CDS Scientific Council meeting 2014
Summary of CDS activities 2013-2014

3 September 2014

TABLE OF CONTENT

HIGHLIGHTS DURING THE PERIOD

Although less than one year passed since the last meeting of the CDS Scientific Council, significant progress has been made in different aspects of the CDS activities, some visible from the users, others pertaining to our internal procedures or linked to our collaborations. Among the highlights:

CDS Services

- The new tool developed for the ingestion of object lists in SIMBAD, COSIM (Comparisons of Objects for SIMBAD), reaches the exploitation phase. It is replacing the "raccord" tool developed by François Ochsenbein many years ago as the basic work tool for semi-automated ingestion of tables in the database. These tools compare the characteristics of the objects in the list with those of the objects already present in the database.
- The assessment of improved access to data other than tables in VizieR has progressed significantly, with an agreement with the AAS on a common description framework, based on the ObsCore VO standard[1] initially developed as the core components of the Observation Data Model necessary to perform data discovery when querying data centres for observations of interest. A prototype for image data is developed, in collaboration with the High Energy Astrophysics team of Strasbourg Observatory.
- A new version of Aladin, Aladin 8, was released in March 2014[2]. It implements, among other new functionalities, the Multi-Order Coverage maps (MOC), the Hierarchical Progressive Surveys HiPS and dynamical proper motions – an important feature in particular for Gaia.

[1] http://www.ivoa.net/documents/ObsCore/20111028/
[2] http://cdsweb.u-strasbg.fr/news.php?fn_mode=fullnews&fn_incl=0&fn_id=399

- Several widgets are now implemented in the services: in addition to VizieR photometry viewer released in November 2013 after a beta-release in October 2013, Aladin Lite is now implemented to visualize objects in SIMBAD and VizieR. Work is going on on a time series viewer.
- The database contents continue to grow fast, as well as the workload on the team, which was already under stress before. This is true for SIMBAD and VizieR, including ingestion of very large catalogues, but also for Aladin image database in which more and more image surveys are implemented.

Projects, Collaborations and R&D

- The excellent collaboration with Agencies continues, with e.g. with ESA on Herschel catalogues and with ESO with an agreement on the ingestion of the Public Surveys in VizieR. The usage of HiPS and Aladin is also progressing.
- CDS continues to participate actively in the Gaia project, and its collaboration with CoRot was re-activated. Contact has been taken with the French LSST team and our capacities raise their interest. We have to assess the R&Ds necessary for the scale-up of our services.
- Our participation in the Arches project is producing a leading edge cross-match software, allowing to compute probabilities for an arbitrary number of catalogues. The tool will be tested with the different scientific applications included in the project, and then its release in the CDS cross-match service is envisaged.[3]
- A mirror copy of VizieR was installed in South Africa at the SAAO/NRF[4], the first in Africa and in the Southern Hemisphere.
- Among the highlights from the recent R&D internships: 3D immersion for visualisation and support to interpretation of astronomical data, including simulations; "Arches Walker" kit developed for the Arches project; exploration of massive data.

For the CDS itself

- The CDS was recently awarded the Data Seal of Approval[5]. The DSA is granted to repositories that are committed to archiving and providing access to scholarly research in a sustainable way. As discussed at the Council meeting last year, it was timely for us in the current context of "open science" to seek formal certification granted by a respected external body to back up the trust in our services demonstrated every day by their high level of usage by the community.
- Excellent participation of the CDS team in the LISA VII Conference[6], covering different aspects of the CDS activities. This was a unique occasion to present and discuss the bibliographic team work and accomplishments, as well as the recent developments of our services.
- Several presentations in different contexts demonstrated again that there is lots of interest for the way astronomy deals with its data, including in particular for the way the CDS deals with "long tail" data in VizieR.
- The computer servers were renewed, and a homogenized back-up system was implemented.

[3] http://www.arches-fp7.eu/
[4] http://viziersaao.chpc.ac.za/vizier/
[5] http://cds.u-strasbg.fr/DSA-CDS
[6] http://eventi.oacn.inaf.it/lisa7/

CDS STRATEGY

The high level strategy drivers presented at the 2011 Scientific Council 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, continuing to update the VO framework and to disseminate the VO knowledge in the astronomical community, and looking for 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 2013 Scientific Council recommendations, for those which were among these recommendations, and in the following section which summarizes CDS activities since the 2013 meeting.

STATUS OF SC 2013 RECOMMENDATIONS

High level recommendations

> **Services:** (1) The core services, SIMBAD, VizieR and ALADIN are the flagship products of CDS. They are extensively used by the astronomical community, and need to be operated and maintained at the highest level with high priority. (2) The services need to be upgraded periodically through research and development, taking into account evolving community needs and feedback. (3) CDS services rely on the provision of high quality, scientifically relevant content. Clearly defined interfaces with content providers, such as large Observatories, are required to draw a roadmap towards the collection of high profile pixel-data from the ongoing and planned large surveys (4) The balance of resources between operations, maintenance and development for these services may have to be reviewed periodically and some basic monitoring of efforts in the different areas is recommended.

Recommendations (1) and (2) are a welcome endorsement of the two first strategic drivers of the CDS, the key elements that empower it to realize its core task, the provision of high quality reference value-added services to the international astronomical community.

Recommendation (3) has two aspects. The building of the high quality, scientifically relevant content of the CDS services relies on the one hand on the work of the CDS team

to extract information from peer-reviewed, and thus scientifically validated, publications, and on our long term collaboration with the academic journals. On the other hand, collaboration with large Observatories and surveys can enrich the CDS data collection, as well as it also leads to the usage of CDS products in services developed by high profile agencies and projects. CDS reputation is now well established for the distribution of catalogue data, and collaborations are set up with ESO for the distribution of the large survey catalogues and with Gaia (CDS is a member of Gaia CU9 and will distribute the mission catalogues). Long term collaboration has been on-going with ESA, with recently for instance the distribution of Planck catalogues in VizieR, and collaborative work for the provision of Herschel catalogues. Positive preliminary discussions have been held in January 2014 with Pan-STARRS and in June 2014 with the French participants in the LSST project, to begin to set the scene for future collaborations.

The power of the HEALPix sky tessellation implemented in HIPs is also better and better recognized, leading to a significant increase in the number of sky surveys stored in Aladin image database[7], for instance the AllWISE HiPS survey[8], and to a more and more widespread interest. This together with the easy implementation of Aladin Lite in a web page leads to additional usage of Aladin as a visualization tool in services outside CDS. One can cite long time partners such as ESA, or STScI for the HST and JWST Astronomer's Proposal Tool, but also LIGO (and hence VIRGO which uses the same tools). CDS is continuing discussions with other partners such as ESO to assess the role it can play in the distribution of images in complement to the Agency one. The new capacity to link images to their progenitors is also a good incentive, since it builds a powerful link to access original images from the original archive or their metadata. We are also working on providing access to data cubes, whereas providing advanced access methods to data cubes is also a current IVOA priority.

Another aspect of data collection is the dissemination of scientifically validated "long tail" research results attached to publications. As shown to the previous SC meetings CDS already hosts non-tabular data in VizieR, and they make a significant fraction (~15%) of the VizieR collection. The point is now, in line with the strong recommendation at the political level that "data obtained on public funds should be publicly usable (except in case of legitimate exceptions)", to make the distribution of data attached to the articles easier and more systematic. *Astronomy & Astrophysics* is fully in line with this strategy, and as explained at the 2013 SC meeting we aim at reaching an agreement with other key players, in particular the AAS journals (ApJ, AJ), on the metadata and access protocols (a similar approach was used around 1973 for the description of tabular data, first established by CDS and then shared with other data centres and the journals). A meeting was organized by the AAS during the January 2014 AAS meeting to discuss this topic, and the agreement was to assess the possibility to use the VO ObsCore data description. This is currently being assessed at CDS, in collaboration with Laurent Michel from the High Energy / XMM-Newton SSC team, using its SAADA[9] VO-enabled database system. The role of CDS in the long-term data curation is not only a technical problem, and the fact that we have been awarded the Data Seal of Approval is a proof, given by an external, respected Certification authority, that we can be trusted to fulfil this role.

An evaluation of the respective weight of the different CDS activities will be presented to the Council (Recommendation 4).

[7] http://cdsweb.u-strasbg.fr/news.php?fn_mode=fullnews&fn_incl=0&fn_id=395,
[8] http://cdsweb.u-strasbg.fr/news.php?fn_mode=fullnews&fn_incl=0&fn_id=404
[9] http://saada.unistra.fr/saada/

> **CDS and the VO:** CDS should continue to play a leading role in the VO, at the developmental level as well as in encouraging the scientific use of the VO products and educating the community about their potential. This is particularly important at the present time as the VO is moving to an operational phase, and has to be fully exposed to the astronomical community. The role of CDS has become even more important than before, as some of the major VO projects face the prospect of greatly reduced funding. To sustain VO activity over the long term, CDS should continue to seek funds through European and international programmes.

CDS continued to play a leading role in the IVOA activities and structure. CDS staff currently chairs three IVOA Working Groups, *Applications*, *Data Access Layer* and *Grid and Web Services*, one chairs the *Data Curation & Preservation* Interest Group, and an associate member of the CDS is vice-chair of the *Semantics* Working Group. The chairs of the *Standing Committee on Science Priorities* and of the *Standing Committee on Standards & Processes* also belong to the CDS. CDS staff authored two of the standards accepted as IVOA Recommendations in 2014, *MOC – HealPix Multi-Order Coverage Map Version 1.0* and *Units in the VO Version 1.0*.

CDS also continues to be as active as resource permits in encouraging the scientific use of VO products and educating the community about their potential. Since the last SC meeting, relevant events have been: a strong participation in the 1st Gutenberg Astrophysics Workshop and Winter School[10] organized in Strasbourg in January 2014, with presentations on *Science with the Astrophysical Virtual Observatory*, *Technical aspects of the VO*, *Metadata in Astronomy: VizieR example*, and *Getting the most out of SIMBAD and VizieR* (plus a talk on *Modelling dust in Local Group galaxies*); a VO Module at the Ile-de-France Doctoral School in March 2014; a VO tutorial in Paris Observatory in May 2014 *Aladin and the CDS services*; the French VO (ASOV) presentation at the Annual Meeting of the French Astronomy & Astrophysics Society in June 2014 (presenting the tutorial on VO usage in high energy astrophysics); and outside of France the participation in a workshop *Virtual Observatory Tools and their applications*[11] in Krakow in June 2014 (which was oriented towards VO usage through the CDS services), a seminar *The CDS, the VO and the whole sky* at Manchester University in February 2014, and a seminar *Insightful usage of CDS tools* at the STScI in May 2014. One can also cite, although it was not aimed specifically at the VO but rather at our services, CDS participation in the 3rd school *Scientific Writing for Young Astronomers* organised by Astronomy & Astrophysics in August 2014 in Tihany (Hungary).

A key evolution of the VO is worth mentioning here: it is now better understood, in particular thanks to IVOA Committee for Science Priorities (led by M. Allen) and to Euro-VO activities, that collaboration with data providers, especially the large current and future projects which shape the astronomical landscape, is mandatory to gather their requirements and make sure that the VO framework is relevant to their needs, so that the VO is populated with the data astronomers need for their research. Another aim is to create VO knowledge, and the capacity to participate in the development of the framework, inside the large projects, which would be an essential step for VO medium/long term sustainability. This is the core plan for the first Euro-VO Horizon 2020 proposal, which is included in a wider overarching project, ASTERICS, aiming at clustering ESFRI infrastructures.

In parallel, the evolution of European programmes is towards "e-Infrastructure Commons", a common framework for the European network, computing and data

[10] http://gawws.u-strasbg.fr/
[11] http://www.ncac.torun.pl/?lang=pl&branch=world&id=36

infrastructures. This supports CDS strategy to be involved in the Research Data Alliance[12] (RDA) at the European and international level. The RDA is a recent but high profile organisation which aims at defining building blocks of the data infrastructure facilitating data sharing.

> **Research and Development:** The Council recommends continuing extensive, varied, versatile R&D activities along the current lines, including addressing big data, information retrieval, multi-use software components and continuing with Aladin Lite as a pathfinder for the evolution of services. We also encourage CDS to continue to perform R&D in a costless way as part of the services.

R&D activities continue to be varied and versatile. A list of the internships overseen by CDS staff since the last Council meeting is provided among the SC documents. They cover a large range of aspects of interest for the CDS, such as usage of virtual reality for visualizing astronomical data and simulations (with possible applications to education and outreach), logs to supervise the service operations, and also inputs to the CDS strategic axes, such as the usage of big data technologies for very large surveys or how to deal better with data attached to publications, which is as explained CDS contribution to the dissemination of "long tail" research data.

> **Scientific Activities of the CDS**: The members of the committee are impressed by the efforts to ensure coherence between the high quality of the scientific research conducted at CDS and the relevance for data and services, despite the heavy load of service duties. The close collaboration with communities in various astrophysical disciplines is also highly appreciated. The members of the committee appreciate the recent increase in number of post-docs and Ph.D. students supervised by CDS staff members. To increase the number of students CDS is encouraged to collaborate closely with the University of Strasbourg. The committee recommends setting up a visiting program to CDS for durations of up to several months in cooperation with the visitors programs of the University of Strasbourg. The committee feels that this would lead to increase in the volume of scientific work and maintain and develop the scientific motivation within an institute where public service is the main pillar.

CDS staff continued to look actively for all opportunities to get support for PhDs or post-docs, through projects as well as by answering to Calls organised by the University of Strasbourg. The following PhD and post-doctoral contracts are on-going, or will begin soon:

PhDs

François Nehlig (Oct. 2012 – Sept. 2014): Galaxy evolution in the Virgo cluster (with Bernd Vollmer)

Maxime Beuret (Oct. 2013 – Sept. 2016): Stellar formation, interactions with the ISM (with Laurent Cambrésy)

Jérémy Chastenet (Oct. 2014 – Sept. 2016): Bayesian analysis of dust grain properties variations in resolved nearby galaxies (with Caroline Bot and Karl Gordon, STScI)

[12] http://rd-alliance.org/

Post doc

David Eden (July 2014 – October 2016): 3D extinction in the Galaxy (with Laurent Cambrésy, on the ViaLactea project)

With respect to setting up a visiting program, the first indispensable step is now done: a new astronomer was recruited for CDS, Ada Nebot, who just begun to work in the team on September 1st. The Council input is welcome on how to organize such a program.

> **Staffing:** The CDS has always been successful in the past years in its efforts to do "more with no more people". However, the Council feels that the tension induced by this trend on the CDS, in the context of growing amounts of data, has reached a limit and has induced a critical risk of loss of competitiveness and of services – which the Council considers unacceptable. Therefore, the Council strongly supports the requests for permanent positions made to CNRS and to Strasbourg University in order to compensate for two upcoming retirements of experienced people. It also recommends increasing the number of CDS staff beyond its current level, and in particular encourages the CDS to search for funding beyond its main sources, for example in LABEX structures, in European research programs or in the regional organizations.

This is a well taken position. CDS got the recruitment of a new scientist in the "Astronomers" body, Ada Nebot, in the framework of the annual competition for these positions. CDS and the other data structures were among the priorities set up by INSU (http://www.cnap.obspm.fr/Textes/ColoriagesCNAPAA2014.pdf). Ada Nebot is a specialist of high energy astrophysics, large survey and time domain, and she will bring a very welcome expertise complementary of the current one. She will also work in support to the visiting programme recommended by the Council.

The situation is unfortunately not so good for the technical positions, and the situation there is more and more critical. The CNRS "documentaliste" position lost by retirement at IAP in 2013 (G. Chassagnard, who retired in June 2013) has not been replaced by INSU. The situation of the "documentalistes" team is really worrying, the volume of data to ingest is continuing to grow whereas the manpower is decreasing, and we have "to do more with less people" rather than "doing more with no more people". Another aspect of the CDS content building is the constitution of the HiPS database (see below) which is for the moment ensured by the software engineers working on Aladin, which is an issue since more and more data is ingested.

We are very grateful to the University of Strasbourg, which opened a software engineer position to replace Marc Wenger (this follows the peculiar rules of University technical positions). The University also opened a temporary contract for the replacement of Chantal Bruneau, CDS Secretary, whose retirement happens too late in the year to get the opening of a permanent position.

Other funding sources cannot fund the permanent positions on core activities which are the key for a sustainable CDS. They come in support in particular to the developments of VO standards and tools or of particular functionalities which are then maintained by the permanent staff. The current round of European proposals is Horizon 2020, with a set of calls which closed on September 2nd. CDS participated in two projects in astronomy, the astronomy "cluster" project ASTERICS, led by M. Garret (ASTRON), which has a work package in support to the VO take-up by the ESFRI projects led by CDS on behalf of the Euro-VO, and a minor participation in a Europlanet proposal to support Aladin usage in planetary studies. In addition, it has minor participations in two projects which go beyond

astronomy and deal with the general scientific data infrastructure, the third RDA Europe project to cover F. Genova participation in RDA, and a small participation in a project aiming at building a "certification framework" for Europe.

> **Partnerships:** The Council lauds CDS for its broad participation in high profile projects. Such participation is beneficial for CDS and should be continued. However, the Council cautions CDS to proceed with care so as not to over stress its staff with the risk of distracting them from core CDS activities.

We agree with the Council about the caution to exert with project participation. We always try to align projects with the CDS strategy. Some projects are linked to the scientific subjects of the CDS astronomers. They enter in the fraction of their time devoted to their own scientific research, with no adverse consequence on the time they devote to the CDS but very positive ones on their scientific activities. This enters in our strategy to encourage high level scientific activities from CDS staff members, which is also essential for the quality and relevance of the CDS services. Some of these projects will also bring the possibility to hire post-doctoral contractors, and/or to distribute high quality result data through the CDS services, which is also beneficial. Other projects, such as the ones linked to the VO development, are closely linked to one of our main strategic axis. The Arches project is exemplar of the synergy which can be built between the CDS capacities and a project: F.-X. Pineau develops a leading edge cross-match software, allowing to compute probabilities for an arbitrary number of catalogues, which will be extensively tested during the project and will likely contribute to the CDS cross-match service.

> **Outreach:** As in recent years we recommend forming strategic partnerships both to strengthen the effectiveness of the CDS outreach and to leverage funding.

We take this recommendation very seriously, but it has not been possible to build such a partnership during the last period. We have given the priority to find a host for the bulk of the Euro-VO activities, including those towards education and the general public, and did not look for the development of partnership specific to CDS. We hope that the situation will be better during the coming years.

CDS participation in the Arches project has been an occasion to develop the *Arches Walker*, based on the use of both a desktop/laptop on the display side to host Aladin and a mobile device (tablet or phablet) on the user side to interact with it through a user-friendly multi touch interface. The demo booth uses an information base gathering typical objects observed by the project with their characteristics. We also engaged discussion with the Jardin des Sciences (the University of Strasbourg outreach framework) and the planetarium to assess whether some of our developments could be of interest for them, in particular those which have both educative and ludic dimension, e.g. virtual reality.

The local activities towards secondary school teachers of the Académie de Strasbourg, supported by the Rectorat, continued, in collaboration with the Jardin des Sciences. A seminar was given in a Lycée on *Astrophysics: a profession?*, as well as a conference aimed at the general public at the Université de Bretagne Sud (Vannes).

> **Strategy:** CDS has developed a long-term strategy, which takes into account changing technology, the advent of large surveys and the changing needs of users. CDS has been successful in orienting its programmes to meet these strategic goals, largely due to the high calibre and dedication of its staff. Maintaining such an elevated level of performance in the future will require rapid replacement of retiring staff and assured funding.

Among the current R&D axes, the management of time data comes in support in particular of our participation in CoRoT and Gaia, and of our interest for the LSST. We are also interested in assessing whether "Big Data" techniques could be used to fulfill our needs, in particular because the methods we currently use may not be adequate for the significantly larger LSST data volume. In phase with the VO priorities, we improve the way we manage and give access to multi-D data. And we continue to develop modular elements for web interfaces, for ourselves to include them in our own interfaces, but also to make them usable by others.

We cannot agree more with the final recommendation. As explained above, the staffing situation is critical.

SUMMARY OF ACTIVITIES, NOVEMBER 2013 – AUGUST 2014

Complementary information can also be found in the two previous sections of this report, and in the companion documents.

The services

SIMBAD

September 1st, 2014, SIMBAD contained 7,556,225 objects, 18,563,653 identifiers, 294,449 bibliographic references, and 10,749,766 citations in papers (7 342 000 objects, 18 162 000 identifiers, 285 000 bibliographic references, 10 000 000 citations of objects in papers on October 1st, 2013, and 6 970 000 objects, 17 100 000 identifiers, 269 000 bibliographic references, 8 719 000 citations on July 1St, 2012).

There has been around 506 000 queries/day in average during the last year, a figure similar to the 520 000 queries/day measured during the first six months of 2013.

SIMBAD content: cross-match tool

The specific software ("raccord") which was used to cross-match lists of objects with SIMBAD taking into account as many parameters as possible to prepare object ingestion has been re-written as COSIM, in close collaboration between the engineer in charge, SIMBAD "documentalistes" and astronomers. The original code was using the C language, and the new one uses Java. It allows the user to choose the parameters to compare in a modular way instead of providing a mean score including all comparison parameters. Problems occurred more and more frequently, with fields containing more and more sources, punctual as well as extended, of various types, at different wavelengths, with more or less accurate coordinates. With COSIM, the documentaliste decides, according to the situation, which offset is acceptable for coordinates, velocity, magnitudes etc., and which object types are compatible. Rapidly and with a very low risk of errors, COSIM sorts objects into good and bad candidates, while bringing to light borderline candidates and possible merges among existing objects in SIMBAD, thus allowing the

documentalistes to concentrate their attention on "difficult" individual cases which require all their expertise.

SIMBAD content: specific operations

The daily update of the database of course goes on. Here only specific, significant programmes are listed. Some of these programmes are long term ones and require manpower, but they are undergone because they improve significantly the database quality and scientific relevance.

Quality flags for redshifts, and for parallaxes and proper motions

An important added-value data made available by SIMBAD is the quality of the measurements stored. Due to the huge variety of techniques used in astronomy for any set of observables, assigning the quality of the data is crucially important, and not always simple. When authors provide errors on a measurement, it is used to set the quality flag. However, often enough the authors will not give errors, and we must use the instrument setup description to assess a quality.

A special operation was performed in order to homogenize SIMBAD redshift qualities and set clear rules for the future. A survey of the existing redshift measurement techniques was performed, exemplified by some of the most prominent instruments in each wavelength, along with their typical accuracy. A table was built associating instruments, redshift uncertainties, and qualities ABCDE, which documentalistes refer to when inserting new data into SIMBAD. The document will be maintained so as to include new instruments/setups.

A similar special operation was performed on parallaxes and proper motions, especially in view of Gaia, based on the whole content of Simbad and instruments analysis. The main result is that the historical Simbad quality letter will be suppressed, replaced by a flag for high or poor data quality.

Stellar associations, moving groups, stellar streams

Objects types Stellar Stream and Moving Group have been added to Simbad. A literature analysis has been performed to clarify the definition of such objects compared to Clusters and Association. All such objects have been updated in Simbad accordingly.

Main identifiers of the most famous/brightest objects

Historically the main name of an object in Simbad was chosen by the Simbad update software as a function of its main object type, or following the alphabetic order. Thus it can change when a new object type or a new name is added. It is however a strong request of the users, and of the CDS team, that at least the name of most famous objects corresponds to the habits in the literature, and does not change. Two years ago the possibility to fix definitely the main name of an object was implemented. It was first applied to the most famous objects, e.g. LMC. This year the name of the nearly 4000 brightest stars (e.g. * bet Cep) has been definitely fixed.

Revision of the list of object types in SIMBAD

The list of Simbad object types has been for long available from the main Simbad web page[13]. It is used as a documentation for both the users and the CDS team. It is a hierarchical list, in fact the algorithm itself used in the update and Raccord/Cosim

[13] http://simbad.u-strasbg.fr/simbad/sim-display?data=otypes

software. However this linear hierarchical list can no longer take into account the complexity of relations between object types, and is not user-friendly as a documentation. A first preparatory step was done since the last CDS Scientific Council to improve this situation. A new reorganized list of Simbad object types was built with an astronomical point of view. It is much more user-friendly, both for the users and the CDS team. Short definitions are already available for quite a few types. This new list is not available yet to the users, but it is available for the CDS team in the internal wiki. A new system of priority and compatibility relations between the object types has been built based on the astronomical expertise of the whole team. The next step will be to implement it in the update and Raccord/Cosim software, followed by a test phase on real lists of objects. The second step is planned to start in October 2014 over a period of 6 to 9 months.

Other software/hardware evolutions

- The sorting of SIMBAD references for an object, released in September 2013, can now be parameterized by the user.
- The set of three servers has been updated to a single, more powerful one, which will be easier to maintain.
- Log management is being improved and automatized.

Future evolutions in preparation

- The improvement of the links between SIMBAD and VizieR remains a priority for the two services and for the CDS.
- The update of the web interface has been delayed but is a high priority for the coming year.

Marc Wenger's retirement

Of course, Marc Wenger's retirement after 41 years of faithful service at CDS, which will be effective before the end of September 2014, is a major milestone in SIMBAD's life. Anaïs Oberto has taken more and more responsibilities during the last months and continuity is assured. However, additional manpower will be needed to ensure some of the software developments, and a contractor has been hired for that purpose on a temporary contract. He will work in particular on the update of the web interface.

VizieR

September 1st, 2014, VizieR contained 12 691 catalogues (11 579 catalogues in October 2013, 10 360 in September 2012). Three very large catalogues were included since the last SC meeting, CMC15, AllWISE[14] and the Initial Gaia Source List (IGSL V3). AllWISE produced 22,000,000 queries since February this year.

There has been 530 000 queries/day during the last year, with a peak of 33 210 000 queries in March 2014, i.e. an average of one million queries/day.

A new mirror was set up in South Africa at the SAAO/NRF in April 2014, the first one in Africa and in the southern hemisphere.

"Additional data" in VizieR

As discussed in 2013, this is a critical aspect of VizieR in the Open Data/Open Science context, since it allows scientists to disseminate data resulting from their scientific projects, validated by a publication in a journal. CDS is in particular collaborating with *Astronomy & Astrophysics* in this domain, and also hosts data attached to papers

[14] http://cdsweb.u-strasbg.fr/news.php?fn_mode=fullnews&fn_incl=4&fn_id=392&cat=4

published in other journals. VizieR already includes a significant fraction of catalogues and tables with "additional data".

The increased importance of "additional" data in VizieR led us to assess direct publication of these data in the VO, so that they are found when someone looks for a spectrum or an image of a given object, and not only through the VizieR information (which appears as information from a catalogue) retrieved by a cone search of the VO. This means providing an image, spectra, time series service in addition to VizieR table service. This will require a fine grain indexing of data in the database (indexing of individual data and not only of full catalogues), and also gathering additional metadata, which can be a huge, difficult, or even impossible work, since many files are involved, spanning a large range of scientific fields and observing techniques.

Since the SC meeting last year, good progress has been made, as explained above. A prototype VO-enabled image server is being developed based on SAADA. A R&D is performed by Laurent Michel, SAADA developer, to extract as many ObsCore metadata as possible from the FITS headers. Another aspect will be to prepare a web form allowing users to provide the metadata when they drop the data. One key problem is that we can extract information e.g. from the data FITS headers, or try to find it in the publication, but the workload related to deep checks is not affordable. The authors are (hopefully) the best source for attaching exact metadata to the data, which is one of the quality challenges of the endeavour. Also, the implementation of plots allowing users to visualize the different data types is being assessed and revised.

Data Seal of Approval

VizieR team and the Observatory system engineer have been at the forefront for the huge amount of work which was required to be awarded the Data Seal of Approval. In particular, implicit procedures have been made explicit, and we demonstrated (also to ourselves, since the DSA begins by a self-assessment!) that we follow an OAIS-like model in the management of data. The DSA is as explained a powerful external support to our data preservation activity, and the work devoted to obtaining it is an integral part of the CDS strategy to provide more systematically "additional data" in VizieR.

Other software/hardware evolutions

- A new interface for data submission is being developed. It provides a pre-filled ReadMe to facilitate the author's job, a frequent request from our providers.
- Implementation of Aladin Lite in VizieR result pages.
- Replacement of the VizieR machine and extension of disk space. VizieR data is now under the homogenized back-up system.

François Ochsenbein's retirement

François Ochsenbein's retirement took place in September last year and continuity has been ensured. He is still present on an emeritus status but does not have operational responsibilities any more.

Aladin

Usage statistics are tricky because a fraction of usage moved from Aladin Java to Aladin Lite, and the HEALPix (HiPS) database is more and more used whereas the comparison with the "traditional" usage is not easy. Also, users can use Aladin without accessing our database and we cannot track this usage.

Nevertheless, the figures are:

- 2 200 sessions/day (+800 since last year, -300 for Aladin Java, +1 100 for Aladin Lite)
- 52 500 HiPS diamonds/day (+23 700, +15 000 for Aladin Java, +8 700 for Aladin Lite)
- "Traditional" queries/day: 10 500 (-10 000 for Aladin Java, N/A for Aladin Lite)

Clearly the usage of the HEALPix images combined with Aladin Java and Aladin Lite is making a breakthrough. This is also clear in our relations with our partners, e.g. with ESAC or with the STScI.

In August 2014, Aladin contains 175 surveys for a volume of 45 TB (128 surveys with a volume of 30 TB in August 2013, 81 surveys for 19 TB in August 2012).

Aladin 8

The new version of Aladin, Aladin 8, was launched smoothly in March 2014. The Multi Order Coverage maps or MOC functionalities clearly had the highest impact on the users, but the new proper motion management e.g. was also well received – it is important in particular in the Gaia context. Aladin 8 also provides a pre-visualization of VizieR photometry, with a link to VizieR photometry viewer.

R&D

R&D on the usage of HiPS for catalogues, cubes and access to progenitor data, datalinks and metadata, and the exploitation of the MultiOrder Coverage maps were very well received both by partners and by users. They will continue during the coming year.

Other software/hardware evolutions

- Significant effort on hardware including backup (2 x 100 TB in two different buildings)
- Significant effort in software development for the HiPS generator (images, cubes and catalogues)

The cross-match service

With an average of 16 jobs per day, the activity on the cross-match service increases slightly (from 12.5 jobs/day during the six first months of 2013, and 10 jobs/day in 2012) while the number of different users is stable. In the last 8 months, 8 billion positions have been submitted to the cross-match API.

New VizieR tables keep being integrated at the same pace as they are ingested in VizieR: tables smaller than 10 million rows are automatically integrated, integration of larger catalogues is part of the large catalogues pipeline.

Access to VizieR data and cross-match

One of the milestone for 2014 is that TopCat and STILTS, popular tools to manipulate tables, now give (since last July) an easy access to the CDS cross-match API allowing one to submit lists of positions to find potential matches in Simbad or any VizieR table with positions. This new access method replaces and improves the deprecated multi-cone search of individual objects in VizieR or SIMBAD. Submission from Topcat of a list of one million objects could previously take several hours and seriously overload the VizieR server whereas the new method performs the cross-match in a few minutes, significantly reducing the server load on our services which were sometimes overloaded and blocked.

One collateral consequence is that the number of queries on SIMBAD and VizieR will likely decrease, eventually significantly. But this is a GOOD thing in terms of rationalizing the service usage. All the Big Data actors are aiming to move to "smart queries" to avoid overburdening their systems. Let's say that in collaboration with TopCat, and using the capacities offered by the VO, CDS is also implementing "smart queries" to serve its users better. But this raises (once again) the question of how to measure the impact of a service. Citations in papers are not relevant since the VO provides seamless access to data, and the number of tools and data services is high. Statistics on access are a way to demonstrate our impact. But this is not a reason to keep outdated and counterproductive access methods when cleverer ones are possible.

Meanwhile, we are investigating new ways of measuring the usage of the services by the community. Currently, the number of requests is easy to gather but difficult to interpret since it does not differentiate between large and small queries. In the future, the new metrics should reflect the amount of data being requested or accessed.

Other aspects of the cross-match API

Access to the cross-match API has been integrated in the Astroquery package of the popular AstroPy[15] Python library.

Cross-match in the Arches project

CDS is taking the opportunity of its participation in the Arches project to assess new, more advanced cross-identification methods. The project aims at providing a multi-catalogue cross-match framework. The challenge was two-fold: first, existing statistical methods had to be generalized to handle more than two catalogues, then a flexible, highly configurable and still efficient tool had to be developed. E.g., taking into account extended objects and proper-motions required to study and use new indexing data structures. The new method will be presented as an oral contribution to ADASS 2014. It is being tested on the scientific cases developed in Arches, and will be included in the CDS cross-match service if the assessment is positive.

R&D

A list of the R&D programmes performed through internships is provided in a separate document. Studies linked to attached data in VizieR were addressed above. Among the other topics, three are selected to appear here.

MOC in SIMBAD

A MOC HEALPix sky tessellation was implemented for SIMBAD. Accesses by a web page and scripts are being developed, allowing on-the-fly generation of MOCs showing SIMBAD sky coverage with objects selected following criteria defined by the user. This will also be available through VO TAP queries. The next step will be the interface with Aladin.

New interfaces and new interactions

To continue our investigation of new interfaces and new interactions, we assess the possibilities offered by virtual reality through the usage of Oculus Rift. This kind of tool will likely democratize in the very short term and offer usage perspectives which go well beyond games. The work includes technical aspects (use of OpenGL, optimization) and also looking for use cases (visualisation of and navigation in data, simulations, etc). This work led to collaboration with the IMCCE to use the Oculus with the Skybot 3D software[16],

[15] http://www.astropy.org/
[16] http://vo.imcce.fr/webservices/skybot3d/

their VO portal to the solar system. The results will be presented in an oral contribution to ADASS 2014.

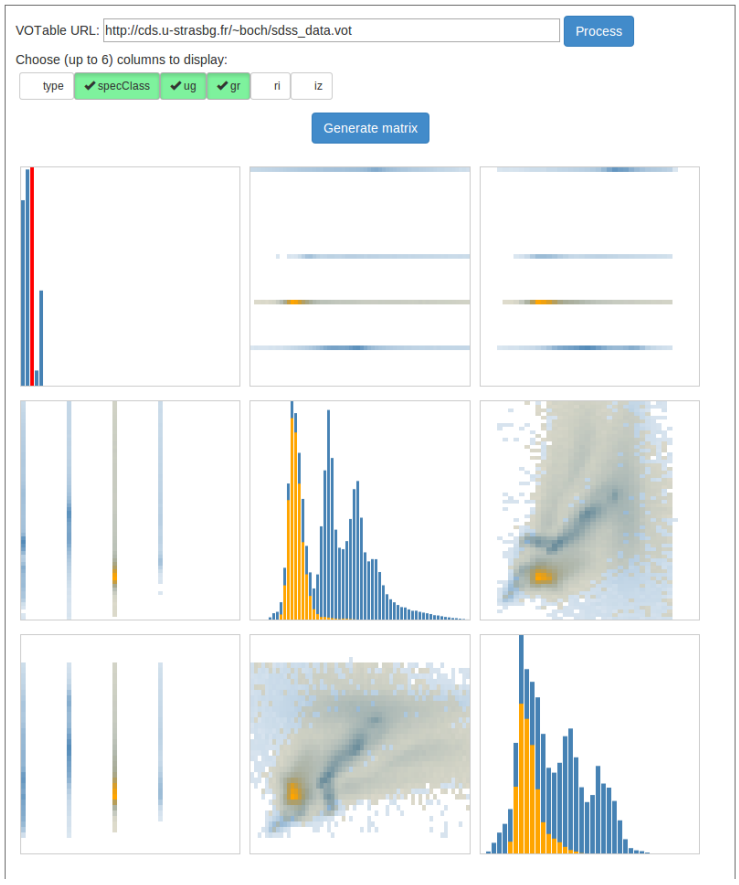
Big Data exploration

We made an assessment of various techniques to easily browse, explore and visualize large tabular datasets.

1. On the client side, we developed a Web application allowing one to create a SPLOM (Scatter Plot Matrix) from any VOTable. The different views of the matrix are linked allowing for easy exploration of the parameter space (cf. screenshot).

2. For larger datasets, we developed our own version of the Nanocubes[17] data structure, allowing fast interactive visualization of a catalogue with hundred million rows for a few attributes. HiPS heatmaps are dynamically created server-side according to user-chosen criteria and visualized in Aladin Lite.

This preliminary study proved promising and we plan to will continue this effort in the frame of the large catalogues.



Virtual Observatory

Implementation of CDS services in the VO

HiPS and MOC have been implemented in Aladin. The progenitor functionality in Aladin makes use in some cases of the Data Link VO standard, which is currently in the Proposed Recommendation status – this has been tested with the CFHT-LS.

[17] http://www.nanocubes.net/

Update of the VO framework

The list of new IVOA standards developed with CDS participation is given in a companion document. As explained above, two standards authored by CDS members reached the IVOA Recommendation status this year, namely *MOC - HEALPix Multi-Order Coverage Maps Version 1.0* and *Units in the VO Version 1.0*.

As explained in 2013, MOC is an important standard which defines the footprint of a catalogue, observation, etc., and can be used to define regions of overlap between data collections. This is the key to allow astronomers to get an answer to a question of this kind: "get me all the Sey1 galaxies which are in SIMBAD and have been observed by HST F555W and SDSS DR9".

The CoSADIE project

The CoSADIE project coordinates European VO activities, with three strands of work: "Increasing awareness and gathering requirements from the user and provider communities" (INTA and GAVO), "Coordinating technical activities and defining the technical needs to maintain the VO Framework" (UEDIN), "Outreach towards education and the general public interested in astronomy" (INAF), and assesses the strategies, governance and financial sustainability of the European Virtual Observatory. The project has been working in close collaboration with the Astronet ERA-NET, which gathers European funding agencies. M. Allen, CoSADIE project scientist, participates in the on-going update of the Astronet Infrastructure Roadmap. The second year of the project has been mainly devoted to the sustainability assessment, continuing the close collaboration with Astronet, and a six month extension (to February 2015) was obtained to continue to accompany Astronet to the completion of their study.

The VO evolution is towards a better integration with astronomy large projects. We hope that the cluster proposal submitted by M. Garrett on September 2nd, on behalf of a consortium which includes the CDS and the other Euro-VO teams, will be selected. This integration is at the core of this project.

The Research Data Alliance

The Research Data Alliance, created in March 2013 with support of the European Commission, NSF and Australia, continued to grow. F. Genova is one of the members of RDA Technical Advisory Board, and liaison is built with the IVOA. Astronomy has historically been at the forefront for the sharing and reuse of data, and thanks to the Virtual Observatory its data holdings are a rare example of a global interoperable data infrastructure. The lessons we learnt and our requirements have to be taken into account by the RDA. F. Genova was invited to participate in the second European project set up in support to the RDA (RDA Europe – the European plug-in to the global Research Data Alliance), which will begin in October 2014 for 18 months, and to the RDA/Europe3 proposal submitted to the September 2nd EC Call for Proposals.