross-identification and bibliography were first made available through two different data bases, the *Catalogue of Stellar Identifications* (CSI – Jung & Bischoff 1971, Jung & Ochsenbein 1971), which aimed at cross-identifying a few fundamental star catalogs, and the *Bibliographic Star Index* (BSI – Cayrel, Jung & Valbousquet 1974). SIMBAD (Set of Identifications, Measurements, and Bibliography for Astronomical Data, Wenger *et al.* 2000) was created in 1981 from the merging of the BSI and of the CSI.

The *Dictionary of Nomenclature of Celestial Bodies outside the Solar System* (Lortet, Borde & Ochsenbein 1994) and the Aladin interactive star atlas (Bonnarel *et al.* 2000) are two other important CDS products which will be referred to in the following.

Involvement of active researchers in the CDS staff has always been recognized as an important factor for success, since it is the only way to ensure the required scientific expertise on the service content and to make sure that the services are suited to the constantly evolving needs of the scientific community. To maintain a high level of scientific activities, the CDS was created inside and as a part of an Observatory, which provides an active research environment. The CDS team integrates astronomers, computer engineers who manage the databases and the user interfaces, and highly specialized bibliographers, who are called *documentalistes* in French – this shows that the CDS librarians are specialized in extracting 'information' from and building metadata for documents. Information and metadata of interest for CDS are not only the usual bibliographic information managed in libraries (authors, publication references, etc.), which constitute the relatively easy part of the work. The point here, as will be shown in the following, is to check the description of the catalogs and tables stored in VizieR or to extract important data about astronomical objects for SIMBAD.

In the following, CDS services and procedures are described with the aim of showing how they are built and how the scientific community can help the data center in its constant quest for quality, along with a few hints to how they can be used. Only a selected set of information about the CDS services is highlighted in this paper. For more details, the reader is referred to Genova *et al.* (2000) and to the other papers of the series published in the April 2000 Special Issue of *Astronomy & Astrophysics Supplement Series* (Volume 143, No. 1, several of which have been cited above): "The CDS and NASA ADS resources: New tools for astronomical research", and for a global more recent view of CDS history and evolution to Genova (2007).

2   The VizieR database

2.1   Content

The link between VizieR and SIMBAD is not always clear to new users since the two databases are indeed very different. SIMBAD results from the cross-identification of major catalogs, tables and individual objects published in the literature. VizieR comes from a different approach: the astronomical catalogs are kept in their original form, in what we call the *FTP archive* (a set of files and directories openly accessible from `cdsarc.u-strasbg.fr`). In this repository, one catalog occupies one directory and is always accompanied by a *standard description*, in the form of a file named `ReadMe` described in the next section (2.2). These homogeneous descriptions are the key to the VizieR system, a relational database gathering the tabular data (catalogs) existing in the FTP archive; in addition, VizieR gives access to very large surveys that are not part of the FTP archive because of their huge size.

VizieR currently provides access to the most complete library of published astronomical catalogs and data tables available online (more than 12000 catalogs in May 2014), organized in a self-documented database. A brief summary of VizieR is given below (Sect. 2.3); the contents of VizieR are illustrated by Fig. 1, which represents the spatial distribution of $\sim 10^9$ sources available in VizieR after removal of the all-sky surveys.
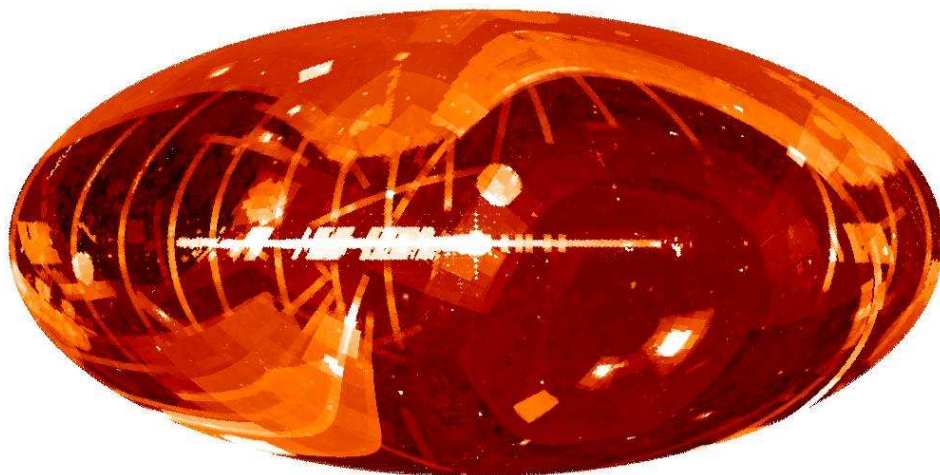

Fig. 1. VizieR surface density map in galactic coordinates, representing $\sim 10^9$ sources after removal of all-sky surveys. The horizontal bar in the central region is due to the GLIMPSE catalog (map built on May 2014 with more than 12000 catalogs).

2.2   The ReadMe file

Information that describes the data – the metadata – is traditionally presented in the introduction of the printed catalog, or explained in one or several published papers presenting

and/or analyzing the cataloged data. Metadata play a fundamental role. First they provide scientists with information about the context of the data so that they can make their judgment about the suitability for their project. Also a minimal knowledge of the metadata is required by data processing systems in order to merge or compare data from different origins – for instance, the comparison of data expressed in different units requires a unit-to-unit conversion, which can be performed automatically only if the units are specified unambiguously.

This need for a description that is readable both by a computer and by a scientist led to a standardized way of documenting astronomical catalogs and tables, promoted by The CDS and adopted by other data centers in the form of a dedicated ReadMe file associated to each catalog (Ochsenbein 1994). The ReadMe description file (see example Fig. 2) starts with a header specifying the basic references of the catalog – title, authors, references – and contains a few key sections introduced by standard titles like *Description* or *Byte-by-byte Description of file*. Such a file is relatively easy to produce by someone who knows the catalog contents – or better by the authors themselves.

The most important part of the ReadMe file is the *Byte-by-byte Description*, which details the table structures in terms of formats, units, column naming, or labels (some conventions are used for label assignments for consistency), existence of data (possibility of unspecified or null values), and brief explanations.

This standardized way of presenting the metadata proved to be extremely useful, especially for data-checking and format conversion. Many errors were detected in old catalogs simply because a general checking mechanism became available.

Astronomy & Astrophysics authors are expected to supply the documentation of their data in this simple form. Template files, as well as a few tips on how to create the ReadMe file, are accessible on the VizieR webpage[1]. The ReadMe files and the data files are then checked by a specialist who contacts the authors if errors are detected or when changes are necessary to increase the clarity or homogeneity of the description.

The data accessible in the *FTP Archive* are not restricted to tabular data. If historically astronomical catalogs have been the first type of data distributed on electronic media by the CDS in the early seventies, and the astronomical data available from this repository now include also spectra, images, data cubes, etc. However, data on the public repositories are in principle restricted to *scientific data*, i.e., data that can be re-used in the data processing tools used in astronomy. This restriction means that FITS images, for instance, can be found (FITS includes essential details about origins, sky location, etc.), but JPEG or PDF images are generally excluded.

2.3 VizieR

Once a catalog, as described according to the above conventions, is part of the *FTP archive*, it becomes visible to any interested scientist. Search engines provided by the CDS help to locate these catalogs from their textual contents with *google*-like queries. Textual searches, however, cannot answer to questions like *"find in all tables of this*

---

[1] http://vizier.u-strasbg.fr/doc/submit.htx

Fig. 2. Example of ReadMe file for catalog I/221

```
J/A+A/466/137       JHKs photometry in Westerlund 2          (Ascenso+, 2007)
================================================================================
Near-infrared imaging of Galactic massive clusters: Westerlund 2.
    Ascenso J., Alves J., Beletsky Y., Lago M.T.V.T.
    <Astron. Astrophys. 466, 137 (2007)>
    =2007A&A...466..137A
================================================================================
ADC_Keywords: H II regions ; Clusters, open ; Photometry, infrared
Keywords: open clusters and associations: individual: Westerlund 2 -
          stars: pre-main sequence - infrared: stars

Description:
    Photometry for the stars detected with SOFI, NTT in J, H and Ks in the
    field containing the massive cluster Westerlund 2 (also know as RCW 49)
    associated with the HII region Gum29. For each star detected in all
    three bands the table contains the ID, equatorial coordinates, J, H
    and Ks magnitude and photometric errors. For the stars detected only
    in Ks or in H and Ks the magnitude and photometric error fields in the
    absent bands are filled with the values 99.99 and 9.999 respectively.

File Summary:
--------------------------------------------------------------------------------
 FileName   Lrecl  Records   Explanations
--------------------------------------------------------------------------------
ReadMe        80        .    This file
w2phot.dat    65      5727   Positions and magnitudes of stars in Westerlund 2
--------------------------------------------------------------------------------

Byte-by-byte Description of file: w2phot.dat
--------------------------------------------------------------------------------
   Bytes Format Units   Label      Explanations
--------------------------------------------------------------------------------
   1-  4  I4    ---     Seq        Identification number
   6-  7  I2    h       RAh        Right ascension (J2000.0)
   9- 10  I2    min     RAm        Right ascension (J2000.0)
  12- 16  F5.2  s       RAs        Right ascension (J2000.0)
      18  A1    ---     DE-        Declination sign (J2000.0)
  19- 20  I2    deg     DEd        Declination (J2000.0)
  22- 23  I2    arcmin  DEm        Declination (J2000.0)
  25- 29  F5.2  arcsec  DEs        Declination (J2000.0)
  31- 35  F5.2  mag     Jmag       ?=99.99 Magnitude in the J band (1)
  37- 41  F5.3  mag     e_Jmag     ?=9.999 Photometric error in the J band (1)
  43- 47  F5.2  mag     Hmag       ?=99.99 Magnitude in the H band (1)
  49- 53  F5.3  mag     e_Hmag     ?=9.999 Photometric error in the h band (1)
  55- 59  F5.2  mag     Ksmag      ?=99.99 Magnitude in the Ks band (1)
  61- 65  F5.3  mag     e_Ksmag    ?=9.999 Photometric error in the Ks band (1)
--------------------------------------------------------------------------------
Note (1): The values 99.99 for the magnitude and 9.999 for the errors
    correspond to sources for which we do not have photometry in the
    corresponding band.
--------------------------------------------------------------------------------

Acknowledgments:
    Joana Ascenso, joanasba(at)astro.up.pt
================================================================================
(End)    Joana Ascenso [CAUP, Portugal], Patricia Vannier [CDS]    25-Jan-2007
```

archive which contains entries related to point sources in a 2-arcmin region around my target". Such queries have to deal with numeric values and to know what a position is and which coordinate system is used in each table, etc.

These queries are possible with VizieR, the database created from the contents of the *FTP archive*. VizieR is basically a relational database that keeps both the original data (each catalog or table of the FTP archive becomes a table of the database) and the associated metadata extracted from the `ReadMe` files (stored in a set of dedicated tables). These metadata play the essential role of giving a meaning to all the numbers and symbols included in the catalogs and tables. VizieR structure permits queries to be constrained by any parameter, i.e. column, of the tables. Some further metadata homogenization is performed to a better interpret the parameters contained in the thousands of tables of VizieR, essentially by assigning a *Unified Content Descriptor* (UCD) to each column of each table. The UCD, originally defined at the CDS (Ortiz *et al.* 1999) is shorthand to qualify the meaning of a parameter in a thesaurus of the existing parameters used in astronomy (e.g. positions, magnitudes, oscillator strength), which is now developed in the framework of the *Virtual Observatory*.

Besides the homogenized metadata, VizieR also introduces *links* to associated data. As an example, when a paper contains images or spectra associated to observed or modeled objects, VizieR may provide a link to the images or spectra related to the corresponding object entry in the table. Such links may also point to other VizieR tables, to other databases (SIMBAD, NED, HyperLeda), or to repositories (SDSS, HST archive, etc.).

VizieR also gives access to large surveys, such as 2MASS or USNO-B1. The access to these huge surveys, which is not possible in the *FTP Archive* because of their volume, is optimized to give fast answers to positional queries.

Currently (May 2014), VizieR contains more than 12 300 catalogs made of 26 600 different tables for a cumulative number of 4×10^5 columns, i.e., an average of 2.2 tables per catalog and 15 columns per table. VizieR responds to an average of 3×10^5 queries/day.

2.4   Usage of VizieR

Access to VizieR by Web is available by script, or by programs. The data published in VizieR also become accessible by the Virtual Observatory, which means that the tabular material, once properly documented and ingested into VizieR, becomes directly accessible by any tool aware of these standards. As a practical example, tables describing astronomical sources become visible in Aladin (Sect. 5). The results available this way are also easily retrieved for the creation of large samples required to derive new astronomical results.

An accurate description of the data related to astronomical objects is also an important step before integrating them into SIMBAD (see Sect. 3.3); therefore, do not miss the documentation of your data!

3   How your data are integrated in SIMBAD

3.1   Acronym creation in SIMBAD

Astronomical object designation, either called name or identifier, is a key point for SIMBAD since it refer unambiguously to an object. Some rules need to be defined for clarity. The general syntax of an identifier is the abbreviated catalog name, or acronym, followed

by one or several fields, which can be numbers, coordinates, strings, etc. An identifier is case and space insensitive.

When no general catalog name can be used, the acronym is directly based on the author's initials between square brackets with the year of publication of the paper. For example, HD 97300 is also named [FL2004] 44 after Feigelson & Lawson (2004).

Object names, such as Vega and Altair, but also Barnard's star, Crab Nebula, Sgr A, HDFN, or HDFS, are stored in the database in a specific catalog called "`NAME`", while star names in constellations, such as Lyra, are stored in the catalog "`*`" (e.g. `* gam Lyr`), and variable stars (such as RR Lyrae) in the catalog "var" (also called "V*"). To some extent, identifiers can be shortened (i.e. NAME can be omitted), or can be typed according to some accepted usage. A set of rules, based on regular expressions and substitutions, try to normalize the typed identifier before querying the database.

Greek letters should be abbreviated as three letters: alf, bet, for α and β, but also `mu.`, `nu.`, and `pi.` (with a dot), for μ, ν, and π. This is required to avoid ambiguities in some identifiers. Constellation names should be abbreviated with the usual three letters: `alf Boo`, `del Sct`, `FG Sge`, `NOVA Her 1991`. The full list is available online.

Identifiers of a multiple system may generate a list of all objects belonging to the system. For instance, `ADS 5423` calls for the four components, `A` to `D`, of the stellar system around Sirius. This is only true for some specific identifiers.

Clusters that have no `NGC` or `IC` number are named under the generic appellation `Cl` followed by the cluster name and number: e.g., `Cl Blanco 1` is the 1st stellar cluster named by Blanco. Stars in clusters follow complex *historical* rules. They may belong to a *main* designation list or to subsequent lists. `NGC 5272 692` is star 692 in the list by Von Zeipel, considered as the main list. Subsequent lists have designations starting with `Cl*` such as `Cl* NGC 5272 AC 968` (list by Aurière & Cordoni), `Cl* Melotte 25 VA 13` (13th star in the list by van Altena for Melotte 25 — the Hyades cluster).

Astronomical objects have several correct designations and a default has to be chosen for practical reason. It is selected following rules based on the object type, or it is set manually, especially for the well known objects for which most users expect the same name.

3.2   Integration of objects cited in the text of papers into SIMBAD

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. This has long been done by *documentalistes* reading journals and checking them for object citations. Additional paths for object entry have been made possible by the development of electronic journals. In particular, the CDS and *Astronomy & Astrophysics* collaborate to provide links between object names in papers and SIMBAD, by asking authors to tag *important* object names. These tags are used to identify objects to be referred to in SIMBAD (Sect. 3.2.1). Most of these objects, however, are entered through a systematic scan of the papers or by a semi-automated processing of published tables. The procedure used for naming objects and cross-identifying them is described in Sects. 3.1 and 3.2.3.

3.2.1   Tagging astronomical objects

The *Astronomy & Astrophysics* instructions ask authors to *"surround any astronomical object in your text, as well as in short tables with the command"*:

```
\object{<objectname>}
```

Objects tagged with this macro will appear linked to the corresponding SIMBAD page in the journal electronic edition. Only the authors are allowed to tag the objects. About 20% of the references in *Astronomy & Astrophysics* contain tagged objects, which explains why many astronomical objects do not have a SIMBAD link in the electronic edition.

The list of publications that cite an object is provided for each entry in SIMBAD. Although the tag does not require having a paper included in the list of references associated with an object, it significantly shortens the delay of this process. Tagging is therefore suitable for rapidly informing the community interested in an object that a specific paper is relevant.

Tags seem an efficient tool for allowing authors to directly interact with the SIMBAD content, however reality is not that simple. Naming an object is indeed a critical step that can often lead to ambiguities. Each tag requires a check by the SIMBAD team at the time of the paper's electronic publication, because it can be a new name, which is not in SIMBAD yet, or because of a potential careless use of nomenclature by the author **and** a lack of attention from the referee. Regarding the authors' part, CDS provides a web service to perform all the verifications by submitting the *.tex or the *.obj files (the .obj is generated by the compilation of the LaTeX file):

```
http://vizier.u-strasbg.fr/viz-bin/Object
http://vizier.u-strasbg.fr/viz-bin/Sesame.
```

The *Astronomy & Astrophysics* instructions to authors provide these two addresses and *"encourage you to test the astronomical objects using the sites and easy tools available at the CDS"*. Journals from the *American Astronomical Society* have adopted similar methods for maintaining links from the electronic publications to SIMBAD. Despite these tools, the number of necessary actions to build a working link is increasing. In 2002, 13% of the papers had tags and 25% of them needed *corrections*. In 2006, 25% of the papers had tags and 45% of them needed *corrections*. *Corrections* mean incorrect designations or creation of new acronyms for SIMBAD (new objects). The latter case is obviously not an author's mistake. The increase in corrections was a consequence of tagging the objects present in long tables. Since these objects, only cited in long tables and not explained in the text, are not individually important for the paper, authors are now asked to tag only the most relevant objects, meaning no more than 10–20 tags per paper. Long tables must be excluded systematically. In the present paper, tags are not used at all because none of the cited objects are studied, and this reference is obviously not relevant to any science topic.

Besides this, correct designations are sometimes ambiguous because they may refer to several different possible objects, for instance, the result of a SIMBAD query for *Perseus* yields 17 possible entries, including a molecular cloud, a star association, a superbubble, a spiral arm, a Seyfert 2 galaxy, or even a galaxy cluster. It is therefore critical

to avoid ambiguous names. As shown in Fig. 3, *Perseus* alone will point the Perseus OB 1 star association, so more specific names must be chosen to reach the correct target, hence the importance of checking links before submitting the paper.

object	OType	J2000 position	Nrefs	Resolver
perseus	Assoc*	02:21:00.00 +57:36:00.0	77	Simbad: Ass Per OB 1-
perseus molecular cloud	MolCld	03:36:00.00 +31:00:00.0	101	Simbad: NAME PER CLOUD
perseus cluster	ClG	03:19:48.00 +41:30:42.0	791	Simbad: ACO 426

Fig. 3. Summary of objects tagged in a test manuscript with the \object macro. This is obtained with the tool available at the URL `http://vizier.u-strasbg.fr/viz-bin/Object`.

3.2.2   Designations of astronomical objects

SIMBAD is not limited to tagged objects. All papers from ~ 30 journals are read by the CDS staff to extract all astronomical objects cited in the text. Taking the example of *Perseus* again, staff have to decide which of the possible objects the paper actually deals with. The first step is to recognize an astronomical name in the text, given the prolific imagination of some authors. One might naively think that any word with special characters or numbers is the signature of an astronomical object, but this is not necessarily true. A field designation for instance will probably have the typical structure of an object name although its inclusion in SIMBAD is not desired. Also some object type names are based on their prototype, such as T Tauri stars or FU Orionis stars. When the star name is used as object type, the star T Tauri or FU Ori is not to be added to the list of objects for the reference, and, symmetrically, the reference should not be added to the star T Tauri or FU Ori. The context of the paper is usually clear enough on this aspect for a well-trained *documentaliste* to find out.

Confusing names such as LW1 are more difficult to track. LW1 is an ISOCAM filter onboard the ISO satellite, and it is also a globular cluster named NGC 1466. LW1 stands for Lynga & Westerlund (1963) source number 1 and comes from a table of 483 sources with a record number as a first column. If this paper were being ingested at the CDS now, the LW1 name would be [LW63] 1. Using a prefix formed by the author's initials with the year of publication within square brackets has the advantage of linking the designation to a publication and prevents any confusion.

For example, if one decides to name his/her object M1, SIMBAD will remove the confusion with the famous Messier 1, the Crab nebula, by using the prefix. Unfortunately, some authors might cite the object but drop the prefix, and it will become more and more difficult to know which M1 it is, except maybe for the specialists of the relevant topic. Confusion will spread over the literature. Another example presented in Fig. 4 shows the content of the dictionary of nomenclature[2] for the identifier L. The second

---

[2]Dictionary link on `http://cdsweb.u-strasbg.fr/`

column gives the actual designation used in SIMBAD. Lynds dark nebulae should be

Acronym	Use	Format	Year	1st Author	Obj. Type
L	L	FFFF-NNNNA	1942	LUYTEN W.J.	PM*
(L)	[AEL81]	{L}NNa	1981	ALLOIN D.+	Part of G
(L)	[H68]	{L}	1991	WAKKER B.P.+	HVC
(L)	[L63]	NNN	1963	LINDSAY E.M.	Em*
(L)	[L65]	N	1965	LYNGA G.	HII
(L)	[L92]	HHMMSS.s+DDMMSS {Sgr D} N {Sgr E} NN	1992	LISZT H.S.	(Rad)
(L)	[L92b]	{SA} FFF LN {TPhe} A {Mark} AN {RU} NNN A {PG} HHMM+DDd A	1992	LANDOLT A.U.	*
(L)	Lal	NNNNN	1950	BAILY F.	*
(L)	LAWD	NN	1949	LUYTEN W.J.	WD
(L)	LBN	LLL.ll+BB.bb	1965	LYNDS B.T.	HII
(L)	LDN	NNNNA NNNNWW	1962	LYNDS B.T.	Dark N
(L)	[LGC91]	{ACO} NNNN NNNN	1991	LUCEY J.R.+	G + * in ClG
(L)	Lin	NNNa	1961	LINDSAY E.M.	Em. Obj.
(L)	Lindsay	NNN	1958	LINDSAY E.M.	OCl
(L)	LLNS	NNNNN	1976	LODEN L.+	*
(L)	Loden	NNNN	1977	LODEN L.O.	Loose and,or poor OCl
(L)	[PWA92]	NN	1992	PRUSTI T.+	*
NLTT	L	FFFF-NNNNA	-	LUYTEN W.J.	PM*
OL	Ohio	NNN.N -NNN.N	1967	SCHEER D.J.+	(Rad)

Note: Usage of acronyms in parentheses like 'L' should be *avoided*

Fig. 4. Dictionary of nomenclature for the acronym L.

called LDN followed by a number, such as LDN 113, for instance. The name encountered is often L113 instead. This happens so often that the ambiguity is actually solved by SIMBAD where L113 is correctly interpreted as LDN 113. Nevertheless, if one writes L113, thinking to the open cluster Lindsay 113, the result will be wrong.

On the other hand, the square-bracket prefix is not suitable for *official* catalogs. It would be meaningless to add a prefix to SDSS or 2MASS sources since the designation is formatted, so unique, and comes from a released catalog. There cannot be any confusion. Actually there are, because some authors decide to alter the official name by truncating the correct designation or by changing one digit (when the format is based on celestial coordinates). This should be absolutely forbidden. Official names must be respected even if based on coordinates, regardless of the slightly different coordinates found by follow-up observations.

Finally, papers would gain in clarity if astronomical objects were systematically accompanied by coordinates or with a second known designation in case of cross-matches. This simple recommendation would avoid a lot of ambiguities in the object identification.

3.2.3 Cross-identifications proposed in papers

Among the recent progress in astronomy, multi-wavelength analysis is certainly one of the most spectacular. This approach has been supported by SIMBAD since its beginning in the parent database CSI. One of the main SIMBAD goals is to ingest the cross-

identifications found in the literature. The scientific impact is considerable for characterizing of astronomical objects and for understanding the underlying physics. Cross-identifications will eventually lead to spectral energy distributions. Because of their crucial role, cross-identifications are introduced as carefully as possible in SIMBAD.

Unfortunately, the literature cannot always be trusted. A nearest neighbor may be a reliable cross-identification if wavelengths, sensitivities, and resolutions are similar. If not, this technique is not sufficient so a scientific analysis is required. For instance, the cross-identification of a 2MASS source (arcsec resolution) with an IRAS source (arcmin resolution) can be very hazardous, except if the object type, say a carbon star, ensures that a single 2MASS object can explain the far-infrared flux measured by IRAS. Also the cross-identification of optical with radio or X-ray sources can be risky if the method is limited to the nearest-neighbor criterion. Since authors are specialists in their domains, we cannot reach the equivalent level of expertise in all fields of astronomy to decide whether a cross-identification proposed in a paper can be trusted, but we try to ignore hazardous matches as much as possible.

Incomplete cross-identifications are often found in papers. A typical author's choice, for instance, is to provide a list of sources in the optical and to add the SDSS magnitudes without providing the SDSS names. There are definitely cross-matches, but they are implicit and we might have trouble finding out why. Ambiguities (i.e. two nearby objects) generally occur, implying a serious loss, sometimes waste, of time for the SIMBAD team when a lot a papers are waiting to be ingested, and loss of useful information known to the authors.

3.3 Integration of catalogs and tables in SIMBAD

While short published tables are integrated in SIMBAD manually by the bibliographers in charge of those articles, longer tables listing more than about 50 objects are integrated in an semi-automated way by other specialized bibliographers. Such tables are first archived electronically in the VizieR database. These VizieR electronic tables are then used and processed to include the objects and new data in SIMBAD in the most consistent way. In the first step, cross-identifications with objects already in SIMBAD are systematically searched, using coordinates and names, and checking the consistency between object types. Most of the work of the bibliographers consists in adjusting the search parameters to optimize the cross-identifications, and in checking ambiguous cases individually, with the support of Aladin facilities and of the expertise of CDS astronomers. Once all problems have been solved, the integration of the tables in SIMBAD is done automatically. Although such a procedure can be time-consuming, it is one the major goals of SIMBAD to perform careful cross-identifications between various catalogs. In case of doubt, the conservative rule is to create a new object rather than to risk an erroneous cross-identification.

In past years the number of tables published in the articles has dramatically and continuously increased. Among all the tables archived in VizieR, about one third does not have to be integrated into SIMBAD, because they are not relevant (e.g. list of atomic transition probabilities) or because it would not improve or would even degrade the level of information: lists of detailed measurements, lines or variability studies for instance;

photometric CCD surveys without characterization of the sources in very crowded fields (like in globular clusters); extremely huge catalogs like 2MASS or the SDSS. Pieces of huge catalogs are integrated via shorter tables of follow-up studies, or sometimes by dedicated operations, to cross-identify the content of SIMBAD with new large catalogs. Still, since 2006 the CDS has faced a total of 400–600 tables a year to be integrated into SIMBAD, with sizes ranging from a few 10 to 10^5 objects.

The global CDS strategy is to take full advantage of the complementarity between VizieR and SIMBAD: all tables are made available in VizieR, and a large fraction of them in SIMBAD – in priority those for which the SIMBAD added value is the most prominent.

To avoid too much delay in the integration of more important tables or catalogs, the strategy is that the CDS team, astronomers and bibliographers, perform a scientific expertise of all tables. The fundamental data and measurements to be entered in SIMBAD, as well as the main parameters to search for cross-identifications, are setup by the team, as is an order or priority. Then tables are classified into 3 groups: group 1 (first priority, 50%), to be integrated in SIMBAD as soon as possible; group 2 (second priority, 25%), to be entered in SIMBAD when possible; group 3 (10%), lists of known objects without new data, which can be done entirely automatically. The remaining 15% are not tables suitable for SIMBAD. Various criteria can yield a first-rank priority. In particular, the level of information input, which includes accurate redshift or velocity measurements, memberships, expert cross-identifications especially between different wavelength ranges, and improvement of data (coordinates, spectral types, magnitudes, etc.). The priority also goes to outstanding fields of research, such as cosmological deep fields, brown dwarfs, X- or gamma-ray researches, and identification of YSOs. The number of citations (relative to the date and journal of publication) and number of catalog queries in VizieR are checked as well.

The content of the published tables is fundamental to gaining or wasting time and information. For instance, a table listing SDSS sources can be integrated into SIMBAD two to three times faster if the authors use the latest SDSS version and list the correct and entire name. Careful cross-identifications, expertized by the authors, are also very valuable, because they save time by including the table and improve the level of information. The content of the SIMBAD database is made not only by the CDS team, but also by the astronomers themselves through the information included in their papers and the tables therein.

Information from tables is thus stored in VizieR and/or SIMBAD. To help users retrieve the information available at the CDS more easily, the team created a general portal able to access all CDS resources, including of course VizieR and SIMBAD, from a single query.

3.4 Hierarchical links

In SIMBAD, some tentative hierarchical links between astronomical objects were first introduced through object types, such as a star or galaxy in a cluster or group. This, however, has turned out to be inadequate, since it did not allow all cases of hierarchy to be included and was often reflecting more an observing method than a real link. In particular, CCDs surveys of clusters easily lead to a field of *stars in cluster* distributed

according to the instrument field of view rather than the actual cluster shape. Finally, the link by object type did not allow keeping track of either the reference in SIMBAD or the parent body, if relevant. A new tool was thus developed to include hierarchical links between SIMBAD objects. The link provides the following information: status (*parent* or *child*), probability of membership (given in the reference or estimated by SIMBAD), and bibcode of the reference providing the link. Hierarchical links can be physical, like objects belonging either to a cluster or a galaxy, multiple systems, or to a shell and or central star of a supernova. They can also be non-physical due to the improvement of instruments and resolution. For instance, an IRAS source can be resolved in several components in MSX; then, the IRAS source is the *parent* of the MSX sources, although it is no longer a single astronomical object but historic artifact worth keeping track of.

Thus, when a link is indicated on a SIMBAD object, it does not mean that the two objects are actually physically linked: a link is also put for foreground or background sources with a low or no membership in order to save this information of non-membership which can be as important as membership for some studies. In some cases, one astronomical object may have several parents. For instance, one candidate Wolf-Rayet (WR) star in a galaxy was later resolved into one B star and one WR star; the parents of both stars are the historical WR candidate and the galaxy. Hierarchical links have been systematically entered in SIMBAD as a test since 2008 for tables and catalogs archived in VizieR (Sect. 3.3). No hierarchical links are put for short lists of objects entered manually by the bibliographers, as it would always require the expertise of an astronomer. We do not consider performing a full update of the SIMBAD database for hierarchical links, which would be incredibly time-consuming. However, users can send a request to CDS[3] proposing to include them for a dedicated catalog or areas, and providing us with as much information as possible.

4   The SIMBAD database

This section focuses on the user's point of view by describing the content of SIMBAD to illustrate the final result of the data entry process described in the previous sections. The query modes are described in the Appendix.

4.1   Content

SIMBAD is continuously updated and contains 7.5 million objects in May 2014, with several kinds of data.

- **Identifiers**: census of the different names found in the literature to designate an astronomical object. The list relies on successive cross-identifications. The total number of identifiers in SIMBAD reaches 18.5 million.

---

[3]mail to: cds-question@unistra.fr

- **Basic data**: mostly data that allow the identification of the object, i.e. coordinates, proper motions, parallaxes, radial velocity and redshift, spectral type, morphological type, dimensions, and magnitudes.

  The object type is also provided, together with a list of complementary, or historical, object types.

  Objects can have *notes* to inform the user of some mistakes found in the literature, or a link with some other object, or simply more extensive information connected with this particular object.

  Hierarchical links are proposed as links between a *parent* object and its *children*, for instance, a double star and its components, a cluster of galaxies and the galaxies it contains, a star cluster and its stars.

- **Bibliographic references**: associated through citation in the paper of any of the object's identifiers. The bibliography gives access to abstracts and electronic articles when available either directly from publishers or through ADS[4]. Reciprocally, SIMBAD bibliography is systematically included in ADS.

- **Observational data**: also called *measurements*, they are always associated with a bibliographic reference. New entries have been recently added, namely distances and star diameters.

  Beside, external links are provided for some identifiers. Currently SIMBAD has a link to HEASARC[5] triggered by the existence of an identifier from a High Energy catalog and a link to some catalogs in VizieR (such as IRAS, 2MASS, HD, GSC, Hipparcos) and a link to the NED[6] database.

4.1.1 Identifiers

The identifiers constitute the basis for the SIMBAD name resolver facility that provides, in response to any object name, the coordinates corresponding to the object position or the list of papers citing the object. The name-resolving power of SIMBAD is used by many archives and information systems (such as the archives of Hubble Space Telescope or European Southern Observatory, the High Energy Astrophysics Science Archive Center, the Astrophysics Data System, servers of the Digitized Sky Surveys).

4.1.2 Basic data

- Object types: from star to maser source, or cluster of galaxies, some 180 different categories, general, or very specific, are proposed. The object type list refers to a hierarchical classification of the objects in SIMBAD derived by the CDS team (Ochsenbein & Dubois 1992). It is evolving by adding new object types when required by the evolution in astronomy. For instance, the development of the ultra luminous X-ray sources (ULX) research domain yielded the creation of a dedicated

---

[4]Astrophysics Data System: `http://ads.harvard.edu`

[5]High Energy Astrophysics Science Archive Research: `http://heasarc.gsfc.nasa.gov`

[6]Nasa/IPAC Extragalactic database: `http://nedwww.ipac.caltech.edu/`

object type in SIMBAD. A complete list is available online on the SIMBAD web-
site. The object types are also a powerful tool for data cross-checking and quality
control.

- Coordinates: they are stored in the International Celestial Reference System (ICRS,
  see Feissel & Mignard 1998) at epoch 2000.0, after the publication of the Hipparcos
  and Tycho catalogs.

- Magnitudes: SIMBAD allows storage of any number of magnitudes and fluxes
  in the Vega or the AB photometric system. Historically, SIMBAD has stored
  B and V magnitudes. Currently this list has been extended to 13 magnitudes:
  U, B, V, R, I, J, H, K, and SDSS magnitudes u, g, r, i, z. They are provided with the
  associated uncertainty and bibliographic reference.

4.1.3   Bibliographical data

The creation of a list of papers citing an astronomical object requires a code to refer to
each publication. A collaboration between SIMBAD and NED has led to the definition
of the so-called *bibcode* assigned to every paper. This 19-digit bibcode contains enough
information to locate the article (including year of publication, journal, volume, and page
number) and has been rapidly adopted by journals and ADS. With the development of
electronic publications, the page number is becoming obsolete, so too the bibcode. Pub-
lishers have defined a generic code, the digital object identifier (DOI), which is more
adapted to current publication processes. SIMBAD will obviously follow this natural
evolution.

The access to the list of references for any object is one of SIMBAD major products
for users, but for many objects, there are many references where it may be difficult to
find those of interest. To increase its usefulness, a flag is being added to represents the
relevance of the citation in terms of astronomical contents for each paper linked to an
object, so different relevance levels are possible: paper entirely devoted to an the object,
object cited in the title, object in the abstract, object cited in a table, or object simply cited
in the paper.

Several types of comments can be associated with the references in SIMBAD and
are displayed after the reference:

- General comments: they are often comments added by the bibliographers about the
  problems encountered while cross-identifying the objects mentioned in the paper,
  typos in object names, etc.

- Notes about the existence of associated electronic tables or abstracts in the CDS
  server. Papers including no object are also flagged.

- Information on how the quoted objects are named in SIMBAD (comments related
  to the Dictionary of Nomenclature of Celestial Objects).

4.1.4   Some statistics on the data contents

The astronomical content of SIMBAD results from adding a selection into the database of *important* catalogs and published tables and a survey of the complete astronomical literature. This can be illustrated by a few histograms of the *V* magnitudes (Fig. 5) and of the 13 magnitudes already available in SIMBAD (Fig. 6). The list of the 30 catalogs for which SIMBAD has the most identifiers is displayed in Table 1.
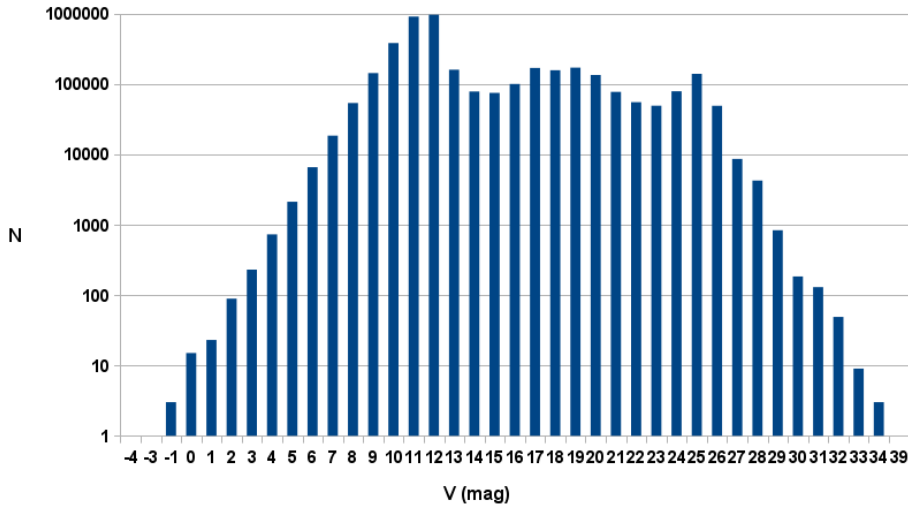

Fig. 5. *V* magnitude histogram. SIMBAD is complete up to $V \approx 10$ mag (since inclusing of the Tycho catalog) and also contains very faint objects from the deepest fields observed so far.

Table 1. Main catalogs in SIMBAD with the corresponding number of identifiers in the database. A single source generally has several identifiers from different catalogs.

|					
TYC	2526100	2MASX	311469	USNO	191343
2MASS	1907430	IRAS	301354	AG	183425
GSC	1003608	SAO	258982	SSTISAGEMA	143804
SDSS	873312	FIRST	256076	YZ	137836
COSMOS	598087	2dFGRS	234466	[VV2010]	133335
PPM	468706	CPD	231600	6dFGS	125038
Cl*	406790	OGLE	229073	1RXS	124793
SDSSCGB	388299	CD	213135	HIC	118204
HD	344234	LEDA	205593	HIP	118193
BD	319447	NVSS	202480	LCRS	92992
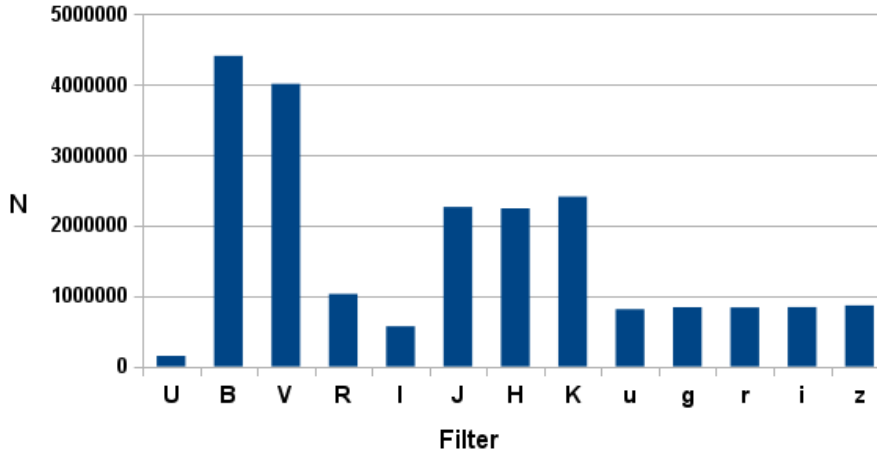
Fig. 6. Histogram of all available magnitudes. Historically only *B* and *V* magnitudes have been stored in SIMBAD but this is now extended to more filters. The high proportion of sources with *JHK* magnitudes comes from a dedicated operation of massive cross-identification between SIMBAD and 2MASS performed at the CDS, which added the near-infrared magnitudes for ∼ 1.6 million sources.

4.2 SIMBAD access

SIMBAD query modes are described in detail in the Appendix, to show how the wealth of available information can be accessed by users by simple to advanced queries, depending on their needs. The main access mode is the web, on which each query mode has its own form. Apart from the regular use of filling out a query form and getting the result in a webpage, the user has the possibility of submiting input files and to get the result in an ASCII file for further processing. SIMBAD can be accessed in different ways: defining an URL with all requested parameters, using the SIMBAD defined web services, or querying with an ADQL command using the TAP protocol. Results can be received as an HTML page, a VOTable file, a JSON output or an ASCII formatted result.

5 Aladin

A web query to VizieR or SIMBAD produces a page containing information about an object or a list of objects. An interesting alternative is to investigate sky positions and to display the results of several queries simultaneously, superimposed on images. A few details on Aladin are given here, as an illustration of the possible use of data from VizieR and SIMBAD in a software tool.

Aladin (see Fig. 7) is an interactive software sky atlas allowing the user to visualize digitized astronomical images, superimpose entries from astronomical catalogs or databases, and interactively access related data and information from the SIMBAD

database, the VizieR service and other archives for all known sources in the field. Created in 1999, Aladin has become a widely-used Virtual Observatory portal capable of addressing challenges, such as locating data of interest, accessing and exploring distributed datasets, and visualizing multi-wavelength data.
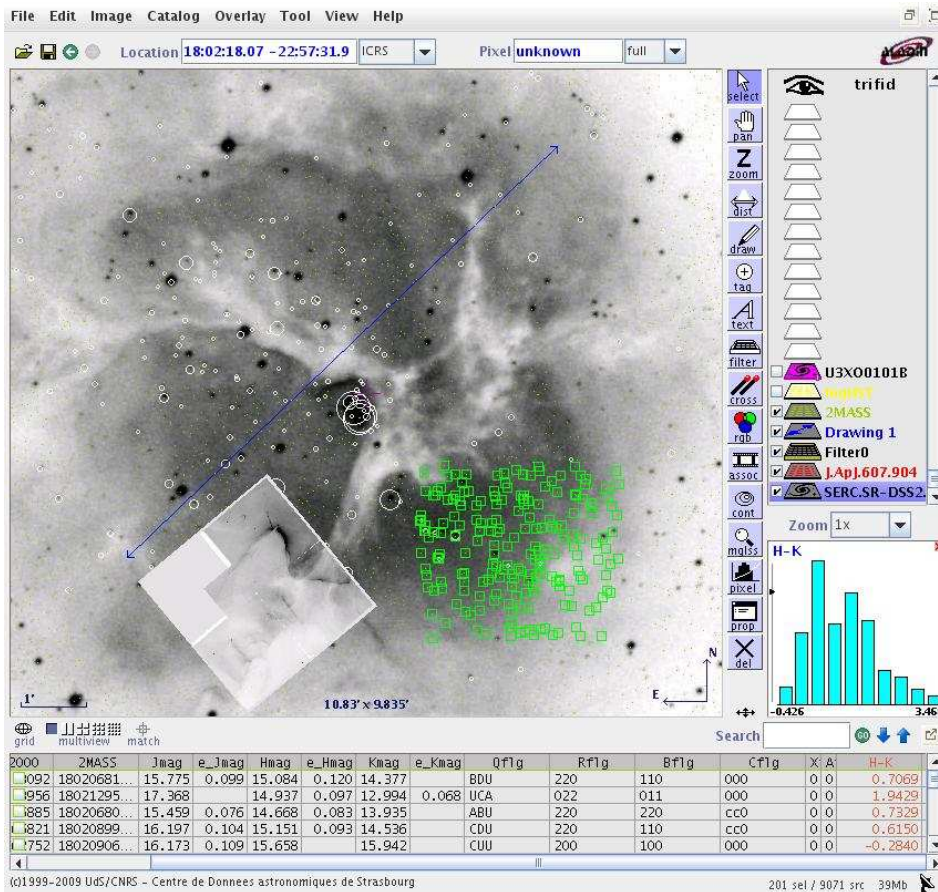

Fig. 7. Aladin : optical and HST images of the Trifid nebula. White circles are Chandra X-ray sources (Rho *et al.* 2004) with radius proportional to counts; squares are selected 2MASS sources for which an histogram of the $H - K_s$ color is displayed in the lower-right hand corner.

Aladin can be launched as an applet running in a web browser or as a standalone application. Installation instructions and documentation are available at: `http://aladin.u-strasbg.fr/`

5.1 Accessing data from Aladin

Aladin gives access to a wide range of data services that delivers images or tables:

- Access to images: the Aladin software can access and query the Aladin image server, a dedicated database running at the CDS and providing optical images coming from the Digitized Sky Survey (DSS), as well as infrared images from 2MASS. Aladin can also load images coming from external services and archives, such as SkyView, the Sloan Digital Sky Survey (SDSS), the Multimission Archive at STScI (MAST) or the NRAO VLA Sky Survey (NVSS), or even local images over 1 GB. For most services, comprehensive metadata including the coverage and the orientation of the image fields help the user choose data of interest.

  Since version 7 (released late 2010), Aladin also provides an access to HiPS (Hierarchical Progressive Surveys), allowing to browse seamlessly all-sky image surveys in a similar way to Google maps. This new functionality is based on a HEALPix-based tiling of the original images, and lets the user freely pan and zoom in the survey of interest. More than 100 image HiPS have already been processed and made available through Aladin.

- Access to tabular data: for a given region of the sky, Aladin can retrieve information coming from the SIMBAD and NED astronomical databases. It can also query any of the ∼12,000 tables with positions available in VizieR, including large surveys such as NOMAD, 2MASS, AllWISE, UKIDSS, SDSS, USNO-B1, observation logs from missions such as Chandra, HST or XMM-Newton, but also tables of individual objects published in journals.

In addition to the predefined list of available servers and to the access to VO-compliant resources, the user can describe additional servers in a configuration file, assuming they are available through HTTP calls. Moreover, local data in the form of FITS images, HEALPix maps, FITS tables, or tabulated format can be loaded into Aladin and compared with data retrieved from servers.

5.2 *Visualizing and comparing data with Aladin*

Aladin offers a whole set of tools allowing easy comparison between data covering the same region on the sky.

- Superimposition: once loaded, catalog data can be superimposed on an image of the same field. The user can then zoom in and out or pan the image by dragging the mouse. When selecting a set of objects, the associated measurements are shown in a dedicated panel.

- Combination of images: images can be combined to create a color *RGB* image or an animated blink sequence. The latter proves to be very useful when looking for objects with high proper motion or for transient events.

- Multi-view: the main window can be splitted into several panels allowing easy comparison of different images of the same region. The different views can be synchronized to show the exact same part of the sky at any time.

- Cross-identification: a cross-match tool lets the user cross-identify catalog data coming from different services, using the separation between the sources.

- Visualization of catalog sources: a tool allows filtering out and modifying the way catalog data are displayed (size, color), using information stored in the loaded catalogs. It allows, for instance, stars to be displayed as circles whose radius is proportional to the value of the flux or to display their proper motions as an arrow whose direction and length reflect the actual values of the corresponding parameters.

6 Conclusion

The CDS has constantly evolved since its creation, taking advantage in particular of the development of the Web, of electronic access to publications, and of course of the technical progress in storage capacity and network bandwidth, among others. The primary goals are to construct content from many information sources, especially papers published in journals, and to make it available to the whole astronomical community.

Metadata and normalization are absolutely critical ensuring easy access to data since they make it possible to select published data relevant to a project without having to read every paper. It also allows astronomers to find appropriate data among the increasing amount of available information (data mining). Such an ambitious goal implies a continuous concern for quality, which can only be achieved with the help of authors **and** referees. Special attention is given to providing data access simultaneously to non specialists and to providing detailed information that may be of use to specialists.

The content of VizieR and SIMBAD obviously reflects the literature, so references are systematically associated to the information stored in these databases. When errors or outdated information are detected by a user, it is advised to inform CDS, using `cds-question@unistra.fr`, and to provide as many details as possible with a reference so that a correction can be applied. Users can also contribute to the database content by posting annotations on any SIMBAD object or VizieR table.

CDS has been at the forefront of networking online information, in close collaboration with the journals, the ADS, NED, and observatory archives. Thanks to the Virtual Observatory framework, CDS services, hence their selection of published information, are integrated better and better with other online resources, e.g., through the Aladin portal, and through all other VO-enabled tools.

References

Bonnarel, F., Fernique, P., Bienaymé, O., *et al.* 2000, A&AS, 143, 33

Cayrel, R., Jung, J. & Valbousquet, A. 1974, Bull. Inf. CDS, 6, 24

Feissel, M. & Mignard, F. 1998, A&A, 331, L33

Feigelson, E.D. & Lawson, W.A. 2004, ApJ, 614, 267

Genova, F., Egret, D., Bienaymé, O., *et al.* 2000, A&AS, 143, 1

Genova, F. 2007, in ADASS XVI, editors, Shaw R.A., Hill F. & Bell D.J., ASP Conf. Ser. 376, p. 145

Jung, J. & Bischoff, M. 1971, Bull. Inf. CDS, 2, 8

Jung, J. & Ochsenbein, F. 1971, Bull. Inf. CDS, 3, 6

Lortet, M.-P., Borde, S. & Ochsenbein, F. 1994, A&AS, 107, 193

Lynga, G. & Westerlund, B. E. 1963, MNRAS, 127, 31L

Ochsenbein, F. 1994, Bull. Inf. CDS, 44, 19

Ochsenbein, F., Bauer, P. & Marcout J. 2000, A&AS, 143, 23

Ochsenbein, F. & Dubois, P. 1994, in Astronomy from Large Databases, II, editors, *Heck A. & Murtagh F.*, 405

Ortiz, P.F., Ochsenbein, F., Wicenec, A., Albrecht, M. 1999, ASPC, 172, 3790

Rho, J., Ramirez, S.V., Corcoran, M.F., *et al.* 2004, ApJ, 607, 904.

Wenger, M., Ochsenbein, F., Egret, D., *et al.* 2000, A&AS, 143, 9

APPENDIX

A   SIMBAD query modes

SIMBAD has several query modes, using different kinds of criteria and allowing different outputs, both for direct reading and use in applications. Also some Virtual Observatory enabling facilities have been implemented.

A.1   Basic query

This is the simplest query mode. The form contains a single field that can be filled with an identifier, coordinates, or a bibcode, and it returns an astronomical object or a list of objects or the reference summary. This query mode can be added to the list of query servers in any browser's tool bar.

A.2   Simple query

A.2.1   Query by identifier

Any identifier understandable by SIMBAD can be submitted to the database to retrieve information known for this object. An option allows a query around the given name, also specifying the search radius. Query with partially known identifiers is possible by embedding wildcards (*=any string, ?=one char, [c1-c2]=one char from c1 to c2, ˆ as first character in the brackets means one char excluding those specified). It is also possible to submit a list of identifiers prepared in an ASCII file, one identifier per line. The result can be a simple list of all the objects or a detailed display for each object found.

A.2.2   Query by coordinates

Queries by coordinates allow to retrieve all the objects contained in a circle defined by its center and a radius, which is 2 arcmin by default, and is limited to 10 degrees. A list of coordinates, each of which can be associated with a particular radius, can also be submitted from a user's local file. The result will contain a set of lists corresponding to each submitted position.

A.2.3   Query of bibliographic references

There are several ways to query the bibliographic references contained in SIMBAD. The basic way is to query a bibcode to get either the paper summary (title, authors, comments, ...) or the list of objects contained in the paper. The bibcode can be written using its regular syntax, with the possibility of using wildcards, or by specifying the different parts of the bibcode: year, journal abbreviation, volume, and page. Any field can be omitted. A reference summary can also be obtained by using several criteria, namely journal abbreviations, publication year range, authors, and words in the title.

Any list of references obtained by such queries can be sent to the ADS for further querying, with the benefit of all ADS facilities.

A.3   Advanced query

A.3.1   Query by criteria.

Queries by criteria allow to create samples of astronomical objects that share several common properties. SIMBAD offers more than 230 different criteria, such as basic data fields, existence of identifiers or measurement catalogs, bibliography in a range of years, and fields in the measurements. It is also possible to select the objects in a sky region, which can be a circle, an ellipse, a rectangle, a declination zone, a box defined by great circles, or a polygon.

Most criteria can be defined by a value, a range of values, or memberships in a list of values. For spectral types and object types, thereare two possibilities: either a strict correspondence with the given value (`sptype = 'F3'` or `otype = 'star'`) or with all the underlying types (`sptypes = 'F3'` will include F3 type with all the luminosity classes and/or peculiarities; `otypes = 'star'` will include all kinds of subtypes of star, e.g., variable stars, pulsars, carbon stars.

Query example: get all galaxies in a 30 arcmin circle having a *B* magnitude less than 16, and belonging to either the UGC or the MCG catalog:

```
otypes = 'gal' & region(CIRCLE,12 30 +10 10, 30m) & magB < 16
        & (cat = 'UGC' | cat = 'MCG')
```

A.3.2   Query by scripts.

A SIMBAD script is made up of a set of commands defining the output, some common parameters like a default frame, and a radius for coordinate queries and any mixture of identifiers, coordinates, bibcode, and samples to be queried.

The output is always a straight ASCII file, with no embedded HTML tag. By defining its own output format (or a VOTable output with the needed fields), the user can easily ingest the SIMBAD output in its own application, maintaining a fair independence of upgrades made in the SIMBAD server software.

Example of a script, displaying, for a list of stars, an identifier (Hipparcos if it exists, HD if not, or any first one if HD is also missing), coordinates, B and V magnitudes, spectral type, and the number of references between 2000 and 2009:

```
format object f1 "%-10IDLIST(HIP,HD,1) | %-27COO(A D) |
      %FLUXLIST(B)[%4.2*(F)] | %FLUXLIST(V)[%4.2*(F)] |
      %-5SP(S) | %#BIBCODELIST(2000-2009)"
hd 1
hd 2
hd 3
```

The output of this query is

```
HIP    422 | 00 05 08.8331 +67 50 24.013 | 8.73 | 7.43 | K0   | 2
HD       2 | 00 05 05.950 +57 46 13.22   | 8.67 | 8.20 | F5   | 0
HIP    424 | 00 05 09.7582 +45 13 44.505 | 6.75 | 6.69 | A1Vn | 7
```

Such a script can also be submitted from a user's file, which can be prepared manually or as the result of some application.

A.3.3   Query by TAP

TAP (Table Access Protocol) is a Virtual Observatory protocol allowing to query a database using the ADQL language (Astronomical Data Query Language). ADQL is an extension of the SQL language implementing some astronomical specific functions like coordinates queries. SIMBAD has implemented the TAP facility as a url access as defined in the TAP protocol, and also in a web page[7] in which an ADQL query can be typed, checked and executed. The result can be received in several formats: VOTable, JSON, CSV, TSV and simple text.

The web page queries can be performed in two modes: synchronous in which the user has to wait for the result; asynchronous in which the user launches the request, and will be later notified when it is done, and get the result from a storage space. All information requested to build queries and use the facility is available from the web page. It includes a schema of the database tables and their fields, as well as few examples to play with, like an object query:

```
-- Basic data from an object given one of its identifiers.
SELECT basic.OID, RA, DEC, main_id AS "Main identifier", coo_bibcode
    AS "Coord Reference", nbref AS "NbReferences", plx_value as
    "Parallax", rvz_radvel as "Radial velocity"
```

---

[7]http://simbad.u-strasbg.fr/simbad/sim-tap

```
FROM basic JOIN ident ON oidref = oid
WHERE id = 'm13';
```

or a coordinate query filtered by B magnitude, using the special ADQL syntax for cone search:

```
-- Display basic data about objects contained in a given circle and
whose mag B < 9.0.
SELECT basic.OID, RA, DEC, main_id AS "Main identifier", coo_bibcode
    AS "Coord Reference", nbref AS "NbReferences",
    plx_value as "Parallax", rvz_radvel as "Radial velocity",
    flux as "magB"
FROM basic JOIN flux ON oidref = oid
WHERE filter = 'B' AND flux < 9.0
    AND CONTAINS(POINT('ICRS', RA, DEC),
         CIRCLE('ICRS', 10.1234, -5.4321, 2.0)) = 1
ORDER BY "Main identifier";
```

A.4   Output options

A dedicated form allows defining many output parameters, such as the kind of output that can be an HTML page, an ASCII output, or a VOTable, a list of data to be displayed or not, selection of a preferred identifier, range of reference years, notes. Most of the output options can be defined separately for full object display and lists of object display.

The output options are kept in a cookie and are reloaded during each session run from the same workstation and the same browser.