CDS Scientific Council meeting 2013
Summary of CDS activities 2012-2013

Final version, 28 October 2013

TABLE OF CONTENT

HIGHLIGHTS DURING THE PERIOD

- The CDS has been confirmed as a « Research Infrastructure » (a label at the national level) by the Ministry of Higher Education and Research in the update of the national Research Infrastructure Roadmap for the period 2012-2020, published in October 2012[1].

Projects

- Five European projects with CDS participation have begun during the period. One, *Collaborative and Sustainable Astronomical Data Infrastructure for Europe*[2] (CoSADIE), is the current project supporting the European Virtual Observatory. Two, *Astronomical Resource Cross-matching for High Energy Studies*[3] (Arches) and AstroDeep[4] (*Unveiling the power of the deepest images of the universe*), have been selected in the 2012 SPACE Call for projects on space data exploitation, and one, VIALACTEA (the Milky Way as a Star Formation Engine), in the 2013 Call. The fourth, *International Collaboration on Research Data Infrastructure*[5] (iCORDI), is the European project in support to the nascent *Research Data Alliance* (RDA)[6]. It has been recently renamed RDA/Europe. The rapid emergence of the Research Data Alliance, funded by the European Commission, the NSF and Australia to "implement the technology, practice, and connections that make Data Work across barriers", is potentially a major evolution of the international research data context.

---

[1] http://cache.media.enseignementsup-recherche.gouv.fr/file/TGIR/29/8/infrasUK_mcgs2_243298.pdf
[2] http://www.cosadie.eu
[3] http://www.arches-fp7.eu/
[4] http://www.oa-roma.inaf.it/astrodeep/astrodeep/Home.html
[5] https://europe.rd-alliance.org/Pages/Home.aspx
[6] https://rd-alliance.org/node

- CDS is now officially involved in the Gaia project. It will distribute Gaia catalogues in the framework of the Gaia Data Processing Archive Access Co-Ordination Unit (CU9), selected in March 2013.

Services

- SIMBAD can sort references attached to objects with respect to where the object is cited in the text (title, keyword or abstract, table row). This new functionality has been officially released on 30 September 2013.

- *Aladin lite*, a spectacular "light" version of Aladin running in the browser and geared towards simple visualisation of a sky region, was released in May 2013[7].

- Implementation of the CDS services in all the aspects of the VO framework is continuing. "TAPVizieR" was released on 15 January 2013. This is a new way to query VizieR using SQL-like queries through the Table Access Protocol VO standard. Its implementation has required a huge work because of the very large number of columns involved (currently ~370 000) and of data heterogenity.

- *VizieR photometry viewer*, a new tool allowing to visualize photometry points around a given object or position, was released on 18 October 2013. Photometric data is currently available for 2050 tables among the 11 580 table of VizieR. The viewer uses a widget which can be included in any web page, including those of the other services, a first step towards customizable web pages, in line with the web 3.0 paradigm.

- The collection of very large surveys in VizieR and of reference images of the sky in Aladin has increased significantly, thanks to the new pipelines. Specific highlights have been the release of Planck maps in Aladin March 2013, and CDS participation in the data release of the final version of the Canada-France-Hawaii Telescope Legacy Survey (CFHT-LS - October 2012).

- A resolute action to improve reliability of CDS services, with a systematic analysis of all the relevant elements, has been launched with the Observatory system engineer and participation of all CDS services. Evaluation following the guidelines of the Data Seal of Approval[8] is on-going with the aim to obtain this label, which is recognized by the European Union.

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

---

[7] http://aladin.u-strasbg.fr/AladinLite/ . You can copy/paste the code given in the document in your web page if you want to make a test.
[8] http://datasealofapproval.org/en/

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 2012 Scientific Council recommendations, for those which were among these recommendations, and in the following section which summarizes CDS activities since the 2012 meeting.

STATUS OF SC 2012 RECOMMENDATIONS

High level recommendations

> **Services:** The services provided by CDS are extensively used by the astronomical community, and these need to be maintained at the highest level and upgraded periodically through research and development. Procedures should be developed to limit the efforts needed for curation and generation of the data. It would be beneficial to collaborate in all these matters with organisations which generate large volumes of data.

Collaboration was sought with ESO to explore the possibility of releasing ESO Public Surveys through the CDS services. ESO agreed to act as single point of contact for the release of catalogues in VizieR. CDS would also like to release images with the HEALPix progressive viewing method. Collaboration with ESA for dissemination of public data is on-going, with the dissemination of Planck public data, distribution of Herschel observation log and result catalogues, and in the framework of the Gaia project. Aladin is also used for data visualisation of the ESA archives. Positive preliminary discussions were also held with LOFAR and the Australians involved in the preparation of SKA.

> **CDS and the VO:** CDS should continue to play an important role in the VO as it moves to its operational phase, at the developmental level as well as in encouraging the scientific use of the VO products.

CDS has continued to be one of the driving forces of the VO development. The activities during the reference period are summarized in companion documents giving the list of talks given by CDS staff during IVOA meetings and the list of IVOA standards with CDS participation. CDS also participates actively in the European actions in support to VO take-up by the science community, in particular in the School organised by CoSADIE Spanish partner in Madrid 5-7 February 2013[9], and in the update of the tutorials. Since the VO success also depends on the availability of data in the VO framework, CDS was also active in preparing and attending the CoSADIE Data Centre Forum[10] and the Focus session which gathered major data providers at the Heidelberg IVOA Interoperability meeting in May 2013 (which was organised by Mark Allen as chair of the IVOA Standing Committee on Science Priorities). More generally, CDS took the responsibility of the Euro-VO web site[11], previously managed by ESO. The web site is used to disseminate information on VO activities and to host tutorials and "sustainable" information.

---

[9] http://www.laeff.cab.inta-csic.es/projects/cosadie/main/index.php
[10] http://www.laeff.cab.inta-csic.es/projects/cosadie/main/index.php
[11] http://www.euro-vo.org/

In France CDS was invited to prepare a tutorial for a school organised at the national level on Cosmic radiation at high energy[12] on 16-20 September 2013 in Annecy, which is an important sign of interest from the high energy astrophysics community, one of the "neighbouring communities" targetted by the astronomical VO for interdisciplinary liaison. At this occasion a tutorial focused on the scientific interests of this particular community was prepared. The tutorial is now available on the Euro-VO web site[13].

> **Scientific activities of the CDS**: The Council encourages CDS to appoint Ph.D. students and post-doctoral fellows to increase the volume of the scientific work and to keep alive the motivation in the science development, augmenting the number of available fellowships by exploring available sources at the French and European level. The Council strongly recommends the establishment of a visiting scientist programmes at CDS.

One PhD supervised by B. Vollmer begun in 2012 (galaxy evolution in the Virgo cluster based on cm-mm radio observations and numerical simulations). Another one, on stellar formation in the galaxy and interactions with the interstellar medium using Herschel/Hi-GAL and Spitzer/GLIMPSE data, supervised by L. Cambrésy, has begun in October 2013. We also expected on PhD cofunded by the STScI to work with C. Bot but the candidate finally gave up. The efforts to attract PhDs and early carrier astronomers will go on. Post-docs will be open in the AstroDeep and Via Lactea project(s) at a later stage.

Visiting scientist programme: the analysis of the recommendation showed that a specific staff would be needed for supporting the visiting scientist programme. This was the core functional task of the profile presented by CDS for CNAP recruitment in 2012 (recruitment held at national level with a strong competition). Recruitment was unfortunately unsuccessful although a suitable candidate was presented. This element of the profile will be kept for the 2014 competition. Currently, and not in the framework of an organised visitor programme, Prof. Quentin Parker from Maquarie University is visiting the Observatoire de Strasbourg and he and his Post-Doc collaborate with CDS.

> **Staffing**: It is essential for CDS to have adequate number of staff members with the right mix of scientific, technical and documentation personnel, and to have a long term policy to compensate for the retirement and attrition of its senior and highly experienced staff.

The situation is not good this year, since it has not been possible to get any recruitment during the period, neither in scientific nor in technical staff, whereas François Ochsenbein (CNAP position) has been on Emeritus status since September 2013 and Gratiane Chassagnard, a librarian working for CDS at the Institut d'Astrophysique de Paris, retired in July 2013.

> **Outreach:** CDS needs to maintain an adequate level of outreach activities for training students, and particularly teachers in the use of data products, tools and services through workshops and schools.

The activities with the Rectorat towards local secondary school teachers have been renewed during the academic years 2012-2013 and 2013-2014, in a context of significant reduction of these activities (only 13 trainings will be organised for physics teachers during the coming year).
At European level, the efforts are put in the CoSADIE project, in close relation with Astronet Education & Public Outreach activities. For the moment one issue is how to organise the wealth of EC funded projects in the domain.

---

[12] https://indico.in2p3.fr/conferenceDisplay.py?confId=8048
[13] http://www.euro-vo.org/sites/default/files/documents/HighEnergyTutorial.pdf

The activities at the international level have been focussed on the European level, since one of CoSADIE aims is to assess how to establish a sustainable framework for Euro-VO activities. We understand the importance of addressing more generally the international level but it has not been possible this year. It is worth mentioning that an IVOA Education Interest Group, led by M. Ramella (INAF) and S. Barway (SAAO), was created in October 2012.

> **Future Directions:** The CDS should (1) maintain its services at the highest possible level in terms of content and functionalities, (2) evolve core services in line with users' needs and scientific requirements, and (3) take into account the change in scale of CDS activities, due to the increase in the volume of publications and the advent of many large surveys.

This is a welcomed endorsement of our high level strategic axes. The addition of a fourth high level strategy driver is proposed to the Council. It is linked to the evolution of the context with the rapid emergence of Open data/Open science concepts, and would include our participation in scientific data curation and in IVOA and Euro-VO, and the dissemination of CDS and astronomy expertise towards other disciplines at the national, European and international levels.

Detailed recommendations

The 2012 Council report contains very interesting recommendations on all the facets of CDS activities. Not all of them could be implemented in one year but they will be kept in mind. A few comments on some of the additional recommendations.

Services

> **Recommendation:** The services provided by CDS are at the core of much astronomical work involving data, and it is essential that these services are maintained at the highest level. It is essential to make major research and development efforts to continuously upgrade the services in phase with ever increasing volumes of data, and the need of the astronomical community to access the data easily and to obtain deeper insights into it. In spite of its cost, the verification work for maintenance and updating of the SIMBAD database needs to be continued without compromise, using mechanisms which would help to reduce the time spent on data ingestion work. Efforts need to be made to improve the links between SIMBAD and VizieR. Further work on data curation and services could be carried out jointly with observatories and research centres which produce the data. The longstanding collaboration with A&A and other journals to archive and provide access to observational data tables should be strengthened to include all data used in published articles.

> SIMBAD database. Despite its manpower cost to CDS, the Council considers that the quality of the verification work must not be compromised, as the credibility of SIMBAD largely relies on it. At the same time, the Council recommends that mechanisms are established during data ingestion projects to ensure an acceptable match between invested resources and expected return, so as to avoid that major efforts are spent in very marginal improvements of projects at the stage at which the outcome already satisfies strict quality criteria.

This comment came from a specific example given during the presentations. It has been noted and is taken into account. On large ingestion projects, when to stop when the improvement becomes marginal is defined by a scientific evaluation of the impact. No large ingestion effort ends up with 100% ingestion.

A few other points:

- Efforts to promote the cross-match service are in particular through participation in scientific projects, for instance those submitted to the SPACE Calls of the European Commission or to the Agence Nationale de la Recherche. Also discussions with the SKA team and booths at the American Astronomical Society Winter meeting and ADASS allowed us to publicize this new facility.

- The progenitor facility is being finalised. HST and XMM data are ready with links to the progenitors (collaborations with CADC and XMM-SSC), a prototype interface is available and a new version, using a progenitor table and the proposed IVOA DataLink standard, could be made available in the next version of Aladin (release foreseen before the end of 2013). The delay in production of the standard working draft, expected during Summer 2013 and finally released on 22 October 2013, may introduce a delay for the release of a compliant service.

- Links between SIMBAD and VizieR are frequently being discussed but a complete assessment remains to be done. The photometry viewer opens the path for inclusion of VizieR data in SIMBAD like in any other service. A link to SIMBAD is included in VizieR tables when it is possible.

CDS and the VO

> **Recommendation**: CDS should continue to play an important role in the VO, at the developmental level as well as in encouraging the scientific use of the VO products. 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. To sustain this activity over the long term, CDS should continue to seek funds through European and international programmes.

Usage of VizieR through the VO is a remarkable example of the VO power and usage (unknown from the users themselves since this is fully transparent, which is compliant with the main aim of the VO to provide seamless access to data).

Participation in the preparation of Horizon 2020 Work Programme is by various ways, including active participation in meetings and consultations organised by the European Commission. For instance, a Euro-VO Integrated Infrastructure Initiative was submitted to the "Consultation on possible topics for future activities for integrating and opening existing national research infrastructures" held by the European Commission in autumn 2012. The Assessment Report states that "The European Virtual Observatory topic has high potential for Horizon 2020 and is relevant to Research Activities under both 4.1.2 'Integrating and Opening national RIs of pan-European interest' and 4.1.3 'Development, deployment and operation of ICT based e-Infrastructures' ". At present the first call of the new framework program Horizon 2020 is being finalized and we will know the details of the AOs before the end of the year.

Staffing

> **Recommendation:** It is essential for CDS to maintain adequate number of staff members with the right mix of scientific, technical and documentation personnel, and to have a long term policy to compensate for the retirement and attrition of its senior and highly experienced staff. This can be done by preparing and encouraging its younger staff members to assume greater responsibilities and by timely hiring of new staff with the requisite skills. CDS has been very successful in providing important services and expertise to the French, European and wider astronomical community, and it is inevitable that the demands on it will only grow over the coming years due to the phenomenal increase in data. It is therefore very necessary to increase the available human resources beyond the present strength.

This is a well taken recommendation but during the period CDS LOST manpower. The retirements of two key staff in 2014, the engineer responsible of SIMBAD and CDS secretary, will make the situation still worse in the very near future. Any additional help from the Council in these critical matters would be very welcome.

Future directions

> The key to continuing success for CDS will be its services to the community, spearheaded by SIMBAD, VizieR and Aladin (see Section 2 of this report). These services will need to continue to evolve and adapt to changing user needs and expectations (not always the same!) and CDS will need to provide expert, dedicated and dogged shepherding. CDS needs to continue to work with its partners on the VO, so that the VO becomes widely perceived as an essential part of the astronomical infrastructure rather than a project in development. In addition, the Council encourages CDS to continue to look for additional opportunities to deploy its expertise and experience to maximise the scientific return from the community's various endeavours. Possible areas include involvement in large projects, for example Gaia, and in supporting the preservation and utilisation of published data, for example through further collaborations with A&A.

> To achieve these goals, CDS needs its staff to continue independent scientific research, to continue R&D activities in areas of technology related to its activities and to remain close to various communities.

CDS activities in these directions are described in this report. One can note that we sought participation in scientific projects, in particular in the framework of the European Commission SPACE Calls, to build collaborations with top level scientific programmes, in addition to participation in Gaia and advances in the collaboration with *Astronomy & Astrophysics*.

SUMMARY OF ACTIVITIES, SEPTEMBER 2012 – OCTOBER 2013

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

The services

SIMBAD

October 20th, 2013, SIMBAD contained 7 342 000 objects, 18 162 000 identifiers, 285 000 bibliographic references, 10 000 000 citations of objects in papers (6 970 000 objects, 17 100 000 identifiers, 269 000 bibliographic references, 8 719 000 citations on July 1[St], 2012).

There has been 520 000 queries/day in average between January and June 2013 (416 000 in 2012).

Sorting of references linked to object

Released 30 September 2013 as explained.

New user interface and query optimization

This development has not much progressed during the last period. A contractor was hired in August 2013 for one year to work on web interfaces, including this aspect, in collaboration with the engineers in charge of the service.

SIMBAD content: cross-match tool

The specific software ("raccord") which is used to cross-match lists of objects with SIMBAD taking into account as many parameters as possible to prepare object ingestion is being re-written, in close collaboration between the engineer in charge, SIMBAD "documentalistes" and astronomers. The first release of the new software should be available before the end of the year.

SIMBAD content: specific operations

The daily update of the database of course goes on. Here specific, significant programmes are listed.

Galactic Planetary nebulae

Though very rare, Planetary Nebulae (PNe) are important astrophysical objects. In view of ingesting in Vizier and in Simbad the new version of the catalogue of Planetary nebulae by Q. Parker and collaborators which should be submitted in the coming months, a systematic check of the content of Simbad has been done. The oldest PNe, from the beginning of the 20th century to 1990, have been entirely revised: improvement of coordinates, check of the real nature of the object in the literature, cross-identifications, spectral type of the central star when known, size, main identifier as found in the literature.

Old/bad coordinate cleaning

In the past, coordinates in Simbad were entered without reference, and by default with a quality D. Estimates of the error ellipse were assigned by default as a function of the number of digits in the coordinates. As the quality of the coordinates is now well defined (from 1 to 9'' for quality D, larger than 10'' for quality E) and used by the cross-identification program to ingest long lists of objects in Simbad, these old coordinates with often incorrect qualities or error ellipses are a real problem. Many objects with such old coordinates have never been updated, and they are difficult to recover because the reference of the coordinates is missing.

A massive coordinates cleaning operation is on-going in Simbad, starting with very large and well known old catalogues. The first target was IRAS, including the PSC, the FSC, the SSC, the catalogues of rejected sources, and the catalogue of small scale structures. To identify whether an object was still with historical IRAS coordinates in Simbad, a comparison between Simbad and original IRAS coordinates was done. Finally the coordinates of about 100 000 objects have been revised in Simbad, putting the right quality, the right error ellipse, the reference, and the wavelength range. It thus becomes easy to replace them by better coordinates as soon as the objects are found with a better identification in the literature. The next target of this cleaning program will be the ROSAT catalogues.

VizieR

October 20th, 2013, VizieR contained 11 579 catalogues (10 360 in September 2013). Between January 2012 and June 2013, 27 tables with more than 10 million rows have been ingested (compared to 6 on three years, from 2009 to 2011), including GALEX, SDSS-DR9 and SDSS-DR9Q and UKIDSS-DR9 during the last period.

VizieR passed this year the threshold of 10 billion "objects" (catalogue row with a coordinate) in non-obsolete catalogues (11.556 billion "objects" including the GUMS model).

There has been 600 000 queries/day in average between January and June 2013 (458 000 in 2012).

Very large catalogues

This aspect was discussed above.

Photometry viewer

Released 18 October 2013.

TAP in VizieR

Released, as explained in the highlights.

"Additional data" in VizieR

This is a critical aspect of VizieR in the Open Data context, since it allows scientists to disseminate data resulting from 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 published in other journals. VizieR already includes a significant fraction of catalogues and tables including "additional data". The possibility to "publish" individual data sets (a spectrum, an image) stored in VizieR in the VO is being assessed, and will be discussed below. Currently the object in VizieR is found in the VO,

but all available information about the object is mixed up and one has to have a look at the information to find that additional data is available for the object. We want to improve the discoverability of these additional data.

Aladin

In August 2013, Aladin contained 128 surveys with a volume of 30 TB (81 surveys, 19 TB in August 2012)

There has been 30 900 queries/day in average between January and August 2013 (18 500 in 2012).

Aladin lite

This new version of Aladin, running natively and without plugin in web browser, is already a success. It is currently used by various projects and collaborations (CADE, LIGO-Virgo, ADS Labs, CARMENES, ADS All-Sky-Survey) for different use cases (preview of data products, objects selection, contours display, visualization of bibliography heatmaps).

Functionalities are currently limited by the performances of Javascript engines, and it is expected that the interface between Aladin lite and the standalone version will move in the future when new capacities will become available.

Aladin lite runs on desktops as well as on touch devices (tablets, smartphones).

A nice example of Aladin lite usage in another service for its own needs can be seen for instance here (Canadian Galactic Plane Survey data provided by the Centre d'Analyse de Données Etendues – CADE – of IRAP, Toulouse):
http://cade.irap.omp.eu/dokuwiki/doku.php?id=cgps.
Another example is the ADS All Sky Survey, developed in the framework of a collaboration which includes the CDS:
http://www.adsass.org/

Usage of the HEALPix visualisation method

The power of the HEALPix tessellation implemented in Aladin is demonstrated by the software capacities. The "Hierarchical Progressive Survey" method promoted by CDS allows one to display a survey progressively, based on the principle that the more you zoom in on a particular area, the more details show up. CDS also provides a package which allows any project to format its own image data in the same way (and will soon provide a package to generate catalogues). We know that this package is used by some projects for their own needs. Other projects prepare their data and provide them to us.

Next public version

The public release of Aladin (V8.0) is foreseen before the end of 2013. It will include proper motion management (implemented in particular for Gaia), pre-visualisation of VizieR photometry with a link to the photometry viewer, and updated implementation of VO standards (VOTable 1.3, MultiOrder Coverage map, Space Time Coordinates).

The cross-match service

There has been an average of 12.5 jobs/day between January and June 2013 (10 jobs/day in 2012). 724 different users have been identified in 2013 until now, to compare to 499 in 2012.

The catalogues ingested in VizieR are immediately available in the cross-match service.

Cross-match Application Programming Interface (API)

Released on 2013, January 30th, the cross-match API [14] allows one to access programmatically to the cross-match service. Scripts or programs can easily submit a list of positions and quickly look for counterparts in any VizieR table with position or in SIMBAD.

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 is two-fold: first, existing statistical methods have to be generalized to handle more than two catalogues, then a flexible, highly configurable and still efficient tool has to be developed. E.g., taking into account extended objects and proper-motions requires to study and use new indexing data structures. The new methods will be tested on the scientific cases developed in Arches, and they will be included in the CDS cross-match service if the assessment is positive.

Improvement of tools offered to users and to the CDS team

To improve the efficiency of CDS work, a significant action has been devoted to update the collaborative space which is widely used by the team in its daily job, in particular to share procedures and software directions for use. This has been done in collaboration between "documentalistes" and software engineers. Also a new system to manage the hot line, keep track of tickets, etc, has been assessed and is being developed. It will be released in the coming months.

R&D

New interfaces

Aladin lite can be seen as a pathfinder of Aladin evolution, anticipating the replacement of Java applets by other technologies.

The photometry viewer in VizieR is the first implementation of widgets which will allow users to customize their interface to CDS, and which can be used by us in different services. This is one of the paths to increase the links between SIMBAD and VizieR, and to provide visualisation of different kinds of information.

SkyTouch (available for both iOS and android) is a study of a bridge between desktop applications (Aladin in our use case) and smartphones/tablets based on the IVOA Standard SAMP (possible use in Education for example).

SkyObjects is an android application using the CDS name resolver to retrieve basic information about an object and a pointing capability.

Data cubes

Collaboration is on-going with the International Centre for Radio Astronomy Research (ICRAR, A. Wicenec) to adapt Aladin to visualize data cubes from ASKAP/MEERKAT/SKA based on JPEG2000 technology. CDS will also explore the extension of the HEALPix method to cube pre-visualization. The two approaches could be complementary.

---

[14] http://cdsweb.u-strasbg.fr/news.php?fn_mode=fullnews&fn_incl=0&fn_id=312

Assessment of the Apache Solr search platform

This R&D begun in April 2013 on an internship and will certainly be pursued because it has an important potential impact. Solr[15] open source search server includes powerful full-text search, hit highlighting, faceted search, near real-time indexing, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search. It would allow in the short/medium term queries with multiple keywords (object names, coordinates, concepts), and on a longer term queries in natural language.

Assessment of the usage of the cloud

As the Cloud is now a usual technology in both public and private domains and not just a fashion we are continuing to evaluate its evolution in term of cost and performances. During the last year we prototyped a HEALPix image server.

CDS role in data curation

Assessment of CDS expanded role in scientific data curation

Astronomy remains at the forefront for data sharing and reuse, in a fast evolving context: more and more governments and funding agencies promote "open data" policies requiring that data obtained on public funds be in general publicly accessible. The G8 Science Ministers, meeting in London with the Presidents of their respective science academies on 12 June 2013 stated[16]:

> We have decided to support the set of principles for open scientific research data outlined below as a basis for further discussions.
> i. To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.
> ii. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.
> iii. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
> iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.
> We decide to build on the existing work to coordinate and enable international data collaboration.

In this fast evolving context of Open data/Open science. CDS can play a major role for astronomy, widely anticipated already. If agreed by the Council, CDS role in the Open data/open science world could be added as a fourth high level strategic objective. This would include IVOA/Euro-VO role.

CDS already provides result data attached to publications, not only tabular data, through VizieR. The recent period confirms the acceleration of this trend. 14% of the 11500 VizieR catalogues contained additional data in August 2013. In July 2012 812, 268 and 87 data sets containing respectively time series, spectra and images were included in

---

[15] http://lucene.apache.org/solr/
[16] https://www.gov.uk/government/news/g8-science-ministers-statement

VizieR, to compare to 1026, 349, 108 respectively in August 2013: 26%, 30%, 24% increase in 13 months respectively.

Most of these data sets are provided in the framework of the agreement with Astronomy & Astrophysics, but we also take into account data sets provided by authors or other journals.

The importance of this function in the current context was recognized by the Astronomy & Astrophysics Board, and discussed with the journal scientific editors, who ask authors to provide the data to CDS for dissemination.

The CDS is by far not the only actor of dissemination of data resulting from astronomical research. Observatory archives are in charge of disseminating observational data, a few national centres (CADC, Heidelberg, Trieste, …) are hosting data produced at national level, and the other journals, in particular the AAS ones, are building a strategy. The US-VO has developed a DropBox service. A meeting will be organized by the AAS journals after the Winter AAS meeting to discuss technical aspects. CDS will of course participate in this meeting.

The increased importance of 'additional' data in VizieR leads to assess possible relevant evolutions of the service:
- visualisation of 'additional data' spectra, time series, luminosity curves, in addition to SED, and images in Aladin
- direct publication of these data in the VO also has to be assessed, 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 about the object 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 may be a huge work since many files are involved. A fist step will likely be to provide an image service, for which metadata should be found in the image FITS header.

The assessment of the required evolution is one of the major strands of work for the coming year.

Recognition of CDS role and expertise in scientific data curation

It is worth noting here that CDS is widely known at the national and international level, well beyond astronomy, as a centre with a long term, successful expertise in scientific data curation and dissemination. For instance, F. Genova was a member of the High Level Expert Group on Scientific Data set up by the European Commission in 2010, which produced the influential "Riding the wave" report. She is also among other responsibilities "chargée de mission" at INSU for scientific data and information systems, and often invited, as well as her colleagues, for audits or to explain lessons learnt by CDS and more generally astronomy in the domain. This is also the core of her involvement in the Research Data Alliance described below.

The rapid evolution in the domain also has possible consequences on the evolution of librarian job profiles, which are well demonstrated by the skills and impact of CDS "documentaliste" team. S. Lesteven was invited to bring our expertise in a recent meeting of the national network of scientific information professionals at CNRS devoted to management and valorisation of research data [17] . And several team members participated in the preparation of descriptions of these new profiles for submission to the job profile registry used by CNRS and other organisations.

---

[17] http://renatis.cnrs.fr/spip.php?article263

Participation in projects

Gaia

The CDS is now officially a member of the Gaia project, and it will disseminate the data from the project. Its capacities to do so rely on the HEALPix-based pipeline of VizieR and Aladin, and they have been demonstrated by the fast ingestion of the GUMS model catalogue and the fast queries it allows, plus the handiness of the progressive visualisation. Possible provision of data beyond catalogues and images has to be assessed.

CoRoT

CDS will preserve and distribute the processed data from the CoRoT project. This had been planned from the beginning of the project, but it is getting more real with the end of CoRoT observations. CDS already distributes the CoRoT "observation log", including light curves. Discussions are on-going with the project on the content of the final data release and on the interface to be provided. We will probably have to work on time series visualisation and publication, well in line also with the foreseen definition of specific VO standards in this domain.

Arches

CDS is in charge of providing a cross-identification tool and to distribute the result data. We also provided support to identify VizieR tables with a sufficient overlap with 3XMM using MOC, a nice example of scientific usage of the CDS services and VO standards (Ada Nebot's talk, presented by L. Michel at the September 2013 IVOA meeting: http://wiki.ivoa.net/internal/IVOA/InterOpSep2013Applications/MocsInArches.pdf)

AstroDeep

ASTRODEEP goal is to develop and test new algorithms, reduce and release data, and perform scientific validation and analysis to achieve the best possible exploitation of the deepest multi-frequency surveys. The CDS is in charge of the dissemination and public release of the catalogues, and will ensure long term availability of the data products. This will be done with VizieR and SIMBAD, but also via a prototype dedicated science portal.

VIALACTEA

CDS participates in the scientific aspects of the project, which aims at exploiting the combination of all the new-generation Infared to Radio surveys of the Galactic Plane from space missions and ground-based facilities, using a novel data and science analysis paradigm based on 3D visual analytics and data mining framework, to build and deliver a quantitative 3D model of our VIALACTEA Galaxy as a star formation engine that will be used as a template for external galaxies and study star formation across the cosmic time.

Virtual Observatory

Implementation of CDS services in the VO

TAPVizieR was implemented and released during the last period, as explained.

Update of the VO framework

The list of new IVOA standards developed with CDS participation is given in a companion document. Since September 2012, VOSpace 2.0, VOTable 1.3 and Photometry Data

Model 1.0 were accepted as IVOA Recommendations in March, September and October 2013 respectively. Several other standards are reaching the end of the Recommendation process.

One important proposed standard is MOC (MultiOrder Coverage map), 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". CDS will release before the end of a year a MOC server including all its catalogue and image data sets. This server was used by A. Nebot for Arches in its prototype form.

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. During the first year, most of the activities have been towards the science and data centre communities, with the organisation of a successful School and Data Centre Forum. A close liaison has also been established with Astronet Education & Public Outreach activities. The second year of the project will be mainly devoted to the sustainability assessment.

It is also worth noting that the Focus Session organised during the Heidelberg IVOA meeting in May 2013 allowed the VO community to discuss in depth their requirements with major projects.

The Research Data Alliance

The rapid emergence of the Research Data Alliance, which was created in March 2013 with support of the European Commission, NSF and Australia, is a new element in the landscape of research data sharing, with a high potential impact if the activity continues to build up at the current pace. 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 (or unique) example of a global interoperable data infrastructure. The lessons we learnt and our requirements have to be taken into account by the RDA.