Observing-Services

Centre de Données astronomiques de Strasbourg (CDS)

The CDS, initially called *Centre de Données Stellaires* until 1983, was created in 1972 by the *Institut National d'Astronomie et de Géophysique* (INAG), the predecessor of the *Institut National des Sciences de l'Univers* (INSU), in agreement with the *Université Louis Pasteur*, now *Université de Strasbourg*. The CDS obtained the national label of *Très Grande Infrastructure de Recherche* from the *Ministère de l'Enseignement Supérieur et de la Recherche* (MESR) as a result of the 2008 evaluation of UMR 7550. This is a recognition of the fact that it is indeed a Research Infrastructure, of its impact and of the quality of its work.

CDS initial charter shows the far-seeing vision of its founders and still governs its activities:

- Collect 'useful' data on astronomical objects, in electronic form
- Improve them by critical evaluation and combination
- Distribute the results to the astronomical community
- Conduct research using the data

The CDS role is to support the international astronomical community in its research tasks, and not only to collect information: it is thus *science driven*. Its core task is to provide added-value services to the community. The key aim of the activity is to maintain the highest possible level of quality and of scientific and technical relevance for its services, which requires the active participation of scientists and a resolute R&D programme. Another important driver since the beginning has been the development of its international role and collaborations, and the networking of expertise and resources with other actors.

The CDS has built over the years a unique expertise on astronomical data, on data dissemination and on exchange standards. The CDS services are major tools for astronomical research. They are widely used and their impact are ever increasing, with an average of 500.000 queries/day on the main services in 2010 (80.000 queries/day in 2006). It also plays a major role, at the international, European and national level, in the development of the astronomical Virtual Observatory, as described in the VO section of this report (see Page 41), and has led several European projects in this domain during the last period. It provides support to projects, such as recently the extension of Aladin capacities to visualize Planck data.

The CDS team is an integrated team of scientists, "*documentalistes*" and software engineers, with a deep symbiosis between the different types of activities. It relies on the general services of the Observatory for support activities (computer and network system, administration, logistics). The organisation of the Observatoire and of CDS in the Observatoire is described in page 63 (Section "Organisation chart").

The CDS services

The provision of high value-added services constitutes the basis of CDS activity and of its international credibility. The major services are **SIMBAD**, the reference database for the identification and bibliography of astronomical objects, **VizieR**, the federation of tabular data, catalogues, tables published in journals, observations logs, and surveys, and **Aladin**, a tool to discover, integrate, visualise and manipulate distributed information (images, databases, catalogues, user data).

One major challenge on the long term has been the change in scale of the volume of information to be ingested. The number of references published in the main journals increases very significantly, which impacts directly the CDS work. Also, there are more and more very large surveys, which can have more than one billion objects, and require a specific strategy to be implemented in the services. One important evolution of CDS in the last period is linked to that challenge: we have defined and put into action a strategy of complementarity between VizieR and SIMBAD, with respectively VizieR as an exhaustive collection of catalogues and published tables, and SIMBAD, with more added-value but not complete. **Sesame**, a common name resolver (including also NED, the NASA/IPAC Extragalactic Database), has been implemented to give a unified access to SIMBAD and VizieR, as well as the **CDS Portal** to give a unified access to the three CDS services. Also, an appraisal procedure has been set up to select tables to include in SIMBAD on scientific criteria, including a weekly meeting of CDS scientists and documentalistes.

For which concerns the evolution of the services, the strategy for adding new functionalities to the core ones is built on the basis of our expertise, of the users' needs, and of R&D results. One major example is the implementation of the « All Sky/Healpix mode » in Aladin, allowing Aladin usage for Planck and WMAP data, new user community for CDS. This development has led to a new way of using Aladin. The excellent capacity of Healpix to deal with hierarchical views of the sky, and the deep knowledge of Healpix gained during the process, will be reused for instance in the soon-to-come **Cross-match service**. Healpix is also an entry point to the Gaia project. Another example is the first implementation of functionalities linked to the advent of the Web 2.0/3.0, which will be described below. This includes the implementation of the CDS Portal which gives access to the three services.

SIMBAD

SIMBAD provides a view of the nomenclature and bibliography of astronomical objects which is homogenised across the sub-disciplines of astronomy. The current version of the software, which is the fourth major upgrade since 1972, has been operational since December 2006. The upgrade was from a home-made, object-oriented database, to the open source PostgreSQL data base system, and from the C language to Java, with the aim to improve maintainability, flexibility and searchability. Specific fast accesses to SIMBAD for the needs of ADS new portal and of Aladin have been developed to speed up simple queries by other services, and the possibility for users to annotate objects has been implemented in 2010. SIMBAD has a mirror copy at CfA, maintained by the ADS.

SIMBAD content is built by a highly qualified team of documentalistes and astronomers. Methods for semi-automated ingestion of data from the text of the articles and from tables have been developed over the years, but validation by a specialist remains mandatory to ensure the required level of quality. A notable evolution has been the daily usage of DJIN (Detection in Journals of Identifiers and Names), a tool which recognizes possible object names in article texts initially studied as R&D, since January 2008 by the documentalistes. Another essential evolution during the last period has been the implementation of a regular appraisal procedure for journal tables' ingestion, as discussed earlier.

In terms of content, SIMBAD passed three thresholds in April 2011: 5.000.000 objects (3.900.000 in 2008), 15.000.000 identifiers (object names – 11.750.000 in 2008), and 8.000.000 reference citations (5.750.000 in 2008). SIMBAD contains in 2011 250.000 references (216.000 in 2008).

SIMBAD has received an average of 240.000 queries/day in 2010 (70.000 queries/day in 2007, 20.000 in 2003).

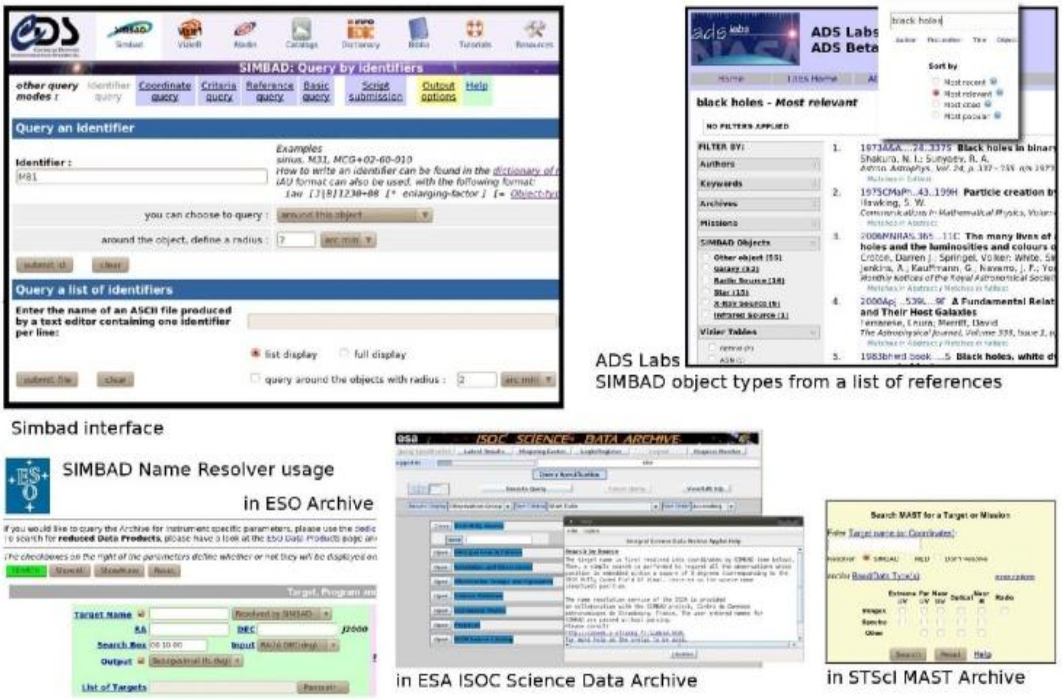

Figure 14: Usage of SIMBAD in different contexts: SIMBAD own web interface (upper left), the SIMBAD name resolver in observatory archives and the ADS (bottom), and advanced usage in the new ADS Labs interface (top right).

VizieR

VizieR is the reference database for tabular data from astronomical catalogues and tables published in scientific papers. The service contains tables and their descriptions, which link their physical and astronomical contents. Tables are available by ftp, and VizieR itself is a relational database system which allows queries by criteria on any of the table columns, except for a small number of very large catalogues (more than several million rows), for which a specific system allows very efficient queries by position. The database has 8 mirror copies (ADS, CADC, INASAN, IUCAA, NAOC, NOAJ, UCam, UKIRT). It uses the Sybase database system, and a PostgreSQL version, which is used for some of the mirror copies, has also been implemented. A new interface has been recently developed to facilitate usage and give access to advanced functionalities such as a search by the content of the tables or easier usage together with SIMBAD using the CDS Portal.

VizieR content is built in close collaboration with the journals, and in particular CDS has been in charge of the on-line publication of "long" tables for Astronomy & Astrophysics since 1993, immediately after the advent of the web – a change in paradigm which transformed printed numbers into usable data! One notable evolution is that many tables come now with "attached" data (images, spectra, time series, results of models, etc.) stored at CDS or elsewhere. Also, additional metadata allowing photometry output has been recently added for several hundred catalogues. In July 2011, there are 9.200 catalogues in VizieR (6.500 in 2008, 3.800 in 2003), containing 20.245 tables, 283.000 columns, and a total of 8.7 billion rows. About 20 catalogues, including observation logs of space and ground-based telescopes and reference databases, are regularly updated.

VizieR has received an average of 246.000 queries/day in 2010 (90.000 queries/day in 2007, 31.000 in 2005).

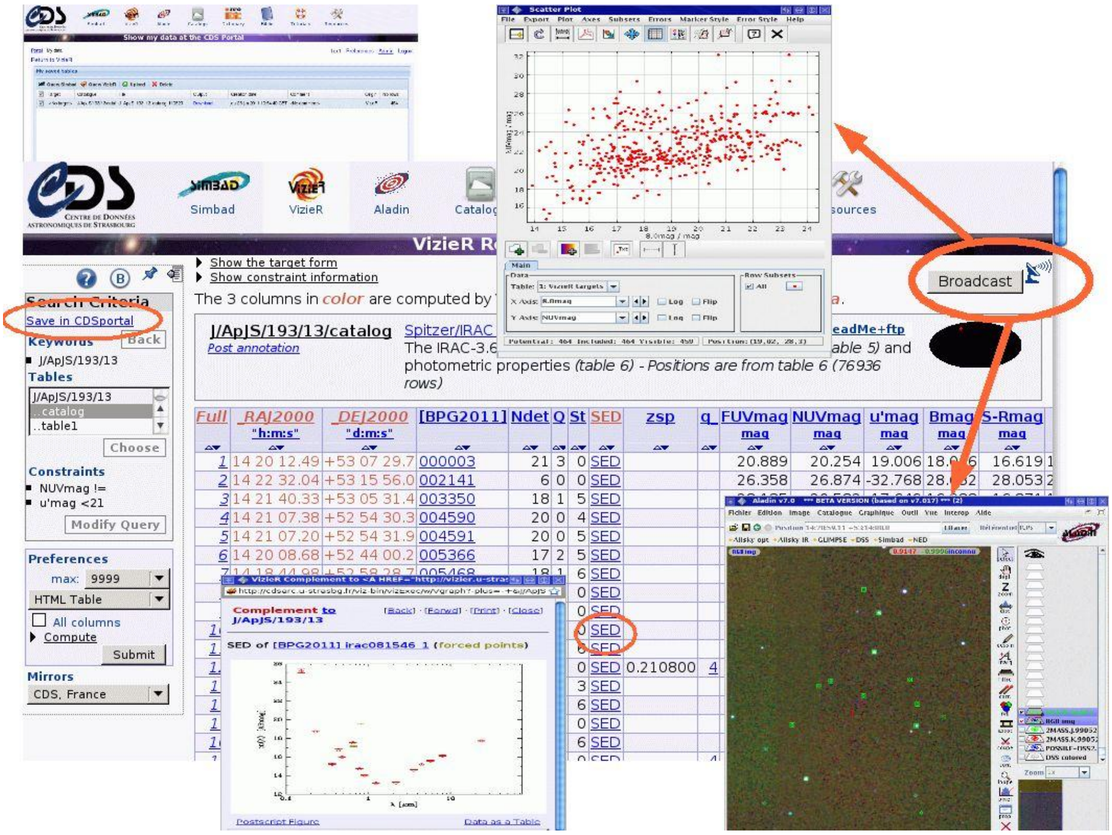

Figure 15: The new VizieR interface, showing calibrated SED using the recently added photometry metadata, links to the CDS portal, and links to TOPCAT and Aladin enabled by the Virtual Observatory.

Aladin

Aladin is a reference software dedicated to the discovery, integration, visualisation and manipulation of images and catalogues. It can access most astronomical data servers (CDS, ESA, ESO, NASA, CADC, …). The tool is continuously evolving, with many recent new functionalities, such as the management of huge images and of huge data cubes, photometry, image convolution, usage in scripts, cross-match, etc. It is used by CADC, ESA, StScI, NED, to provide visualization of their images. Aladin has been a precursor of the VO, able to access distributed data bases before the emergence of the VO concept. It has become the VO image portal, able to interact seamlessly with other VO tools such as TOPCAT.

Aladin has recently undergone a major evolution, with the implementation of the usage of the Healpix sky tessellation, which provides a hierarchical view of data with efficient zooming capabilities. The initial reason was the usage of Healpix by Planck, and our wish to offer an efficient tool to a community which was not so accustomed to use CDS. The success of this implementation opens a new way of using Aladin, with a very easy navigation between views at different scales, from the whole sky to close view of objects. This also added a new dimension to the CDS image database: the “historical” reference image database (5TB), in which image data sets are fully documented (“hierarchical data tree”) is now joined by a rapidly growing collection of Healpix views of the sky from different projects covering a wide range of wavelengths (currently 5 TB and 7 more to come soon). The historical image database is evolving into a metadata database, and it is expected that Aladin users will start from the Healpix view, and get at a later stage full information about the images of interest.

The possibility is also offered to projects to build their own Healpix local database from their images, which they can choose to open for usage to all Aladin users, or to reserve to themselves or their own community, but with the whole capacities of Aladin as an access tool.

Aladin has received an average of 16.500 queries/day in 2010 (9.000 in 2007, 2.100 in 2003). A large fraction of usage is now through local installations of the software, and is not counted in our statistics.

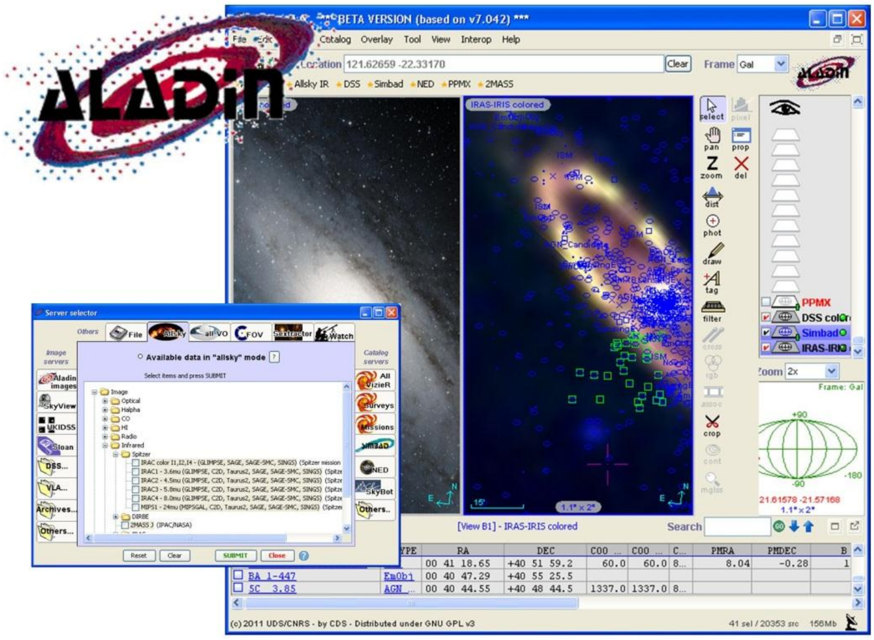

Figure 16: List of datasets available in the Healpix collection, and views of an object in two different wavelengths, with a catalogue and SIMBAD superimposed on one of the images.

Technological watch and R&D activities

Technological watch and R&D activities are fundamental for the medium and long term sustainability of the CDS, in a domain in which technology evolves continuously, and sometimes very significantly on a short time span – one excellent example is the advent of the WWW, but many less prominent advances potentially impact our activities. Technological watch and R&D have thus to be maintained continuously, in spite of the heavy operational constraints, but the strategy for these activities has to be finely adjusted to optimize them.

One key element of the strategy is that it has to be driven by the data centre needs: the aim is to improve our service to the community. We cannot afford to be technology driven, testing all appealing new technologies of possible usage for us, and because of the tight constraints, we have to carefully choose the activities we engage in. The point is to deal with the evolution of astronomy, to take advantage of the evolution of technology, and to fulfil our users' expectations: users want to find in their professional environment capacities they use in their everyday life. We thus have to identify promising new technologies and assess their possible usage. Two important aspects of the assessment are on one hand to evaluate the capability of the technology to bring really new capabilities in our services, and on the other hand to survive for a sufficient time, which is a key in a domain where buzz on short-lived supposedly "miracle" technologies pop up again and again.

At CDS technological activities are organised in-house, they are managed by permanent staff with often temporary contractors and trainees. A significant fraction of the time of the engineers and of "instrumentalist" researchers is devoted to them.

We take advantage of projects to build our R&D programme: the series of VO-oriented European projects in which the CDS participated since 2001 has had a very positive impact on the development of the VO, as discussed in the next section (Page 41), but has also strongly influenced CDS services. The CDS R&D products are new services and new functionalities in the services, and also input to VO standards and tools. Research aspects of some of these programmes have been described in the Section devoted to the research activities of the CDS team (Page 11). The long list of VO standards in which CDS staff has participated is given in the bibliography part of this document (see Page 192 in Section *PT: Transfert of knowledge publications*) .

A few example of the outcome of some of our R&D activities (some of them have been described above, and others will be discussed in the prospective section):

- the DJIN tool to extract possible object names from articles;
- the Aladin HEALPIX mode;
- the cross-match tool;
- the rapid emergence of Web 2.0/3.0 functionalities: annotations, the CDS Portal ("mashup"), the first implementations for mobile and multitouch devices;

One has also to note the collaboration with DICE[2] and IN2P3 on the implementation of iRODS, the Integrated Rule-Oriented Data System (http://www.irods.org) which manages access to distributed storage.

There have been fewer activities directly in collaboration with IT labs during the last period, but many of the results have used what we had learnt from collaborations built during the previous period through the two ACI projects coordinated by CDS in 2001-2006, in particular on semantics and image metadata. This knowledge was heavily reused in different aspects of VO standards and of their test and implementation in the CDS services.

Strategy for participation in/collaboration with projects

Participation in projects is driven by the CDS needs and expertise. Support can be given at different levels:

- Customized usage of services (e.g., VizieR customization for the XMM Survey Science Centre; Aladin update to make it able to deal with Planck data; or a specific fast access to SIMBAD for the needs of ADS new portal)
- Data distribution (e.g., discussions are on-going with CFHTLS for distribution of the public version of the catalogue and images)
- Interoperability (e.g., seamless VizieR access in TOPCAT)
- Counselling (e.g., CoRoT)

In addition, we have chosen to participate fully in the VO endeavour, a strategic choice of CDS at the beginning of the VO (see Page 41) .

European projects

CDS had been in charge of Work Packages in the first Euro-VO projects, the RTD project *Astrophysical Virtual Observatory* (FP5, 2001-2004, Coordinator: P. Quinn, ESO, Work Area *Interoperability*), and the VO-TECH Design Study (2005-2009, Coordinator: A. Lawrence, U. Edinburgh, Design Study *Intelligent Resource Discovery*). Since then it has led the three projects which have been selected in e-Infrastructure Calls:

- Euro-VO Data Centre Alliance (EuroVO-DCA[3], FP6, Call Communication Network Development, 2006-2008, 1.7 M€) identified the population of European astronomical data centres, and coordinated their

[2] Data Intensive Cyber Envirомnents, University of North Carolina

[3] http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/WebHome

first integration in the VO framework. The project produced a census of European data centres. It also performed "pre-standardization" activities to assess the possibility to include theory data in the VO framework.

- Euro-VO Astronomical Infrastructure for Data Access (EuroVO-AIDA[4], FP7 Integrated Infrastructure Initiative, Call Scientific Digital Repositories, 2008-2010, 2.7 M€), with several strands of work covering all the aspects of the astronomical VO for its transition to operations, with among the results (the lead partner is indicated into brackets) :
  - For Networking activities : a Medium Term plan for Euro-VO (CNRS), support to European technical collaboration on VO standards and tools (bi-yearly technology Forums, UEDIN), support to European participation in IVOA, Workshops for astronomers (2, ESO) and data centres (1, ESA)
  - For Service activities : the European VO Registry of Resources (ESA), two Calls for scientific proposals (« Euro-VO Research Initiative », ESO + partners) ; dissemination and outreach towards education and the general public, the first time Euro-VO and the European partners got support for this type of activities (INAF + partners)
  - For Joint Research activities (R&D) : support to the development and prototype implementation of IVOA standards and tools ; assessment of the usage of new technologies : Web 2.0 for data centres (CDS) ; Semantics and ontologies (INAF + CDS) ; Data Mining (UEDIN)

  CDS has participated very actively to nearly all these activities, except the provision of the Euro-VO Registry.
- Euro-VO International Cooperation Empowerment (EuroVO-ICE[5], Coordination Action, Call INFRA-2010-3.3, 210 k€), is a small "bridging" project initially foreseen on one year (an extension to one additional year is being negotiated), which begun on 1 September 2010. This bridging project continues to co-ordinate critical activities by providing support for meetings at European level and for participation in international VO activities.

The European projects have had a major impact on CDS services: many of the new functionalities and tools, including DJIN, the CDS Portal, annotations, etc., have been studied through the projects. On the other hand, the European projects have been a major tool to build European-wide cooperation and a common view on VO standards among the European VO teams. They have helped CDS to develop excellent technical collaborations in particular with INAF on Semantics and Education/Outreach, with ESAC (which uses Aladin for visualising the images of astronomical missions, including Herschel and Planck, but also the SOHO solar mission), and with ESO on dissemination and support to scientific programmes.

One very important result of Euro-VO efforts is the recognition by the European astronomy strategy exercise of the Astronet ERA-NET that the VO is one of the Research Infrastructures of astronomy (Astronet Roadmap *A strategic plan for European Astronomy*, 2008). This validates the CDS strategy to participate fully in the VO. Nevertheless the implementation of the Roadmap recommendations is not so easy, although they were endorsed by the funding agencies. One problem is that there is no sustainability in European funding. These aspects will be discussed in the section which deals with the CDS prospective.

User support, dissemination of knowledge

During the last period, CDS has strongly developed its actions towards the astronomical community, for disseminating knowledge on the usage of CDS services and of the VO. With the maturation of the VO,

[4] http://cds.u-strasbg.fr/twikiAIDA/bin/view/EuroVOAIDA/WebHome
[5] http://www.eurovo-ice.eu

these actions have evolved from a more “technical” point of view to a predominantly “scientific” point of view.

Many of these activities have taken place in the framework of the European projects, with an active participation to “Hands-on” Workshops and to VO days in many European countries. European Workshops provide a template for activities at the national level, and CDS has also organised several VO dissemination actions at the national level. CDS also participated actively in the preparation of tutorials on the usage of the VO, which are maintained on the Euro-VO web pages, and in turn these tutorials heavily use the CDS services.

Several events were organised in Strasbourg, including the second EuroVO-AIDA School in January 2010, the EuroVO-ICE School in March 2011, and the *Ecole Observatoire Virtuel* for French astronomers in June 2010. Events were also organised in Bordeaux and at the 2011 meeting of the French astronomical society.

The list of events in which CDS participated is as follows:

2007

- Virtual Observatory Masters Level Lecture Course, Groningen, January 9-11, 2007
- "EURO-VO Workshop on how to publish data in the VO", ESAC, Villafranca del Castillo, June 25-29, 2007.
- Workshop on Astronomy with Virtual Observatories, IUCAA, Pune, India, October 15-19, 2007
- " Virtualios observatorijos ", Moletai, Lituanie, October 17-18 , 2007

2008

- EuroVO-DCA Workshop 2008 on how to publish data in the VO, ESO, Garching bei München, June 23-27, 2008

2009

- EuroVO-AIDA School 2009, ESO, Garching bei München, 30 March - 2 April 2009
- Euro-VO AIDA Workshop on how to publish data in the VO, ESAC, Villafranca del Castillo, June 22 – 26 2009
- Black Hole Universe: 1ST School on Multiwavelength Astronomy, Paris, July 9, 2009
- ESO VO Day, September 15, 2009
- Journée OV, Observatoire de Bordeaux, September 24, 2009

2010

- EuroVO-AIDA School 2010, Strasbourg, January 25-28, 2010
- Ecole Observatoire Virtuel, Strasbourg, June 2-4, 2010
- Swedish VO Day, Stockholm, June 8-9, 2010

2011

- EuroVO-ICE School, Strasbourg, March 21-24, 2011, Strasbourg
- VO sessions, SF2A, Paris, June 21-22, 2011

- NEW CDS Tutorial, (step-by-step) **[Mar 2011]** *Uses the CDS Portal and Aladin*
- NEW Study of the Coma Cluster, with a ster-by-step description and a more expanded presentation; **[Mar 2011]** *Uses Aladin and TOPCAT* **VizieR**
- NEW A TOPCAT tutorial, with a section on multi SSA queries (step-by-step) **[Mar 2011]** *Uses TOPCAT, SPLAT-VO*
- H-alpha emitters in X-ray surveys (ster-by-step) **[June 2010]** *Uses Aladin and TOPCAT*
- Proper motion of unstudied open clusters (ster-by-step) **[June 2010]** *Uses Aladin and TOPCAT*
- A study of NGC1068 using TOPCAT for data retrieval (step-by-step) **[Apr 2010; UPDATED Mar 2011]** *Uses Aladin, TOPCAT and SPLAT-VO*
- Quasar candidates in selected fields (step-by-step) **[Mar 2009; UPDATED Mar 2010]** *Uses VODesktop, TOPCAT, VO services, VOSED and VOSpec*
- Classifying the SEDs of Herbig Ae/Be stars (step-by-step) **[Jan 2010]** *Uses TOPCAT, VOSpec and VOSED* **VizieR**
- The nature of a cluster of X-ray sources near the Chamaeleon star-forming region(step-by-step) **[Jan 2010]** *Uses VODesktop, TOPCAT and Aladin* **VizieR**
- Confirmation of a Supernova candidate (step-by-step) **[2009, UPDATED Jan 2010]** *Uses Aladin, TOPCAT, SPLAT-VO or VOSpec* **VizieR**
  - And a lighter version for undergraduate students **[Apr 2010]**
- Search for ULX sources(step-by-step) **[Mar 2009; UPDATED Mar 2011]** *Uses Aladin and TOPCAT* **VizieR**
- Study of Exoplanets(step-by-step) **[Oct 2009]** *Uses the VizieR and Simbad services and TOPCAT*
- Searching for Data available for the bright galaxy M51 (step-by-step) **[Mar 2009, UPDATED Sep 2009]** *Uses Aladin, Simbad, VizieR, TOPCAT and VOSpec*
- Discovery of Brown Dwarfs mining the 2MASS and SDSS databases (step-by-step) **[Mar 2009]** *Uses Aladin, VizieR and TOPCAT*
- The Pleiades open cluster (step-by-step) **[Mar 2009]** *Uses Aladin and TOPCAT*
- Using VOSpec: a VOSpec typical session (movie) **[2009]**
- From SED fitting to Age estimation: The case of Collinder 69 (step-by-step, includes illustrations) **[2008]** *Uses VOSA*
- Individual objects: 3C295 (step-by-step, includes illustrations) **[2007, OUT OF DATE]**
- IMF of massive stars (step-by-step, includes illustrations) **[2007, OUT OF DATE]**

Figure 17: List of VO tutorials from the Euro-VO web site (June 2011), with the usage of CDS services indicated, demonstrating that the services are among the major building blocks of the VO

Expertise

CDS staff is often asked to participate in reviews and experts groups, in France and at the European level.

In France, this concerns astronomy projects, but also other disciplines of INSU (seismology, marine data bases) and other disciplines of CNRS (SHS, IN2P3).

In Europe, CDS staff has participated in particular in a working group in charge of the Astronet Roadmap, and in the *High Level Expert Group on Scientific Data* set up by the European Commission in 2010, which has produced the report "Riding the wave – How Europe can gain from the rising tide of scientific data".