CDS Scientific Council meeting 2016

Summary of CDS activities 2015-2016

31 October 2016

1. Introduction

The CDS has pursued a very active program in the 2015-2016 period in support of the core mission of the CDS. Progress has been made on a number of fronts including CDS data curation procedures, the publication of important data sets, and major improvements to CDS services and tools. The CDS has been very visible in the community via strong collaborations with partners, and training activities for scientific users. Many collaborative activities have been undertaken via our participation in, and leadership of projects. These actions prepare the CDS to continue its role as a reference astronomy data centre, and maintain CDS at the forefront of the development and scientific use of global astronomy e-Infrastructures.

The period has been the first year under the direction of Mark Allen. A smooth transition was prepared with the previous director Françoise Genova, and her support throughout the year has greatly assisted the new director to manage this first annual cycle of CDS activities. The CDS staff have equally made an enormous effort to ensure continuity of CDS activities and contributing to future planning of CDS strategies and work plans. The transition has also been strongly supported by Hervé Wozniak (Director of the Observatoire Astronomique de Strasbourg) and the administrative and technical staff of the observatory. Assistance has also been provided at the CNRS-INSU level by Denis Mourard (CNRS-INSU Director Adjoint for Astronomy and Astrophysics), and at the Université de Strasbourg by the office of the Vice President for Research.

1.1 Evaluations and Status of the CDS

During the period the CDS has prepared for a number of high level evaluations and renewals of the status of the CDS. In early 2016 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). The roadmap was presented at MENESR in April 2016, and the accompanying compendium of research infrastructures is available in book-form at the MENESR pages[1] in French and English. This status attributed to CDS is a high level recognition of its role within the French national research infrastructures.

In April 2016 the CDS was informed of an official visit of the CDS by the CNRS Committee for very large research infrastructures (Comité des TGIR). The visit of the committee took place in September 2016, with the participation of the Observatory Director, a representative of the Université de Strasbourg Vice President for Research, and the Scientific Director of CNRS-INSU. The CDS was presented to the visiting committee with the help of the CDS service team leads, where we described a high level summarised prospective plan for the 2018-2022 timescale. The visit received a favourable official response, with the remarks noting the quality and international visibility of the CDS.

The CDS has prepared 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 have been prepared as a part of the response by the Observatoire Astronomique de Strasbourg. All such units of the University de Strasbourg are concerned by this evaluation and there has been coordination of the preparations at the university level. The CDS parts of the documents include an activity report for the CDS as one of the National Observing Services (SNO) of the Observatoire Astronomique de Strasbourg, and also an activity report for the scientific research of the CDS team. Also included are reports on the prospective plans of the CDS services, and of the CDS team scientists. These documents have been submitted in the summer of 2016 and the visit of an evaluation committee is expected between January and March of 2017.

[1] http://www.enseignementsup-recherche.gouv.fr/cid70554/la-feuille-de-route-nationale-des-infrastructures-de-recherche.html

At the 2015 Scientific Council meeting it was announced that an updated strategy for the CDS would be presented at the 2016 meeting. Hence we include here in this report (in Section 6) the ***CDS Prospective and Strategy*** that has been submitted for the HCERES evaluation (in its original formatting).

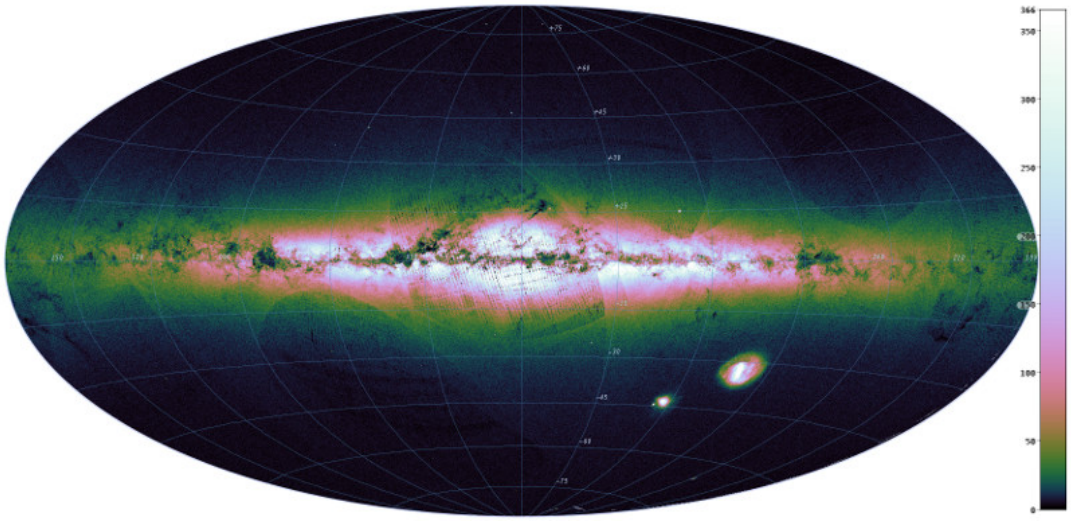
The other parts of the reports prepared for the HCERES evaluation are also provided as appendices. These are Appendix 1, ***"CDS 5 year Report 2011-2016 as presented to HCERES"*** which provides a summary of the highlights and evolution of CDS services over the past 5 years, and may serve as a background briefing document for new council members. Appendix 2, ***"CDS Science Team Achievements 2011-2016 as presented to HCERES"*** describes the scientific research achievements of CDS team scientists in their independent research.

Sections 2 presents the Highlights 2015-2016, and Section 3 presents the Activity Report for CDS services for the period September 2015 - September 2016. Section 4 concerns the projects to which CDS participates. The status of the Scientific Council 2015 recommendations are addressed in Section 5, and the (HCERES) prospective and strategy are presented in Section 6.

2. Highlights 2015-2016

Gaia DR1 in CDS Services

The CDS is a partner data centre for the publication of the Gaia Data Release 1. Following many months of preparations, on September 14, 2016 the Gaia DR1 was made available in the CDS services Vizier, Aladin and the CDS catalogue cross matching service. The inclusion of these data into the CDS makes them interoperable with all of the CDS reference data including more than 14000 catalogues and hundreds of image surveys. Aladin and Aladin Lite visualisers provide an all-sky hierarchical view of the entire GDR1 data set, using the capabilities of CDS HiPS (Hierarchical Progressive Surveys) technologies. Gaia DR1 is also made available from the CDS via Virtual Observatory protocols (IVOA Cone Search, and Table Access Protocol). The Gaia Data in the CDS represent a showcase of the services, and have generated some 3 million queries in the first month since the data release. The top level page for access, 'Gaia at CDS' is given below[2].



Density map of 1 billion Gaia sources generated by the CDS

[2] http://cdsweb.u-strasbg.fr/gaia

VizieR Associated Data Service released

A new service has been released to provide easy and interoperable access to data associated with journal publications. These associated data (images, spectra, time series) have always been part of VizieR in addition to the catalogue collection, but they are now made much more easy to find and use. The new service includes a dedicated interactive data ingestion tool for authors, and uses standardised metadata from the IVOA ObsCore Data Model. The Saada database engine is used in collaboration with L. Michel (OAS XMM-SSC). The data can be queried by an interactive interface and the service automatically publishes the data via VO protocols. Important data sets already in the system include 170,000 CoRoT time series and 2.4 million LAMOST DR1 spectra.

Aladin Lite Improvements and HiPS network

The Aladin Lite embeddable widget continues to be a runaway success with some 33 professional astronomical data centres and projects implementing it in their services. A major improvement has been made to the accuracy of the image display computation, minimising distortions, and allowing for accurate overlay of catalogues. This was a critical issue for use of Aladin Lite for CDS applications in particular the CDS Portal, and also an essential improvement required for ESA to use Aladin Lite as the basis or their archive new interface called ESASky[3]. The implementation in ESASky provides a highly visible use of CDS HiPS technologies, which is fully in line with CDS promotion of the network of HiPS data set providers.

Françoise Genova received the Étoiles de l'Europe prize for the CoSADIE project

The EU-funded project CoSADIE was one of the 12 research teams to be awarded an Étoile de l'Europe prize[4] on December 16 2015. Najat Vallaud-Belkacem, Minister of National Education of France and Thierry Mandon, French Secretary of State for Higher Education, presented the prize to Françoise Genova in the presence of Carlos Moedas, European Commissioner for Research, Science and Innovation. The Étoiles de l'Europe Prizes that were created in 2013 by the French Ministry of National Education, Higher Education and Research to reward French teams that coordinate European projects as part of the research and innovation framework programme.

ASTERICS - multi-messenger astronomy progress

The ASTERICS cluster (Astronomy ESFRI and Research Infrastructure Cluster[5]) project started in May 2015 and is in full operations. 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. One of the DADI highlights in the first year is the rapid up-take of VO and interoperable systems for handling the probability maps associated with gravitational wave events. This field came to prominence in 2015 with the first detection of gravitational waves. The European gravitational wave community, partners of ASTERICS, and a representative from the US Laser Interferometer Gravitational-Wave Observatory (LIGO) came together to collaborate on sharing data, software and tools. The all-sky capabilities of Aladin Lite have been used to visualise the probability maps of the gravitational wave locations, with Aladin visualisations appearing in the media. The tools are being used for the sky coverage planning of EM follow-up observations.

---

[3] http://sky.esa.int

[4] https://ec.europa.eu/digital-single-market/en/news/eu-funded-e-infrastructure-project-cosadie-wins-etoile-de-leurope-prize

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

AENEAS Horizon 2020 Project approved

CDS is a partner in a newly approved Horizon 2020 proposal called AENEAS (Advanced European Network of E-infrastructures for Astronomy with the SKA). In AENEAS the CDS will contribute high level expertise on the development of interoperable systems for astronomical data services and on the use of the Virtual Observatory framework.

Scientific Training Events

The CDS has run a very active program of scientific training events in 2015-2016. The include:

- The CDS Student Workshop on “Accessing and Analyzing Multiwavelength Astronomical Data” at the South African Astronomical Observatory in Cape Town, preceding the IVOA Interoperability Meeting in May 2016.
- CDS tutorials at the A&A/EDP school Science Writing for Young Astronomers in Puerto Varas, Chile in April 2016.
- CDS tutorials at the ASTERICS DADI VO School in Madrid, December 2015, and preparations for the upcoming ASTERICS DADI School in Strasbourg to be held in November 2016.
- VO tutorials in various Doctoral Schools in Paris in march and May of 2016

Collaborations and Interactions with Partners and the Astronomy Community

Establishing and reinforcing contacts with CDS partners and up-coming projects has been a priority in 2015-2016. Discussions with the American Astronomical Society and the ADS were held at the Winter AAS meeting in Florida in January 2016. Discussions with ESO and ESA were held at the SCIOPS 2015 conference in November 2015 where the CDS presented an invited talk, with the focus of the discussion being on the role of HiPS and Aladin Lite in ESA and ESO archives. An invited CDS talk (Vollmer) was presented at the ASKAP meeting Sydney, Australia in June 2016. An invited CDS talk (Allen) was presented at the LSST@Europe2 Meeting in Belgrade, Serbia in June 2016. Visits were made to CNES, and discussions with many other collaborators (NED, CADC, ALMA, CfA, STScI/MAST, NOAO, NRAO, Caltech) at ADASS conferences.

CDS Permanent Staff

In November 2015 Mihaela Buga was successful in the CNRS competition for a documentalist position.

In August 2016 François Ochsenbein departed after being at the CDS from the very beginnings in 1971-1972. François was instrumental in the creation and operations of many aspects of the CDS, in particular the catalogue services.

3. Activity Report for CDS Services 2015-2016

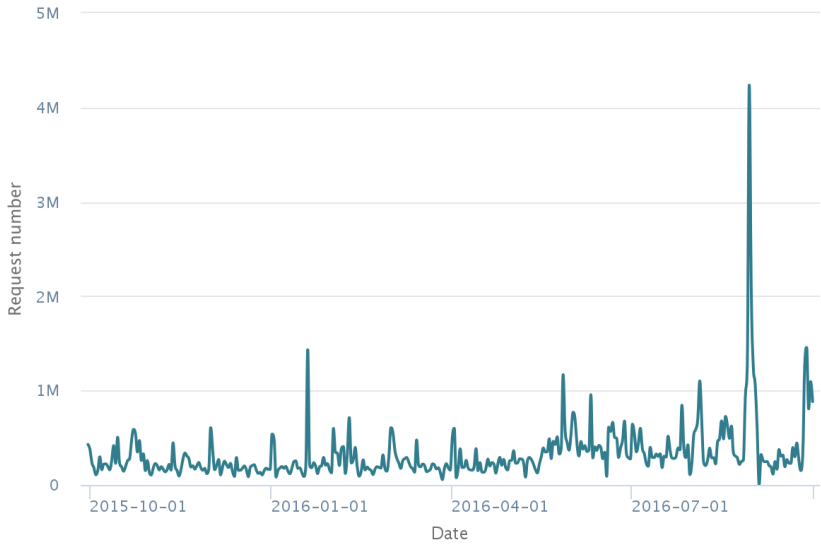
In 2015 we have used the ADS interface to track text citations of CDS services. The ADS interface allows counting of the number of papers in which the CDS services are cited in the text of the paper. **In 2015, 695 refereed papers cited the word SIMBAD, 384 the word VizieR, and 81 the word Aladin** in reference to the respective CDS services. This represents approximately the same level of citation for SIMBAD, but increases of 25% and 40% for Aladin over the results in 2014. These text citations will continue to be monitored.

Here brief reports on the services and more details are provided in presentations by the service teams.

3.1 SIMBAD

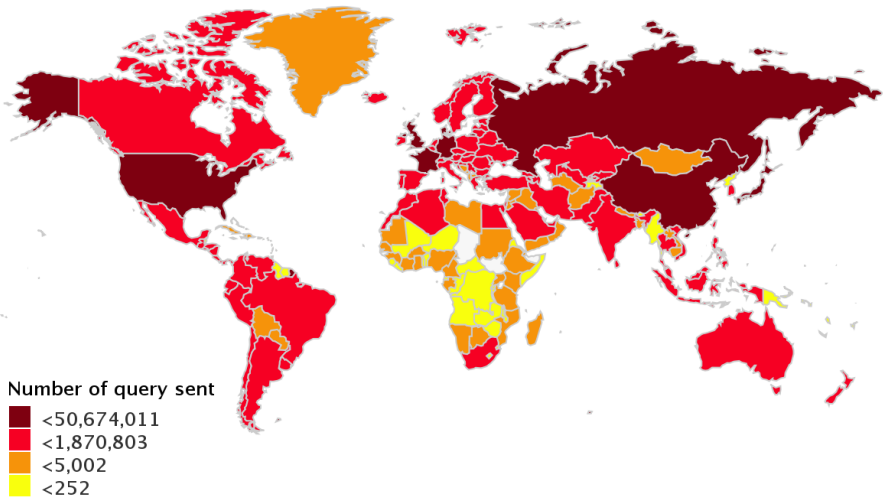
The SIMBAD database continues to grow by about half a million astronomical objects per year. The table indicates the status on October 26, 2016, compared to the previous three years.

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



The mean number of queries on SIMBAD each day during the last year have been around 319,000. A new prototype monitoring tool has been implemented to provide easy use of SIMBAD logs. The queries over the course of 2015 are shown above, and a world map showing the origin of the queries is shown on page 7. (The notable spike in August 2016 was identified as large batches of inefficient queries from a user and was acted upon by providing advice on better querying strategy.)

2015 was a big year in terms of the number of in-coming articles that were identified (by earlier steps in the CDS process) to be appraised by SIMBAD. This number increased in 2015 by ~30% to 928 articles, and it is notable that it is the number of small tables that are increasing most rapidly. These small tables are not usually treated by VizieR, so there is also more work for the SIMBAD team to extract the table from the article. Despite this increase, *all* of the in-coming "priority 1" tables were processed. This result is attributed to the greater efficiency enabled by the COSIM tool that is now in full operations. In 2013 and 2014 there was significant time spent on testing and adapting this tool, but now the benefit of the investment in COSIM is paying off. There is still a back-log of articles to be processed by SIMBAD but the status is that the back-log is being steadily reduced.



Other improvements to SIMBAD procedures and scientific content include:

- Revision of scripts for input of bibliographic information received from journal publishers (Titles, Authors, Abstracts, page numbers, etc.) into the SIMBAD bibliographic database. This is a systematic upgrade of old scripts that were written in various languages, to more manageable and consistent python scripts.
- Modifications to procedures for the Dictionary of Nomenclature to enable a shared workload on the creation and management of acronyms. This included training of a second documentaliste to work on the dictionary, and modifications of the software tools.
- SIMBAD includes basic data on astronomical objects as part of the database. New fields were added for V_{LSR} and for G band magnitudes in preparation for Gaia. Gaia will be cross-matched with SIMBAD and for TGAS the objects are already included.
- Continuing special operations - objects without or with inaccurate coordinates (summarise from 2015 report)
- A full cleaning of historical White Dwarf star objects in SIMBAD was performed.
- there are some 24,000 White Dwarf objects in SIMBAD and 5000 of these were updated over a ~3 month period in 2015.
- Important progress was made on revision of object types in SIMBAD, following details presented in the 2015 report. This effort is now in a phase where new object types list has been defined and the software is being adapted and tested.
- Work between the Scientists and documentalists in the team included the transfer of expertise on specific topics. A course for documentalists was provided on how to treat the double stars (4 hours dedicated training).

- Following a general trend of increased integration of CDS services, more links were made between SIMBAD and Vizier. Object identifiers in SIMBAD are now linked to thier catalogue of origin in Vizier. The links take you to the relevant Vizier catalogue page. This highlights the depth of CDS data curation, in that the provenance of the information is preserved and can be used in multiple ways.
- The ***SIMWATCH*** service was released. This is a monitoring service that sends notifications to subscribed users when their favorite object is cited in a new paper processed by SIMBAD.
- Bibliographical data (Abstract, keywords, copyrights ...) previously hosted on a separate server have been integrated directly in SIMBAD
- Distinct improvements have been made to the maintenance of SIMBAD mirrors making them more independent and easier to manage.
- A new server has been provided by ADS for the CfA mirror.

3.2 VizieR

As presented above, the VizieR service has opened a major new feature for non-tabular data. The VizieR Associated data Service includes the ingestion pipeline and metadata mappings necessary to take into account the description of spectra, time-series and images. Authors may provide these data, and the ingestion process (supervised by documentalists and engineers) maps the description to the standard IVOA ObsCore data model. The data are made discoverable in a web interface, and also through IVOA protocols. It is part of the CDS contribution to the rapidly evolving field of data publication, where journals and publishers are exploring new technologies for adding value to journal articles.

The standard VizieR treatment of catalogues has been in smooth operations throughout 2015-2016. The table below indicates the growing content of the database. Changes have been made to modularise and share the scientific validation steps of catalogue ingestion, whereby an additional astronomer contributes a fraction of their time to validate catalogues ‘close’ to their area of expertise.

An important improvement has been made for managing the original column names provided by authors, alongside the standardised VizieR column names. This has been implemented in the main Vizier interface (where one can switch between ‘standard’ and ‘original’), and also for Vizier TAP interface where the columns can now be referred to by their original names in the actual ADQL query.

	2013	2014	2015	**2016**
Number of Catalogues in the VizieR database	11,579	12691	14065	**15485**

The CDS process for managing large catalogues (> 5-10 million rows) has been revised to be more tightly integrated in VizieR. While large tables require special processing, the common tasks have been identified and are now managed in a uniform way. The large catalogues treated in 2015-2016 include AAVSO Photometric All Sky Survey (APASS) DR9, the XPM Catalog of positions and proper motions, the VPHAS+ DR2 survey, and in particular the Gaia DR1 catalogues. The time taken revising the process has introduced some delays and priority is currently focused processing catalogues including SDSS data releases, the Hubble Source Catalogues, and the ESO Public survey catalogues.

A dedicated effort has been made to process the ESO Public Survey catalogues. These fall into both ‘normal’ and large catalogue categories. Of the 31 public survey catalogues identified, 9 of these are already in VizieR. Processing of these is a priority for 2016-2017.

VizieR includes observing log catalogues of observatories which are regularly up-dated. Work has been done to improve the update process using TAP queries to the observatory archive services.

The first steps for CDS assignment of DOIs to catalogues have been discussed internally. This requires further coordination with the journals and ADS, and these discussions are currently underway.

3.3 Aladin and Aladin Lite

The Aladin project encompasses the Aladin java desktop application, the Aladin Lite javascript embeddable widget and the Aladin databases of image and cube data. As with all CDS services there is movement toward greater integration. The evolution of Aladin continues to be dominated by Hierarchical Progressive Survey (HiPS) systems. The number of HiPS surveys is growing and importantly many data providers (ESA, CADC, JAXA, CADE, …) have adopted this system and are generating HiPS data.

There are now 325 HiPS data sets and these amount to 105 TB of data, with the evolution since 2012 shown in the table below. Procedures for the curation of HiPS data sets have been established and a documentalist has been trained for this task. A new computation server has been installed for the efficient calculation of HiPS from the original FITS image and cube data to support the process.

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

Many CDS collaborations are now based on the developments and potential of HiPS. This has provided important support for standardising HiPS within the IVOA which is in progress. Aladin is also being improved by adding features related to VO protocols, in particular TAP and DATALINK with this work supported by the ASTERICS DADI project.

As mentioned in the highlights, significant improvements were made to Aladin Lite, and a large effort was made to help ESA implement HiPS data sets for many missions. This is a strong collaboration where CDS has provided the *hipsgen* tools for computing HiPS and expertise on its use, as well as the HiPS client Aladin Lite and the definition of HiPS and the HiPS network. ESASky provides a high visibility of Aladin Lite, and the data are shared with CDS via the HiPS network.

Aladin Desktop version 9 was released in November 2015. Aladin Desktop has been used as an important component of tools for the preparation of observations for major observatories. In particular it is used in the NASA 'Astronomer's Proposal Tool (APT[6]) for HST and JWST. We also welcome the decision in June (discussed with Pascal Ballester, ESO) to use Aladin as an integrated part of the ESO GuideCam[7], in replacement of the Skycat tool.

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

[7] https://www.eso.org/sci/observing/phase2/SMGuidelines/GUCT.generic.html

3.4 The CDS Catalogue Cross-Match Service

The X-Match service continues to operate smoothly with increasing visibility in the community. The service has a web interface front end, and also programatic (HTTP API) interfaces. Both of these modes are heavily used, with the large jobs being mostly managed by the web interface making use of the associated user storage space available to temporarily (~10 days) host the large result files.

The total usage is of the order of ~600 jobs submitted per day, with ~88 million associations computed per day. The table below shows the growing use of the two interfaces, and in particular the large increase in use of the HTTP API in 2015 due to collaborative use of the service via TOPCAT.

	2013	2014	2015	**2016**
Web interface (jobs/day)	15	16	20	**25**
HTTP API (jobs/day)	47	50	580	**574**
Associations/day (Web Interface)	~13,000,000	~70,000,000	~55,000,000	**~82,000,000**
Associations/day (HTTP API)	~298,000	~1,600,000	~6,600,000	**~5,800,000**

In 2016 this service was migrated onto new servers, which contribute to a increase in speed of order 25%. The lead technical responsibility was changed following the successful recruitment of an engineer as mentioned in the 2015 report.

A major paper on the an algorithm for multi-catalogue cross-matching was accepted by A&A (Pineau et al., 2016[8], with 3 other CDS authors). This will be the basis for future long term development of the CDS X-Match service. The paper has also received critical acclaim from principal scientists associated with the Gaia mission.

3.5 Service Integration and Portal Development

The CDS portal is a single entry point to CDS services and plays an important role, in particular for first time users. It is used in tutorials of CDS services and provides an integrated view of the services. The current portal is however ageing, so in 2015 a major revision of the portal was initiated. The new portal fully integrates HiPS and Aladin Lite and the 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).

This development also provides a test bed for CDS service interfaces and greater integration of the services, that will be important for planning major future upgrades to the SIMBAD and Vizier interfaces. Developments for the Astrodeep project, where a project-specific portal was built, have have also aided the development of the general CDS portal. The development is also designed to be modular, with each component being independent and with interactions between components through being managed via a message bus. The new CDS Portal is to be released in Winter 2016/2017.

[8] http://arxiv.org/abs/1609.00818v2

3.6 R&D

R&D activities continue to be varied and versatile. A list of the internships overseen by CDS staff since the last Council meeting is provided with the set of documents. They cover a large range of aspects of interest for the CDS

We continue to test Big Data technologies to address the future scientific needs. The CDS X-Match has been used as a test bed for services that will require high performance computing. A dedicated project of implementing the X-Match using big Data tools Hadoop and Spark on a rented cluster of 11 servers (from a French provider OVH) was done. Via this project we gained valuable experience in understanding the issues related to efficient use of these systems.

- A monitoring system to track and visualise the number of queries of CDS services
- Indexation of document for contextual searches
- Study "DevOps" for the deployment of data services in virtual machines and containers (Docker, etc.).
- The use of all-sky images in visualisation systems such as Google Carboard, and in virtual reality environments
- Development of Python code for spectral analysis of CALIFA IFU data cubes
- Analysis of photometric flux conservation in the use of HiPS for images and cubes
- 3D visualisations and fish-eye projections
- Development of a system for ‘alerts’ of changes in Virtual Observatory services
- Conception of a user interface for authors to provide data to be ingested into CDS VizieR (images and spectra)

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 2015-2016, and the CDS contributions to the IVOA interoperability meetings in Sydney, October 2015, and Cape Town, May 2016.

M. Allen and F. Genova contributed to a paper presented at ADASS 2015 by Christophe Arviset (ESA, and chair of IVOA) 'The VO: A Powerful tool for global astronomy’ re-motivating the concepts of IVOA, and in particular highlighting the need for engagement with large projects. An excerpt is provided below.

Since its inception in the early 2000's, the Virtual Observatory has become a major factor in the discovery and dissemination of astronomical information worldwide. It has been developed as a collaboration of many national and international projects. The International Virtual Observatory Alliance (IVOA) has been coordinating all these efforts worldwide to ensure a common VO framework that enables transparent accessibility and interoperability to astronomy resources (data and software) around the world. The VO is not a magic solution to all astronomy data management challenges but it does bring useful solutions to many. VO interfaces are broadly found in astronomy's major data centres and projects worldwide. Astronomy data centres have been building VO services on top of their existing data services to increase interoperability with other "VO-compliant" data resources. The continuous and increasing development of VO applications (Aladin, Topcat, Iris etc.) greatly facilitates multi-instrument and multi-wavelength science.
More recently, several major new astronomy projects are directly adopting VO standards to build their data management infrastructure, giving birth to "VO built-in" archives (eg CADC, Gaia, CTA).

Embracing VO framework from the beginning brings the double gain of not needing to reinvent the wheel and ensuring from the start interoperability with other astronomy VO resources. Some of the IVOA standards are also starting to be used by neighbour disciplines like planetary sciences. There is still quite a lot to be done on the VO, in particular tackling the upcoming big data challenge and how to find interoperable solutions to the new data analysis paradigm of bringing and running the software close to the data.

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 optimize 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).

It is important to note that the project gathers astronomy and astroparticle physics, in particular to tackle the aspects linked to new messengers, and that pathfinder projects are associated with the ESFRIs to be able to deal with real data (HESS, MAGIC and VERITAS for CTA, LOFAR and JIVE for SKA, ANTARES for KM3Net, and VIRGO/LIGO for the ET). ESO (VLT-E/ELT) is associated with the project, and ASTERICS WP4 also works closely with ESAC

ASTERICS is now 16 months into its work program. The first annual report was delivered in April 2016 (based on a long list of deliverables in year 1). The DADI report shows all WP4 activities, deliverables and milestones to be on schedule (some ahead of schedule, and some justified changes).

Important events have included:

- IVOA Meeting, 14-19 May, Sesto, Italy - DADI introduced to IVOA with major role of providing connections to ESFRI and pathfinder projects as participants in the VO
- RDA 6th Plenary Meeting, 23-25 September 2015, Paris
- First ASTERICS DADI Technology Forum, 17-18 September 2015, Strasbourg
- ADASS XXV, 25-59 November, Sydney, Australia
- IVOA Interoperability Meeting, 30 October - 1 November, Sydney, Australia
- First ESFRI Forum and Training Event, 3-4 December, Trieste, Italy
- First DADI VO School, 15-17 December 2015, Madrid, Spain
- 7th RDA Plenary Meeting, 1-3 March 2016, Tokyo, Japan
- Second DADI Technology Forum, 7-8 March 2016, Edinburgh, UK
- Discussions with ESFRI projects about specific support activities
    - CTA , EGO-VIRGO, ANTARES data, Time Domain Astronomy, Authentication and Authorisation
- Collaboration with EUDAT

The next ASTERICS DADI VO School[9] has been prepared and will take place at CDS 15-17 November 2016. Other major items in the year 2 plans include the Third DADI Technology Forum to be held at CDS in March 2017, and a mid-term delivery of a repository of training products.

ASTERICS will have its mid term review on 13-14 March 2017.

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

ARCHES

The Astronomical Resource Cross-matching for High Energy Studies (ARCHES) project is a three-year long FP7-Space funded programme that started in 2013, and ended in December 2015. The project involves the Observatoire Astronomique de Strasbourg including the CDS (France), the Leibniz-Institut for Astrophysik Potsdam (Germany), the University of Leicester (UK), the Universidad de Cantabria (IFCA, Spain) and the Instituto Nacional de Tecnica Aeroespacial (Spain). The project aims at providing the international community with well-characterised multi-wavelength data in the form of spectral energy distributions (SEDs) and catalogues of clusters of galaxies for large samples of both resolved and unresolved sources extracted from the 3XMM DR5 X-ray catalogue of serendipitous sources. SEDs are based on an enhanced version of the 3XMM catalogue and on a careful selection of the most relevant multi-wavelength archival catalogues (GALEX, SDSS, UCAC, 2MASS, GLIMPSE, WISE, etc). For these purposes ARCHES has developed advanced methods providing probabilistic cross-identification of several catalogues in one go as well as a multi-wavelength finder for clusters of galaxies. Importantly, these tools are not specific to the X-ray domain and are applicable to any combination of well-described multi-wavelength catalogues. Both tools and data will be made available through specific interfaces, through CDS services and through the Virtual Observatory. These enhanced resources are tested in the framework of several science cases involving both Galactic and extragalactic researches. A number of ARCHES products have been ingested into the CDS VizieR service as planned, and other products will likely become available in 2017.

EuroPlanet

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). The CDS work in 2015-2016 includes progress toward updating the WCS projections system in Aladin to support planetary data (AZP, SFL, MER projections).

ASTRODEEP

The project, funded by the European Commission for 4 years starting in January 2013 will conclude in December 2016. It is coordinated by Adriano Fontana (INAF Roma). The project aims at getting the best scientific return from the exploitation of the deepest sky surveys available to date, studying the birth and early phase of galaxy evolution. The role of the CDS is to develop a dedicated portal for the internal validation and manipulation of the data, and publication via CDS services. A presentation of the portal will be provided at the Council meeting.

VIALACTEA

The VIALACTEA project came to a conclusion in October 2016. This 3 year project was funded by the European Commission and was coordinated by Sergio Molinaro (INAF Rome). By using a novel data and science analysis paradigm based on 3D visual analytics and data mining framework, the project built a quantitative 3D model of our Galaxy as a star formation engine that will be used as a template for external galaxies and study star formation across the cosmic time. Among the various products, there are catalogues of sources with physical properties (mass, temperature, distance, etc.). The catalogue of filaments and bubbles makes innovative use of the VO standard MOC to describe the morphology. As part of the final model of our Galaxy developed by the VIALACTEA collaboration, we, at CDS, built three-dimensional maps of interstellar extinction in the Milky Way Plane. These 5 and 10 arcmin spatial resolution and 100 pc distance resolution maps are based on 2MASS and GLIMPSE photometry of giant stars and revealed the dust 3-D distribution in our galaxy out to a distance of 15 kpc. The wealth of structures exhibited by the maps at all scales make them great tools to study the morphology of the Milky Way. Besides, by comparing the 3-D extinction maps to other interstellar medium indicators such as gas, or dust emission derived column density, it is possible to extract valuable information on how dust

properties vary on galactic scales and how this evolution occurs with respect to the gas content (Arab et al. in preparation).

The Research Data Alliance

The evolution of European programmes is towards “e-Infrastructure Commons”, a common framework for the European network, computing and data infrastructures. The European Commission also launched in 2016 the European Open Science Cloud 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. The RDA is a recent but high profile organisation which aims at defining technological and sociological building blocks of the data infrastructure facilitating data sharing.

The Research Data Alliance, created in March 2013 with support of the European Commission, NSF and Australia, continued to grow. It has now nearly 4500 members from more than 100 countries. A summary of its current status can be found at the link below[10]. F. Genova who has been one of the members of RDA Technical Advisory Board since the beginning, was elected as co-chair in October 2015. She participates in the RDA Council meetings and is a member of the Council's Strategy subcommittee. This builds a liaison with the IVOA, also through regular participation of IVOA members from different countries in the RDA meetings. Astronomy has historically been at the forefront for the sharing and reuse of data, and thanks to the Virtual Observatory its data holdings are a rare example of a global interoperable data infrastructure. The lessons we learnt and our requirements have to be taken into account by the RDA. F. Genova was invited to participate in the series of European projects set up in support to the RDA. The third one began on 1 September 2015, and the next one should be submitted in March 2017.

Lessons learnt when building the IVOA are shared with the RDA through F. Genova’s participation in the RDA Technical Advisory Board. She 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, DOIs and data publication are among the topics of common interest. 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.

Also, the IVOA Registry of Resources was entered in the B2FIND “generic” registry of the European EUDAT project, which is important to demonstrate that the astronomy well developed, global and interoperable data infrastructure is not an isolated island but can interoperate with the generic data framework.

[10] http://rd-alliance.org/sites/default/files/attachment/RDA_in_a_nutshell_October2016.pptx

5. Status of 2015 Recommendations of the Scientific Council

Core Services: CDS should continue to maintain and upgrade its services taking into account the needs of the users.

The CDS continues to follow its mission as a science driven astronomy reference data centre, serving the international astronomical community. The core of the CDS work involves maintaing and improving the services. Major improvements in 2015-2016 include i) the VizieR Associated Data Service for non-tabular data, ii) the image display accuracy of Aladin Lite, iii) interlinking of CDS services in particular between SIMBAD and VizieR and iv) publication of the much awaited Gaia data in the CDS services. In 2015-2016 there have been many occasions to interact with the CDS users, in particular at hands-on training events, and also via CDS booths at the AAS and ADASS conferences.

CDS and the VO: It is very important for CDS to maintain its leadership role in the VO.

CDS staff play major roles in the VO at the French (OV-France), European (Euro-VO) and International (IVOA) levels. CDS leads the DADI work package of the European Horizon 2020 ASTERICS project that is making the Astronomy ESFRI projects participants of the VO. CDS staff currently chair three Working Groups of the IVOA ("Applications", "Data Access Layer", and "Semantics") as well as the "Data Curation and Preservation" Interest Group. Mark Allen is the chair of the Committee for Science Priorities, and in October 2016 the Vice Chair of the IVOA Executive Committee. In addition to leadership by CDS staff, the CDS services are also maintained and developed as leading VO services.

Research and Development: R&D efforts should be maintained at least at the present level and gradually increased to accommodate the more complex requirements which are expected to arise over the coming years.

R&D is carried out in support of evolution of the CDS services. The strong CDS participation in the ADASS conference provides a record of many of the results of this work. the list of CDS contributions is provided in an accompanying document. The internship program allows for testing of many different technologies as required for planning of CDS service evolutions. Important considerations concerning the requirements for Big Data are briefly considered below along with the associated staffing issues.

Scientific Activities of the CDS: The Council recommends that the fraction of time spent on research by CDS staff be maintained at least at the current level. Diversity in scientific interests is very important for support to be provided to data related activities, which involve very diverse content.

The scientific activities (2011-2016) and prospectives of the CDS scientific staff research staff are outlined in Appendix 2. This is a more detailed view of these activities than is usually provided in the reports for the Scientific Council, and is possible this year due to the coincidence with the 5 year evaluation and planning for the HCERES campaign. The research activities are diverse, as in necessary to support CDS data curation. The research time for the 7 CNAP personnel at CDS is fixed at 50%.

Staffing: Taking into consideration the increasing R&D activities at CDS, and the important role that CDS is expected to play in this epoch of large surveys and big data, it is necessary to gradually ramp up the staffing at CDS to maintain well its leading role and also to allow for future retirements and possible staff attrition. The highly specialised and unique nature of work at CDS requires a long apprenticeship and it will not be easy to replace key staff who may leave for some reason.

Recognition of the unique nature of the CDS work and that it requires long training times, is appreciated. The CDS work is best carried out on permanent positions due to the continuing nature of the reference services. Astronomy is growing, and the CDS will need to grow in order to maintain its leading position. One part of the growth is due to the higher number of reference papers to process which is steadily growing at ~15% and which can be addressed by increases to the documentalist staff (in addition to tools to improve efficiency).

We note that currently, the core CDS data curation work relies on contractor staff in addition to the the permanent documentalist staff, in order to process the load of incoming catalogues from journal publications.

Another part of the growth is related to the exponential growth of the data flows, which will become extreme on the 5-10 year timescale with projects such as LSST and SKA, and requires a strategic approach for the CDS infrastructure. The staffing requirements here are in relation to the effort needed to manage the evolution of CDS services into the Big Data era, and in particular to be able to plan and test Big Data systems while at all times maintaining full operations.

Two of the main CDS services, SIMBAD and VizieR, now rely on a single engineer for each service. During the transition period where the original creators of these services were still present at CDS, there was a level of redundancy with 2 persons with full knowledge of the system, to ensure operations of these services. Now there is only very limited redundancy. The increased integration of the CDS services is one element of addressing this, but it will need to be addressed with eventual staffing increases.

In terms of science staff, it should be noted that a senior CDS member will retire ~2020.

Contract staff 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 2015-2016 period CDS has had the following contract staff:

- Three contract staff working on core CDS functions(1 documentalist, 2 Software Engineers)
- Two contract staff on specific projects (ASTERICS - Software Engineer, ASTRODEEP - Software Engineer)
- One Postdoc (VIALACTEA)
- Two PhD Students, one of whom successfully defended their thesis in September 2016
- Twelve short term interns, as detailed in the ‘Trainees at CDS’ document

(In October 2016 a new postdoc arrived for the ASTERICS project, and contract engineer was hired to work on the bibliographic DJIN tool. A Contract engineer is currently advertised to work on CDS Gaia services, and we expect to hire another contract documentalist)

Participation in Projects: It is important for CDS to participate in large projects and data alliances, where its experience with data makes CDS a key contributor. Such projects enable CDS to apply its expertise at the highest level. However, care must be taken to ensure that these projects do not divert attention of the CDS staff from its own key services or and over-stress them.

The CDS is a leading participant in many VO, scientific, and data-related projects as outlined above. These projects bring resources that are applied to their specific goals, and are carefully managed within their scope. CDS participation in IVOA has been strong in 2015-2016 and indeed much effort has gone into the IVOA priority area of Multi-dimensional data taking more time than initially expected. These efforts do not take away from the core CDS work, and are a long term investment in the global interoperability framework.

The Scientific Council: The terms of reference of the Scientific Council should be looked into and redefined if that is found necessary. The frequency and nature of the meetings should also be reconsidered.

The terms of reference are managed by the CDS authorities. In terms of frequency, we prefer to remain on an annual cycle.

6. CDS Prospective and Strategy Presented for HCERES

1.1 Centre de Données astronomiques de Strasbourg

CDS strategy

The main strategy of CDS in the coming period is to continue its mission at the highest level as a premier astronomy reference data centre that serves needs of the international astronomy community. It will continue to develop its leadership role on added-value services, tools and interoperability.

The CDS is science-driven with strategic planning that adapts to the changing reference service needs of astronomy research. Astronomy is making a transition into a Big Data science characterized by large all sky surveys with inter-connected use of astronomy data. This includes the strong emergence of time domain and multi-messenger astronomy and data science, and the evolution of the wider research environment into a more 'open' scientific endeavor. These are all factors that guide the strategic planning of CDS activities.

The themes for the CDS strategy in 2018-2022 are: i) Reinforcement of the CDS role as a Trusted reference data centre, ii) Enabling science with the CDS services, supporting specific scientific projects, direct support of astronomers, and development of the CDS science team, iii) Engagement with the astronomy community to ensure relevance of CDS services, and integration of the CDS in the French, European and international infrastructures of astronomy, iv) Adaptation and innovation of the CDS services to meet the challenges of the increasing volume of data, and to define a path for the sustainable technical evolution of the overall technical framework of the CDS, v) Building on the success of the CDS by maintaining specialized staff profiles and scientific-technical work environment that fosters innovative solutions and teamwork.

CDS core mission

CDS services will continue to provide easy access to the information and data that scientists need, with tools and trusted reference services for visualization, comparison, computing and analysis of the data. The CDS services, SIMBAD, Vizier, Aladin and X-Match will be continually adapted to scientific needs with the highest attention to quality and sustainability.

The core CDS mission is to develop the services in terms of content, functionalities and operations. This relies strongly on the maintenance and development of the scientific and technical expertise of the CDS. CDS scientists provide the essential astronomy expertise that ensures the scientific quality and relevance of the services. Software engineer staff provide the high level operational framework and development of the services, which must be kept at the forefront on new technologies by R&D programs driven by scientific needs. The CDS documentaliste staff enter content into the system with well-maintained procedures and under the guidance of the scientific and technical expertise. The core mission in 2018-2022 will be supported by continuing this interlinked scientific and technical development and operations.

Engagement with the community

The CDS operates in the astronomy research environment composed of the various infrastructures at French, European and international levels, and it is necessary for CDS to maintain high level engagements at all of these levels. As an Infrastructure de Recherche on the French national roadmap CDS will continue its role as a pillar of the French astronomy data centre community with a leading role in the national SO5 organisation and the Action Spécifique Observatoire Virtuel (ASOV). CDS will also continue to lead the French and European participation in the global Virtual Observatory (VO) development.

CDS will maintain close engagement with the astronomy and astrophysics Journals, the observatories, and other astronomy data centres and archives in France and internationally. Journal publishing is evolving quickly in response to strong demand for new-technology features. CDS as the main astronomy data centre responsible for providing access to data associated with astronomy publications, will play a part in this new era of data publication with the Vizier associated data service which will provide access to spectra, time series, images, and other data associated with journal articles.

Continuing relationships with the French Space Agency (CNES), ESO, ESA, NASA, the Astronomy & Astrophysics journal, the American Astronomical Society (AAS) journals via SAO/NASA will keep CDS connected to the most important sources of reference data. CDS is also connected to the large future astronomy ESFRI projects (SKA, E-ELT, CTA, KM3-NET) via the Horizon 2020 ASTERICS cluster project, in which CDS leads a major work package on Data Access, Discovery and Interoperability. CDS will also have a role to play for LSST data dissemination.

Scientific Aspects

The CDS's role as a reference data centre will be as important to astronomy in 2018-2022 as it has ever been because of the qualified reference nature of professionally curated information that gives it scientific value. The next years will involve the publication of important reference data sets such as the Gaia mission data, that will be made interoperable in the CDS services with added-value capabilities in particular the ability to

perform fast cross-matches between Gaia and the thousands of Vizier catalogues plus user-provided tables. Gaia will bring a significant global astrometric improvement over the whole sky, and the major operation of updating SIMBAD with Gaia will benefit the coordinate accuracy of all CDS services.

Data from major projects, missions and surveys are scheduled for ingestion in CDS services using a coordinated approach with the data providers, including ESO (in particular the public surveys) and space mission data (ESA, NASA, JAXA, CNES). There are many major surveys that are relevant to CDS that will be active or have data releases in the period: Pan-STARRS, SkyMAPPER, ASKAP, MeerKAT, LOFAR, WEAVE, LIGO/ EGO to name just a few. Major projects in preparation include Euclid, LSST, SKA, and CTA and our early contributions to the planning of interoperability systems for the expected data will be important for the eventual public data releases.

The diverse range of active and future projects highlight the emphasis on multi-wavelength, multi-messenger, and time domain astronomy. The CDS all-sky HEALPix based hierarchical approach and the development of standards and tools for interoperability of multi-dimensional and time domain data, as prioritized at the IVOA level, are of key relevance for the scientific use of the data from these projects. CDS will continue to actively participate in the definition of the interoperability standards and to develop our innovative all-sky systems as part of the global astronomy initiative to share common solutions.

Technical Aspects

In the coming period CDS will face the challenges of very large data volumes. CDS is well positioned to build upon its recent progress for managing large volumes of heterogeneous data, as will be needed for Vizier and SIMBAD, and particularly for Aladin. The CDS approach is based on hierarchical methods defined by our Hierarchical Progressive Surveys (HiPS) scheme and Multi-Order Coverage (MOC) maps, which provide the underlying basis for the indexation of astronomical information over the whole sky. This approach is scalable to the largest data sets and can be implemented on a wide range of Big Data computing and storage infrastructures.

Other technical aspects involve the visualization of astronomical data; multi-dimensional data sets, server-side visualization of large data sets of billions of points, and support of new platforms such as mobile devices and virtual environments which have become much more accessible, e.g. Occulus Rift systems. Server-side computation is also required to extend the capabilities of the CDS X-Match service. Integration of text mining technologies is planned to improve CDS ingestion process from journal articles, and also for science user

applications. CDS service user interfaces are also subject to strong evolution to take advantage of new modular technologies, for example in the CDS portal that integrates information from all CDS services.

The computing systems and operational service server infrastructures of the CDS services must be kept at the forefront in order to provide quality services. Stability, control and flexibility are key elements for operating the system (24h/7days a week), and continuation of high level of technology awareness and testing is necessary. Engagement with computing infrastructure developments at Strasbourg University and within the national French level planning for computing centres will be necessary in 2018-2022.

Data Sharing

Taking a wider view of data in research and society, we recognize the strong move toward Open Data, Data Sharing, and highly interconnected science. Key stakeholders are responding to the changing research environment. The VO is astronomy community's contribution to open and interoperable systems. Journal publishers are moving toward models of Open Access to publications and research data. Researchers across all fields expect seamless access to data, with user interfaces and collaboration platforms that are as easy to use as the digital services and tools that we now enjoy in many aspects of everyday professional and personal life. Educators promote the use of real research materials and tools in university courses, and science is becoming more public with 'citizen science' and a greater emphasis on public participation and appreciation of publicly funded science. The European Commission considers that such evolutions are driving a systematic change in "the modus operandi of doing research and organising science", and they label this approach as "Open Science", a view also supported by the G8 Science Ministers Statement in 2013.

CDS will continue to contribute to participate in global development of generic data structures, bringing the experience of astronomy to the Research Data Alliance (RDA) created in 2013 by the EC, the US NSF and the Australian Government. Renewal of CDS certification (DSA) will help to maintain the position of CDS in this wider context, along with our continued strong connections with the journals and data providers of the discipline that drive the innovative and interconnected use of open astronomy research data. Educational versions of CDS tools will continue to contribute to public participation in astronomy, as will our participation in the citizen science, as planned for example in the ASTERICS project.