CDS Scientific Council meeting 2012

Summary of CDS activities 2011-2012

3 September 2012

TABLE OF CONTENTS

HIGHLIGHTS DURING THE PERIOD

- The periodic evaluation of Strasbourg Astronomical Observatory, including the CDS, by the *Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur* (AERES), which took place on January 11th and 12th, 2012. An activity and strategy report was produced beforehand, and the 2011 Scientific Council (SC) report was provided to the evaluators. CDS contribution to the Observatory report (data centre and scientific activities 2007-2011, strategy) is attached to this document. The evaluation was very positive and supportive of CDS and of its needs, which had been already identified by the SC during its June 2011 meeting.

- Participation in four proposals submitted to different European Calls in November 2011, one under CDS responsibility. The four proposals are currently in negotiation with or accepted by the European Commission. In addition, a R&D proposal was submitted to the CNES "Call for ideas" in September 2011, and another one to the ANR "White Call" in January 2012. Unfortunately they both failed.

- The development and commissioning of the efficient handling of very large catalogues in VizieR, as well as of the CDS cross-match service which was made publicly available on November 7th, 2011. This comes in addition to the efficient handling of image data bases ("all-sky" and more generally very large surveys) in Aladin implemented since the end of 2010. These developments have been the basis of a demonstration of CDS capabilities for the Gaia project, and allows a quick implementation of CFHT-LS images and catalogue in the CDS services.

- A complete refurbishing of the CDS web pages, which was much needed after 15 years. The new page is easier to use and should give a better feeling of CDS technological capacities compared to the previous version.

- The implementation of the possibility to sort references attached to objects in SIMBAD by their relevance to the object, a functionality which had been requested by users for many years. It is an important step to facilitate usage of bibliographic information of objects often cited in the literature.

CDS STRATEGY

Five-year evaluation by AERES

CDS contribution to the report produced by the Observatory for the five-year evaluation, which includes a description of past activities and a statement of the strategy for the future, is attached to this document. The Council members will see that the discussions at the SC June 2011 meeting have nurtured the five-year strategy document.

Here are the conclusion and the recommendations of the Evaluation Committee. The Council members will recognize most of them, others concern specific situations, in particular professional recognition of the specialized *documentalistes* who work for CDS (Rec#3). Progress has been made on this recommendation since a set of adequate profile descriptions is currently proposed to the CNRS Human Resources Direction.

Conclusion:

CDS has a unique scientific position in the French astronomical community with an extremely wide international visibility, obtained with a relatively small team. CDS has a very high visibility at the international level, which makes the Observatory of Strasbourg known by basically every astronomer in the world, through the SIMBAD and VizieR services.

One of the weaknesses of CDS is that it relies on a very small number of key people. The activities at CDS are part of a long-term project that requires long learning. The core business is very specialized, even within the astronomical community, which makes it difficult to recruit scientific leaders and technical staff with the adequate profile. Adequate replacement of the key people must be planned long ahead and requires full support from the funding agencies and the University.

Recommendations:

1. CDS should attract more PhD and postdoc students, and explore all possible means like Marie-Curie grants, ANR, EIFFEL, ERC grants, PhD grants in collaboration with Industry or major institutes and organizations.

2. CDS should enhance collaboration with external teams, in particular by the exchange of senior scientists or co-direction of students.

3. As in the 2008 AERES report, the committee renews the recommendation to properly define and recognize the activity of those people working in CDS on data ingestion. This task definition should include some means to ensure adequate continuing professional development.

4. The Funding Agencies and the University should maintain their involvement in the support of CDS, in particular give priority to maintaining the staffing of CDS at a level that take into account the increase of workload, in order to guarantee the continuity of the services.

5. Observatory management, funding agencies and the University should maintain the balance between the three profiles: Science, Computer Science, Documentalists, which is essential to the good functioning of CDS.

These recommendations will be guidelines for the five coming years. It is noticeable that the Evaluation Committee, like the Scientific Council, made strong recommendations on CDS staffing. We hope that these recommendations will be heard by our funding agencies.

Concerning more specifically one concern expressed during last year meeting, and again during the evaluation, about risks of "mutualisation" of support functions such as software engineers or librarians at CNRS and/or University level, for CDS sustainability, the answer of the President of Strasbourg University is the following (sorry, this text is in French):

- La collaboration de longue date entre l'Université de Strasbourg et le CNRS donne lieu depuis plusieurs mois à un important travail visant à mettre en place une plateforme de services partagés. L'objectif est d'apporter plus de services et d'efficacité dans la gestion quotidienne. Les réflexions qui sont menées autour de ce projet ne perdent jamais de vue la qualité des services de proximité déjà existants et ne sauraient en aucun cas désorganiser des services qui ont prouvé leur efficacité ;

The last sentence says that the current establishment by the University and CNRS of a "shared service platform" never loses from sight the quality of existing proximity services and should in any case never disorganise services which have proven their efficiency.

Status of the strategic axes presented in 2011

The high level strategy drivers presented at the 2011 meeting remain:

- Maintain the services at the highest possible level in terms of content and functionalities;

- Add functions to the core services in line with our expertise, the users' needs and R&D results;

- Take into account the change in scale of CDS activities due to the increase of publication volume and to the advent of many very large surveys.

The strategic axes identified in 2011 were of different types. (i) Those linked to the *evolution of astronomy* were to accompany the very large survey era; to put our expertise at the service of Gaia usage by the community; the construction of Spectral Energy Distributions; data cubes and polarimetry. (ii) The main driver for *technological evolution* was identified to be the new Web 2.0/3.0 paradigm. (iii) For the *VO aspects*, CDS strategy was VO implementation in the CDS services seen as a priority because they are major building blocks of the VO, to continue to update the VO framework and to disseminate the VO knowledge in the astronomical community, and to find a framework to pursue outreach towards education. (iv) The possible new role of CDS in the fast evolving landscape of *scientific data curation* was to assess.

The status of these strategic priorities will be discussed in the following. It will be commented in the next Section which summarizes our answer to the 2011 SC recommendations, for those which were discussed among these recommendations, and in the following Section which summarizes CDS activities since the 2011 SC meeting.

STATUS OF SC 2011 RECOMMENDATIONS

The previous meeting of the CDS Council, which was held on June 29-30, 2011, produced a set of recommendations which have been taken very seriously by CDS. They have for example been presented and discussed in a meeting gathering the whole staff, and are kept in mind in all "Strategy" discussions. In this relatively short period of time several recommendations have been put to action, but not all. They will remain guidelines for CDS actions in the coming years.

In the following a short comment to each of the recommendation is provided to summarize its current status.

High level recommendations

(i) Appoint PhD students and post doctoral fellows so as to increase the volume and quality of scientific work

The number of possible PhD students with University support is currently severely limited for Strasbourg Observatory (one or two/year for the whole Strasbourg Observatory, which is made of 3 teams among which CDS). One PhD position will be funded by CDS starting in 2012 for 3 years. The Observatory envisages joining a different "Doctoral School" during the coming years, which may provide a slightly higher number of PhD supports to be shared within the Observatory.

One of the European projects from the "Space" Call which will begin in 2013 will provide a post-doctoral position. This is not possible with the e-Infrastructure projects which are the usual framework of CDS European projects, and in particular of the future one which is under CDS responsibility, CoSADIE.

Efforts will continue to define and fund post-doctoral projects of interest both for CDS scientific topics and for the data centre activities. One possibility is open for 2012-2015.

(ii) Consult major data centres and cooperate with them while further developing its services

The effort has been mainly dedicated to the development and commissioning of efficient handling of very large catalogues in VizieR, which was put into operation at the beginning of 2012. It takes advantage of the developments using the HEALPix tessellation performed for the cross-match service, and allows relatively quick installation of the very large catalogues in VizieR, compared to the previous case-by-case method. UKIDSS-LAS (58 million objects) and WISE (563 million objects) have for instance been added to the catalogue collection using this method. In parallel, a large number of sky surveys have been ingested in the Aladin image collection using the efficient HEALPix method described during the last meeting of the Scientific Council. The data collection using this method started at the end of 2010, and in July 2012 it contains 81 surveys and 19 TB of data.

The available manpower does not allow CDS to canvass all major data centres. Collaboration with large projects was concentrated on Gaia and CFHT-LS during the period. The CDS expertise has been demonstrated on GUMs, the Gaia Universe Model Snapshot (2 billions rows), which should be released soon in VizieR. CDS now appears among Gaia participants. CFHT Legacy Survey Version T0007 will be released simultaneously at CADC and at CDS (likely around the end of August 2012), the images being available in Aladin and the catalogues in VizieR.

Now that the ingestion method of catalogues and images is well established, other data centres will be approached for possible collaboration.

It is also worth mentioning that the Aladin "All sky" mode provides a "do-it-yourself" method allowing projects to create their own HEALPix sky map, and to share it with other Aladin users if they wish so. This can be used by data centres (for instance it is being tested by CADC for HST data) as well as by scientific teams willing to share their data, either inside the team through a protected web site or with the community.

(iii) Continue to play an important role in the Virtual Observatory

The CDS has continued to play an important role in the Virtual Observatory. A summary of contributions to the two IVOA meetings and the list of IVOA standards with CDS authors are provided in companion documents, and some details are given in the next section. Another key contribution is the provision of CDS services in the VO framework. A

major progress has been the provision of SIMBAD in the Table Access Protocol[1]. The very complex implementation of more than 310 000 VizieR columns from 22 000 VizieR tables in this standard, which has required a huge effort because of the number of columns to be described, will be released soon.

CDS also keeps a significant influence in the IVOA structure, whereas chair positions are rotating after 3 or 4 years, as shown below.

Status of CDS role in IVOA structure (July 2012):

- F. Genova, Executive Board Member representing France, Chair of the *Standing Committee on Standards & Processes*
- M.G. Allen, Secretary of the Executive Board, Chair of the *Standing Committee on Science Priorities*
- S. Derriere, Chair of the *Semantics* Working Group until May 2012
- H. Wozniak[2], Chair of the *Theory* Interest Group until May 2012
- A. Schaaff, Vice-Chair of *Grid & Web Services* Working Group since May 2011
- M. Louys[3], Vice-Chair of the *Semantics* Working Group since May 2012
- P. Fernique, Vice-Chair of the *Applications* Working Group since June 2012

(iv) Prepare to handle a very large data base and catalogues which will become available for the forthcoming services

This point is discussed above in (ii) since it has been a major development of the services linked to cooperation with major projects.

(v) Seek strategic partnership with suitable organizations for its outreach activities

This point is the one which is not as advanced as far as it should have, because the available manpower has been overloaded by other tasks. Several steps forward have however occurred:

- A set of astronomy courses for teachers of secondary school has been organised for the *Académie de Strasbourg*, in official partnership with the Observatory, the Planetarium and the Jardin des Science (the University structure which is in charge of Outreach). These courses included an introduction to astronomy and to Aladin and Simbad provided by C. Bot, and the provision of "pedagogic sheets" using Aladin for the last three levels of secondary schools, developed by the Jardin des Sciences. The action has been very successful and will be renewed in 2012/2013

[1] The table access protocol (TAP) defines a service protocol for accessing general table data, including astronomical catalogs as well as general database tables. Access is provided for both database and table metadata as well as for actual table data. A multi-position query capability permits queries against an arbitrarily large list of astronomical targets, providing a simple spatial cross-matching capability. More sophisticated distributed cross-matching capabilities are possible by orchestrating a distributed query across multiple TAP services.

[2] Hervé Wozniak, director of Strasbourg Astronomical Observatory, is formally not a member of CDS

[3] Mireille Louys, member of the LSIIT, is associated to CDS

(although the actions for teacher formation are strongly reduced), with two identical sessions instead of one to allow more people to participate.

- An encouraging informal discussion about possible collaboration has taken place with Daisy Selematsela, Executive Director Knowledge Management and Evaluation at the National Research Foundation (NRF) of South Africa, at the World Data System Conference in Kyoto (September 2012), but no firm follow-up action has been taken for the moment, except a mention of South African interest in the CoSADIE European proposal described below.

- Most activities in this domain apart from the local activity with the *Académie de Strasbourg* have gone to find a way to continue the European effort for VO outreach towards education, which is led by our INAF colleague Massimo Ramella and in which CDS plays an important role, in particular through the provision of Aladin. Assessing the sustainability of this programme in the very complex European landscape is one task of the CoSADIE project, which should begin in September 2012 for two years. The assessment will be led by M. Ramella/INAF, with of course support from CDS, and in close collaboration with the Astronet ERA-NET, which is more generally closely associated with CoSADIE as explained elsewhere.

(vi) Seek the help of CNRS-INSU and the University of Strasbourg to deal with retirement of key staff and the requirement for additional staff and resources which would be needed for improving and developing services

This recommendation has been strongly backed by the AERES Evaluation Committee, which identifies among weaknesses and risks:

> To maintain the quality of CDS services and development is at risk. The exponential increase of published papers in astronomy creates a similar increase of the workload, which has not been compensated by a staff increase. In addition, several key engineers and astronomers will retire soon. Without a proper reaction by the funding authorities of the CDS, these 2 effects can create a situation where the CDS would not continue to fulfil its commitments. Due to the fact that the retiring people belong to different organizations, University of Strasbourg, CNRS and CNAP a coordinated plan should be agreed on and worked out by the 3 partners, each of them having different rules and priorities.

and recommends:

> The Observatory, the University of Strasbourg and the CDS should work together to ensure continuity in the quality of services provided by the CDS to the astronomical community. This implies a staffing plan that takes into account the retirement of key personnel and the increase of workload. It also implies that the CDS keeps its capability to define internally, with the help of its own advisory committees, the best strategy for new developments.

The foreseen retirement of key staff, which will be effective in 2013 and 2014, remains a huge concern, and the funding authorities are insistently made aware of it. Concertation between CNRS-INSU and the University has not yet formally happen. One can note the continuing support of CNRS-INSU, which takes seriously the need to at least maintain CDS work force. CNRS-INSU has opened a competition for a permanent *Assistant Ingénieur* documentaliste position in replacement of the departure we suffered in 2011. They have also included one recruitment at CDS among the 5 priorities of the 2012 CNAP competition. Pierre Ocvirk won an *Astronome-Adjoint* position in a very competitive context and he will join the team on September 1st, 2012 with statutorily 30% of his time devoted to functional duties.

Additional recommendations

- *The SC encourages CDS researchers to publish as first authors*

The recommendation is noted. It is reminded that IVOA standards follow a recommendation process which is much tougher than "refereed publications" and that they should validly be considered as refereed; that the ADASS plays the same role in our community as the SPIE papers for the researchers with "instrumentalist" profile; and that in the order of signatures in a publication, being the first author is not the only way to show leadership, in particular for publications in common with students.

Lists of publications in refereed journals and in ADASS conferences, and of IVOA standards, are provided in companion documents.

- *Set up a regular process of consultation of major data centres and service providers, as well as of individual users*

Contacts with project. During the past year most of the concertation has been with ESA for Gaia and Planck and with TERAPIX for the CFHT-LS. It is worth noting, since participation in CoRoT was discussed at the 2011 SC meeting, that CoRoT observation log is available in VizieR (catalogue B/corot), including time series, and that it is regularly updated, in collaboration with the Institut d'Astrophysique Spatiale.

Discussions have been on-going for a while on possible support given by CDS to ALMA users. As a first step, ALMA footprints are available in Aladin.

It is worth noting that Aladin is used in observation preparation tools in the STScI Astronomer's Preparation Tool, the CFHT Queue Service Observing, and also the future JWST preparation tool and the ING current one.

This set of collaboration should be expanded soon by contacting formally in particular ESO.

Contacts with users have continued to be mainly through demonstrations of the services, including a demonstration of VO tools in a plenary talk of the annual meeting of the French astronomy society and a "VO school" performed in Grenoble. We have begun to discuss a procedure to consult the user community through a questionnaire, which would be tested first through the French community. Suggestions are welcome on how to distribute the questionnaire.

- *Not to give up on developing joint programs with the Computer Science community*

A proposal around Web 2.0 technologies for data centres has been submitted to the ANR "White Call" in November 2011 in collaboration with Telecom ParisTech. It has unfortunately failed and the evaluation report shows fundamental misunderstandings. The Call itself was not about IT research but opened the possibility to propose R&D programmes of interest for astronomy together with more mainstream scientific projects in astronomy. The evaluators would have liked to see IT research whereas we submitted a R&D programme involving application of IT technology. The evaluation committee can probably evaluate instrumental astronomical R&D, which is very well known by the astronomical community, but seems to be not well organised to evaluate our kind of R&D. And our aim is certainly not to do research in IT.

In view of these misunderstandings, which were already met last year, of the pressure on the Call (a factor of 8), and of the amount of time needed to prepare proposals, we conclude that it is not worth pursuing our efforts to submit to these Calls for the moment.

The contacts established to prepare the project will be continued informally in support to the development which will begin in September on part of the programme foreseen in the ANR project.

SUMMARY OF CDS ACTIVITIES, JULY 2011 - AUGUST 2012

The services

The CDS web site

The previous version of the CDS web site had been created in 1998-1999, and the concept had remained unchanged since then, with only minor updates. The objective was to present and organize the different types of information which might be of interest to visitors. One important characteristic was that it was managed using the CDS GLU: this allowed to define a symbolic name for each URL present in the pages, the GLU Dictionary building the correspondence between this symbolic name and the real URL. This gave an excellent robustness to the pages, since when the URL of a local or remote URL changed it was sufficient to modify it is the GLU dictionary to have the new URL appear in all pages, and this robustness, plus the lack of dedicated manpower, explains the longevity of those web pages.

The page was getting more and more obsolete, and it passed in particular a wrong message on the technical capabilities of CDS. This is why the "general" CDS pages have been completely refurbished using state-of-the-art technologies, with a team work on the content which also led to a strong improvement in the content of the "Support" pages.

The evolution has not touched to pages of the services for the moment, but it builds a framework for future evolutions of user interface of the individual services.

SIMBAD

Sorting of references linked to objects

One of the main usages of SIMBAD is to retrieve the list of references which cite an astronomical object. But using this information becomes tricky when the number of references is large – the Large Magellanic Cloud is linked to more than 9 500 references. Users have requested for a very long time that references are somehow sorted by their relevance for the object.

Thanks to the methods implemented a few years ago to detect possible object names in references, with in addition a check by the documentalistes when they deal with the article, we are now able to order the references associated to an object by the importance the reference has for this object: it will for instance have a larger score if it is cited in the title, the keywords or the abstract. Objects cited in tables (for which data may be available in the paper) will also be indicated. The sorting will not apply to the whole collection of SIMBAD papers because semi-automated object detection has been implemented only a few years ago, except for objects in titles and keywords which have been systematically verified.

New user interface and query optimization

SIMBAD current web interface is based on a thin client which role is just to display the HTML page created and sent by the server. Every interaction implies the server.

The new web interface will receive the data in a manageable format and all processing,

from creating the web page to displaying an interactive plot of a list of objects and dynamically filtering a list of references, will be done in the user web browser without further data exchange with the server. This will dramatically improve the speed of many functionalities and allow new ones. The interface is also developed to facilitate usage, in particular to make advanced functionalities more conspicuous. It makes usage of the technologies around HTML5, CSS and Javascript.

The new interface is developed in parallel with an optimization of the database queries made possible by the new format used to send the data to the server.

SIMBAD usage in the VO : TAP interface

A SIMBAD TAP interface (using the VO Table Access Protol standard) has been developed and is available. It allows queries using the ADQL language (another VO standard) on a large subset of the Simbad table schema and can perform synchronized or batch queries using the UWS (Universal Worker Service), a VO protocol managing jobs. A web page providing TAP access is available at http://simbad.u-strasbg.fr/simbad/sim-tap in addition to the possibility to perform TAP queries from a program.

SIMBAD content – Specific operations

As the number of objects in Simbad has dramatically increased in the last years, and in view of the forthcoming data from Gaia, it appeared useful to include the whole Tycho2 catalogue in order to be complete for the brightest stars until magnitude 12. This dedicated operation has been performed by one astronomer of the team. The great majority of the Tycho2 objects could be processed in Simbad by script. The missing 0.7% (about 18000 objects) have to be checked individually (mostly due to missed cross-identifications in Simbad or close binary stars). This last part of the work will still take about one year.

The whole IRAS Point Source Catalogue was included in Simbad in the 1990ths. Still, many IRAS sources have never been studied (no references attached) or cross-identified. A dedicated operation on faint-warm IRAS sources (detected at 12 and 25 microns only) in Simbad has been performed by a documentalistand an astronomer. More than 79000 such objects, with no references in Simbad, were searched for in the japanese AKARI-IRC survey (9 and 18 microns), taking into account the IRAS error ellipse up to 3 sigmas. Then a search in Simbad, using the much better AKARI coordinates (typically 3''), allowed to cross-identify about 12000 IRAS sources with known stars. The remaining unidentified IRAS sources at least have a much better position in Simbad now.

The update an maintenance of binary stars in SIMBAD, a long overdue task, was finally tackled. The adopted strategy was to use the Washington Double Star catalogue (WDS) as a tool to clean, sort out and update the database. More precisely, all the binary stars in SIMBAD were checked for a WDS acronym, which was added when appropriate. The components listed in the WDS were then searched for in the database, the acronym added, and the position corrected if necessary. There were 6613 WDS objects at the beginning of the operation, and now they are 78742, out of a total of 84443 double stars in SIMBAD. This dedicated operation, which took over 8 months of full-time work, has been performed by one astronomer of the team.

VizieR

The number of catalogues available from VizieR crossed the symbolic line of 10,000 in the Spring 2012. The number of astronomical objects stored in the VizieR tables is now 9.3billions, a very significant increase from the the 6.9billions present in June 2011.

Very large catalogues

As explained, the new procedure allows swifter ingestion of very large catalogues: these are now 41 vs 28 in June 2011. The list of large catalogues and surveys available in VizieR is available at http://vizier.u-strasbg.fr/vizier/surveys.htx. The total number of objects from VizieR catalogues which are not obsoleted is 8.4 10^9.

SED display

Details about the photometry are now included in VizieR's metadata, such as accurate filter/photometric systems used, which magnitude scale is used, and the frequency or energy ranges of the photometric data delivered by VizieR. A generic "cone search" around a position through all vizier catalogs can deliver "photometric points" (position, frequency, flux density and the necessary metadata to retrieve the origins of each point). Currently about 1500 catalogs contain these metadata, representing about 4billion sources. A prototype which presents a plot of the resulting SED is currently available at http://cdsarc.u-strasbg.fr/viz-bin/SEDplot ; other interfaces using the HTML5 techniques are currently being developped and tested.

TAP in VizieR

As explained, a TAP interface to VizieR is being developed. The prototype in its beta release is currently available from http://tapvizier.u-strasbg.fr/.

Aladin

A new version of Aladin was released in July 2012. The new version extends a lot the all-sky facility (better performances, "build your own sky" facility, HEALPix coverage map,...), supports more formats (regions (DS9/IDL), tables (IPAC,Excel), planetary images (PDS), cubes (colored)) and incorporates an ALMA dedicated footprint utility.

As explained above, the multi-wavelength collection of surveys was greatly enriched thanks to the HEALPix-based ingestion method, with in July 2012, 81 surveys implemented and 19 TB of data.

The cross-match service

The CDS cross-match service is a new service allowing astronomers to efficiently cross-identify sources between very large catalogues (up to more than 1 billion rows) or between a user-uploaded list of positions and a large catalogue. SIMBAD is among the "tables" which can be cross-matched. The beta version was demonstrated to the 2011 SC meeting, and the service was released on 7 November 2011.

The service has been designed for large jobs; for example, the entire SDSS DR8 can be cross-matched with the 2MASS point source catalogue in less than 17 minutes, resulting in ~55x10^6 cross-matched associations with separations <2.5", and producing a ~15GB output file. This is a unique service globally, with no other cross-match service able to handle such large cross-match.

Between November 2011 and August 2012, more than 2500 jobs have been submitted by more than 400 users, representing 2.8 TB of results for 16 billion associations. Discussions are on-going with Mark Taylor to link TOPCAT directly to the cross-match service.

R&D

New interfaces

We are witnessing a rapid evolution of new human-machine interfaces based on the widespread use of multitouch screens. This evolution is not just a replacement of the mouse-keyboard couple but requires a recast of the interfaces to take advantage of the new features (example: simultaneous selections in different parts of the screen). "Traditional" operating systems (mostly Windows and Linux) are also moving towards the integration of multitouch. The user interfaces of existing applications should be deeply impacted, as it is not just an adaptation of the existing ones: it is a transition from a "selection in menus, click on button, …" to an intuitive based interaction. In this context the use of the semantics could help to "understand" what the user wants to do and to simplify the interfaces. The number of mobile devices (Smartphones based on iPhoneOS, AndroidOS and others, tablet computers such as iPad, Galaxy Tab, etc.) is growing exponentially with a sustained frequency of replacement (18 months for a device). Smartphones provide an access to Web services but also to dedicated applications (available on App Store, Android Market, etc.). Investment in human resources to provide services on mobile devices could be limited in the first case (a simple adaptation of existing Web pages), but is higher in the case of dedicated applications (software development for a given operating system and the porting to other systems to achieve sufficient diffusion).

Aladin has now an All sky mode based on the use of HEALpix. It runs well on a desktop or a laptop with good performances. We reused a part of the Aladin sources to create an Android application (called SkySurveys) and we made a few tests on an Android Internet tablet but it was too slow for a normal use. Therefore, it was a real challenge to improve this application. That version did not use OpenGL[4] but it is available on a large part of the mobile devices (tablets and smartphones). A new prototype has then been developed with the OpenGL libraries and the performances becomed better (thirty to fourty times less time to refresh a screen during the navigation in the surveys). All the surveys available in Aladin can also be shown in SkySurveys.

Another mobile application, SkyObjects, has been developped both on iOS and Android. This application is based on the Sesame name resolver and has several functionalities like the bookmarking of objects (with notes), and the pointing of the objects through the GPS/Gyroscope features.

We are also focusing the study on the use of HTML5, an emerging technology supported by recent versions of Internet browsers, which can provide rich content. HTML5 has the advantage of allowing developments independent of the mobile platform ('write once, run everywhere'). A prototype of Aladin Allsky mode is under development using HTML5 Canvas.

We also expect that this evolution will broaden the user base of the services to new audiences and in particular to the educational community through new interface user-friendlier in terms of usability and interaction.

Evolution of CDS services: Assessing the possibly expanded role of CDS in scientific data curation

This point had been briefly exposed to the 2011 SC meeting among the strategic axes, and was presented in the Strategy document provided for AERES evaluation as follows:

[4] http://www.opengl.org/

Astronomy has been at the forefront for the sharing and widespread re-use of scientific data. CDS has been an early precursor and is one of the important actors of this endeavour. One question to be assessed in the coming years is whether the CDS has new roles to play linked with respect to the Agencies' objective to enforce that *data produced on public funds should in general be made publicly available*. One of the key issues is that the physical storage of data is not the only question to address, because, to be re-usable, data must be properly described and easily retrievable. CDS relevant expertise is the management of tables and, more and more, of "additional data" (spectra, time series, models, etc) linked to publications. A possible new dimension of its service to the community could be in the direction of distributing more systematically these additional data. There are many open questions (perimeter of possible CDS action, funding, sharing of tasks with other actors, etc), but the landscape is moving very fast, and a detailed assessment of this possible major evolution is needed.

The current understanding of this possible expansion of CDS role would be to add more final processed data linked to publication in VizieR. The assessment at this stage has been to revisit the current situation. It is important to note that VizieR already contains "additional data" of different kinds (spectra, images, time series, data cubes, filters, models, profiles), with among the ~10.300 catalogues available in July 2012 as much as 812, 268 and 87 containing respectively time series, spectra and images (more than 120 000 time series and 40 000 spectra). There are established procedures to deal with these data, extract information from FITS files, etc, but this requires specific knowledge from the documentalistes who create the data description. The next step of the assessment will be to understand the kind of data and data volumes which would have to be taken into account to be able to evaluate the additional workload. Collaboration with journal editors is required to do this evaluation. Also, the possible sharing of the tasks between actors still to be defined has to be assessed.

Virtual Observatory

Implementation of CDS services in the VO

As explained, the main activity has been the implementation of the Table Access Protocol on top of SIMBAD and VizieR, which is operational for the former and will be finalized during the coming months for the latter.

In addition, a test for displaying VizieR SED data in the VAO SED display tool has been successfully completed during the May 2012 Interoperability meeting. This allows for instance the display of both NED and VizieR photometric points.

All Aladin updates benefit the VO, since Aladin can be considered as the VO image portal. Interoperability between Aladin and TOPCat and TOPCat and VizieR are among the powerful and widely used research tools enabled by the VO.

Update of the VO framework

As shown in the companion document listing the VO standards with CDS authors and CDS contributions to the IVOA meetings, several standards have been completed with CDS participation during the reference period: two with one editor from CDS, *Observation Data Model Core Components and its implementation in the Table Access Protocol* (IVOA Recommendation 28 October 2011) and *Simple Application Messaging Protocol* Version 1.3 (IVOA Recommendation 11 April 2012); three with one author from CDS: *Spectrum Data Model Version* 1.1 (IVOA Recommendation 20 November 2011), *Simple Spectral Access* Version 1.1 (IVOA Recommendation 10 February 2012) and *Simulation Data Model* Version 1.0 (with also Hervé Wozniak as editor, IVOA

Recommendation 3 May 2012). Several additional standards should hopefully be near completion: *Units in the VO*, *Photometry Data Model*, *VOSpace specification*.

As shown in the companion document listing CDS talks presented at the Interoperability meetings, CDS staff have managed the Working Groups and Interest Groups they are leading and have actively participated in particular in the discussions of the Applications, Data Access Layer, Data Model, Grid & Web Services and Semantics Working Groups, and of the Data Curation and Preservation and Theory Interest Groups.

Dissemination in the community

A training school on the scientific use of the Virtual Observatory Tools has been organised at the Institut de Planétologie et Planétologie de Grenoble, in collaboration with the Jean Marie Mariotti Centre, from November 30th to December 2nd, 2011[5].

A plenary talk/demo on the Science usage of the VO was presented during the 2012 meeting of the French Astronomical Society (June 6th, 2012). This was more successful that the lunch demos organised during the 2011 meeting (June 21st and 22nd) which were not so well attended.

Find a framework to pursue outreach towards education

The different aspects were described above. The collaboration with the *Académie de Strasbourg* and the *Jardin des Sciences* seems promising at the local level.

European projects

Euro-VO projects

The *Euro-VO International Collaboration Empowerment* (EuroVO-ICE), the small Coordination Action initially foreseen for one year starting 1 September 2010, was extended for one year and ended on 31 August 2012.

The next step for Euro-VO will be the *Collaborative and Sustainable Astronomical Data Infrastructure for Europe* (CoSADIE) project, a Coordination Action submitted to Call Infrastructures 10, which has completed negotiation with the European Commission. It aims at designing a sustainable version of the already successful European Virtual Observatory Euro-VO. Astronomy has had a bottom-up, incremental, pragmatic approach to the construction of e-Infrastructure, to build a world-wide, discipline-wide e-infrastructure complementary to the huge physical infrastructure of the discipline – the telescopes and their instruments. CoSADIE closes the loop by producing a top-down study of the strategies and coordination structures to be put in place, through a feasibility study of a sustainable Euro-VO, including the findings of previous projects and gathering requirements from all stakeholders. Interfaces with the generic e-Infrastructure will also be studied. Continuing European coordination, dissemination of results and consolidation of the education activities are also major elements of the project.

CoSADIE action plan is:
- to increase awareness among the community of users, the European astronomical scientific community, and the community of data providers, the European astronomical data centres, and to gather their requirements ;

[5] http://ipag.osug.fr/ov2011/

- to coordinate European technical activities and to understand the elements needed to reach technical sustainability, in the national, European and international contexts, and to assess how to interface with the generic e-Infrastructure ;
- to consolidate and co-ordinate outreach activities towards education and the general public interested in astronomy, and to assess the elements to reach sustainability in this domain ;
- and to define strategies and coordination structures, in collaboration with Astronet, to forge a governance and financial sustainability model for Euro-VO, taking into account the input from all the other activities.

The project is led by CDS on behalf of CNRS, with INAF, INTA, the University of Edinburgh and the University of Heidelberg as partners. It is again a small project (475 k€ for 2 years) which started on 1 September 2012. The project includes a close collaboration with the Astronet ERA-NET[6], which gathers nearly all the funding agencies of European Astronomy as partners, associates and "Forum members", which has specific recommendations in its 2008 Roadmap for European Astronomy about the VO, which is identified as an important infrastructure for astronomy.

Participation in scientific "Space" projects

CDS participates in two projects submitted to the European Commission "Space" Call in 2011, which will start on 1 January 2013:

- ARCHES (Astronomical resources Cross-matching for High Energy Studies), led by Christian Motch (Strasbourg Astronomical Observatory). In this project the CDS will work on improving multi-wavelength cross-identification tools and will distribute the final products. This will allow us to continue the collaboration with the Observatory High Energy Team on cross-identification tools, which had begun a few years ago by the common supervision of François-Xavier Pineau PhD.

- ASTRODEEP (Unveiling the power of the deepest images of the Universe), led by Adriano Fontana (INAF – Roma), which concentrates on deep data from ground and space based observatories. CDS will provide in particular dissemination of the results, provide expertise including on VO standards and work on the inclusion of objects from deep surveys in SIMBAD.

Building the Collaborative Data Infrastructure – the iCORDI project

The current European vision of the data infrastructure includes "vertical" (disciplinary) as well as "horizontal" (generic) aspects. The European and international astronomical virtual observatory should obviously be a strong and sustainable "vertical" pillar of the *Collaborative Data Infrastructure*. It is then important to share lessons learnt from the pioneering work of astronomy and to understand how a community-driven data infrastructure which is already well organised can interface with the Collaborative Data Infrastructure.

With this purpose, CDS/CNRS is a partner of the *international Collaboration on Research Data Infrastructure* (iCORDI) project submitted to the November 2011 Call 10 to organise EC/NSF collaboration on the Collaborative Data Infrastructure. This participation will in particular allow us to participate actively in the so called international *Data Access and Interoperability Task Force* (DAITF) which is being contituted, and to set up a collaboration to share lessons learnt on medium term maintenance of the disciplinary IVOA interoperability framework and to follow developments led by Edinburgh ECC to

[6] http://www.astronet-eu.org

interface the OGSA-DAI generic framework with astronomical data using TAP. It is well understood that EC/NSF collaboration is only a facet of a required global view of the data infrastructure, and astronomy has expertise and good practice to share in this domain.

ANNEX

AERES Evaluation 2012 : CDS Contribution To the *OBSERVATOIRE ASTRONOMIQUE DE STRASBOURG* Report

DATA CENTRE ACTIVITIES 2007-2011
SCIENTIFIC ACTIVITIES 2007-2011
STRATEGY
CHALLENGES

THE CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG (CDS)
ACTIVITIES 2007-2011

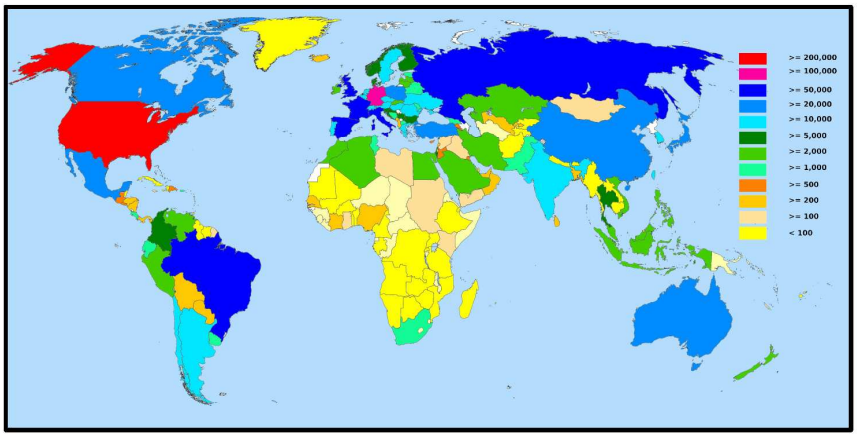
The CDS, initially called *Centre de Données Stellaires* until 1983, was created in 1972 by the *Institut National d'Astronomie et de Géophysique* (INAG), the predecessor of the *Institut National des Sciences de l'Univers* (INSU), in agreement with the *Université Louis Pasteur*, now *Université de Strasbourg*. The CDS obtain 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, which 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 keywords of the activity are 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, and has led several European projects in this domain during the last period. It provides support to projects, such as recently with the extension of Aladin capacities to visualize Planck data.



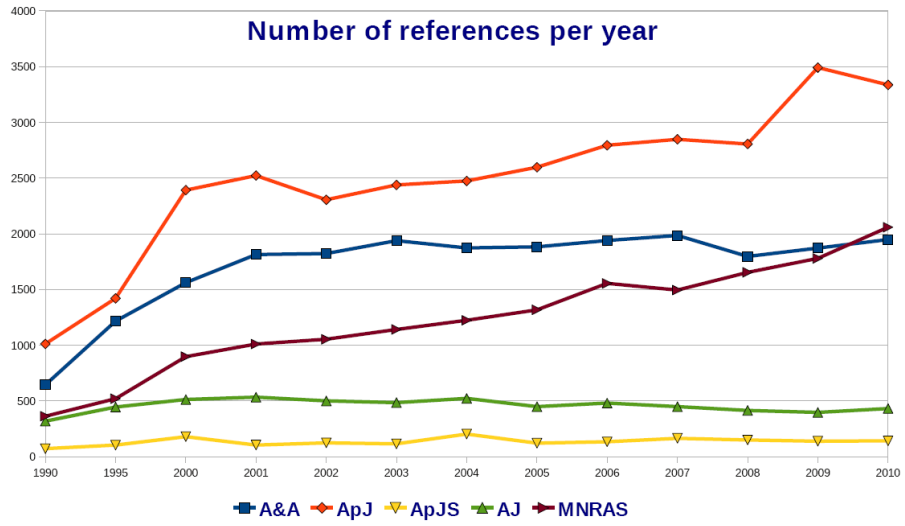
The CDS world-wide impact: the number of IP addresses which accessed SIMBAD on 30 months (Jan. 2007-Jun. 2008) in each country is shown by a color code. Usage of SIMBAD has been detected in 223 countries. Only Western Sahara, Kosovo, Azerbaidjan and North Korea did not use SIMBAD during that period

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).

1. 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 major 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 accepted that we have to put into action the 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. A common name resolver including also NED, Sesame, has been implemented, as well as an appraisal procedure to select tables to include in SIMBAD on scientific criteria.



The Content challenge: evolution of the number of references published each year by the major journals

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 (a 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 Gaia. 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 a CDS Portal which gives access to the three services.

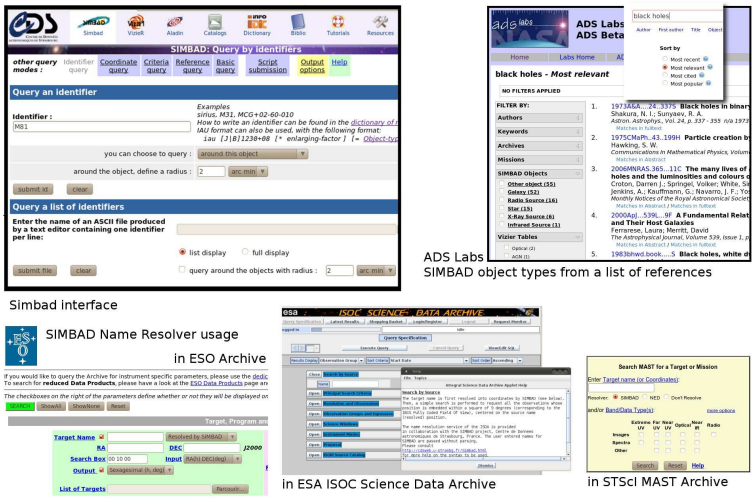
1.1. SIMBAD

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 PostgeSQL data base system, and from the C language to Java, with the aim to improve maintenability, flexibility and searchability. Specific implementations have been developed to speed up simple queries by other services (ADS, Aladin), 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.

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).



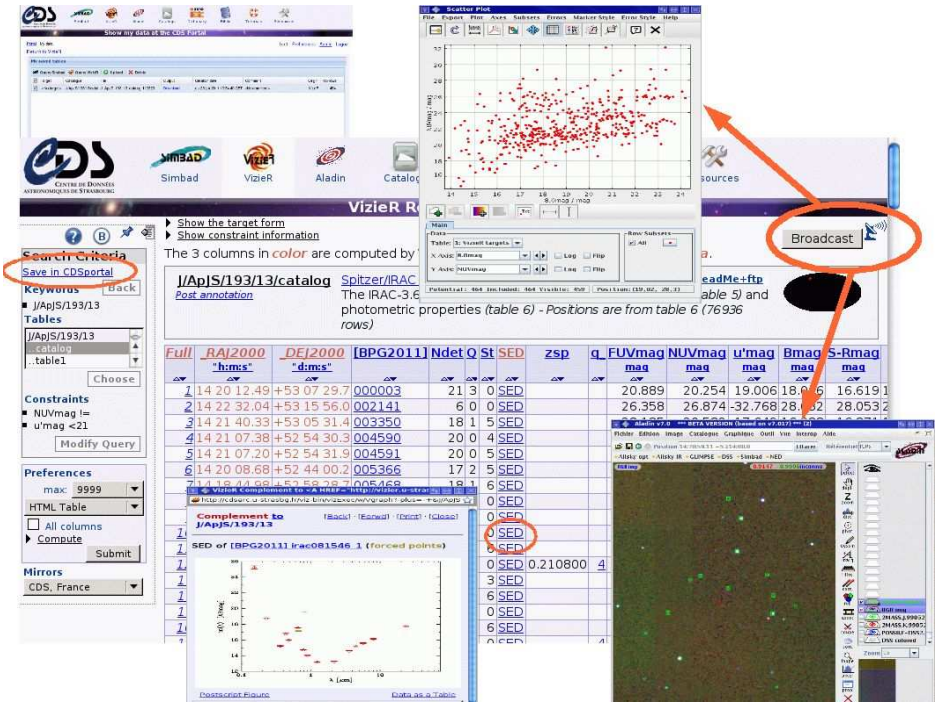
Usage of SIMBAD in different contexts: SIMBAD own web interface (upper left), the SIMBAD name resolver in observatory archives and the ADS (bottom), and advanced usage in the new ADS Labs interface (top right).

1.2. 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 the physical and astronomical contents of the tables. Tables are available by ftp, and VizieR itself is a relational database system which allows queries by criteria on any of the table columns, with the exception of 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 system, and a PostgreSQL version for some of the mirror copies. 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 useable 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).



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.

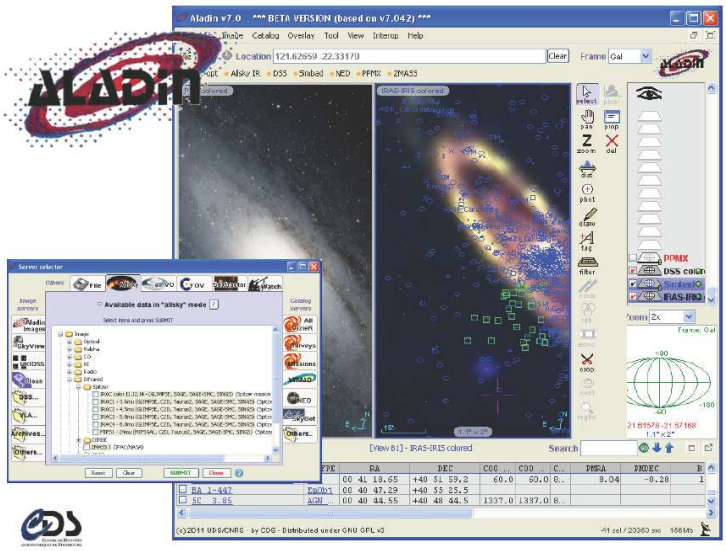
1.3. Aladin

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 one of the VO precursors, 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 pixellisation, 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 user community which is 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 image database is evolving into a metadata database, and it is expected that usage will start from the Healpix view, and get then 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.



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

2. 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, 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: user want to find in their professional environment capacities they use in their everyday life. We have to identify promising new technologies and assess their possible usage. One important part of the assessment is to evaluate the capacity of the technology 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 with 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 very positive impact on the development of the VO, as discussed in the next section, 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. The long list of VO standards in which CDS staff has participated is given in the bibliography part of this document.

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 [7] and IN2P3 on the implementation of iRODS, the Integrated Rule-Oriented Data System (http://www.irods.org) which manages access to distributed storage

There has been less activities in collaboration with IT labs during the last period, but many of the results have used what we had learnt from collaborations built through the two ACI projects coordinated by CDS in 2001-2005, in particular on semantics and image metadata (which were heavily reused in different aspects of VO standards and of their test and implementation in the CDS services).

[7] Data Intensive Cyber Enviromnents, University of North Carolina

3. Participation in and collaboration with projects

3.1. 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:

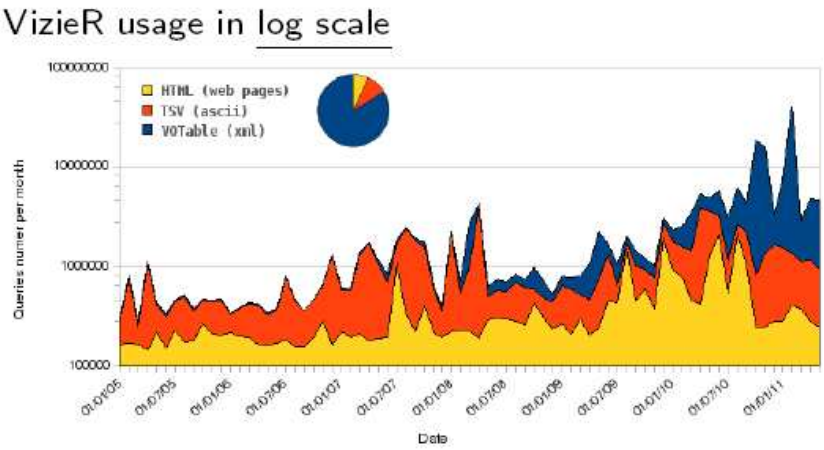
- Customized usage of services (e.g., the XMM Survey Science Center, Aladin update to make it able to deal with Planck data, or a specific fast access to SIMBAD for the needs of ADS new portal)
- Data distribution (discussions are on-going with CFHT-LS for distribution of the public version of the catalogue 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.

3.2. The CDS and the VO

The CDS has been a precursor of the VO in many respects, and it has been one of the major actors of the VO since the emergence of the concept in 1999-2000:

- It has a major role in the activities of the International Virtual Observatory Alliance (IVOA), participating actively in the Executive and Working/Interest Group leadership, in the definition of interoperability standards, and in the re-definition of the IVOA architecture which took place in 2010. CDS staff participated in the definition of 13 of the 35 standards which, in July 2011, are IVOA recommendations or on the recommendation path, in the domains of Applications, Data Access Layer, Data Models, Grid and Web Services, Semantics, VOtable, and on the standardisation process itself. VOTable, the first VO standard adopted in March 2002 (before the start of IVOA), which was initially a collaboration between CDS and the US National Virtual Observatory project.
- CDS services are important building blocks of the Virtual Observatory, Aladin as VO portal, seamlessly interacting with other VO-enabled tools such as TOPCAT, SIMBAD as a reference for astronomical objects, and VizieR for catalogues, as shown in the VizieR Figure above. For the CDS, one can say that the VO is not virtual at all, as shown in the Figure below, which demonstrates the very significant increase of the usage of VizieR through VO-enabled tools (principally TOPCAT.



Usage of VizieR through the VO: The significant and increasing usage from the VO appears in blue, which shows output in the VO format VOTable.

- CDS has a leadership role at the European and national level on VO development and the creation of a community of data centres.
    - At the national level, CDS has been the cornerstone of the *Action Spécifique Observatoires Virtuel France* since its creation in 2004. It has disseminated knowledge about the VO among the other French laboratories, and is now at the forefront for dissemination in the scientific community (a complex task since it is difficult to find the most efficient way to reach the community).
    - At the European level, the CDS had been at the origin of the first international Working Group on Interoperability (an OPTICON Working Group), which has been superseded by the IVOA in 2002. It has participated in all the successive European projects of Framework Programmes 5, 6 and 7 which have contributed to build the European implementation of the VO, Euro-VO.

3.3. 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. Edinburg, 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, FP6, - Project RI031675, Call Communication Network Development, 2006-2008, http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/WebHome, 1.7 M€), identified the population of European astronomical data centres, and coordinated their first integration in the VObs 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 VObs framework.

- Euro-VO Astronomical Infrastructure for Data Access (EuroVO-AIDA, FP7 Integrated Infrastructure Initiative, Call Scientific Digital Repositories, 2008-2010, http://cds.u-strasbg.fr/twikiAIDA/bin/view/EuroVOAIDA/WebHome, 2.7 M€), with several strands of work covering all the aspects of the astronomical VO for the 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), support to European participation in IVOA, Workshops for astronomers (2, ESO) and data centres (1, ESA)
    - For Service activities : the European VO Registry (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 ; assesment of the usage of new technologies : Web 2.0 for data centres (CDS) ; Semantics and ontologies (INAF + CDS) ; Data Mining (UEDIN)

- Euro-VO International Cooperation Empowerment (EuroVO-ICE, Coordination Action, Call INFRA-2010-3.3, http://www.eurovo-ice.eu, 210 k€), is a small "bridging" project initially foreseen on one year (an extension to an additional year is being negotiated) has 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. 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), 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 implementation is not so easy, although they were endorsed by the funding agencies. One problem is that there is no sustainability in European funding.

4. User support, dissemination of knowledge, expertise

4.1. 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, 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 similar 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

List of VO tutorials from the Euro-VO web site (June 2011), with the usage of CDS services indicated, demonstrating that there areamong the major building blocks of the VO

4.2. 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"

SCIENTIFIC ACTIVITIES OF THE CDS TEAM

One key of the success and sustainability of CDS has been that it has from the beginning been fully included in a scientific structure, the Observatoire Astronomique de Strasbourg, and that the team includes active scientists. Their expertise is mandatory to ensure the quality of content, and the relevance of the services on the long term, taking into account the constant evolution of astronomy and of users' needs. The founding fathers of CDS, when it was the *Centre de Données Stellaires*, were stellar astronomers. Taking into account the extension of SIMBAD to other objects and the development of multi-wavelength astronomy, the expertise of CDS astronomers now covers a palette of domains which will be described in the following: stellar astronomy - circumstellar matter, AGB/Post AGB stars (C. Loup); galactic astronomy - ISM, formation of stars (L. Cambrésy), formation of the galactic disk, large surveys, simulations (A. Siebert); extragalactic astronomy – ISM in nearby galaxies (C. Bot), cluster galaxies, galactic centres, galactic evolution (B. Vollmer), AGNs (M. Allen), "VO science" (M. Allen, I. Chilingarian). Several scientific topics are also linked to the CDS activities, in the domain of ontologies, semantics, information discovery (S. Derriere); cross-identification methods (S. Derriere, T. Boch, F.-X. Pineau, B. Vollmer); image processing (F. Bonnarel, M. Louys, B. Vollmer).

For the CDS, the requirement is the diversity of scientific topics, and not as usual to build a critical mass on the subjects. CDS scientists have built many collaborations with colleagues working in their fields in other laboratories, and there are also very good interactions with the other teams of the Observatory, with common scientific projects and publications, a common seminar with the Equipe Galaxies, and exchange of expertise with the XMM SSC (for instance common supervision of F.-X. Pineau's PhD). CDS astronomers participate in large surveys such as RAVE, but also e.g. to Spitzer surveys such as SINGS, SAGE-SMC, or S^3MC. They are using many ground and space-based telescopes, including AAOmega on the AAT, VLT/FLAMES, Herschel (HERITAGE), Planck, and LABOCA on APEX.

The research activities of the team will be briefly summarised in the following sections, plus a few words on the prospective on some of the topics, since it has been chosen to gather the description of the Observatory scientific strategy under the main themes of the two other teams and there is no specific prospective section for the CDS scientific activities.

1. Stellar astronomy

The activity of Cécile Loup is concentrated on the study of the Asymptotic Giant Branch stars of the Large Magellanic Cloud (LMC). The AGB is a short (100 000 years) evolution stage of stars below 6 $M_\odot$ at the end of their lives, which follows the red giant phase: these stars loose a large part of their mass and build an expanding circumstellar envelope, which is finally ejected into the interstellar medium. These cold and luminous stars are pulsating. Depending on their evolution phase and metallicities they belong to 4 different period-luminosity relations, and depending on the envelope density the emission maximum goes from optical RI bands to mean infrared.

These stars have been first catalogued in the LMC from objective prism observations or from optical photometry surveys detecting long period variables. About thirty catalogues of the whole or part of LMC are available, but they are not directly usable because, except one, the authors did not perform cross-identification between the different types of surveys. Also, the modern large photometric surveys such as 2MASS, DENIS and Spitzer SAGE, plus OGLE, EROS and MACHO for the Long Period Variables, should allow one to detect most of the AGB stars, but they do not give access to spectral information. A long term project is on-going, to cross-identify the whole information available since 1960 into a reference catalogue of LMC AGB stars. This has been a difficult task for many

reasons, including problems with the astrometry in early catalogues, which requires verification by eye. The catalogue is being finalised, and will be made easily accessible to the community. It will be used to revisit the period-luminosity relations, and to establish the luminosity function of AGBs in the LMC, which is an important parameter in galaxy modelling.

2. Galactic astronomy

ISM and star formation

Laurent Cambrésy is interested in the interstellar medium of our galaxy, with the main objectives to understand the structure of dense regions, to measure dust grain properties and to study the variations of the gas-to-dust ratio. The aim is in particular to study molecular clouds by comparing extinction in the near and mid-IR, dust thermal emission, or gas emission. He also addresses the history of star formation in the clouds.

Among the results, the study of the molecular cloud of the Trifid nebula (Cambrésy et al. A&A, 2011, 527, A141) has led to an unexpected result on the extinction law: the flattening of extinction beyond 3µm when visual extinction is larger that 20 magnitudes. This had never been observed before. In addition, the emissivity of the Trifid Cores is twice the one observed in other dense regions, and the Trifid distance has been re-evaluated to 2.7 kpc, instead of the current value 1.7 kpc. On the other hand, the gas-dust relation has been studied in details in the Taurus Cloud, and the value of X=[CO]/[H2] in this cloud is similar to the galactic value.

The next step will be to use data from Planck, Herschel and WISE to determine more precisely dust temperatures, and then their emissivity, to understand better the process responsible for their variation. A 3D study of ISM distribution, using submillimetre data and inversion techniques requiring the knowledge of extinction, has begun using Herschel HI-Gal data in collaboration with IPAC/CalTech. WISE data will be also used to develop studies of young stars and stellar formation, in particular in the Rosette Nebula for which lots of X-ray data is also available.

Galactic dynamics and kinematics

Arnaud Siebert's research is focused on the Milky Way, trying to understand the mass distribution and history of our galaxy using large scale surveys, in particular spectroscopic surveys. He has been in particular deeply involved in the RAVE survey (pipeline computing the radial velocity and the atmospheric parameters, data validation, input catalogue construction – Siebert et al., 2011, AJ 141, 87; Zwitter, Siebert et al, 2008, AJ, 136, 421; 19 papers since 2007 based on the RAVE catalogue). A PhD under his supervision at the AIP has built a pipeline to measure the abundances of 14 elements from the RAVE spectra (Boeche, Siebert et al, 2011, submitted). The RAVE pipeline has been adapted to handle AAOmega and VLT/FLAMES, with applications to study respectively the dynamics of globular clusters of the Milky Way (Lane et al. 2009, MNRAS 400, 917, 2010 MNRAS 401, 2521, MNRAS 406, 2732, ApJ 711 L122, 2011 A&A 530 A31) and the Palomar 5 and Sagittarius streams, in close collaboration with the Galaxy team. Work is ongoing to release a stable version of the pipeline for community usage.

Using the RAVE catalogue, the inclination of the velocity ellipsoid at 1 kpc below the galactic plane was measured for the first time (Siebert et al, 2008, MNRAS 391, 793). The knowledge of the amplitude of the tilt at various heights above/below the plane provides a unique tool to test the mass distribution of the Milky Way, and we showed that the measured value further confirms that the dark halo cannot be significantly flattened. Analysing the global motion of the stars in the RAVE survey, we also detected significant non-circular motion in the extended Solar neighbourhood (Siebert et al. 2011 MNRAS 412 2026), likely associated to non-axisymetric pertubations in the Galactic disc,

the influence of local spiral arms being so far the most likely explanation. The importance of the spiral arms and of the Galactic bar on the local dynamics is now well established and can be observed via the clumping in velocity space (Famaey, Siebert & Jorissen 2008, Minchev et al. 2009,2010). Comparing the two effects should help constrain the amplitude of the spiral arms in the near future.

For the future, theoretical work on the perturbations of the galactic disc are planned in collaboration with the Galaxy team, as well as participation in the Gaia/ESO survey and in 4MOST on the VLT, HERMES on the AAT and Gaia.

3. Extragalactic astronomy

Interstellar medium in nearby galaxies

Caroline Bot studies infrared to millimetre dust emission in galaxies, especially in the Small Magellanic Cloud (SMC). The Magellanic Clouds are excellent candidates to study interstellar dust in an environment different from our Galaxy, because of their proximity and low metallicity, and to make the link with more distant galaxies.

The mass of dust can be used to trace the gas which is not observed by other tracers such as HI or CO. This "dark" gas has been observed in our galaxy and in nearby galaxies, and could correspond to molecular cloud envelopes. If so, the "dark" gas quantity should increase at low metallicities. C. Bot has observed that this is the case in the giant molecular clouds of the SMC, by showing that the quantity of molecular gas traced by dust emission is higher than the virial masses computed from CO observations (Bot et al. 2007 A&A 471 103, Bot et al., 2010 A&A 524 A52). Another aspect is the excess in dust emission observed in far infrared and millimetre wavelengths at low metallicities. By extending the spectral energy distribution of the Magellanic Clouds with WMAP and TopHat, it was shown that the excess observed in far infrared and submillimetre extends to the centimetre domain (Israel et al. 2010, A&A 519, A67 , Bot et al 2010 A&A 523 A20). This cannot be explained by the reasons usually put forward, in particular it cannot be very cold dust since it should be as cold as the CMB. Using Planck observations it has been shown that CMB fluctuations can explain the excess observed in the LMC, but not in the SMC where a significant excess remains (Planck collaboration, 2011, arXiv: 1101.2046) .

Among other studies, the search for dust emission in intergalactic clouds, e.g., the Leo intergalactic Cloud (Bot et al, 2009, AJ, 138, 452). Projects involve the analysis of data obtained by Planck and Herschel on the Small Magellanic Cloud and foreground cirrus, which can significantly bias extragalactic observations, and participation in an international team to inter-calibrate dust models.

Cluster galaxies, galactic centres and galaxy evolution

Bernd Vollmer develops different studies in these domains, with a wide set of collaborations. One is linked to the evolution of cluster galaxies, with the development of a code simulating the interactions of a spiral galaxy with its environment, and observations of a significant number in CO, HII and continuum. The comparison between observations and simulations has led to the first model-based ram pressure stripping time sequence for spiral galaxies of the Virgo cluster (Vollmer et al. 2007 A&A 462, 93, 2008 A&A 491, 455, A&A 483, 89, 2009 A&A 502 427, A&A 496 669, A&A 512 A36, Pappalardo et al 2010 A&A 514 A33, Abramson et al 2011 ApJ 141, 164). He also studies the physics and dynamics of gas in the Galactic centre and in AGNs. An analytical and dynamical model of gas in the central 20pc of the Milky Way has been developed, which has then be generalized to active galactic nuclei (Vollmer et al A&A 491, 441). In addition, he is studying the theory of clumpy turbulent accretion disks, by extending it to

describe galactic disks, with application to nearby galaxies (Vollmer & Leroy, 2011 AJ 141 24).

The cross-identification of radio catalogues, realised as a service task for CDS (Vollmer et al., 2010 A&A 511 A53), has also allowed the discovery of new sources with peculiar spectra. Follow-up radio observations have confirmed 65 new Gigahertz-Peaked Spectrum (GPS) and High Frequency Peaker (HFP) sources to add to the known <200 sources from existing samples (Vollmer et al. A&A 489, 49). These are young AGN sources with peculiar properties, which are not so well studied for the moment.

Physics of ionized gas in galaxies

Mark Allen's research work in the field of Active Galaxies is focused on understanding the physical processes in galaxies in order to study the effect of Active Galactic Nuclei (AGN) on their host galaxies, and to establish the role of AGN in the formation and evolution of galaxies.

A major theoretical modeling effort to compute the MAPPINGS III Library of Fast Shock Models was completed in 2008 (Allen et al. 2008 AJSuppSer 178, 20). Combined with AGN photoionization and starburst ionizing models, the shock models have been made accessible in a uniform way via a program called ITERA, the IDL Tool for Emission-line Ratio Analysis (Groves & Allen 2010, NewA, 15, 614). These models are being applied to a range of astrophysical phenomena (66 refereed citations since 2008). Recent low velocity shock models (v<200 km/s), along with all the ITERA models, have been developed to support analysis of ionized gas in nearby interacting galaxies with HST. One can also cite two studies performed with PhD students: an extensive study of the environments of High Redshift Radio Galaxies, and AGN in galaxy clusters (Galametz et al. 2009 A&A 507, 131); the X-ray properties of LINER galaxies (Younes et al. 2010, with the High Energy Astrophysics team).

VO science

One can cite here the results obtained from one of the projects selected in the framework of the 2008 Announcement of Opportunities of the EuroVO-AIDA project for advanced exploitation of astronomical digital archives through VObs tools. This project, *Is a galaxy's disk scale-length universal?*, was supported by CDS. The project has derived key parameters for the largest ever sample of galaxies in a robust and homogeneous way, using Virtual Observatory methods and tools (Fathi et al., MNRAS, **406**, 1595). These parameters have subsequently been used by Fathi to show that the stars populating the discs of galaxies are distributed in the same fashion throughout the entire Hubble sequence.

Another example is the series of studies of compact elliptical and ultracompact dwarf galaxies by I. Chilingarian, which combine data mining using the Virtual Observatory, optical spectroscopy, stellar population synthesis and the usage of workflows.

4. Information research and processing

This domain is in particular the domain of research of the CDS staff with 'instrumentalist' profile, but other researchers with a more classical profile are also involved. In the previous period (2001-2006), several projects in the *Action Concertées Incitatives* framework were performed in collaboration with IT laboratories, in the domain of semantics and ontologies, of image processing, and of 'system architecture'. The period of the present report has rather been for exploitation of the knowledge gained from these collaborations, in particular in R&D programmes linked to European projects (VO-TECH and EuroVO-AIDA), and in the definition of IVOA standards. It is important to note here that IVOA standards follow a tough review process in successive steps by the whole

community of IVOA developers, which is much harder than the one of refereed publications. These standards should be considered as the equivalent of refereed publications for 'instrumental' research. In addition, papers presented at the ADASS conferences are similar to the SPIE papers for more classical 'instrumental' research. The details of these publications can be found in Sections PT, C-ACTI and C-AFF of the publication list.

Ontologies

The aim was to experiment on 'real' usage of ontologies, and for this a full ontology of astronomical object types was built, which contains concepts and the relations between them. This domain was chosen because of CDS knowledge about astronomical objects, gathered in particular when building SIMBAD content, and because the number of concepts remain manageable. An IVOA standard for Vocabularies in the Virtual Observatory was also worked out,. It specifies a standard format based of the W3C's Resource Description Framework (RDF) and Simple Knowledge Organisation System (SKOS), which permit different groups to create and maintain their own specialised vocabularies while letting the rest of the astronomical community access, use, and combine them. The use of current, open standards ensures that VO applications will be able to tap into resources of the growing semantic web.

Cross-identification

CDS collaborated with the High Energy Astrophysics Team in the development of a cross-match algorithm to cross-match the XMMi catalogue with the 7th release of the SDSS (PhD of F.-X. Pineau, Pineau et al 2011 527, A126). One also has to recall here the cross-identification of radio catalogues for SIMBAD, taking into account the physical characteristics of sources and observations, which produced a new catalogue, SPECFIND, and a set of VO tools (Vollmer et al., 2010, A&A 511 A53). Since then, a more general algorithm is being built for the CDS cross-match service, with the aim to be able to cross-match efficiently very large catalogues, taking into account positional errors (Pineau, et al., Efficient and scalable cross-matching of (very) large catalogues ADASS XX). This has led to the development of methods relying on multi-threaded modified kd-trees and HEALPix tessellation to manipulate and allow fast query by position for tables up to several billion rows. We also developed a method to compute fast likelihood ratios on a cross-match result.

Image processing

The collaboration with the LSIIT continued, on object detection using a Markovian approach based on hierarchical neighbourhood relations and Bayesian inference. Images pixels are sorted into different classes according to their surface brightness and spatial continuity. This method has been successfully applied to the detection of Low Surface Brightness Galaxies, and its usage in the NGVS project is being assessed.

THE CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG (CDS)
STRATEGY

The main prescriber of CDS strategy for the coming years will of course be to continue its mission at the highest level. The CDS is a French data centre of reference, which serves the international community. It will continue to develop its leadership role on added-value services, tools, and interoperability.

The starting points to define the strategy are the CDS mission to develop its services, and the domains of expertise in which it can bring decisive advances. The analysis of constraints and opportunities includes the different elements of the context of CDS activities, in particular the evolution of the discipline and of technology. For serving at best the astronomical community, the CDS has to evaluate, on the one hand, the relevant aspects of the evolution of astronomy – new topics, new telescopes and instruments – and of the national, European and international contexts. Because it is supported mainly on national funds, the strategic priorities of French astronomy have to be taken into account[8] (for instance, as explained above, the starting point of the development of the Healpix mode for Aladin was the strong involvement of the French community in the Planck project, which uses this sky pixellisation). The European and international contexts are of course other primary drivers, for the development of new telescopes and instruments, which provide new kinds of data, and also for the development of the astronomical Virtual Observatory. On the other hand, technologies relevant to the CDS activities evolve very quickly, and the new possibilities have to be assessed and eventually implemented to fulfil users' expectations.

New constraints and opportunities can also appear from the rapid emergence of the field of *Scientific Digital Repositories*, of which CDS has been a very early precursor. Among the elements to be taken into account, are the more and more widespread requirement that *data produced on public funds should in general be made publicly available*, and the discussions and possible requisite around the notion of *certification of data repositories*.

1. CDS core mission and strategy

The core mission of CDS is to continue to develop its services in their different aspects: content, functionalities and operations. It has also to use at best and to continue to develop its expertise as a reference data centre, for which quality and sustainability are main keywords. This implies the ability and necessity to perform R&D programs driven by the CDS needs. Another important facets of CDS expertise are its long term role in the development of international standards, which has been well exemplified in recent years by its major participation in the international astronomical Virtual Observatory, and the fact that it is one of the major operators in the network of astronomical on-line information services, with observatory archives, journals, and other added-value services such as the ADS or NED.

The core strategy is to continue the services at the highest level and to face the increase in data volume and complexity. This implies requirements on the human resources, but not only. To maintain and improve the quality of content, we have to continue the internal organisation efforts on the sharing of expertise among the staff (documentalistes, software engineers and astronomers) and on the definition and update of procedures. The *Bibliography* and *Infusion* meetings, which deal respectively with the content and computer/software aspects, will continue to be organised regularly. On-line

[8] The strategic priorities of French astronomy are defined taking fully into account the European priorities established by the Astronet and ESA exercises, but the regular national strategic exercise set up the French community view and its policies recommendations.

documentation will continue to be regularly enriched by the collaboration of all (Wiki-type documentation).

A peculiar effort will be devoted to progressive formalisation of the continuity of service and preservation aspects, with the medium term objective to candidate for a "data repository certification", if these certifications develop and finally prove to be do-able, affordable and useful for a data centre such as CDS.

2. Astronomy aspects

Among hot topics of the discipline which are relevant to CDS activities:

- the many very large surveys in operation or in project, and in particular their catalogues, from e.g. RAVE or Pan-STARRS (in which Observatory staff are involved) to the LSST, with a particular accent on Gaia, as one of the major projects of the coming years with a wide participation of the European and French communities;
- multi-wavelength astronomy, which continues to develop, with in particular in the field of CDS the provision of reference image data sets and the construction of Spectral Energy Distribution (SED), one of the IVOA scientific priorities;
- data cubes and spectro-polarimetry, with ESPADONS/NARVAL/SPIROU, MUSE, ALMA, LOFAR, …

For very large catalogues, their inclusion in VizieR brings interesting functionalities, but for the moment this is done case-by-case, implying an often lengthy development of a specific procedure for each catalogue. One CDS objective is to facilitate the inclusion of very large catalogues in VizieR, and also to facilitate their scientific usage. One relevant tool is the *CDS Cross-match service*, which will locally perform or provide pre-computed results of cross-match, in particular of very large catalogues. This new development, which uses the local expertise on access to and cross-identification of very large catalogues, will soon be made available as a beta-version. It will be developed further, in particular to investigate the possibility of taking into account more criteria on the physical characteristics of the sources into the cross-match.

For images, the inclusion in the CDS servers of all-sky, wide and deep surveys will be performed using the HEALPIX mode. For the building of SEDs, one aspect of CDS contribution is the update of VizieR metadata to make it able to provide normalized photometry data. This is particularly important since VizieR is a reference service for tabular information. Additional metadata have been defined and implemented in several hundred catalogues (which has been a huge work). This will be made available on-line to our users and to other services as a VO-compliant functionality. On the other hand, CDS will continue to participate actively in the definition of interoperability standards for photometry.

For which concerns data cubes and polarimetry, they can already be visualized in Aladin. More advanced functionalities could be provided through plug-ins provided by specialized projects, and work has to go on for interoperability at the IVOA level.

A specific case has to be made on the strategy of CDS with respect to Gaia. The scientific exploitation of the mission is a national (and European) priority. How could CDS help? The answer should be built on two relevant CDS expertises: one is the added-value distribution of very large catalogues in a very visible and widely used service, VizieR; the other one is the mastering of Healpix. In practice, CDS is, on the one hand, participating in the preparatory activities for the Gaia archive access (Gaia Archive Preparation – GAP – group recently set up by ESA), and, on the other hand, developing the usage of Healpix for the management of catalogue data. The starting point for a possible participation in the project is the distribution of the public version(s) of the Gaia catalogue, which can be seen as a win-win strategy for the mission because of the

excellent visibility of VizieR. Some support from CNES could be needed, but only on Gaia specific aspects since the basic platform will be already available.

3. Technology aspects

Since its creation in 1972, CDS has gone through several technological evolutions to the advantage of its users. The current challenge is the rapid societal evolution driven by the Web 2.0/Web 3.0 techniques, especially, but not only, in the younger generation, far beyond the usual fandom of new technologies ("tecchies"). The user-centric approach of the Web 2.0, which considers the user as a co-developer, is increasingly combined in the Web 3.0 approach with the usage of the semantics web, mobility, and universality. In addition, the current rapid development of multi-touch, mobile interfaces is another inescapable challenge that CDS must rapidly face: users want to find in the work environment the functionalities they use in their everyday life.

CDS has already tested a few evolutions of its services with a Web 2.0 flavour, with the successful implementation of Web 2.0 functionalities such as user annotations, "mashup" (the CDS Portal), and user space. It has also tested the implementation of mobile/multitouch interfaces (http://cdsportal.u-strasbg.fr/mobile). The goal is now to assess, to prototype, and, if this is successful, to implement, the migration of the whole CDS services towards a "CDS 2.0/3.0", keeping the solid ground of the high quality value-added services and tools, but moving towards a more user-centric paradigm. This involves a move to modular, flexible, sharable user interfaces; usage of the recent advances in semantic technologies to allow for instance "fuzzy" queries close to the user's natural language, to identify links between different contents or to exploit usage statistics; more intuitive human-machine interaction, taking advantage of the new devices such as smartphones, tablet computers, multitouch screens; possible usage of the user space to keep track of a customized user profile. The integration of these new concepts into all the services would be a deep evolution of the CDS, allowing a high level of flexibility and customization, richer links between and combination of information, and a more intuitive usage.

4. CDS and the VO

As explained earlier, one important milestone for the VO is the recognition by the Astronet European strategy exercise that it is one of the Research Infrastructures of Astronomy. This, and also the start of the US Virtual Astronomical Observatory (VAO) project, validates the CDS strategic choice to participate fully in the VO development.

The VO is now in operational phase, with the essential building blocks in terms of interoperability standards ("VO framework") set up by the IVOA. The activities to be performed are of different kinds:

- Implementation of the VO framework by the data centres;
- Continuous updates of the standards to take into account the evolution of astronomy and to allow more advanced functionalities (e.g., as described above, SED building from distributed sources of information, seamless usage of data cubes, etc);
- Support to users;
- Outreach towards education and the general public.

This has local, national, European and international aspects. Concerning more specifically the CDS:

- CDS services, SIMBAD, VizieR, and Aladin, are major building blocks of the VO: SIMBAD and VizieR as information providers, and Aladin as a portal for the visualisation of images, seamlessly interoperable with other VO tools. The services are of course already available in the Virtual Observatory, but the implementation of ALL relevant interoperability standards is a priority.

- CDS will continue to participate very actively in the IVOA in its expertise domains. This activity is expected to be less demanding than during the previous period, since the basic standards are now defined, but it has to continue to maintain the pertinence of the VO with respect to the users needs.
- Dissemination of the usage of the VO in the community is now the most critical aspect for the VO. CDS will continue to disseminate the knowledge of the VO in the national community, and to explore different types of actions to find the best method to do so. European activities are currently used as a template.
- Outreach towards education and the general public: CDS has taken its share of the European activities (Aladin outreach version, translation of user cases in French) led by our colleagues from INAF in EuroVO-AIDA, but it is not easy to find the appropriate framework for these activities at the local and national level. Here also the critical mass for developments is rather at the European level.

As noted above the implementation of the Astronet recommendations in that domain is not straightforward, although they were endorsed by the funding agencies. There is no sustainability of the European framework, and some of the major European partners concentrate on their core tasks. CDS will continue to explore all possibilities to get European funding. One important need is, among other tasks, to assess the different elements of Euro-VO sustainability, since this is a critical level in particular for the coordination of European technical activities, for dissemination and for outreach.

5. Dissemination, education and outreach

CDS will continue to do its best to disseminate knowledge about its services and the VO among the astronomical community, especially, but not only, at the national level, and continue, as explained above, to explore all possibilities to maintain funding at European level for this kind of activities.

One specific difficulty is with outreach beyond the scientific community, towards education and society (including the amateur astronomers), which requires resources which are not available at present. The possibility to involve partners at the local level (University), at the national or international level, will be explored.

6. The new importance given to scientific data

Astronomy has been at the forefront for the sharing and widespread re-use of scientific data. CDS has been an early precursor and is one of the important actors of this endeavour. One question to be assessed in the coming years is whether the CDS has new roles to play linked with respect to the Agencies' objective to enforce that *data produced on public funds should in general be made publicly available*. One of the key issues is that the physical storage of data is not the only question to address, because, to be re-usable, data must be properly described and easily retrievable. CDS relevant expertise is the management of tables and, more and more, of "additional data" (spectra, time series, models, etc) linked to publications. A possible new dimension of its service to the community could be in the direction of distributing more systematically these additional data. There are many open questions (perimeter of possible CDS action, funding, sharing of tasks with other actors, etc), but the landscape is moving very fast, and a detailed assessment of this possible major evolution is needed.

CHALLENGES

Work Force issues

The core need is to maintain the CDS services, during a period which will see the retirements of several key staff, and taking into account the increase in the volume of information CDS has to deal with. CDS has had for many years a resolute programme of actions for automation and appraisal to deal with the work load and to concentrate on added value tasks (e.g. in the recent period DJIN and the selection of tabular data to include in SIMBAD). It is not possible now to go beyond without loosing quality. There is no margin to deal with departures, with the increase in volume or with new tasks such as the possible extension of CDS role for scientific data. Difficulties expected with retirements and departures have also been anticipated, by duplicating the expertise on the software aspects of the services, and by evolving the software to facilitate maintenability. A new repartition of tasks and responsibilities has been worked out in the recent years to deal with retirement of several key staff (documentalistes and astronomer) in the bibliography domain; this will have to be done again for the departures to come.

It is important to understand that the very specific CDS expertise is built on a basis of permanent positions for the core tasks (documentalistes, software engineers and astronomers), with a long formation and transmission of expertise. The profiles are specific and rare, with people able to keep the motivation for serving the community on the long term. All departures of research and technical staff have to be replaced – retirements can somehow be foreseen (although the precise retirement age is a personal decision), but unexpected departures alter the planning of replacements. The difficulty is increased because the staff is on different statuses (CNRS and University), and for the CNAP astronomers because the replacement of one astronomer with 100% functional duties requires three positions with the statutory 30% functional duties. In addition one astronomer and one documentaliste work in other laboratories (E. Davoust, IRAP; G. Chassagnard, IAP). Moreover, the increase in data volume requires a small increase in manpower. New tasks (astronomical data curation, outreach) have to come with new human resources. Specific tasks, e.g. linked to European and other projects, will continue to be dealt with on temporary contracts.

The new French research system

The recognition of CDS as a *Très Grande Infrastructure de Recherche* by the MESR has been an important result of the previous period. The CDS has been from the beginning a joint endeavour of INSU and the University. With the new organisation of the French system the University autonomy is now the rule. CDS is by essence working in a national/international context, and it has been very positive that it has been recognized as an *Action Structurante* by the University of Strasbourg. However, the situation is much more difficult for human resources, and several key staff who will retire soon are BIATOSS. Also, the new funding structures are not well adapted to the CDS case, which means that we will continue to rely on CDS, UdS and CNES for its basic funding.

The local situation

The will of INSU, to move towards a grouping of all INSU laboratories into a single local OSU structure, has been taken very seriously and its different aspects have been assessed. The two current OSUs are organised very differently, and one of the critical points for CDS is that it is, and has to remain, fully embedded in the research environment of the Astronomical UMR. Because of its international role, there is an absolute necessity, if the move towards a single OSU in Strasbourg is concretized, that the CDS, its organisation, and its capacity to get recruitment and to manage its resources, is preserved.