CDS Scientific Council meeting 2017

Summary of CDS activities 2016-2017

30 September 2017

1. Introduction

The CDS has actively pursued its core mission in the 2016-2017 period with significant progress made in many areas of the operations, scientific support of the services, release of new capabilities, publication of important data sets, and planning of major upgrades to the CDS service interfaces and architectures. There have been a number of special events in the period which were related to evaluations and the hosting of a major conference, plus workshops and schools. The CDS has also been highly visible in astronomical community and has made visits to a number of CDS partners. The CDS has made important contributions to global data sharing initiatives, and has taken on new roles in the IVOA. Many collaborative activities have been pursued via our participation in, and leadership of projects. All these activities prepare the CDS for its continuing role as a reference astronomy data centre in support of the international astronomy research community.

An important change in the immediate environment of the CDS is the arrival of a new Director of the Strasbourg astronomical observatory (Observatoire astronomique de Strasbourg, ObAS[1]), Pierre-Alain Duc who began a 5 year mandate in January 2017. A smooth transition from the previous Director (Hervé Wozniak) has ensured that CDS activities continue to be well integrated into its hosting institution.

1.1 Evaluations and Status of the CDS

HCERES

In 2016 we reported that the CDS was in the process of preparing for the HCERES *(Haut conseil de l'evaluation de la recherche et de l'enseignement supérieur)* evaluation campaign which occurs on a 5-year timescale. The documents were prepared as part of the response by the Observatoire Astronomique de Strasbourg, with CDS being evaluated as a science team of the Observatory, and also as a "National Observing Service". In 2016 we provided the CDS Scientific Council with the CDS contributions to these documents, and now the full final submitted documents for the whole unit are made available[2] to the Council.

The visit of the evaluation committee occurred on January 30-31, 2017. The 2-day visit included presentations of the 2011-2016 activities, and plans for the 2018-2022 period. The committee met with the various scientific, engineering and documentation staff of the CDS. They visited the server room and had a demonstration of CDS documentation activities.

The evaluation committee provided a report on the whole unit, including a global assessment and a detailed analysis, followed by a team-by-team analysis with specific recommendations. CDS has obtained excellent results in this evaluation with the keyword "outstanding" being used multiple times. The conclusion of the HCERES global assessment of the unit, with CDS aspects in bold:

> OAS has an excellent reputation and visibility both nationally and internationally. **CDS has an outstanding activity in terms of services and R&D on data mining, data-matching and capability to set international standards in this field.** The GAL and HE teams are very active, they produce highly and sometimes very highly rated science in various fields of astronomy.

[1] ObAS - new official acronym. Note that some associated documents and text in this report use the older OAS acronym.

[2] In PDF format and also as printed books available on request or at the Council meeting. Report on UMR7550 (173 pages), and report on the Observing Services (37 pages).

The full report of the evaluation committee is provided to CDS Scientific Council[3].

The CDS features in the *Global Strengths* of the detailed analysis of the unit as an *"indisputable flagship for OAS"*, and the detailed assessment recognises that the *"CDS provides world-class and essential scientific services to the international community"*.

The *Global Weaknesses* section, and other parts of the report identify the serious risks associated with the necessary staffing and support of the CDS. These concerns are also expressed in the specific recommendations for the CDS. The most prominent issue identified by the evaluation committee is need for the urgent replacement for the Observatory System Engineer — fortunately we can report that this highest priority need has been addressed following the HCERES visit with a replacement starting in October 2017.

Overall it is clear that the evaluation committee has appreciated the high visibility of the CDS, its productivity and international impact, as well as its leadership role in IVOA and the widespread integration of CDS tools in other organisations. The committee has recognised the integrated nature of the team, and the specific role of the CDS documentalists. The recommendations and comments focus on issues that concern the long term sustainability of the staffing and the pressures of ever increasing data volumes, and the necessary evolutions of the IT infrastructure. The specific recommendations from the HCERES committee are included for information in the Appendix of this report.

MENESR

In 2016 we reported that the CDS had its status renewed as a "Research Infrastructure" on the National Research Infrastructure Roadmap established by the Ministry of National Education and Research (MENESR). An update of the Roadmap will be published in 2018, and in 2017 there have been a number of activities associated with this status.

An updated description of the CDS as a Research Infrastructure has been prepared for the next Research Infrastructure Roadmap document, the "Feuille de Route". CNRS/INSU has helped us to prepare the CDS submission to the on-line survey (September 2017). In the context of this survey CDS is considered as a "Virtual Infrastructure", with strong European and International dimensions related to the International Virtual Observatory Alliance (IVOA) and our international partners. The CDS mission, strategy, and resources are reported, efficiently re-using where possible the texts prepared for the HCERES evaluation. The survey also includes general questions on data management, service use, and number of users, where we highlight the essence of CDS as a reference data centre. We expect feedback and finalisation of the survey before the end of 2017.

In 2017 the MENESR led the 95 French national Research Infrastructures (TGIR & IR) in an exercise of calculating their complete costs. The CDS and observatory administration prepared all of the information, including calculation of the material/investment costs via the "ammortissement" option that considered all of our major hardware investments back to 2010. The reported annual cost of the CDS is 2.8 M€, made up of personnel costs of 1.95 M€, functioning costs of ~750 k€ (including indirect costs), and material costs of ~100 k€.

In this report we provide highlights of 2016-2017 in Section 2. The activities of the CDS services are described in Section 3, and Projects in Section 4. Section 5 provides the CDS responses to the 2016 recommendations of the CDS Scientific Council.

[3] Included in the documents sent to the Scientific Council - HCERES_evaluation.pdf (22 pages)

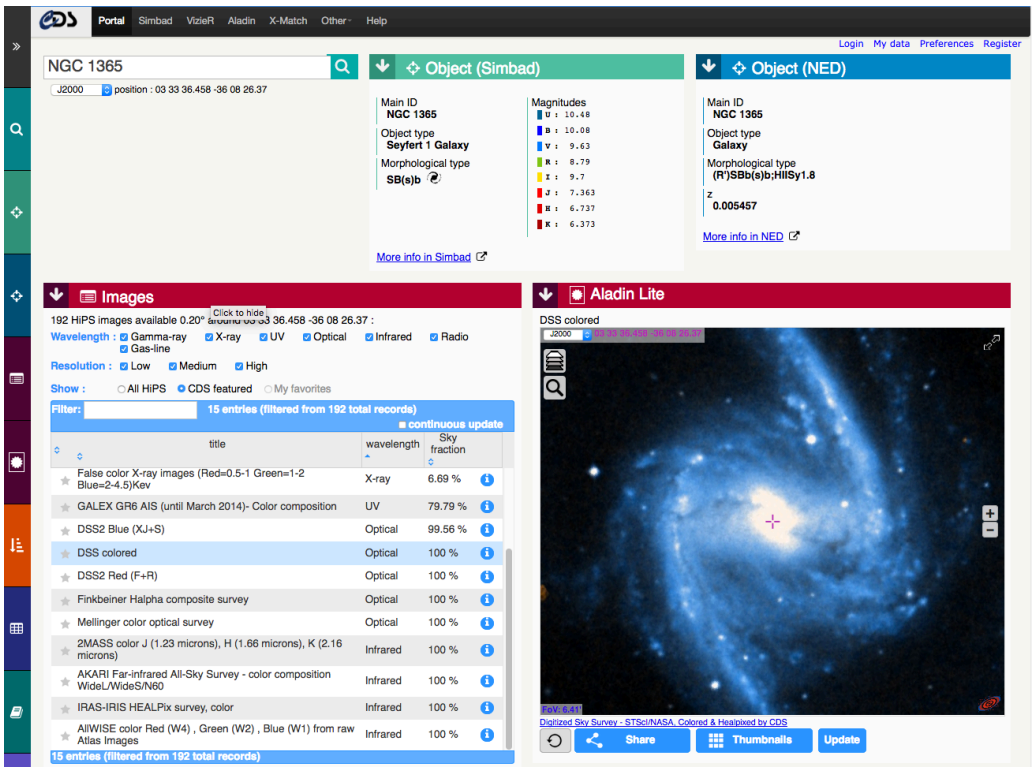
2. Highlights 2016-2017

LISA conference hosted by CDS

During June 6-9, 2017 the CDS hosted the *Library and Information Services in Astronomy*[4] (LISA) conference with 110 participants. This international conference is held every 3-4 years as a "scientific meeting for librarians and scientists that aims to provide a platform to discuss the state of the art of information maintenance, retrieval, delivery, and preservation and to learn from invited experts the directions in which our profession is moving". This conference facilitated high level interactions between major information providers in astronomy, as well as journal publishers and editors, and scientists. The interactions supported by this conference have strengthened the CDS coordination with other reference data centres in particular with the ADS and NED. CDS staff were prominent in the program, including the keynote speaker and 6 selected contributed presentations, with associated papers that provide a detailed and current record of CDS data curation practices. The enormous effort of the CDS LOC and staff ensured a very productive conference, with a high level of sponsorship, and it will be remembered as an important moment in the life of the CDS. The 30th anniversary of cooperation between ADS, CDS and NED was celebrated during LISA.

CDS Portal Released

A major revision of the CDS portal[5] has been released in mid-December 2016. The CDS portal is a single entry point to CDS services and plays an important role, in particular for first time users, and it is central to tutorials, providing an integrated view of SIMBAD, VizieR and Aladin. The new portal fully integrates HiPS and Aladin Lite and exploits the development of the MOC server that provides very fast sky coverage comparison of thousands of data sets (images, catalogues, and any data set with sky-coverage). The portal development project now moves into a user feedback phase.



[4] http://cds.unistra.fr/meetings/Lisa8/

[5] http://cdsportal.u-strasbg.fr

Major data sets implemented in the CDS services

Recent high priority data sets that have been ingested into CDS include: Gaia in SIMBAD - resulting in an improvement of the astrometry for about 2 million stars in SIMBAD ; Ten very large catalogues in the VizieR service, and the inclusion of a number of ESO Public survey catalogues. Notable additions to the Aladin collection of HiPS data sets are: the DECam Legacy Survey (DECaLS) DR3; the 3-d data cubes of the MaNGA (Mapping Nearby Galaxies at APO) project which is the newest survey component of the Sloan Digital Sky Survey; the revision of the whole HST collection (in coordination with CADC); the HI4PI data release of 21cm neutral atomic hydrogen data cubes of the Milky Way; and ATLASGAL, the APEX Telescope Large Area Survey of the Galaxy at 850 micron. SkyMAPPER and Pan-STARRS are currently being processed.

HiPS Standardisation

The Hierarchical Progressive Survey (HiPS) system developed by the CDS, which underpins the CDS approach to Big Data, and the success of the distributed HiPS network, has become an international standard. HiPS Version 1.0[6] was approved as an IVOA Recommendation in May 2017, led by a CDS editor, 3 CDS authors plus authors from STScI/MAST, CADC, JAXA, Observatoire de Strasbourg, ESA and ESO/ALMA.

ASTERICS - successful mid term review

The ASTERICS cluster (Astronomy ESFRI and Research Infrastructure Cluster[7]) project which started in May 2015 reached an important milestone with a successful mid-term review in Brussels on 14 March 2017. CDS leads WP4, Data Access, Discovery and Interoperability (DADI), which gathers VO teams and teams from the large ESFRI and ESFRI-like projects to optimise the usage of the data through the Virtual Observatory. The overall assessment by the reviewer indicates that the “Project has delivered exceptional results with significant immediate or potential impact”.

Research Data Alliance (RDA)

The next project aimed to support the Research Data Alliance in Europe, *RDA Europe 4.0*, was submitted to a dedicated Call for Proposals of the European Commission in March 2017. It was selected to enter into negotiation, with the Grant Agreement signature expected in November 2017 for a start in February 2018.

Scientific Training Events

The CDS has run a very active program of scientific training events in 2016-2017. We have either led or made significant contributions to the following events:

- The ASTERICS DADI VO School[8], Strasbourg, 15-17 November, 2016
- The GAVO-CDS Gaia Data Access Workshop[9], Heidelberg 21-24 November, 2016
- A Doctoral school in Paris, 31 January, 2017

---

[6] http://www.ivoa.net/documents/HiPS/index.html

[7] https://www.asterics2020.eu, https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:start

[8] https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:school2

[9] http://mintaka.ari.uni-heidelberg.de/gaia-workshop-2016/

- The “Detecting the Unexpected - discovery in the Era of Astronomically Big Data” workshop[10], Baltimore, 27 February - 3 March, 2017. Tutorial: *“Advanced use of HiPS and MOC”*
- 3rd Cosmology School[11] in Cracow, Poland, 10-23 July 2017
- Cosmology School in the Canary Islands[12], 18-22 September 2017
- Observatoire Côte d’Azur Session Observatoire Virtuel, Nice, 26-27 September 2017

ADASS 2016

A record of 5 CDS oral presentations (1 invited and 4 contributed) were presented at the ADASS meeting in Trieste, October 2016. This included a dedicated effort to communicate and document the scientific basis of the best known CDS service SIMBAD with the paper “What is SIMBAD, and what is it not?”. The full list of CDS ADASS contributions is provided in the documents provided to the council.

Collaborations and Interactions with Partners and the Astronomy Community

Supporting and reinforcing CDS collaborations with partners and up-coming projects has been continued as a high priority in 2016-2017. A number of dedicated visits were made to CDS partner sites by groups of CDS engineers and scientists, with an emphasis on ensuring high levels of interoperability and coordinated developments in areas of mutual interest. A visit to STScI/MAST (Allen, Boch) in January 2017 was instrumental in coordinating CDS ingestion of Pan-STARRS data, and also the STScI use of Aladin in the JWST Astronomer Proposal Tool. A visit to NASA/HEASARC (Allen, Boch) in January 2017 fostered the use of HiPS by HEASARC and an arrangement for sharing mission log catalogues maintained by HEASARC. A visit to ESAC (Allen, Boch, Fernique, Oberto, Pineau) in May 2017 focused on ESA and CDS plans for visualisation and sharing data and use of HiPS and Aladin Lite in ESASky. An extended visit to SAO/ADS (Oberto) in August 2017 ensured a new installation of the SIMBAD mirror at CfA. Each of these visits also included an invited seminar (Allen, Oberto). CDS also welcomed a visit of the Australian SkyMAPPER project in May 2017 resulting in participation in the DR1 data release by preparing the HiPS version of the image data.

CDS also participated in LSST meetings organised by CNRS/IN2P3, and has engaged in specific discussions with the LSST-France project about long term use of LSST data. An invited presentation was made at the Australian ADACS Data Intensive Astronomy Workshop[13] (Nebot, Melbourne 7-9 August 2017). P. Ocvirk and G. Landais participated in the 9th Summit of Information Providers in Astronomy, Astrophysics and High Energy Physics (AAHEP9) May 4-5, 2017 in Hamburg.

CDS also interacted with partners and scientific users as usual at the American Astronomical Society (AAS) Winter AAS meeting in Texas in January 2017. A CDS Booth was installed at the European Week of Astronomy and Space Science meeting (EWASS) in June 2017 and will become a regular event. We also participated in the French national astronomy meeting (SF2A, July 2017, Paris), JDEV (July 2017, Marseille), IVOA meetings, VO France Days (March 2017, Paris), and the ASTERICS-OBELICS Workshop (December 2016, Rome).

---

[10] http://www.cvent.com/events/detecting-the-unexpected-discovery-in-the-era-of-astronomically-big-data/event-summary-0db6808d548b4a9ea6466b43046a1ff5.aspx

[11] http://cosmoschool2017.oa.uj.edu.pl/index.html

[12] http://www.iac.es/congreso/cosmo2017/pages/scientific-program.php

[13] https://adacs.org.au/index.php/2017/06/15/adacs-data-intensive-astronomy-workshop-melbourne-7-9-august-2017/

CDS Permanent Staff

It is with great relief that we can report that our highest priority need, for a system engineer at the level of the Strasbourg Observatory, has been addressed by the CNRS (with also support from the Université de Strasbourg). Christophe Saillard will begin as a system engineer in October 2017 to take charge of the IT infrastructure of the Strasbourg Observatory (nominally shared at the 50% level with the CDS). The timing of this arrival allows for overlap with the current system engineer in order to ensure a smooth transition.

3. Activity Report for CDS Services 2016-2017

In 2016-2017 we have continued to use the ADS interface to track text citations of CDS services. The ADS interface allows counting of the number of papers in which the CDS services are cited in the text of the paper. **In the calendar year 2016, 660 refereed papers cited the word SIMBAD, 382 the word VizieR, and 92 the word Aladin** in reference to the respective CDS services. This represents approximately the same level of citation as in 2015.

Here brief reports on the services and more details will be provided in presentations by the service teams. A presentation on the scientific activities of CDS research staff will be made in the meeting.

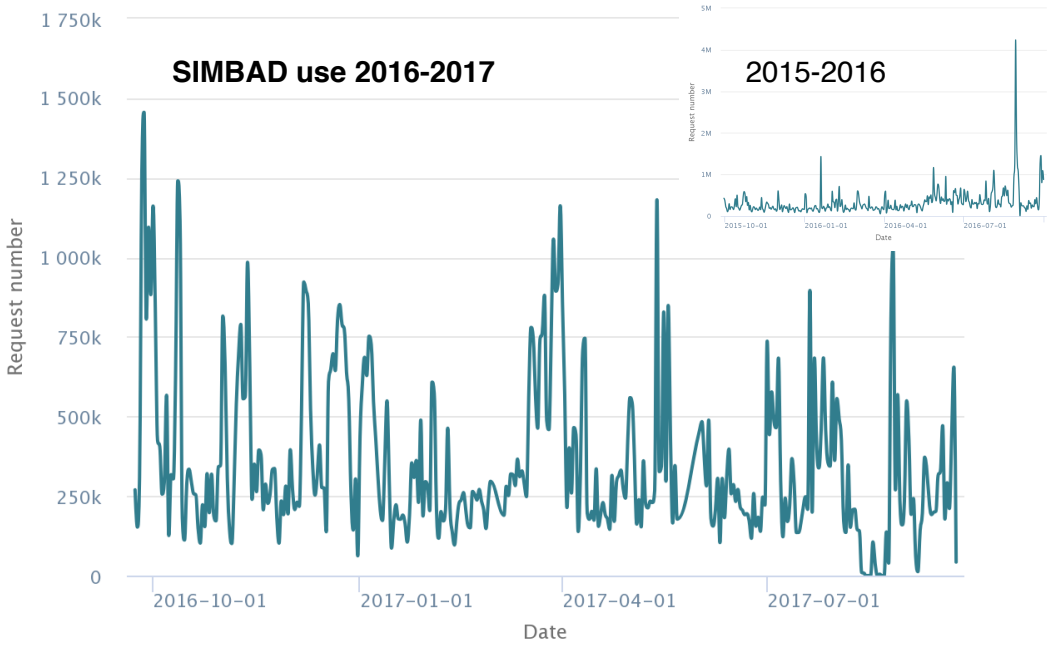
3.1 SIMBAD

The content of the SIMBAD database grew by about 800,000 astronomical objects in 2016-2017, an increased rate compared to ~500,000 per year in the previous 3 years. The table indicates the status on September 24, 2017.

SIMBAD Content

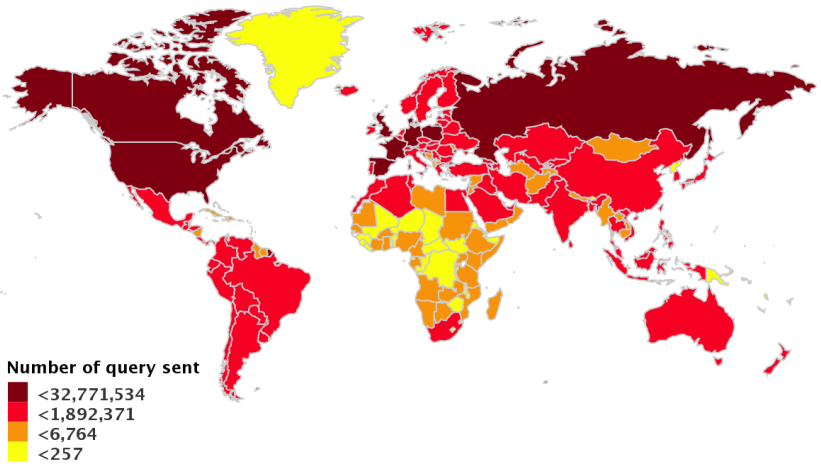
	2013	2014	2015	2016	**2017**
Objects	~7,342,000	7,556,225	7,998,221	8,493,230	**9,298,005**
Identifiers	~18,162,000	18,563,653	22,322,732	23,553,608	**26,799,877**
Bibliographics references	~285,000	294,449	308,588	323,689	**336,179**
Citations of objects in papers	~10,000,000	10,749,766	12,126,329	14,352,859	**16,169,095**

The usage of the SIMBAD service remains very high with an average of 552,000 queries per day in 2017, with a global distribution (see map). The query rate is an increase over previous years (see SIMBAD Service Use table), and compared to 2016 there were fewer large spikes in the usage as shown in the graph below.



SIMBAD Service Use

	2013	2014	2015	2016	**2017**
Queries/day	520,000	506,000	500,000	510,000	**552,000**



In this period we have undertaken a deep review of the SIMBAD architecture in order to ensure that the core parts of this service are set up in a way that will support future evolutions. We have used a “whole CDS” approach in the architecture review to draw on the expertise of all of the staff. The main structural issue that was identified is the way that the various components of SIMBAD communicate. These components include the data ingestion tools and the various client programs that communicate to the main SIMBAD “server” via a mixed range of interfaces. A number of important changes were found to be necessary to simplify and improve the communication interfaces via implementation of a more simplified set of APIs. Also a number of the functions of the current server should be split out to separate modular processes. One example is that the generation of the SIMBAD results pages is currently done by the server, limiting the flexibility for changing the SIMBAD user interface. Addressing this structural issue will enable the use of recent technologies for improving the SIMBAD user interface and the way the results or queries are presented. This review has led to a list of tasks for updating the SIMBAD architecture.

Work on a major update of the DJIN tool was started in October 2016. DJIN is used by the documentalists to process the astronomy journals with the main task of identifying astronomical objects (names, acronyms and identifiers) in the text and tables of articles, an essential step at the beginning of the chain of CDS processes. DJIN operates on PDF files, and is linked to the objects already known in SIMBAD. The update takes into account updates in PDF, and uses text mining capabilities. Following tests in our R&D program in recent years, we have chosen to use GROBID[14] which is a machine learning library for extracting, parsing and re-structuring raw documents such as PDF into structured TEI (Text Encoding Initiative) documents with a particular focus on technical and scientific publications. A contract engineer was employed in October 2016

[14] P. Lopez: “GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications”, in proc. of ECDL 2009, 13th European Conference on Digital Library, Corfu, Greece, 2009. (link - https://lekythos.library.ucy.ac.cy/bitstream/handle/10797/14013/ECDL069.pdf?sequence=1)

(V. Kaestle) on this project and we expect the first internal release of DJIN2 before the end of 2017. (The creator of GROBID, Patrice Lopez has visited CDS and participated in the LISA conference, providing additional expertise for the CDS use of this library.) There still remains a big challenge to make this into an efficient tool that the DJIN documentalists use as the core of their work.

Other improvements to SIMBAD procedures and scientific content include:

- A new database system has been implemented for the Dictionary of Nomenclature to facilitate multiple staff members to update the Dictionary.
- A revision of the format for data structures in SIMBAD was made in order to take into account the high precision of the Gaia data.
- Gaia DR1: The TGASS DR1 has been ingested in SIMBAD. Some complications occurred as the Gaia catalogues did not take into account HIP double stars resolved in Tycho2. This has been resolved and this operation resulted in an improvement of the astrometry for about 2 million stars in SIMBAD together with the addition of new parallaxes and the Gaia G magnitude.
- LEDA (PGC): The last version of the LEDA database of galaxies has been ingested in SIMBAD. About half a million of new galaxies have been created, contributing to the larger than usual increase in the number of SIMBAD objects in this period.
- New object types : ***"Gravitational Wave Event"*** has been added. There are only 4 cases right now, but this will obviously increase. ***"Hot subdwarf"*** has been added with historical objects being recovered, and new objects coming mostly from the SDSS and LAMOST pipelines. SIMBAD now contains about 3600 hot subdwarfs and 34500 white dwarfs.
- The SIMBAD TAP service has been improved so that it supports MOC (Multi-Order Coverage maps - IVOA standard) as an output option.
- Improvements to scripts for processing of bibliographic data feeds from publishers (supported by contract engineer, T. Delacour)

3.2 VizieR

The standard VizieR treatment of catalogues has been in smooth operations throughout 2016-2017. The VizieR Content table below indicates the growing number of catalogues in the database. A new contract documentalist started in March 2017 (T. Pouvreau) to help with the ingestion. The goal of this contract is to allow the service to cope with the increasing publication rate of the catalogues relevant for the CDS services.

VizieR Content

	2013	2014	2015	2016	**2017**
Number of Catalogues in the VizieR database	11,579	12,691	14,065	15,485	**16,528**

In terms of use of the service there was an average of 326,000 queries/day in 2017. This is a decrease compared to 380,000 in 2016, but there may be a several factors that affect this number. With the increase in the number of ways of querying VizieR, namely via TOPCAT and also via the VizieR TAP (Table Access Protocol) service, we are making large queries more efficient and this

may be contributing to the difference, just as the X-Match service changed the query statistics in 2014-2015. We continue to assess these changes, and to try to determine whether this is a fluctuation or a trend.

VizieR Usage

	2013	2014	2015*	2016	**2017**
Web Service Queries/day	600,000	530,000	300,000	380,000	**326,000**
Associated Data Service Queries/day					**470**
VizieR TAP service Queries/day					**5,250**

VizieR Associated Data Service

In 2016 VizieR opened a major new feature, the VizieR Associated Data Service for non-tabular data. This service includes an ingestion pipeline and the metadata mappings that are necessary to take into account the description of spectral, time-series and image data. Authors may provide these data, and the ingestion process (supervised by documentalists and engineers) maps the description to the standard IVOA ObsCore data model. These data are discoverable through a web interface, and also through IVOA protocols.

This service is now operational and we have been ingesting “associated data” that comes with newly arriving catalogues, but with a priority on ingesting the existing associated data already in the 16000+ catalogues already in VizieR. This is a huge undertaking and we started by processing the catalogues with the largest volumes of associated data and working our way downwards in size. This cautious approach allows us to build experience with the process and to understand the resources needed to operate it. At the moment however, the added workload of the ingestion procedure is still holding back the systematic ingestion of newly arriving associated data. Nevertheless, the new service receives almost 500 queries / day, and was presented with an oral contribution at the LISA VIII conference organised by the CDS. It comes as a good surprise that astronomers are already using the service, even though it is very recent and it has not yet been advertised much to the astronomical community. This will be a point of focus for the coming year: we aim to build a higher awareness for this tool in the community.

Large Catalogues

10 large catalogues have been ingested in VizieR in the 2016-2017 period. This includes the Hubble Source Catalog, SDSS DR12, UCAC5 and a number of ESO public surveys catalogues: KiDS DR2 and DR3, Viking DR2, VVV DR2. The priority of ingesting ESO Public surveys announced last year has been followed, and a test page showing the ESO Public surveys in the CDS system has been created[15], and these catalogues can also be found by using the ESO keyword in the VizieR interface.

The generation of HiPS catalogues is gradually becoming part of the large catalogues pipeline. 15 new HiPS catalogues have been generated for large surveys in the later year. The long-term goal is to make every VizieR table with positions available as a HiPS catalogue.

[15] http://cds.u-strasbg.fr/eso

Internal Improvements

A catalogue follow-up interface (not public, for use only by CDS staff), similar to a ticket system, has been developed and deployed to track the ingestion procedure and follow the different steps. Through this new interface, astronomers (currently P. Ocvirk and C. Bot), are able to see and access the catalogues ingested by the documentalists, and requiring their expertise for scientific validation. The interface also allows the validators to signal problems, and intervene in the catalogue publication when necessary. This ensures focused interaction between the astronomer and the documentalists working on a catalogue. This is an important improvement in the sense that it streamlines some of the process of ingestion and validation, makes it clearer and traceable, and regroups the relevant information in a single place. It clearly helps organising our work.

The communication within the large catalogues team has also been improved by the use of the Redmine project management application. It is used to report and track the status of the ingestion of a given catalogue, while keeping a detailed history of the interactions and work done.

VizieR catalogue metadata are now routinely extracted from the VizieR database to populate the MOCServer. This allows for rich metadata retrieval and facets-like filtering by clients like Aladin and Aladin Lite.

Digital Object Identifiers

We have sought agreements from the AAS and A&A journals , and also the ADS for our proposed mechanisms for managing DOIs associated with VizieR catalogues. We are taking into account the detailed comments received and will propose a definitive process as soon as possible.

3.3 Aladin and Aladin Lite

A new major version of the Aladin desktop application, Version 10, has been prepared for release at the ADASS meeting in October 2017. This new version has a more modern “look and feel”, and many new features have been enabled via the implementation of IVOA standards: TAP, MOC, Datalink. One of the major revisions is the Data Discovery Tree that enables exploration of all the data available in the Virtual Observatory, a major improvement on the earlier “All-VO” capability in previous versions. It presents the data available via CDS services, and via the IVOA Registry in a uniform way. These efforts have been supported by ASTERICS DADI for the implementation of IVOA standards, and by CNES for a specific Gaia TAP access interface.

The Aladin Desktop interface is intended to be “full-featured” and allows for complex queries and supports our “power-users”. A tutorial “Advanced Usage of HiPS and MOC” (available on the Euro-VO science tutorials pages[16]) has been developed and exploits many of the new capabilities, in particular those for managing complex sky regions.

Aladin Desktop is being used for the next generation observation preparation tools, notably the STScI Astronomers Proposal Tool[17] for HST and JWST, and also the ESO GuideCamTool++.

As described in the highlights, HiPS is now an IVOA standard, and there is much interest from scientists and projects for the use of this data, with a corresponding large increase in the HiPS data that has been ingested into our system. The number of HiPS data sets has now grown to 380, representing 137 TB of data (see the Aladin Content table below). A further major increase is

[16] https://www.asterics2020.eu/dokuwiki/lib/exe/fetch.php?media=open:wp4:hipsmocs_4asterics.pdf

[17] http://www.stsci.edu/hst/proposing/apt

underway as we are currently ingesting the Pan-STARRS[18] DR1 imaging (50-145 TB depending on resolution) (in coordination with STScI/MAST) as well as the SkyMapper[19] DR1 images (in coordination with the Australian SkyMAPPER project) into the system.

Aladin content: HiPS image data sets

	2012	2013	2014	2015	2016	**2017**
HiPS data sets	81	128	175	236	325	**380**
HiPS Volume	19 TB	30 TB	45 TB	50 TB	105 TB	**137 TB**

In terms of usage we note an increasing popularity of Aladin and Aladin Lite, with the number of queries (or “actions”) of order 300,000/day, and a marked increase in the “hosts/day” which is driven by Aladin Lite.

Aladin usage: Aladin and Aladin Lite

	2012	2013	2014	2015	2016	**2017**
Queries/day	13,208	56,053	160,103	266,047	283,177	**298,356**
Audience: hosts/day	266	271	1285	1797	1968	**2630**

We also keep track of the implementation of Aladin Lite with a list[20] of web sites and services that use this CDS modular widget as part of their own service. ESASky is the most prominent of these and has been strongly supported, with as mentioned in the highlights, a visit to ESAC in May 2017. Another impressive implementation is in the NASA LAMBDA toolbox[21].

HiPS Python library

A first version of a Python library to handle image HiPS has been developed in the frame of a Google Summer of Code program by T. Boch and external student Adeel Ahmad, and co-mentor Christoph Deil (Heidelberg). This package allows for FITS cutout generation from HiPS tiles for an arbitrary WCS. The source code is available on github (https://github.com/hipspy/hips) and documentation on https://hips.readthedocs.io . We rely on this development to increase HiPS visibility and usage in Python astronomers community. Eventually, our goal is to have this library integrated as an Astropy affiliated package.

ipyaladin

ipyaladin is an IPython widget for Aladin Lite. It allows for easy integration and control of Aladin Lite in Jupyter notebooks, which have gained a lot of traction and interest in the Python scientific

[18] https://panstarrs.stsci.edu

[19] http://skymapper.anu.edu.au

[20] http://cdsbug.u-strasbg.fr/redmine/projects/aladinlite/wiki/Implementations-aladin-lite

[21] https://lambda.gsfc.nasa.gov/toolbox/footprint/aladin/aladinLAMBDA.cfm

community lately. A first working prototype has been developed by a contractor under the supervision of Aladin Lite's developer. Source code, installation instructions and usage examples are available at https://github.com/cds-astro/ipyaladin .

3.4 The CDS Catalogue Cross-Match Service

The X-Match service continues to operate smoothly, with another large jump in 2016-2017 in the number of queries submitted via the programmatic interface (HTTP API). See the table of X-Match usage below.

Following the paper published last year on multi-catalogue cross matching (Pineau et al., A&A 2016[22]), we have monitored reactions in the community and have interacted with potential collaborators (and competitors!).

A number of tests of different technologies for computationally expensive cross matching are reported in the R&D report in Section 3.6.

CDS X-Match Service Usage

	2013	2014	2015	2016	**2017**
Web interface (jobs/day)	15	16	20	30	**33**
HTTP API (jobs/day)	47	50	580	889	**1254**
Associations/ day (Web Interface)	~13,000,000	~70,000,000	~55,000,000	~104,000,000	**~180,000,000**
Associations/ day (HTTP API)	~298,000	~1,600,000	~6,600,000	~6,700,000	**~12,000,000**

3.5 Service Integration and Portal Development

CDS Portal

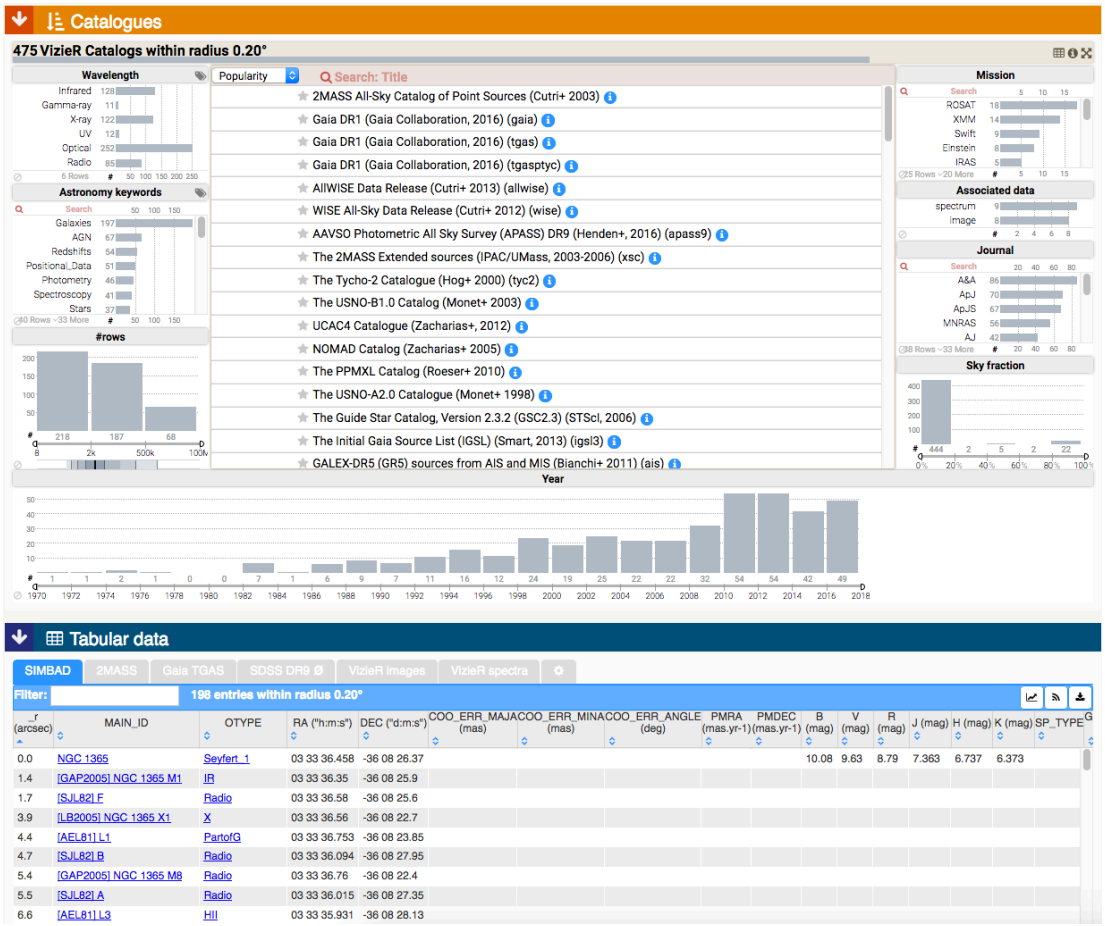
As shown in the highlights, a newly rejuvenated and improved CDS portal[23] was released in mid-December 2016. The development of this portal was set up as the first step for the new CDS Service Integration activity, a dedicated effort to enable and improve integrated use of CDS services.

The new portal has an architecture that is based on a set of modular independent interconnected widgets. It provides available information and data for a given object or position on the sky. The widgets can be reused as standalone tools, embedded in other contexts (eg: the Astrodeep portal) or even integrated in third-party pages. Aladin Lite is a central part of the new portal as it provides interactive browsing and preview of available images and catalogues. Another innovative aspect is the ability to filter the available catalogues through various *facets* (wavelength, astronomical keywords, mission, journal) to narrow down efficiently the list of relevant VizieR catalogues. This facet filtering is enabled by the MOCServer which is an essential part in the new portal

[22] https://www.aanda.org/articles/aa/abs/2017/01/aa29219-16/aa29219-16.html

[23] http://cdsportal.u-strasbg.fr

architecture. It provides very fast sky coverage comparison as well as metadata on the available resources. The development of the portal required various exploratory tests and studies, for which we have used two trainee students.



HTTPS

Another important development across all CDS services is the deployment of HTTPS support (a web protocol for secure communication). This is especially important for programmatic access for Web applications. It has been motivated both by Aladin Lite needs and requests from our partners who are in some cases required by their authorities to implement this protocol. HTTPS has been deployed on the SIMBAD main site, the Sesame server at Strasbourg, the CDS Portal and one of the HiPS mirror servers. This first phase allowed us to gain experience in properly setting up our servers. HTTPS deployment will continue in the coming year as to enable HTTPS access to VizieR.

CDS login

Service Integration also concerns the CDS login service which provides the (non-mandatory) authentication for the CDS Portal. The CDS Annotations, the Cross-Match and SimWatch services also use the CDS login. Its backend has been refurbished and now uses a relational database to store the user information to make it more robust and easier to maintain. While non-authenticated use of our services remains an important aspect of making CDS services easy to use, authentication is very likely to become more important in the future when providing computational services close to the data that require management of user spaces.

3.6 R&D

In the 2016-2017 period we have pursued a varied R&D program with both operational actions and exploratory work to prepare the future. This work was carried out by CDS software engineers with the help of 10 interns and 2 short contracts.

We have made studies aimed at improving the availability and reactivity of VizieR and SIMBAD, in particular for the duplication of these services, and for better indexing of the data. A first operational implementation took place in TAP VizieR.

Both the VizieR and SIMBAD services have mirrors at external sites. The maintenance of the mirrors requires a lot of time because the installations are dependent on the host machines. Numerous tests have been carried out to use Docker components and to facilitate this maintenance work. A first VizieR Docker version is currently used for local testing of the VizieR update and will soon be implemented for some of the external mirrors. A Docker version of the X-Match has also been created, it will be used to gain time during the test phase of the service before the upgrade of the service in production.

An evaluation of NoSQL and column-oriented solutions has been carried out to improve the response times of queries by criteria in SIMBAD (currently a relational database involving a large number of joins).

We continued the study of Apache Spark in the frame of the X-Match service, through collaborations, in particular with the IN2P3. The aim of this study is to simplify the X-Match and to scale it up as easily as possible. This work is a part of a larger study that aims to enable code execution as close as possible to the data.

With the arrival of Big Data it becomes crucial to be able to carry out treatments as close as possible to the data: it is a radical change for a data center that switches from a data provider to also being a provider of "computing" resources with more advanced user management (customer account management, storage spaces, security / privacy aspects). The relevant R&D work focused on Notebooks and especially Jupyter Notebooks has been done.

We have started the development of a library for the IVOA VOSpace protocol in Python which is mainly dedicated to the internal needs of our services.

We continued also the work around 3D Visualisation in a Web browser. The objective is to progressively visualise large simulation data or large catalogues whose volume (several TB) exceeds the capacities of a browser.

We have competitively obtained Research Credits from Amazon AWS to carry out tests with our services. We plan to test two aspects of HiPS surveys: Access of HiPS from the cloud; and HiPS generation compute time (which can be up to a month using local resources). We will also test the use of Spark for the X-Match and SIMBAD.

A “communication and design” intern worked on the definition of a method for producing video tutorials. The objective is to enable CDS people to easily generate video material for users of our services. The first video of the new tutorials have been produced and have been used at the EWASS meeting.

4. Projects

Virtual Observatory and ASTERICS

CDS continues to play a leading role in the development of the Virtual Observatory. The contributions to the IVOA are listed in the document provided “CDS Participation in IVOA” which has been updated with a list of the standards developed in 2016-2017, and the CDS contributions to the IVOA interoperability meetings in Trieste, October 2016, and Shanghai, May 2017.

CDS staff currently hold a high number of Chair positions in the IVOA. These are listed below, with the most recent addition being Ada Nebot who took on the role of the Chair of the Time Domain Interest Group in May 2017. Other recent changes include M. Allen and F. Genova ending their terms as Chair of the Science Priorities Committee (CSP), and of the Standing Committee for Standards and Processes (SCSP) respectively.

- Deputy Chair of the IVOA Executive Committee - M. Allen
- Chair of the Applications Working Group - P. Fernique
- Chair of the Data Access Layer (DAL) Working Group - F. Bonnarel
- Chair of the Semantics Working Group - M. Louys
- Chair of the Data Curation & Preservation Interest Group - F. Genova
- Chair of the Time Domain Interest Group - A. Nebot

In addition to the highlight of the standardisation of HiPS, other important achievements are the finalisation of the first set of standards to address the IVOA priority area of enabling VO access to multi-dimensional data, largely guided to completion by the DAL WG chair. We also note good results from an effort to include more scientists at the IVOA meeting, for example J. Sorce made a presentation on the experience of the ASTERICS-DADI VO Schools.

The Astronomy ESFRI and Research Infrastructure Cluster, ASTERICS, brings together astronomers and astroparticle physicists of 23 European institutes to help world-leading facilities, such as SKA, CTA, KM3NeT, and E-ELT, work together to find common solutions to their Big Data challenges, their interoperability and scheduling, and their data access. CDS leads WP4, Data Access, Discovery and Interoperability, which gathers VO teams and teams from the large ESFRI and ESFRI-like projects to optimise the usage of the data through the Virtual Observatory. WP4 is in charge of about one third of the budget (4.5 M€ over four years).

ASTERICS is 30 months into its work program. As shown in the highlights we had a successful mid-term review. The submitted review documents included the following description of the CDS led DADI Work Package:

> Progress is made on the Virtual Observatory (VO). This group continued the practice they established before the project, organising workshops and schools for developers and scientists. ASTERICS Technology Forums are devoted to discussions of the partners’ relevant technological activities and of potential collaborations, and to the preparation of the global standards discussions in the International Virtual Observatory Alliance (IVOA) meetings. The ESFRI and European Data Provider Forums, including Training Events, are meetings dedicated to gather the ESFRI and pathfinder requirements and feedback, and allow us to check the wider relevance of the VO developments. Schools are organized on the scientific usage of the VO and are open to participants from all over Europe, mostly early career astronomers. ESFRI staff are invited to participate in the schools, to improve their understanding of the usage of VO and to be able to refine their requirements. ESFRI staff who had participated in the First School as students were involved in the preparation of the second one, updated or prepared tutorials, and participated as tutors, making them active participants in the dissemination of knowledge about the VO framework.

ASTERICS has established new collaborations through the workshops. This has concrete results such as the publication of ANTARES data in the VO and the usage of the VO tool Aladin Lite (CDS) for gravitational wave follow-up in the so-called GWSky tool. The code is stored in the official LIGO/Virgo gravitational wave astronomy repository to support the effort of electromagnetic follow-up of gravitational wave detections. Building blocks from the Virtual Observatory framework of standards and tools, for instance, the IVOA Provenance concepts, have been adopted in the CTA pipeline.

The Virtual Observatory is an ongoing development organised in a cyclic way, where the state of the art is demonstrated in a series of Forums and schools to developers, data providers and users. The feedback and new requirements gathered during these meetings inspires new development. ASTERICS is actively working towards the inclusion of the data produced by the ESFRI facilities and their pathfinders into the Virtual Observatory. This will open up these data products, first of all to the experts in the field, then to the astronomers working in adjacent fields for multi-messenger astronomy, and finally to the general public through their inclusion in citizen science projects.

Important events for the ASTERICS DADI work in 2016-2017 are shown in the list below, with CDS hosted events shown in bold:

- RDA 8th Plenary meeting during the International Data Week, Denver, Colorado, 11-17 September 2016
- ADASS XXVI Conference, Trieste, Italy, 16-20 October 2016
- IVOA Interoperability Meeting, Trieste, Italy, 21-23 October 2016
- **ASTERICS DADI Virtual Observatory School, Strasbourg, 15-17 November 2016**
- **ASTERICS DADI-LOFAR meeting, Strasbourg, 18 November 2016**
- **DADI Provenance Day, Strasbourg, 13 December 2016**
- **ASTERICS DADI Technology Forum 3, Strasbourg, 22-23 March 2017**
- RDA 9th Plenary meting, Barcelona, Spain, 5-7 April 2017
- IVOA Interoperability meeting, Shanghai, China, 14-19 May 2017
- RDA 10th Plenary meeting, Montreal, Canada, 19-21 September 2017

VESPA / EuroPlanet Project

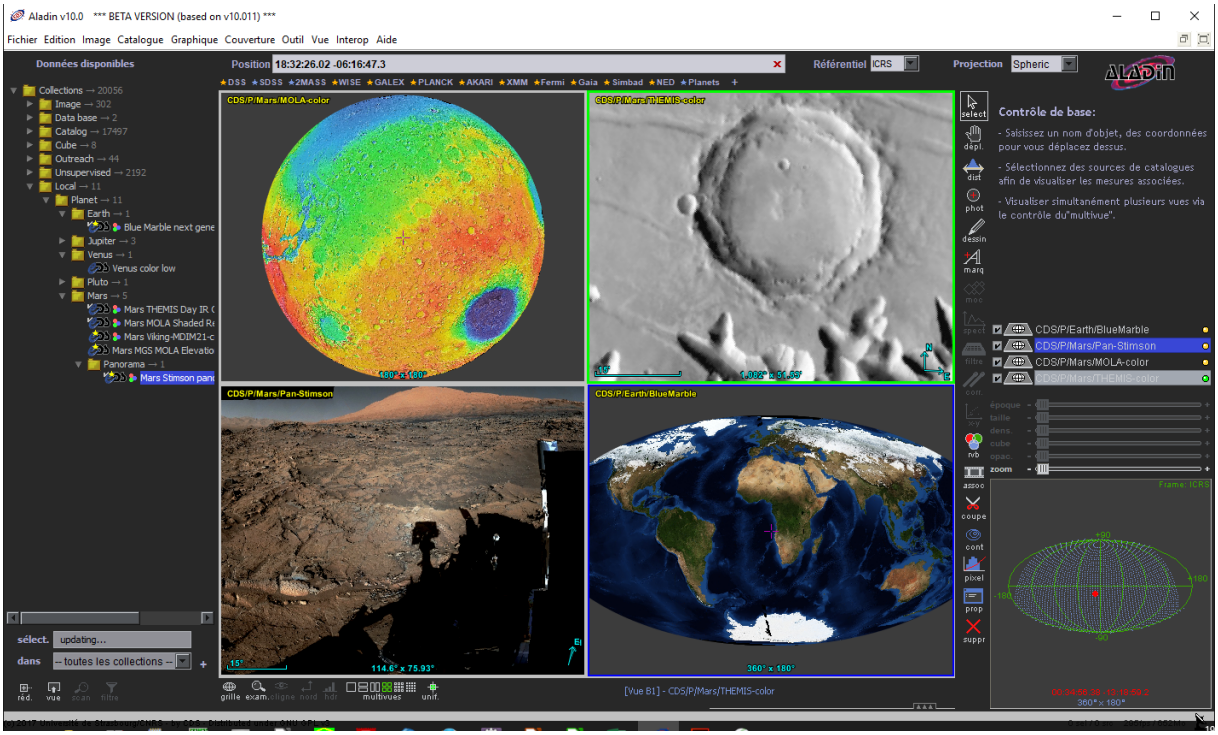
CDS is participating in the VESPA (Virtual European Solar and Planetary Access) activity in the Europlanet 2020 Research Infrastructure programme funded under Horizon 2020 (2015-2019). Significant progress of the CDS contribution to this project has been achieved in 2016-2017. This work has been done in close collaboration with VESPA colleagues at the Observatoire de Paris and at the IMCCE (Institut de mécanique céleste et de calcul des éphémérides).

The main elements of the work are listed below:

1. The first step was a *proof of concept* that showed that Aladin, and associated IVOA technologies could be used for planets
2. HiPS technology adaptation for planets. In progress:
   i) Hipsgen code improvements for supporting planet mosaic maps in cartesian projection
   ii) Collaboration with Chiara Marmo (GEOPS) for using FITS planet image collections as input.
   iii) Generation of several high resolution planet HiPS (ex: Mars Themis IR)
3. Aladin Desktop v10 adaptation for HiPS planets (in progress - see figure below):

i) Planet projection supports
ii) List of Planets plugable in the Aladin data discovery tree
iii) EPN-TAP support

4. Planet panorama R&D



The next steps will involve:

- deciding which institute(s) should generate / distribute HiPS planet data
- continue the Aladin v10 adaptations (projection control, dedicated planet tool , for instance measurement in metres)
- Adapt Aladin Lite (reverse longitude display)
- Participation in IVOA discussions about adjusting IVOA standards to Planets (notably VO registry, HiPS, and TAP)

We also note the creation of the IVOA Solar System Interest Group led by B. Cecconi (Observatoire de Paris / VESPA)

ASTRODEEP

The ASTRODEEP portal[24], developed at CDS during the ASTRODEEP project (Jan 2013 - Jan 2017) was released in mid-2017. This portal provides access to the catalogues of 4 galaxy clusters plus 4 parallel fields corresponding to the Frontier Fields studied in Merlin et al. (2016), Castellano et al. (2016), and Di Criscienzo et al. (2017).

HiPS images have been generated in 10 bands, both for original and processed images, together with colour compositions. A set of interacting widgets allows exploration of the image and catalogue data, with SEDs of each galaxy and data can be exported to other VO tools. The clusters are : Abell 2744, MACS J0416.1-2403, MACS J0717.5+3745 and MACS J1149.5+2223

[24] http://astrodeep.u-strasbg.fr/ff/

The Research Data Alliance and other generic data sharing initiatives

The Research Data Alliance, created in March 2013 with support of the European Commission, NSF and Australia, continued to grow. It has now more than 6000 members from 130 countries, with 88 Working and Interest Groups working on different aspects of data interoperability challenges. A summary of its current status can be found at the link below[25]. F. Genova who has been one of the members of RDA Technical Advisory Board (TAB) since the beginning, was elected as co-chair in October 2015. Her term as a TAB member finishes in September 2017. She applies for re-election and for a second chairpersonship mandate of two years if she is re-elected. She participates in the RDA Council meetings and is a member of the Council's Strategy subcommittee. She was invited to participate in the series of European projects set up in support to the RDA. The third one began on 1 September 2015 and will end in February 2018. She is in charge of research community engagement, and participates in the other project activities towards the international level and technical communities. The next project, RDA Europe 4.0, was submitted in March 2017, and recently entered in negotiations with the European Commission, for a start expected in February 2018 and 27 months duration. This project will focus on the establishment of RDA national nodes, and CNRS/CDS is in charge of the French one, plus the coordination of the programme to be set up to identify and support Ambassadors to science disciplines. Françoise Genova is already identified as the RDA ambassador for astronomy.

Lessons learnt when building the IVOA are shared with the RDA through F. Genova's participation in the RDA Technical Advisory Board. Other people involved in astronomical data sharing also participate regularly in the RDA activities, including A. Schaaff and M. Louys. One aim is to share astronomy expertise with RDA, and to make sure that our requirements are taken into account. F. Genova brings RDA topics in ASTERICS and the IVOA through regular discussions in the IVOA Executive and through the Data Curation and Preservation sessions in the IVOA Plenaries. Certification, provenance, disciplinary repositories, the long tail of data, DOIs and data publication are among the topics of common interest. It is worth noting that the IVOA Registry of resources is taken as a model by the RDA Group which works on interoperability of materials data, thanks to R. Hanisch and R. Plante, who previously worked in IVOA-US and are now working at NIST.

F. Genova was in particular actively involved in the RDA DSA/WDS Certification of Data Repository Working Group, which produced in 2016 a set of criteria for certification aligning the DSA and WDS frameworks. The experience gained by CDS when applying for certification successively in the two frameworks was very useful in that context. She has been a member of the Data Seal of Approval Board since January 2016. DSA and WDS now work together to evaluate applications for certification. They just launched the Core Trust Seal[26], based on the work of the RDA Working Group, which is envisioned as the initial level of a global framework for repository certification.

The European landscape is evolving fast with the 'Open innovation, open science, open to the world' vision driven by Commissioner Moedas, who is in charge of Research, Science and Innovation. The European Commission launched in 2016 the European Open Science Cloud (EOSC) endeavour, which aims at providing a set of services for data management, data sharing and computing. This supports CDS strategy to be involved in the Research Data Alliance (RDA) at the European and international level. Also, the IVOA Registry of Resources was entered in the B2FIND "generic" registry of the European EUDAT project, which is expected to be one of the elements of EOSC. The RDA is able to be one of the key support organizations to the EOSC. On the other hand, F. Pasian (INAF) ensures the liaison between ASTERICS and EOSC. F. Genova is also a member of the European Commission Expert Group[27] on Turning FAIR data into reality.

---

[25] https://www.rd-alliance.org/sites/default/files/attachment/RDA_in_a_nutshell_September_2017_1.pptx

[26] https://www.datasealofapproval.org/en/news-and-events/news/2017/9/11/coretrustseal-certification-launched/

[27] http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464

5. Status of the 2016 Recommendations of the Scientific Council

1. Astronomy Research: The balance between service, research and teaching for the scientific staff is important and should be maintained. This is crucial for retaining astronomical staff who are both knowledgeable and committed to CDS.

Response:

We thank the Council for this recommendation as it recognises the multiple roles of the CDS scientific staff: service, teaching, and research. CDS scientific staff in the CNAP system have a ~30% service task, 66 hrs/year teaching (~17%), plus research research time. The organisation and balance of these components is managed by the individual staff members under the responsibility of the Observatory Director. CNRS staff also must balance their contributions to CDS and other research work.

Retaining the CDS expert scientific staff is a high priority, as is the recruitment of new staff. CDS seeks to provide a good scientific environment for the research staff with support for scientific work (travel, collaborative visits, computing, observing), and the staff themselves are also successful in obtaining funding for their scientific activities. It is absolutely essential that CDS science support not fall below the current level, and that any changes are balanced with replacements. We note that a senior CDS science staff member will retire ~2020.

2. Staffing: It is important to fill key positions as they become vacant and to plan for a gradually expanding workload in future. The following posts are critically important for the immediate future:

   a) A system manager to take charge of the IT infrastructure of the Strasbourg Observatory (nominally shared at the 50% level with the CDS).

   b) A documentalist. The data ingestion for the VizieR service is currently critically dependent on a contractor position. Training requires 1-2 years to reach a sufficient efficiency and quality level.

   c) Two Software engineers for the CDS services. In order to manage operations and development, and to ensure sustainability, it is essential that Vizier and SIMBAD are each supported by two software engineers; this is critical to risk mitigation. Major elements of the VizieR service must be re-written and updated in 2018-2022 and this requires a software engineer.

Response:

These are the top priorities for the Software Engineer and Documentalist positions.

a) Fortunately the top priority has been addressed by CNRS (with support from the Université de Strasbourg) with an new system engineer starting in October 2017.

b) The Documentalist tasks are currently dependent on 2 contractors, and another contractor is currently being recruited so we will soon have 3 contractors with this role. This simply reflects the staffing needs in order to manage with the increased volume of journal articles being processed for catalogues and astronomical objects. The training times are long, and given the large investment in training, these tasks are better suited to permanent staff. We seek a documentalist position as a first priority to address the needs.

c) The long term sustainability of SIMBAD and VizieR requires duplication of the effort, and it should be recognised that these services have expanded scopes in response to scientific needs (e.g. VizieR associated data, complex queries, high volume demands). There is an immediate need to support major works on the core of SIMBAD (as described in this report) and also to re-new major elements of VizieR. We have recruited a highly experienced engineer on a contract (to start in May 2018) to help with these immediate tasks so that operations and development can be managed.

Contractors are an important element of the CDS staff, and in practice having contractors provides a steady stream of people, some of whom can eventually be recruited to the CDS staff.

In the 2016-2017 period CDS has had the following contract staff:

- Four contract staff working on core CDS functions (2 documentalists, 2 Software Engineers)
- Three contract staff on specific projects (ASTERICS - Software Engineer, ASTRODEEP - Software Engineer, CNES supported Software Engineer for specific Gaia work)
- Two Postdocs (VIALACTEA/CDS, and ASTERICS)
- Two PhD Students, one who has successfully defended his thesis in September 2017
- Ten short term interns, and 2 short contracts as detailed in the ‘Trainees at CDS’ document

  (In late 2017 we will welcome another CDS postdoc, and also the ASTERICS postdoc is re-advertised with the current person finishing in December 2017)

3. Major Journals: The CDS has a very good relationship with most of the important astronomy journals, which facilitates the efficient transfer of published data into the databases. The Director should engage with the editors of MNRAS with a view to streamlining the CDS obtaining their published data as reliably and efficiently as possible.

Response:

This recommendation has not been fully addressed since the last Council meeting, in part due to time, and in part due to the confusion created by problems with the MNRAS publisher’s (OUP) transition to a new platform for their data feed. An initial formal communication with the MNRAS Board in 2016 did not receive a response. Contact has been made in person during the LISA conference (June 2017). A request for a scientific level discussion has been made directly to the Editor in Chief (September 2017).

4. Solar System Objects: The Council understands that the provision of data for solar system objects has not been part of the CDS mandate. We suggest that they undertake an analysis of the changing landscape in this regard following recent space missions, appreciating that it will continue to evolve into the LSST era. They should consider the implications of their services to stellar, Galactic and extra-galactic astronomy of not keeping track of solar system objects (which are a major source of "noise"). Note, however, that Council does not intend this undertaking to stress CDS core operations.

We have analysed the situation and provide the following detailed response:

The mandate of the CDS is for astronomical objects outside the solar system. The general approach taken by the CDS is to recognise the international data centres that do focus on solar system objects, and to ensure interoperability with these data via standards and practical agreements with selected solar system data providers.

To give some historical perspective, at the French national level there exists another organisation whose mandate includes solar system bodies, the "Service des calculs et de mécanique céleste of the Bureau des longitudes[28]". The practical work is currently done by the "Institut de Mécanique Céleste et de Calcul des Éphémérides (IMCCE[29])" which is in Paris Observatory, and provides and publishes ephemerides of solar system bodies among other tasks. The original set up of the CDS as the "Stellar" data centre, and the expansion of the scope to galaxies and non-stellar objects (Dubois et al. 1983) would likely have been influenced by the existence of this organisation. The key point for the latter is the fact that data on extragalactic objects were available (for example at the IAP Library) which began to provide staff to CDS at that time. We note that the Europlanet VESPA[30] (Virtual European Solar and Planetary Access) activity also provides services for solar and planetary data.

With the CDS mandate, solar system objects are not collected in SIMBAD, and journals specialising in solar system objects are not processed by CDS. However a large number of catalogues of solar system bodies are included in the VizieR catalogue service; there are currently 909 catalogues in VizieR with the keyword "Planets+Asteroids". In addition there has been a long term collaboration with IMCCE on the interoperability of solar system data, in particular the interactions of VizieR and Aladin with the IMCCE SkyBoT[31] (Sky Body Tracker) service. With the combination of the most important solar system body catalogues in VizieR, and the expert services provided by IMCCE, we do actually keep track of (or at least make interoperable) solar system objects so that they are easily compared/combined with other CDS data.

The IMCCE SkyBoT service has the main function of searching and identifying solar system objects in astronomical images. SkyBoT uses the ASTORB database of "high-precision osculating orbital elements for all numbered asteroids and the vast majority of unnumbered asteroids", which is maintained by Lowell Observatory[32]. It contains 734626 computed orbits based on astrometric

[28] created in 1795

[29] https://www.imcce.fr/langues/en/

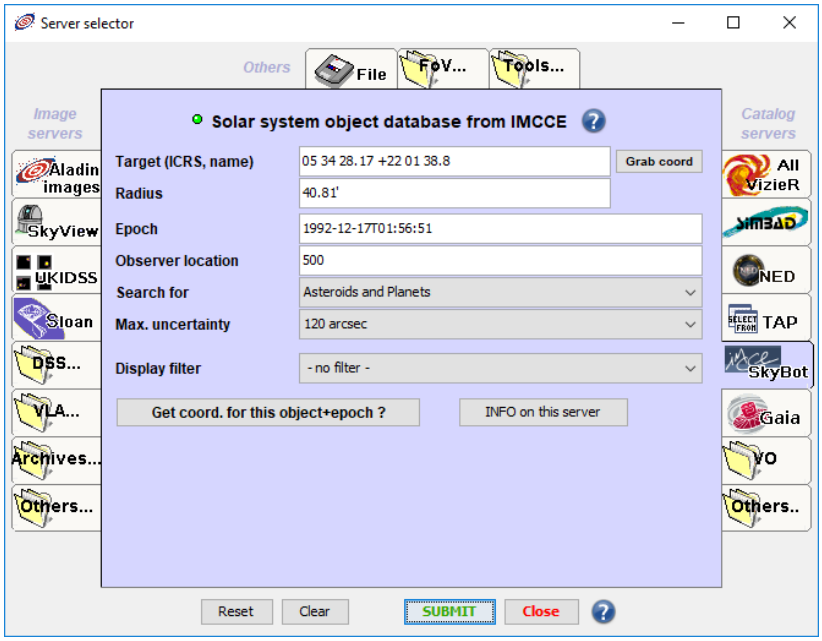
[30] http://europlanet-vespa.eu

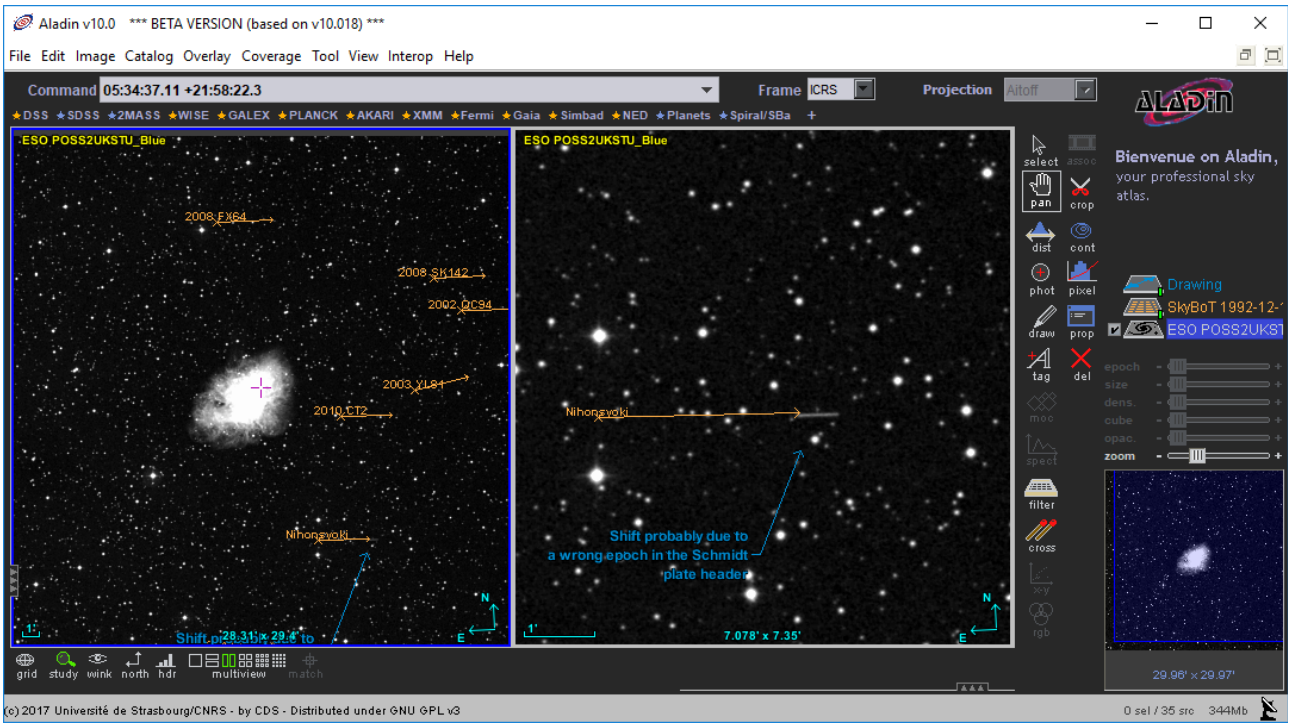
[31] http://vo.imcce.fr/webservices/skybot/

[32] ftp://ftp.lowell.edu/pub/elgb/astorb.html

observations from the IAU Minor Planet Center[33]. As described in Berthier, et al. 2006 (ADASS XV, 351, 367) the SkyBoT database is updated weekly on the basis of the ASTORB database downloaded through the CDS VizieR service (which is itself automatically updated from the Lowell Observatory site once per week).

The SkyBoT service can be queried and visualised through the Aladin desktop application to explore the solar system bodies present in a patch of sky within a specified time window. The figures below show the query form and the result in the Aladin tool with the positions and movement of the bodies shown with arrows.





[33] http://www.minorplanetcenter.net, The Minor Planet Center is hosted by the Smithsonian Astrophysical Observatory, a part of the Harvard-Smithsonian Center for Astrophysics. The Minor Planet Center is funded by NASA.

With the ever larger surveys we expect a significant increase in the known number of solar system bodies. Lynne Jones, Jurić & Ivezić (2015, IAU Symposium, Volume 318, pp. 282-292) in their paper “Asteroid Discovery and Characterization with the Large Synoptic Survey Telescope (LSST)” indicate that the catalogs generated by LSST will increase the known number of small bodies in the Solar System by a factor of 10-100 over all populations. It will be a challenge to integrate this into the current system with the chain of Lowell Observatory - CDS VizieR - IMCCE so we undertake to study this in more detail with these partners.

The issue of solar system bodies being “noise” in scientific studies that need to remove these foreground objects, is not regarded as crucial to the CDS services, provided that our systems continue to allow for the identification of known objects. It will of course be a challenge for the projects (LSST, Euclid etc.) to generate the catalogues from the image data, and we understand that this is a major consideration in the pipelines being designed for such surveys. As such this separation of solar system signals is something we expect will be the responsibility of the processing teams, and CDS would expect to have the science-ready processed products of large surveys, with object classifications ideally provided.

5. Data Ingestion: The CDS has done well at increasing the rate of data ingestion to keep pace with the ever-increasing publication rate by improvements in efficiency and technique. However, there are limits to how much further this can be improved without major innovations. With this in mind they should continue their research into alternative approaches, including considerations of machine learning and other information technology improvements.

Response:

The R&D program at CDS is designed to explore and test new technologies so that we keep up with the latest developments in computing and information processing, and integrate these into CDS services when appropriate. Section 3.6 of this report lists some of the topics currently underway in the program.

One of the topics in recent years has been study of text mining tools that would increase the efficiency of extracting information on astronomical objects described in journal articles. This has led to our use of the GROBID[34] machine learning library in the new version of our journal article processing tool DJIN2.

6. Terms of Reference for the Scientific Council: The revised terms of reference for the CDS Council, which we understand are under consideration, should be circulated when convenient, but before the next meeting of the Council.

Response: The CDS Director has requested the authorities CNRS/INSU and the Université de Strasbourg to check the Terms of reference.

---

[34] P. Lopez: “GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications”, in proc. of ECDL 2009, 13th European Conference on Digital Library, Corfu, Greece, 2009. (link - https://lekythos.library.ucy.ac.cy/bitstream/handle/10797/14013/ECDL069.pdf?sequence=1)

Appendix: Recommendations from the HCERES evaluation committee report

1. Urgently replace the position of system engineer, to allow the training of the new personnel and continuity of service. During the next period, hire staff to account for the increased supports and activities of the Center.

2. Provide a fully replicated data center infrastructure supporting the full range of CDS activities (data ingestion and data services).

3. Define a clear strategy for going to Big Data. CDS is going to « Big Data », requirements in hardware, training with regard to database technology, scientific collaboration, what are the needs with regard to storage technologies, storage requirements?

4. As recommended in the AERES report 2012, the Observatory management, and the funding agencies should maintain the balance between the three profiles: Science, Computer Science, “Documentalists", which is essential to the good functioning of CDS. The acknowledgement of the specific profile of “Documentalists” should be followed by concrete actions to ensure a more efficient hiring process.