STRASBOURG ASTRONOMICAL DATA CENTRE (CDS)

*F Genova[1]**

*[1]*Observatoire Astronomique de Strasbourg (UMR7550 UNISTRA/CNRS), 11 rue de l'Université, 67000 Strasbourg, France*
Email: francoise.genova@astro.unistra.fr

ABSTRACT

The Centre de Données astronomiques de Strasbourg (CDS), created in 1972, has been a pioneer in the dissemination of digital scientific data. Ensuring sustainability on several ten years has been a major issue since science and technology evolve continuously and the data flow increases endlessly. The paper briefly describes CDS activities and main services, and its R&D strategy to take advantage of new technologies. The next frontiers for CDS are the new Web 2.0/3.0 paradigm, and at a more general level global interoperability of astronomical on-line resources in the Virtual Observatory framework.

Keywords: Data Centre, Astronomy, Web 2.0/3.0, Interoperability, Virtual Observatory

1 INTRODUCTION

The *Centre de Données astronomiques de Strasbourg* (CDS, http://cdsweb.u-strasbg.fr/, Genova, Egret, Bienaymé, Bonnarel, Dubois, Fernique et al., 2000) has been a very early player in the dissemination of digital scientific data: it was created in 1972 by the French astronomy agency, INAG (National Institute for Astronomy and Geophysics), which is now CNRS/INSU (National Institute for Universe Sciences), in agreement with the University Louis Pasteur, now University of Strasbourg. The mandate it has been given showed a far-seeing vision, since it included, in these early times, the collection of "useful" data on astronomical objects, in electronic form, their improvement by critical evaluation and combination, the distribution of the results to the international community, and also conducting research using the data. The whole idea of electronic data collection, curation, dissemination and scientific re-use, which is the guideline of current policies about scientific data, has thus been present from the very beginning at CDS. The data centre had originally been created as the Centre de Données Stellaires (Stellar Data Centre), with the initial aim of gathering stellar data for studying the galactic structure, but it was renamed to its current name, keeping the already well known acronym, in 1983 when its domain of action was extended to all astronomical objects (outside the solar system).

CDS main role is to support the international community in its research tasks, not only to collect and to curate information. Its core task is to provide highly used value-added services (Section 2), and the main keywords of the activities are quality, scientific and technical relevance, collaboration with other actors or the field and networking of expertise and resources. Its strategy, including its R&D strategy (Section 3), is *user* and *science driven*, not technology driven. The CDS has built along the years a unique expertise on scientific data, data dissemination and exchange standards. It plays a major role in the astronomical Virtual Observatory, which aims at providing seamless access to the wealth of astronomical on-line resources.

2 CDS SERVICES

CDS has developed highly successful added-value services: SIMBAD, which summarizes information about astronomical objects – it reached 5,000,000 million objects in 2011; VizieR, the reference service for astronomical catalogues and tables published in academic journals; and Aladin, an image visualizer, conceived to access images and catalogues stored locally or remotely. These services are daily used by the international astronomical community, and their usage is constantly increasing (500,000 queries/day on average in 2010).

SIMBAD (see Figure 1)is the reference database for identification and bibliography of astronomical objects, providing a homogenized view across astronomy sub-disciplines (Wenger, Ochsenbein, Egret, Dubois, Bonnarel, Borde et al., 2000). SIMBAD data is selected from articles published in academic journals and astronomical catalogues. The first version of SIMBAD was developed at the beginning of the seventies, and the current version of the software, the fourth major update since 1972, has been operational since 2006. This last update to date has been from a home-made, object oriented database to the open source platform PostgeSQL (Wenger & Oberto, 2007). Another interesting trend has been the inclusion of "less controlled" information in addition to measurements, bibliography and data: notes from the CDS team (2002) and more recently (2010) the possibility for users to post annotations, the first CDS Web2.0 implementation. The database content is built by a team of highly qualified librarians, working closely with CDS scientists. SIMBAD contains in December 2011 5,400,000 objects, 15,200,000 object identifiers, 250,000 bibliographic references (respectively 3,000,000, 8,300,000 and 140,000 in 2003). To cope with the rapidly increasing flux of data, methods have been developed for semi-automated entry of information from the article texts (Lesteven, Bonnin, Derriere, Dubois, Genova, Oberto et al., 2010) and tables, but the results are validated by a specialist to keep a high level of quality.

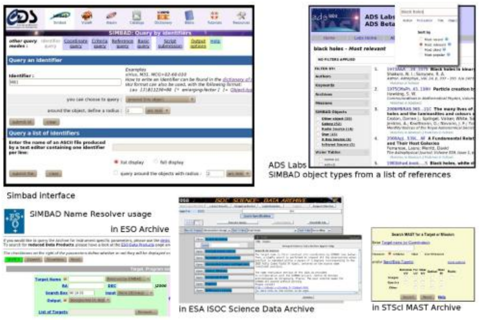

Figure 1. SIMBAD user interface (top left), and usage of SIMBAD in other services, providing object types to ADS and "name resolving" to coordinates to various telescope observation archives.

VizieR (see Figure 2) is the reference database for tabular data from astronomical catalogues and tables published in scientific papers – CDS is the data curator of these tables at the international level. Tables are completed by their description, which links their physical and astronomical content (Ochsenbein, Bauer & Marcout, 2000). Tables are available through a ftp service, and are also stored in a relational database system which allows users to browse them, to discover and extract information. The master database is a Sybase system, and queries are distributed on a local cluster. Some of the seven mirror copies are using PostgreSQL. A specific system allowing very efficient queries by position is implemented for very large catalogues (more than a few 10 million objects, the largest ones contain more than 10^9 objects). The catalogue and table collection has been built in close collaboration with several of the major astronomy academic journals, beginning with the "on-line only" publication of "long" tables from *Astronomy & Astrophysics* by CDS as early as 1993, as explained in Section 3. VizieR contains 9.500 catalogues in December 2011, compared to 3.800 in 2003. More and more tables now come with "attached data" such as images, spectra, time series, etc, well in line with the increasing requirement from the funding agencies to make data produced by research available for checking the research process and for re-use.

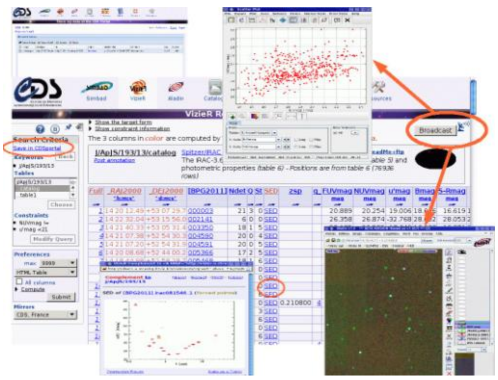

Figure 2. VizieR user interface (centre), with display of data through VO tools, Topcat and Aladin, and an object spectral energy distribution built from the newly implemented photometric metadata.

Aladin (see Figure 3) is a reference software dedicated to the integration, visualization and manipulation of images and catalogues provided by CDS, or the user, or remotely by astronomical data centres around the world. It has been continuously evolving, with functionalities allowing users to manipulate huge images and data cubes, to deal with photometry, to convolve images, to carry out cross-match, to use the software in scripts, etc. Aladin is used by ESA, the Space Telescope Science Institute, the NASA Extragalactic Database NED, the Canadian Astronomical Data Centre, to provide visualization of their images. Aladin is also the astronomical Virtual Observatory image portal, giving access to all data provided in VO-enabled services and able to interact with the other VO tools thanks to the VO interoperability framework. A major recent evolution is the usage of Healpix sky tessellation (Gorski, Hivon, Banday, Wandelt, Hansen, Reinicke et al., 2005), which provides a hierarchical view of data with fast zooming capabilities and is used by the Planck and Gaia projects. This allows a new way of using the tool, also adapted to building views of the full or part of the sky from data obtained by individual projects, which are offered the possibility to build a local Healpix database which they can open for usage by all Aladin users, or by their collaborators, or keep for themselves (Fernique, Oberto, Boch & Bonnarel, 2009). The CDS reference image database is rapidly growing by the addition of “reference skys” in different wavelengths using this method.

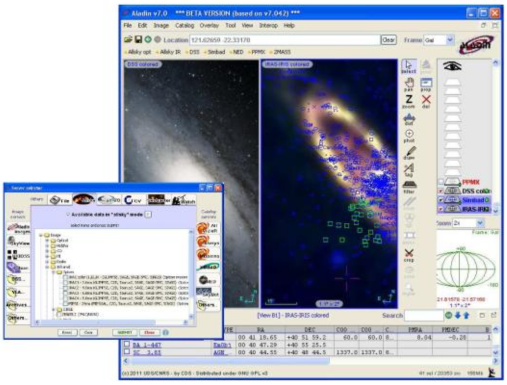

Figure 3. Aladin displaying the list of available All Sky views (left) and images of a galaxy at optical and infrared wavelengths, with objects from SIMBAD and a catalogue displayed on the right view.

3 RESEARCH & DEVELOPMENT AT CDS

Ensuring sustainability on several ten years whereas technology can evolve very quickly is not an easy task. It requires in particular a continuous and significant effort on technological and methodological watch. E.g., very soon after the advent of the WWW, CDS has been at the forefront for networking of on-line astronomical resource, in close collaboration with the academic journals, the ADS bibliographic service and observatory archives. For instance, the CDS and Astronomy & Astrophysics, which was then a European journal (it now includes also several South American countries among its partners), agreed to publish long tables electronically only at CDS, instead of printing them, as early as 1993. This was a true change in paradigm, since information which since then had been available in print only became usable and searchable data, and the system also allows navigation between publication and databases (Ochsenbein, Bertout, Lequeux & Genova, 2003).

R&D is thus a fundamental activity for medium/long term sustainability, and it has to be maintained in spite of the heavy constraints linked to the core data centre role (inclusion of the ever increasing data flow in the databases, software development and maintenance, operations). At CDS it is an in-house activity, which takes a significant fraction of the time of engineers and “instrumentalist” researchers.. R&D actions have to be properly focussed: they are driven by the data centre needs, and not technology driven, i.e. new technologies are assessed only when there is a serious promise that they could improve CDS services or functioning, and not because they are trendy. Relevant technologies have to be implemented early enough to fulfil users’ expectations, but one critical requirement is that they are “sustainable enough” for usage for a certain number of years, in a technology landscape where buzz and bandwagon effect tend to be dominant and highly praised technologies can disappear within a few years.

One current frontier is the implementation of the so-called Web 2.0 user-centric approach, and of the Web 3.0 framework, with the usage of the semantic web, mobility and universality. This is mandatory since users are expecting to find in their work environment the kind of functionalities they are using in their everyday life. CDS has already implemented the first steps, with a portal which provides “mash-up” of its services (Boch & Derriere, 2010), and the possibility for users to post annotations in SIMBAD and VizieR. The implementation of a personal user space opens the way for personalized customization of the user interface, and allows one e.g. to store preferences or results. On the other hand, the evolution towards a CDS Web 3.0 will require a deep evolution of the user interface towards a more intuitive human-machine interaction. A first version of the CDS portal for mobile phones is available.

Another frontier for astronomy service providers is global interoperability of on-line resources, the so-called Astronomical Virtual Observatory. The CDS has been a precursor of the VO in many respects. It has also been a major player of the astronomical Virtual Observatory endeavour since the emergence of the project circa 2000. It has been participating actively in the definition of interoperability standards under the auspices of the International Virtual Observatory Alliance, and implements the standards in its services, which are important building block of the astronomical information system. The CDS services are thus seamlessly available to the VO tools, and Aladin has become the image portal of the Virtual Observatory, able to interact with other tools to manage images, tabular data, spectra or data cubes, to fully explore astronomical data.

4 CONCLUSION

Since its creation 1972, CDS has successfully fulfilled its mission, to provide support without borders to the astronomical research community, and has played an important role in the networking of on-line astronomical resources, with observation archives, academic journals, and other data centres. This has required over the years an agile strategy, to deal with the constant evolutions of astronomy, users' expectations, and technology, and with the endless data flow that the data centre has to manage. Among lessons learnt is that quality, relevance to user needs and partnership with other actors are critical to ensure long term sustainability.

5 REFERENCES

Boch, T., & Derriere, S. (2010). The CDS Portal: a Unified Way to Access CDS Services. *Astronomical Data Analysis Software and Systems XIX ASP Conference Series* 434 (pp. 221-224). Sapporo, Japan.

Fernique, P., Oberto, A., Boch, T., & Bonnarel, F. (2010). Another Way to Explore the Sky: HEALPix Usage in Aladin Full Sky Mode. *Astronomical Data Analysis Software and Systems XIX ASP Conference Series* 434 (pp. 163-166). Sapporo, Japan.

Genova, F., Egret, D., Bienaymé, O., Bonnarel, F., Dubois, P., Fernique, P. et al. (2000). The CDS information hub. On-line services and links at the Centre de Données astronomiques de Strasbourg. *Astron.Astrophys. Suppl.* 143, pp. 1-7.

Gorski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M. et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophy.J.* 622, pp. 759-771.

Lesteven, S., Bonnin, C., Derriere, S., Dubois, P., Genova, F., Oberto, A., et al. (2010). DJIN: Detection in Journals of Identifiers and Names . *Library and Information Services in Astronomy VI: 21st Century Astronomy Librarianship, From New Ideas to Action. ASP Conference Series 433* (pp. 317-323). Pune, Maharashtra, India.

Ochsenbein, F., Bauer, P., & Marcout, J. (2000) The VizieR database of astronomical catalogues. *Astronomy & Astrophysics Suppl.* 143, pp. 23-32.

Ochsenbein, F., Bertout, C., Lequeux, J., & Genova, F. (2003). *Navigating from Publications to Astronomical Databases. Library and Information Services in Astronomy IV: Emerging and Preserving: Providing Astronomical Information in the Digital Age* (pp. 257-262). Prague, Czech Republic.

Wenger, M, & Oberto, A (2007) SIMBAD4: Experiences Gained from the Development. *Astronomical Data Analysis Software and Systems XVI ASP Conference Serie* 376 (pp. 527-530). Tucson, Arizona, USA.

Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S. et al. (2000). The SIMBAD astronomical database. The CDS reference database for astronomical objects. *Astronomy & Astrophysics Suppl.* 143, pp. 9-22.