CDS Scientific Council meeting 2021

Summary of CDS activities 2020-2021

24 November 2021

1. Introduction

The period 01 October 2020 - 30 September 2021 has been a challenging year for CDS where the continuing situation of the COVID-19 pandemic has required many adaptations to the way of working. Ten months of this period (30 October 2020 - 02 September 2021) were under predominantly work-from-home conditions, and there was much uncertainty in the first half of 2021 about when a return to in-person working would be possible. A return to 50% in-person work was implemented during the summer period (July - August 2021) and a return to regular in-person work occurred with the opening of the academic year. We have been in the process of re-grouping since then. The possibilities for regular work-from-home have become more flexible and have been adopted by many staff, with up to 2 days per week work-from-home. The investments made by the ObAS IT team in 2020 in response to the first wave of the pandemic, for video and chat communication tools have been indispensable for maintaining contact within the team, and these tools continue to be important for the many hybrid/virtual meetings. The return to the ObAS site has been very welcome and important for the CDS teamwork, but this also involves managing with the sanitary conditions, such as wearing masks.

CDS activities have continued throughout this period and much effort has gone into ensuring the operational stability and continuity of the services. The overall production and impact of the CDS has remained very high over this period. The core operations for the ingestion of data from publications has processed a large volume of data, and in some processes the volume has surpassed the previous year. High priority data sets have been ingested into the system (e.g. Gaia EDR3). Important upgrades have been made to the underlying technical infrastructure of the CDS information system, and the first physical movements of selected CDS servers to the UNISTRA data centre have been very successfully done.

CDS has maintained a strong engagement with the astronomy community via participation in scientific, technical and data stewardship/librarian conferences. Community engagement has also included many contributions to the development of the Virtual Observatory at the international (IVOA), European (Euro-VO) and national (ASOV) levels. CDS has played a leading role in European projects, and new projects have started in 2021. A number of activities have followed the rapid development of Open Science with CDS contributions being made at the national and European level for the development and governance of the European Open Science Cloud (EOSC).

In terms of changes in the permanent CDS staff, there have been two unexpected departures in 2021 of the CDS Administrative Assistant and of a senior documentalist. As announced in 2020, E. Collas started a permanent documentalist position in November 2020, and G. Mantelet started a continuing contract position (CDI) in January 2021.

In this report we present the status of CDS in Section 2, the highlights of the 2020-2021 period in Section 3. The activities of the CDS services are in Section 4, and Projects in Section 5. Section 6 provides updates and responses about the issues identified in the 2020 recommendations of the CDS Scientific Council.

2. Status of the CDS

The current term of the directorship of CDS commenced on 01 September 2020 for a period of 5 years. The official decision was published in the CNRS Bulletin in December 2020 (DEC201613INSU).

In February 2021 CDS applied to renew its status as a “Research Infrastructure” on the French National Research Infrastructure Roadmap, “Feuille de Route”, established by the Ministry of National Education and Research (MENESR).

CNRS-INSU helped us to prepare the CDS application as part of its coordination of the INSU Research Infrastructures. The application forms included sections on the identification and nature of the infrastructure, its European and international roles, as well as the scientific and technical concepts, data management, scientific production, innovation and socio-economic impact. It also included a new section on Open Science. The CDS application highlighted its European and International dimensions in terms of the partnerships, and also in terms of the IVOA as an international infrastructure. The special role played by CDS in terms of Open Science, namely the application of FAIR principles and participation in the development of the Open Science infrastructure such as EOSC was highlighted. (A new image for inclusion in the published document is show in Figure 1.)

CDS was notified of the successful application in October 2021, and as such is explicitly included in the national strategy for research infrastructures. This ‘labellisation’ covers a ~5 year period, with the next call for updates expected in 2025. This status attributed to CDS is a high level recognition of its role within the French national research infrastructures and is expected to help with maintaining the support of CDS.

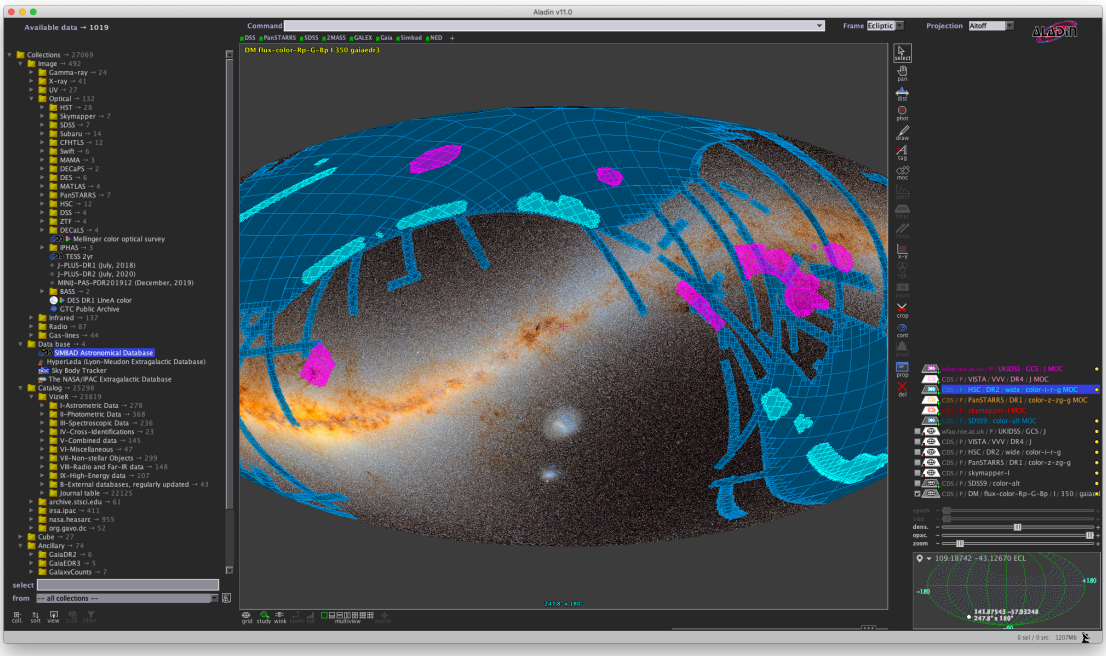

Figure 1. - An image provided for the French National Roadmap of Research Infrastructures. It shows Aladin Desktop and its data access and all-sky capabilities .

3. Highlights 2020-2021

SPECFIND V3.0 released

The updated **SPECFIND V3.0** radio continuum source catalogue, described in Stein et al. 2021[1] (A&A, 655 (2021) A17) was released in VizieR[2], and the press release[3] was highlighted on the A&A site with the image as shown in Figure 2.. Using 204 radio continuum input catalogue tables from VizieR we cross identify 1.6 million radio sources. Some **340 000 radio spectra were created** (tripling the number from earlier SPECFIND catalogues). Furthermore, the paper identifies a sample of objects that may have a special break in their spectral energy distribution, and require follow-up observation for better identification. It is an example of CDS team-work on a specific science topic led by CDS Postdoc (Y. Stein) and involving CDS scientists, engineers and documentalists.

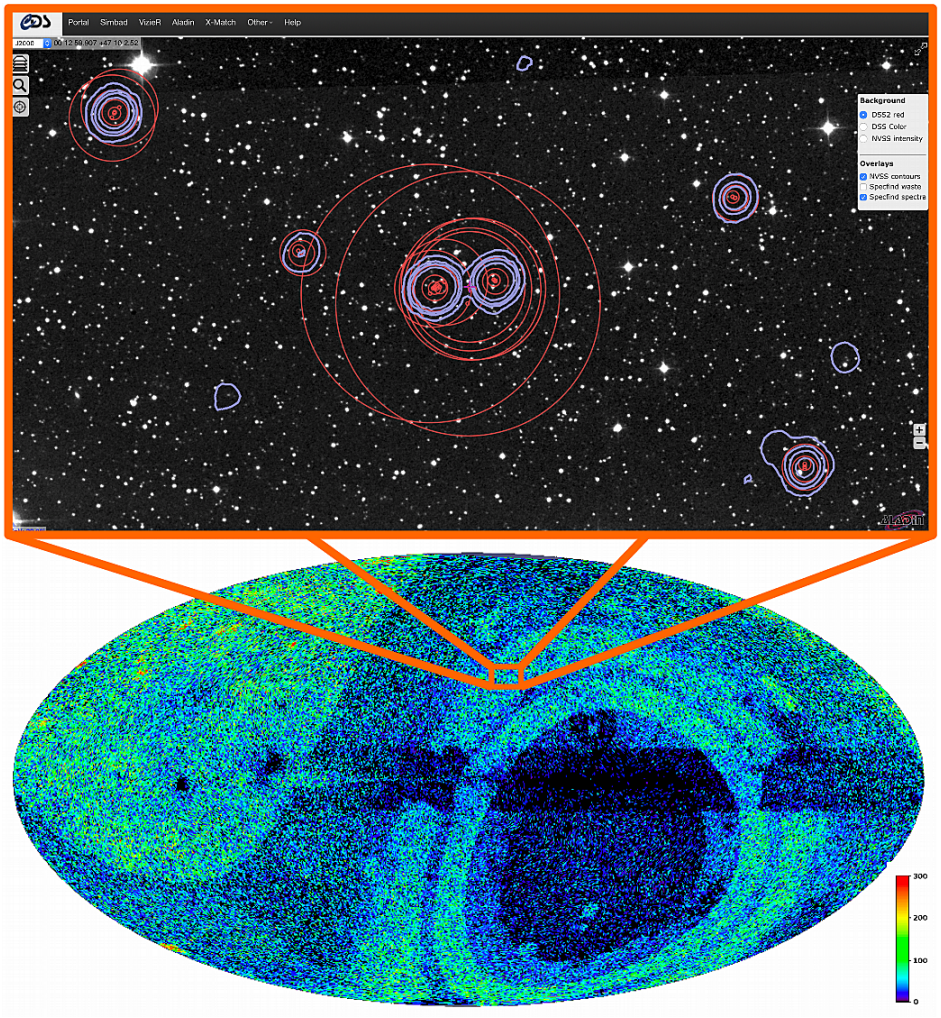

Figure 2. - Sky coverage of the 1.6 million radio sources of the SPECFIND V3.0 catalogue. The colour represents the number of sources in each 0.2 square degree pixel. An Aladin Lite visualiser connected to the VizieR catalogue shows the detail for each source.

[1] https://www.aanda.org/articles/aa/full_html/2021/11/aa39659-20/aa39659-20.html

[2] https://cdsarc.cds.unistra.fr/viz-bin/cat/VIII/104

[3] https://astro.unistra.fr/en/2021/10/29/new-specfind-v3-0-catalogue/

New CDS publishing registry

CDS makes its services available via Virtual Observatory standard protocols and provides a 'publishing registry' that is harvested by Virtual Observatory registries (via the IVOA 'Registry of Registries') so that CDS services are accessible in many tools and interfaces. **A new version of the CDS publishing registry has been released in June 2021.** The update makes it compliant with the most recent IVOA registry recommendations. The upgrades include: A description of the VizieR mirrors, MOC footprints, spectral bands, catalogue DOIs, better management of related resources, and also a keyword mapping using the Unified Astronomical Thesaurus (UAT).

Gaia Early Data Release 3 (EDR3)

On December 3, 2020, the **CDS published the ESA Gaia mission 'Early Data Release 3'** 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 20000 catalogues and hundreds of image surveys. Aladin and Aladin Lite visualisers provide an all-sky hierarchical view of the entire data set, using the capabilities of HiPS (Hierarchical Progressive Surveys) technologies developed at CDS. CDS also gives access to the EDR3 cross-id tables, and Gaia EDR3 is also made available from the CDS via Virtual Observatory protocols (IVOA Cone Search, and Table Access Protocol). Gaia EDR3 is the result of 34 months of observations (July 2014 – May 2017), and contains the positions of 1.8 billion stars. This major data release includes measurements of parallaxes, proper motions and colours for nearly 1.5 billion stars.

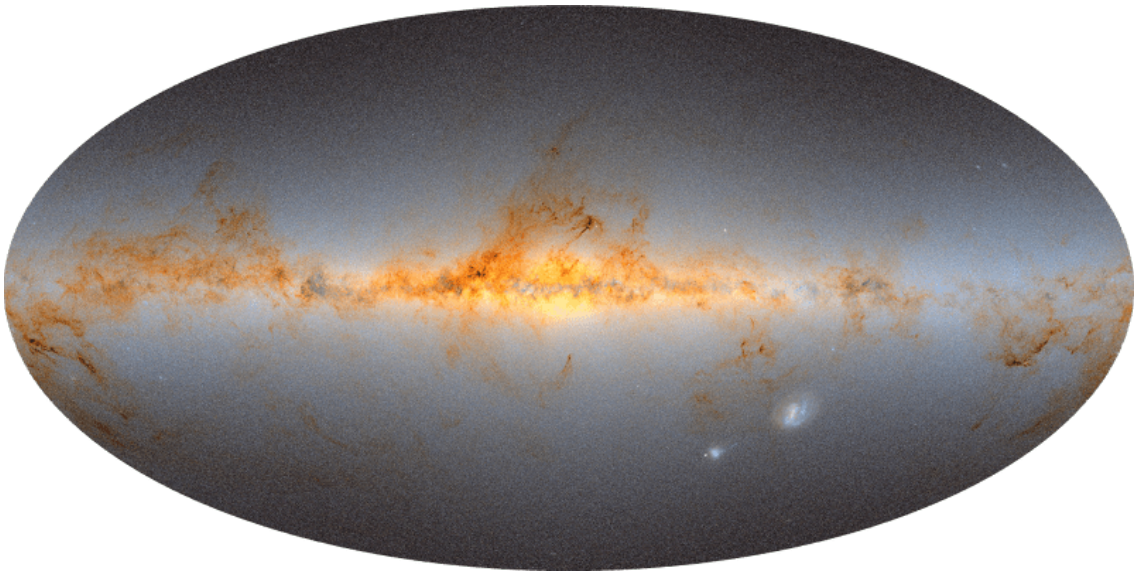

Figure 3. - All-sky view of the flux map of the 1.8 billion Gaia sources generated by CDS

3.1 Highlights of CDS in the Community

The CDS has actively participated in a number of astronomy community events where we have interacted with researchers, partners, journals/publishers, other data centres, observatories, missions and projects. All these events have been virtual over the past year, and we note high levels of participation due to greater accessibility of virtual meetings. Many invited talks have been presented by CDS staff at these large events over the past year (indicated below).

One of the highlights is the Library and Information Services in Astronomy (LISA IX) conference that was held on-line in June 2021 (after being cancelled in 2020). The data stewardship aspects of the CDS work were presented in talks and posters. CDS was also a major contributor to the "Best Practices for Data Publication in the Astronomical Literature" paper[4] (ApJS submitted) which was presented at LISA IX, led by the NASA Extragalactic Data centre (NED).

The highlighted community events are listed below:

Astronomical Data Analysis Software and Systems (ADASS) Conference, Virtual (Hosted by IAA-CSIC, Granada), 8-11 November 2020
Invited talk: *Data interoperability – The CDS experience. (Pierre Fernique and the CDS team)*
(see the CDS-ADASS document for all CDS contributions).

IVOA Interoperability Meeting, Virtual, 17-19 November 2020

American Astronomical Society (AAS) winter meeting, Virtual, 10-15 January 2021
Booth & webinars (M. Allen, M. Brouty, S. Derriere, G. Landais, C. Loup, K. Lutz, M. Neuville, A. Oberto, E. Perret).

New Year Lectures 2021 organized by the NAOC (China National Astronomical Data Center), **18 January 2021**. *SIMBAD, VizieR and Aladin: The CDS Astronomical Suite - P. Fernique*

LINEA workshop On the Future of Data Centers and eScience Institutes, 13-15 April 2021
Invited talk: *CDS services for reference astronomy data - supporting Open Science,* M. Allen.

IVOA Interoperability Meeting, on-line, 25-28 May 2021

Library and Information Services in Astronomy (LISA IX) Conference, June 14-18, 2021
Contributed talk: *Impact of Astronomy Evolution on the Documentalists' Activities at CDS*
E. Son, S. Lesteven, F. Marquis, F. Genova, and M. Allen (video link[5]), + posters

American Astronomical Society (AAS) summer meeting, Virtual, 7-9 June 2021
Special session — Astronomical Data Visualization in the Age of Science Platform
Invited talk: *Aladin Lite, ipyaladin and the HiPS ecosystem,* T. Boch and the Aladin team

European Astronomical Society annual meeting. 28 June 28 - 02 July 2, 2021
Virtual Exhibit - S. Derriere, M. Allen, S. Amodeo, M. Buga, E. Collas, H. Heinl
Special Session - Legacy and stewardship of astronomical archives towards multi-instrument, multi-wavelength and multi-messenger science.
Invited talk — *Spatio-temporal exploitation of astronomical archives,* S. Derriere.

[4] https://arxiv.org/pdf/2106.01477.pdf

[5] Video of the presentation: https://youtu.be/FNfIAgkIoRM

Highlighted events in the French community:

UNISTRA Open Access Week, Virtual, 5 November 2020
Invited lecture: *The FAIR principles in action for Open Science* (M. Allen)

Journées EOSC France 2021, Virtual, 4-5 February 2021

Action Specifique Observatoire Virtuel (ASOV) annual meeting, Virtual 22-23 March 2021

AstroInfo School 2021 (SOC - A. Schaaff, ObAS participants)

Social media communications channels: Facebook (@CDSportal) and Twitter (@CdSportal).

Video materials and hosting

We took advantage from the UNISTRA video hosting facility to request the creation of a dedicated channel[6] called "Centre de données astronomiques de Strasbourg". We have successfully tested that this service can be accessed from countries that restrict access to the YouTube servers (in particular with our China-VO colleagues). By hosting mirrors of the videos on this platform, we will also be able to embed them on CDS web pages.

In 2020-21 we have generated a lot of new video material from our participation in virtual events. Work will continue to make a more consistent and organised collection on the UNISTRA site and then we will advertise it more widely.

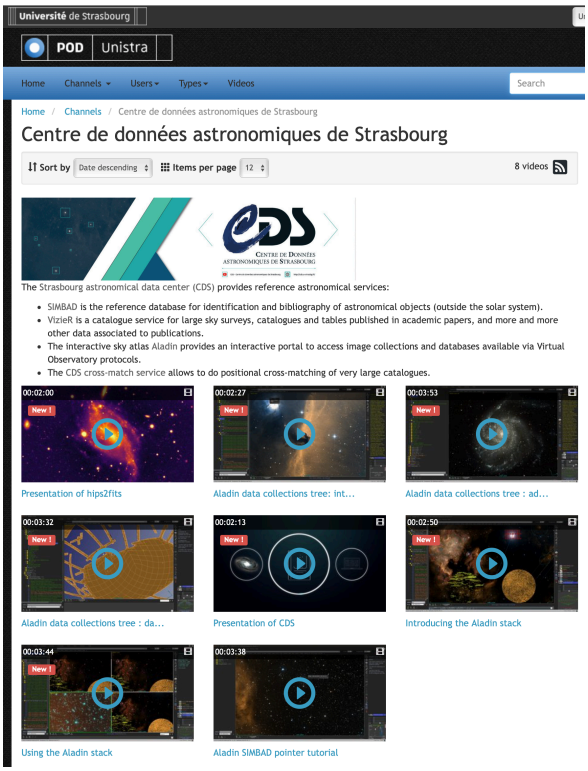

Figure 4. - The CDS video channel on the UNISTRA platform

[6] https://pod.unistra.fr/centre-de-donnees-astronomiques-de-strasbourg/

4. Activity Report for CDS Services 2020-2021

4.1 CDS Information System

The overall statistics for the main services are shown in Table 1. These show that the CDS services continue to be heavily used by the community. The total number of queries per day is ~1.9 million (compared to 1.75M in 2019-20). There are some decreases in the usage statistics compared to the previous years, some of which can be attributed to the fluctuations as described in the 2020 report, and some of which may reflect changes in the behaviour of users and require further investigation.

Main Services →	SIMBAD	VizieR	Aladin	Total (main services)
users / month	**146k** (-13%)	**30k** (-10%)	**230k**	**>230k**
queries / day	**393k** (-46%)	**569k** (+9%)	**925k**	**1.9M**
load / day	**5.3 GB** (-34%)		**100 GB**	**>105 GB**
data volume	**27 GB** (+12%)	**53 TB** (+14%)	**390 TB**	**435 TB**
data content	**12.6 M obj.** (+7%)	**21.3k cats** (+3%)	**990 HiPS**	
reliability	**99.44 %**	**99.65 %**	**99.61 %**	**> 99.44%**

Table 1. CDS statistics Oct 2020 - Sept 2021

The decrease in the number of queries per day on SIMBAD (-46%) follows a very large increase (+61%) reported in the previous period, and shows that significant fluctuation can occur year to year. Investigating further with statistics going back to late 2019 we find that the number of objects 'fetched' from SIMBAD (Figure 5) remains relatively constant (although we note a big peak in March 2020). The fluctuations may indicate some changes in user behaviour. We have already discussed with some of the people/projects involved to understand the reasons for changes in the way they send queries. In some cases a decrease is simply due to the end of a large project. In others (e.g. STScI) they have taken our suggestions to use more efficient query modes, reducing the number of individual queries. A deeper investigation would be necessary to understand some of the other fluctuations.

The use of the Aladin HiPS service is high, and its audience continues to grow, mainly driven by the Aladin Lite client and external public tools.

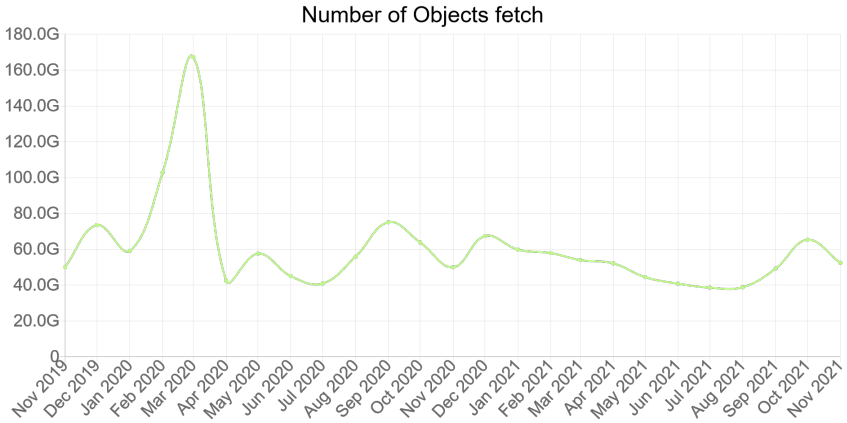

Figure 5 - Number of astronomical objects 'fetched' from SIMBAD Nov 2019 - Nov 2021

The operations of the CDS information system have been relatively smooth over the past period. Close and efficient coordination is done with the ObAS info-sys team.

The effort to update the CDS information system has continued in the direction initiated two years ago (and described in the 2019 report): rationalisation of the system components and removal or evolution of obsolete components. This action must be driven by a precise knowledge of the functioning of the information system. This knowledge is based on two axes: operational supervision, and monitoring of the use and content:

- **Service supervision:** Overall supervision is provided by an in-house tool: *GluSupervisor*, which tests all the system's components every 10 minutes and sends alerts when issues arise. The detailed monitoring of the components is based on classic tools such as *nagios* or *zabbix*. These tools, coupled with strong redundancy, have made it possible to improve and now maintain the availability rate of our services at more than 99.4% last year for all of our systems.
- **Service monitoring for usage and content statistics:** The majority of the monitoring of content and usage of CDS services is provided by a single in-house tool for all the components — the *CDS dashboard* (see the 2020 report and the current results in Table 1 above).

The evolutions of the CDS information system that have been implemented in the past year include:

- **Deployment of the new Bibliographic Centre Supervisor (BCS)**— the internal CDS management system for the processing of the incoming journal articles, and the extraction of bibliographic information to support the CDS service workflows for SIMBAD DJIN, and for VizieR.

- **Retirement of a number of older services** — which been made redundant by new developments. This includes:
  - The Aladin legacy image server which has now been totally replaced by the Aladin HiPS systems on the All-Sky-Data system hardware.
  - SIMPLAY, a widget for displaying simplified SIMBAD results, now replaced by use of Aladin Lite.
  - The thumbnail image service, which is now replaced by the generic cutout service provided by hip2fits.

- **Implementation of a unique domain name cds.unistra.fr**. CDS currently uses a mix of domain names for historical reasons. We note that previous changes such as the change in the UNISTRA domain (previously u-strasbg.fr) and the naming of the University itself (previously Université Louis Pasteur) have meant that there was some uncertainty. Now we are in a position to migrate to a unique name which will provide better consistency, indexing and also control of the certificates that need to be managed. An extended migration over 1-2 years, and maintenance of aliases will enable this to be implemented with minimal impact on users.

- **A re-design of the SIMBAD object-type hierarchy**. This will imply a number of information system changes in SIMBAD and in DJIN and these changes are in progress.

- **Dictionary of Nomenclature workflow review**. A deep review has been done and a new 'SIMBAD-centric' architecture has been proposed and is currently in a prototyping phase.

Other developments that support various CDS services include:

- **Deployment of MOCServer 2.** This is a fundamental development that is based on the MOC standard for the description of spatial-sky-coverage of data sets (images, catalogues, and indeed any thing related to positions on the sky) that has recently been updated with the ability to also characterise temporal coverage of data sets using a similar hierarchical method.

- **Renewal of the CDS web pages.** This is currently in progress with Hugo/gitlab as the proposed technical solution that matches the needs for maintaining the CDS informational pages and also the templates used for the CDS service functional web pages.

- **Rationalisation of the CDS source code tracking**. An effort is in progress to increase the consistency in the way the CDS source code is managed. We aim to use:
    - Gitlab CDS for all of the code.
    - GitHub for a number of public libraries and tools (MOC libraries, Aladin Lite, …)

- **CDS hotline.** A study was conducted this year to assess the impact of replacing the CDS hotline with a ticketing tool (GLPI). This study has already led to the rationalisation of the operation of the current hotline (redefinition of roles and automatic redirection of requests to the relevant people).

4.2 SIMBAD

Workflow

SIMBAD now contains 12.7 million astronomical objects, 44.8 million identifiers, and 396,000 bibliographic references, corresponding to an increase of 1.2 million, 8.2 million, and 15,000, respectively, in one year. The large increase of the number of objects is mostly due to the release of large astrometric (Gaia), spectroscopic (e.g. GAMA, SDSS, RAVE, GALAH, APOGEE), and stellar variability (Gaia, Kepler, TESS, ATLAS, ASASSN) surveys. Despite the second period of home office due to the COVID pandemic, the processing was maintained at the best possible level for the creation and processing of new references, as well as for the appraisal and ingestion of tables of objects.

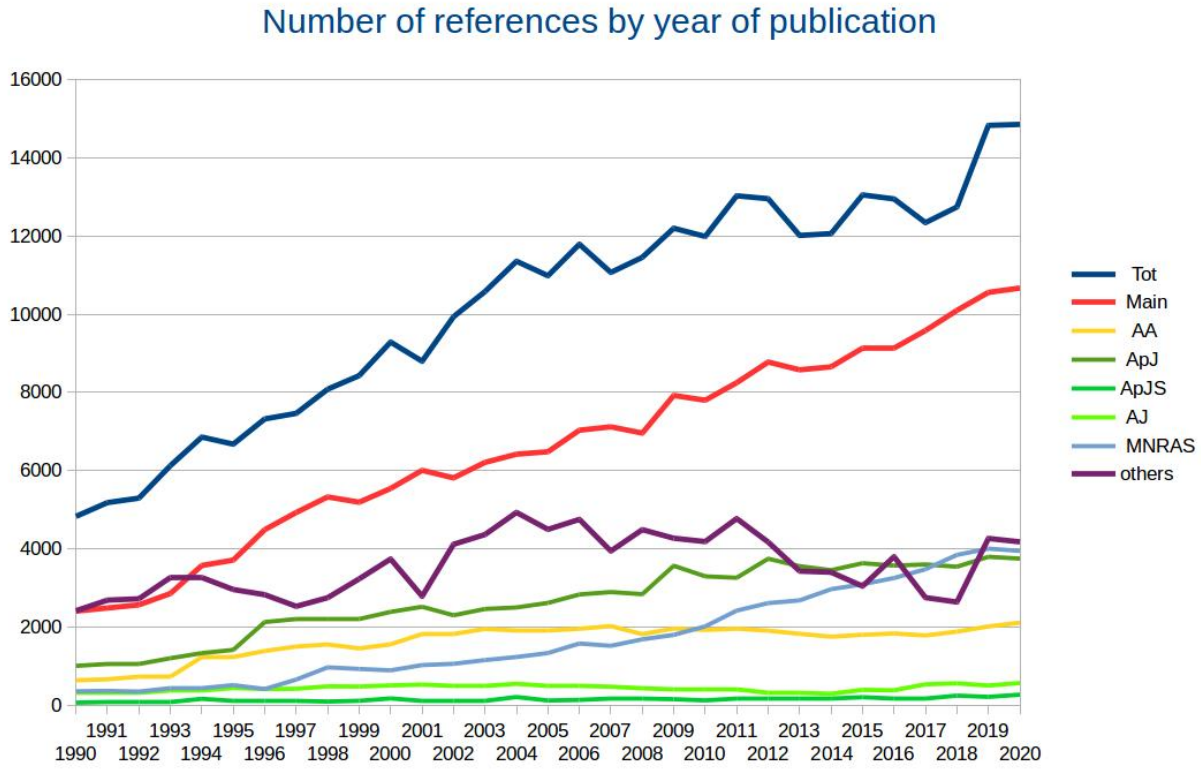

Figure 6 - Number of references per year processed for objects in the text of the article

The full processing of a reference in SIMBAD requires 2 to 6 steps. First step, the new reference is created in the database (with title, authors, abstract, but no linked objects yet). Second step, the reference is processed by a specialised team of documentalists with the help of the DJIN software; objects in the text, figures, and to some extent small tables are then updated or created in the database and linked to the reference; in addition, flags are added if there is a need for a new acronym (third step), if a large table has to be ingested in VizieR (4th step), or has to be appraised (5th step) and processed by another team by script (6th step). The two first steps follow closely the publications and are up-to-date. The evolution of the number of references per year of publication is shown in Figure 6. It is steadily increasing since 1991.

If the reference contains large tables of objects, there are 3 additional steps : ingestion in the VizieR database, appraisal in bi-weekly meetings to asses the relevance and the priority for ingestion in SIMBAD, processing by script by another team of specialised documentalists with the

help of the software COSIM. These 3 steps require much more time than the two first steps, typically 1.5 ± 1 year after publication. Figure 7 shows the status of tables of objects ingested in SIMBAD as a function of the year of publication. Not only the number of tables increases, but also the number of objects in tables increases. The ingestion is nearly complete up to 2018 (the yellow curve contains articles with a low priority for ingestion). For years 2019 to 2021 the processing is in progress; there are still tables to be ingested in VizieR, to be appraised, and then to be processed. The drop of the curve after 2018 is thus normal, it just shows how much time is required to fully process a reference with large table(s) of objects.

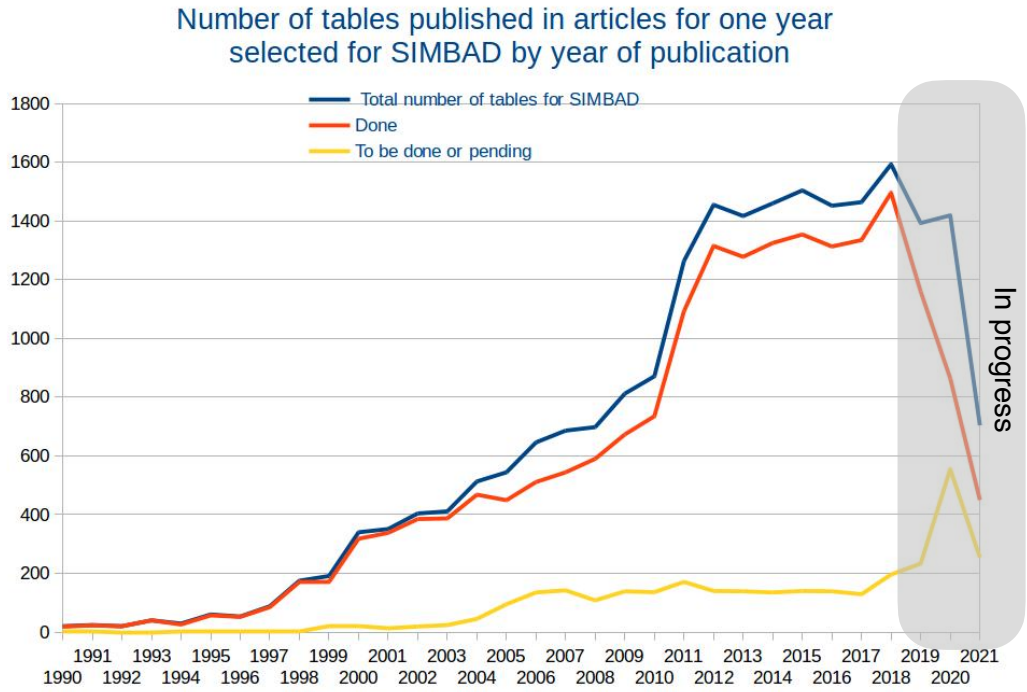

Figure 7 - Number of tables in articles that are selected for inclusion in SIMBAD

Reorganisation and maintenance of object types

The reorganisation of object types in SIMBAD has been successfully implemented. It improves the quality and speeds-up the cross-identification process when ingesting tables of objects with COSIM. When the main object type in SIMBAD is not compatible with the one as given in the table (e.g. QSO versus White Dwarf), while all others parameters give a positive match, the object is automatically stored in a file. These cases are then sent to a CDS astronomer for scientific expertise. This allows us to maintain the accuracy of the main object type in SIMBAD at a high quality level with much less work than in the past.

Bibliographical Centre Supervisor

A new process for ingesting papers directly from publishers has been deployed. It started as an internal repository used for bibliography in SIMBAD, processing the content of papers for SIMBAD ingestion (using DJIN), and also for table ingestion into VizieR service. More discussions and evolutions are planned in order to harvest and process articles coming from different publishers in an (as most as possible) homogeneous way.

Technical evolution on server-side

A major release was made of the SIMBAD core system. It mostly includes: an upgrade to a more recent version of Java (from 6 to 11), and the removal of an old internal communication mechanism between server modules. This evolution is a first step to move SIMBAD toward a micro-services oriented system, and it makes client software (mainly for documentalists) more independent.

Dictionary

The object identification system (Dictionary of Nomenclature) in SIMBAD is undergoing a revision (as mentioned above). The link with the IAU and also its role for internal CDS usage is being taken into account in parallel. C. Loup has recently been elected to the IAU Commission B2 on 'Data and Documentation' and efforts are underway to continue the activities of the IAU WG on Designations by a proposal for this be continued as a IAU Functional Working Group. CDS continues to maintain the IAU Dictionary of Nomenclature.

Future plans

On a short time scale, the new hierarchy of objects types will be implemented in the query services of SIMBAD (criteria and TAP queries); then it will become available to our users.

We shall pursue our effort to cross-identify the content of SIMBAD with the successive releases of Gaia, not only for the scientific interest, but also because the improvement of the astrometry in SIMBAD speeds-up the ingestion process. We also plan to ingest new objects of interest selected in Gaia, especially the closest stars, high proper motion stars, and after the DR3 release, objects with a good measure of the heliocentric velocity. At the moment all stars with a parallaxes larger than 30 (and a measurement uncertainty smaller than 10%) have been ingested.

On a longer time scale, we should develop strategies and technical tools to systematically add validated cross-identifications with very large surveys, especially SDSS, Pan-STARRS, SkyMapper, UKIDSS, and All WISE. We could then provide to our users Spectral Energy Distributions based on cross-identifications.

4.3 VizieR

VizieR content

The VizieR team has been able, after an adjustment period, to perform most, but not all, of its activities under remote working conditions. The standard VizieR treatment of catalogues has resumed smooth operation, as can be seen by growth of VizieR content in the table below. Following new priorities for the ingestion of A&A catalogues, their total number for the period has come down slightly to a more manageable 439 catalogues, close to the desired target of 40 refs/month and a total of 450 refs/yr, in agreement with the A&A editors.

VizieR Content

	2016	2017	2018	2019	2020	**2021**
Number of Catalogues	15 485	16 528	17 673	19189	20289	**21412 (+1123)**

Table 2. VizieR Content 2016 - 2021 Oct 2020 - Sept 2021

Since September 2021, VizieR meetings are held in person again. We also plan to reactivate the “Associated data crunch sessions” dedicated to the ingestion of associated data, which allows to process a fraction of the associated data in the waiting queue.

Very large *versus* **long** *versus* **thick catalogs**

G. Monari, recruited in October 2019 as astronome adjoint, is the main scientific supervisor of the large catalogues section of VizieR. Under his supervision, and in collaboration with engineers F-X. Pineau and T. Boch and the documentalists, 9 large catalogues have been ingested, including the Gaia Early Data Release 3, the ESO VISTA Hemisphere Survey DR5, and the Extended Gaia-Pan-STARRS1-SDSS (GPS1+) proper motion catalog. CatWISE, which was ingested last-year, was re-processed following publication by the consortium of corrections for their measured positions. We also added a few columns with respect to the original entry. The depth of the re-processing (essential because of the coordinate corrections that are important for indexation in our pipeline) required a quasi complete re-ingestion.

To be expected in the 2021-2022 period, are Gaia Data Release 3 (May 2022), the Guide Star Catalog 2.4, SDSS DR16 (in progress), and several ESO phase 3 catalogues. Moreover, we have re-established contact with the Pan-STARRS consortium, who seems open to sending the data to CDS, though this will require a sustained action and collaboration on their side. Not all hope is lost on this item, but it has been on standby for a long time now due to lack of communication from the Pan-STARRS contact at MAST.

The trend of “long” catalogs continues in 2021: small collaborations, even a single author, working with a large input dataset such as Gaia, are able to produce large catalogues of up to a billion sources and more, which we must process through the large catalogs pipeline, even though they may have just one added quantity with respect to the reference catalogue. These are “long” catalogues, i.e. they can have a billion lines but only a handful of columns. An example this year is Bailer-Jones 2021, which provides distance estimates for 1.47 billion stars using Gaia EDR3. There is no doubt that their competitor StarHorse will also release a EDR3 version of their own distances.

Another continuing trend is that of “thick catalogues”, such as RAVE DR6 (Steinmetz et al. 2020), with only 518387 objects but 19 tables, for a grand total of several 100 columns, and RELICS (Coe et al. 2019), with only 190000 objects but more than 150 columns homogenized from 46 tables that were merged at CDS. Such complexity peaks naturally require documentalists to spend much more time than average on these catalogues.

Catalogue created in collaboration with Observatoire de Paris (and EuroPlanet)

There is an ongoing collaboration with Observatoire de Paris to build a compiled solar system catalogue according to the VO EPNCore Data model. The curation is operated by CDS and Obs Paris to populate a table where each records describes a VizieR solar System catalogue, and is connected to the CDS participation in the EuroPlanet project. The TAP VizieR server has been updated for the workflow (and was presented in a recent ADASS poster led by S. Erard — Obs Paris). The B/planets catalogue is a product of this collaboration, and it is linked to the Europlanet VESPA portal[7].

VizieR usage

In terms of use of the service there was an average of 524.8k queries/day in 2021, in continuity with previous year. The huge majority of queries are directed to VizieR, and are performed mostly via python, other APIs, and the web forms. However, the associated data service saw a ~40% rise in usage, while TAPVizieR usage increased by about 65% over the year, drawing an overall stable and encouraging picture of VizieR usage in the community.

VizieR Usage

	2016	2017	2018	2019	**2020**	**2021**
All queries, /day	380 000	326 000	368 000	696 000	520 000	**524 800**
Associated Data Service Queries/day		270	80	543	845	**1212**
VizieR TAP service Queries/day		5 250	3 700	2094	14 000	**23 130**

Table 3. VizieR usage (queries/day) 2016 -

Digital Object Identifiers

Since 2019, the VizieR ingestion pipeline routinely includes the DOI of articles published in A&A, and CDS creates a DOI for the catalogues (via our DOI authority managed via INIST), pointing to the catalogue landing page. Last year, we reached an agreement with AAS to extend this framework to their publications, and are now generating DOIs for their catalogues. We have yet to seek an agreement with MNRAS. No formal agreement exists yet with ESO, and therefore VizieR refrains from creating DOIs for ESO/Phase3 catalogs, but we will seek a dialogue on this topic with ESO. On the ESA side, a specific action was planned following suggestion made at the 2019 Scientific Council meeting to create a DOI for GAIA catalogues and this action is completed.

[7] http://vespa.obspm.fr/planetary/data/

Time domain

Metadata for the time columns are now routinely created in VizieR by the documentalists. As a result, it is possible to compute the epoch of observations in the reference time system used by VizieR, i.e. TCB/Barycenter. This output is available in the VizieR web interface under the “compute” menu, as well as in the temporal SED tool. The time metadata in VizieR also enables the spatial-temporal (TMOC and STMOC) coverage maps of catalogues to be computed, and this has been done for 1400 tables so far.

Workflow and technical evolutions

- Supervised automatic description of catalogues with colmeta: this tool, developed by F-X. Pineau in close collaboration with E. Perret, partially automates the generation of the standardised catalogue documentation using commonly encountered column descriptions, measurement names and units in the ReadMe file, allowing documentalists to focus on producing a higher quality description and presentation of the data. This tool is in production and is updated on a regular basis to increase its functionality.

- In the direction of improving the tools used by VizieR documentalists: BCS (Bibliographic Center Supervisor) is a new tool developed and operated by G. Mantelet for SIMBAD. It centralizes the retrieval of articles and their data and metadata from the publishers. Although originally a development for SIMBAD, it has been integrated into the workflow of VIzieR for the ingestion of AAS papers, and is an illustration of existing synergies between Simbad and VizieR at the workflow level.

- Improving the security of the data depot: the anonymous ftp depot available to users to send their data to CDS is being replaced by a new interface and process: upon providing an email address (https://cds.unistra.fr/ftp/token/) and filling a captcha, the author receives a password usable for 3 days to deposit their data on the CDS servers. This eliminates a security loophole.

- Mini-VizieR installations at observatories: ESO and CFHT have expressed an interest in having a scaled down (in content and functionalities) version of VizieR running at their observatories. G. Landais has conducted the required developments for the ESO installation, however this has been delayed by 1-year on the ESO side (and some changes in requirements will need to be discussed). The project with CFHT is currently on stand-by.

- Improving the large catalogs query interface: F.X. Pineau is working on expanding the positional query functionality of the interface (more general positional constraints are available, such as box region, and query by MOC is in development). Also, on-the-fly computation of columns will be possible, as well as more general indexation.

- The TAPVizieR categories have been re-organized following the VIzieR category convention (http://vizier.u-strasbg.fr/vizier/welcome/vizierbrowse.gml?designation) for clarity.

- R&D efforts have been undertaken to manage large tables and distributed databases, in the form of a wrapper between CDS large file technology and postgres database (Foreign Data Wrapper: 6 month engineer trainee from September to February), as well as a 3 month internship to investigate the cluster postgres database.

- CoreTrustSeal: The certification must be renewed in 2022 (every 3 years). The next iteration will require a more accurate description of the licensing aspect of the data distributed by VizieR.

Strategic elements for Vizier:

Strategic developments for VizieR are currently being discussed. Some of the elements are mentioned below and others related to the future large data sets will be presented at the meeting.

Towards a merge of the historical pipeline and the large catalogues?

While the “normal” and large catalogues can be queried through the same VizieR classic interface, their pipelines, servers and ingestion procedures are different. A fundamental reflection is underway to try to merge them. The colmeta tool is part of this attempt, as it started out as a large catalogs tool but is now a routinely used for the other catalogs as well.

A strategic emergency: the need for an additional VizieR engineer:

The VizieR code is now more than 30 year old, was built with previous generation frameworks/ architectures and hardware in mind. It consists of about 300000 code lines, and was built by a team of 2 (F. Ochsenbein and I. Bourekeb). Each migration / patch puts it at risk of a malfunction in the case of unsupported or evolving dependencies and the cost in maintenance is high. The only way out is a modern rewrite of the code. This can not be achieved by G. Landais alone as he is already in charge of the operational, everyday maintenance and stability of VizieR. The rewrite must be done with the future of the field and CDS / VO workflows and tools in mind. Also, the additional engineer will increase the redundancy in the service, which is very low at the moment.

A new strategy for recruiting VizieR documentalists?

The last VizieR documentalists recruited (contractually) were trained documentalists with no background in astrophysics. They had to learn the specifics of linux, VizieR tools and ingestion and acquire an understanding of the field at the same time. This year, we are trying a new strategy by targetting M2 astrophysics students (from Strasbourg and other observatories/universities in France). We can expect them to have good computer skills (including a working knowledge of python and linux) as well as a fresh background in astrophysics. The expected recruitment (1-2 year contract) will take place in 2022.

4.4 CDS X-Match

The X-Match service continues to operate smoothly with the current established code. We have invested in the disk capacity of the service, doubling the capacity to 42 TB of fast SSD disks. This increase has been done in anticipation of forthcoming large surveys to be ingested in the service such as Pan-STARRS DR2, SDSS DR16 or Gaia DR3, plus the pressing user demands to increasing the number of available columns which requires greater disk capacity. An update of the service which provides more capabilities for storing user uploaded files and cross-match results has been prepared and will be released in 2022.

Usage Statistics

The number of jobs submitted through the interactive web interface continues to grow modestly year after year and is 64 jobs per day with an average of 77 million associations per day. The API is much more heavily used, the usage doubled between 2019 and 2020, and has jumped again with the number jobs tripled in 2021 to ~25732 jobs per day. This increase is dominated by a large number of jobs for cross-matches of relatively small numbers of input entries. As such the total number of associations is not changed very much. This is an interesting change in the user behaviour which will be closely followed to understand how it is evolving.

CDS X-Match Service Usage

	2013	2014	2015	2016	2017	2018	2019	2020	**2021**
Web interface (jobs/day)	15	16	20	30	33	40	43	49	**64**
HTTP API (jobs/day)	47	50	580	889	1256	2687	2556	7244	**25732**
Associations/day (Web Interface)	~13M	~70M	~55M	~104M	~164M	~179M	~44M	~72M	**~77M**
Associations/day (HTTP API)	~298k	~1.6M	~6.6M	~6.7M	17.8M	50.2M	43M	90M	**78M**

Table 4. CDS X-Match service usage statistics 2013-21 2020 - Sept 2021

Developments

During the year, we continued to pursue the development of libraries in the Rust language, and tools which are useful for the CDS activities (in particular VizieR large catalogues) and which will be included in the new CDS X-Match prototype. The components we have already developed (or were already well advanced) are: a HEALPix library, an expression evaluation library, and a framework to easily support multiple file formats and a column indexation library.

This year, we also put a large effort into extending the original Rust core of MOCPy and making a standalone MOC library in Rust. In addition to new functionalities and performance improvements, we adapted it to the newly proposed version of the MOC 2.0 IVOA standard, and it serves as a reference implementation for the standard. This library can be used in Rust projects and is still at the very heart of MOCPy (the CDS MOC manipulation library in python,which is an astropy affiliated package). From the MOC Rust library, we also developed a MOC manipulation command line tool made of a single executable which is pre-compiled and available for various platforms.

Finally, still from the MOC Rust Library, we made MOC manipulations available in Javascript (from Web Browsers and tools like Aladin Lite) by developing a WebAssembly library. While the scope of those developments is much larger than the X-match service, the MOC Rust library will take a place in the new prototype. In addition, we started an astrometric library in which we already implemented the epoch changes of ICRS coordinates (including error propagation) from a full set of Gaia-like astrometric parameters. It will be used to possibly cross-match catalogues with astrometric parameters and different epochs.

Future

Here we list the points we would like to achieve during the coming years:

- Continue the development of the next generation croc-match service prototype ‘ExXmatch’.
- Redesign the current interactive Web Interface.
- Redesign the structure, API and code of the X-Match service.
  - provide all VizieR columns while allowing the user choose the desired column set.
  - allow post filtering to reduce the size of output file.
  - remove the distinction between large and small jobs by automatically splitting large jobs into smaller, individual jobs.
  - let the user upload large tables (and automatically transform the file on the server to prepare it for efficient cross-matches).
- Provide a command line client for a better programmatic access.
- Allow for cross-matches taking into account proper motions.
- Explore distributed cross-matches for better performances.
- Explore the possibility to let the user make plot from the server data to avoid large file transfers.

4.5 Aladin

The Aladin service encompasses the curation of scientific content (images and cubes) and development/operations of client software Aladin Desktop and Aladin Lite. The scientific content continues to increase and the usage has remained stable at a high level. One of the highlights this year was making the move of part of the HiPS storage to UNISTRA data centre. We also gave special attention to different aspects of the interactions with the user community.

Interaction with user community

Interaction with the user community on the Aladin software was pushed at different levels:

> **Documentation:** the long awaited translation of the Aladin manual to English has now been done as a coordinated effort between H. Heinl, C. Bot, M. Buga, K. Lutz, T. Boch, M. Allen and P. Fernique. The manual for Aladin v10 was ready in time for the ESCAPE VO school "*Science with interoperable data"* in February 2021, with the latest manual[8] for Aladin v11 in English following shortly after.
>
> **Participation in conferences, workshops, school:** In November 2020, 4 talks and posters related to the Aladin project were presented at ADASS and resulted in 4 papers (C. Bot on Hipsgen, F. Bonnarel on the SIA service, T. Boch on HiPS related tools, M. Baumann on Aladin Lite v3). T. Boch was invited to the AAS data visualization workshop in June 2021 and gave a presentation on the "HiPS ecosystem" and P. Fernique presented the Aladin project to a China VO conference in January 2021. Aladin was heavily used in the ESCAPE VO school on 'Science with interoperable data' in February 2021.
>
> **Case-by-case assistance:** Help was provided by interacting directly with key/large projects.This is the case for example for the LIGO/VIRGO project for which we extended the implementation of the statistical maps. We also want to mention the collaboration with M. Zamani (ESA/Hubble) which led to Aladin Lite being integrated in NOIRLab and ESA/Hubble outreach pages.

HiPS

Since the creation of Hierarchical Progressive Surveys (HiPS), the importance of this format has increased to become one of the cornerstones of the Aladin project strategy and HiPS are now the dominant scientific content in Aladin. This year, the HiPS ecosystem keeps on going strong with a high level of queries.

Since last year, many new HiPS have been published for a total of 990 available HiPS and a total size of 390TB. Significant additions include the MSX survey in all 4 bands, the Gaia EDR3 HiPS flux maps, Zwicky Transient Facility (ZTF) survey, Transiting Exoplanets Survey Satellite (TESS) images, the CFHT/MATLAS images in 3 bands, Atacama Cosmology Telescope, and the HI/OH/Recombination line (THOR) survey images.

As planned in our disaster recovery plan update, one of the All-Sky-Data system duplicates has been moved last May to the UNISTRA data centre. The operation was thoroughly prepared and went smoothly over the course of one day. The All-Sky-Data system is now setup as intended: one part at the UNISTRA data centre and the other in ObAS server room. This configuration should ensure a better HiPS data availability in case of issues at one of the sides.

[8] https://aladin.u-strasbg.fr/java/AladinManual.pdf

Due to the large number of HiPS surveys and the history of how they were built, some maintenance operations on HiPS have now to be done regularly. As an example, this year, we recomputed all 2MASS HiPS to include the true pixel values (the 2MASS HiPS were some of the earliest ones done and only *quick-look* images were available at the time). The SIMBAD bibliographical HiPS was kept up to date as well and a significant work was done to clean and update wavelength metadata keywords was done by Mihaela Buga.

Among the HiPS usage within Aladin, the **hips2fits** service allows easy generation of HiPS cutouts. Since last year, its usage has significantly increased, with more than 10k queries per day on average which corresponds to a 2.6 fold increase. Highs at 900k per day have been recorded. Also, a new SIA v2 service based on hips2fits is now available and facilitates access to cutouts for VO clients.

Aladin Desktop

The audience for Aladin Desktop remains at a high level of 6500 users/months (+0.6%), 950 launches/day (+3%) and 130K actions/day (-15%). This is rather constant compared to last year.

Time domain capabilities have been the focus of many developments, and are at the centre of Aladin Desktop V11. Work has been done on a prototype that enables a search for available datasets with time constraints. This was followed by the development of a new beta version implementing this possibility.

A new avenue has been the evolution of statistical pixel analysis capabilities, i.e. the possibility to easily access quantities like the average, the sum or the median of a set of pixels defined by the user. This functionality led to a collaboration with the CASSIS team in Toulouse to develop a CASSIS plugin for Aladin. This plugin will enable the visualisation of spectra as extracted from data cubes available in Aladin and for regions selected by the user. A demo of the Aladin and CASSIS Plugin interaction was showcased at ADASS 2021. The plugin will continue to be developed the year to come and be released to a wide audience.

Finally, miscellaneous functionalities were implemented, following needs from the user community or the evolution of CDS services. This includes a change of object categories for the outreach version along with the evolution of object types in SIMBAD; improvement for compatibility with new displays at very high resolution; the use of Aladin Desktop for the real-time display of the LUCI instrument at LBTO and help to users on HiPS colour composition of surveys.

A new Aladin v12 corresponding to the evolution of the current beta version may be available as soon as Spring 2022.

Aladin Lite

The Aladin Lite audience is still increasing. Aladin Lite previewer is launched 724 times per day on average (+9%). At the same time, the number of startups from sites embedding Aladin Lite is strongly increasing: these sites add up to 25k startups per day (+71%). We identified 3 main reasons for this increase:

- Some new sites are now integrating Aladin Lite. For instance, the esahubble.org pages represent nearly one quarter of all Aladin Lite startups. Noirlab.edu is another new site. These 2 sites using Aladin Lite are the result of the collaboration mentioned earlier.
- Audience increases from sites that were already using Aladin Lite.
- And unfortunately there is an increase in the audience sites that sell stars who have implemented Aladin Lite. This usage is currently not a major load on our servers but is increasing (and actions need to be taken to address this undesirable use of the service).

Many Aladin Lite functionalities were developed this year by M. Baumann (CDS ESCAPE contractor engineer) as part of the ESCAPE project. These include support of multiple projections (Aitoff, Mollweide, …), support of HiPS FITS tiles, density-map visualizations of catalogue data, improved rendering pipeline thanks to GPU usage through WebGL. These functionalities are currently integrated in Aladin Lite version 3, which is forecast to be published at the end of this year.

Links to planetaria

Several planetarium systems designers have expressed interest in accessing and using HiPS data for their sky simulators. The planetarium software Digistar 7 (a commercial product of Evans & Sutherland) is the system chosen for the upcoming new Strasbourg planetarium. Thanks to collaboration with CDS, it enables the native display of HiPS online sky surveys distributed by CDS in digital planetariums. This software will be used in tens of planetaria worldwide as well. Another company, RSA Cosmos also contacted CDS to implement visualisation of HiPS surveys in their SkyExplorer planetarium software. A workshop, “Astronomie en partage”, dedicated to scientific mediation in planetarium thanks to research data will be attended by CDS in November.

Future Strategies

- Aladin Desktop: Follow-up prototype evolutions corresponding to the implementation and the validation of VO standards (MOC, HiPS, DataLink,...)
- Aladin Lite: publication and follow-up of M. Baumann’s developments.
- Pursue the use of Aladin Lite in jupyter notebooks (ipyaladin).
- Selection of reference surveys and creation of the corresponding HiPS. Follow the development of LSST, Euclid, SKA etc. for their use of HiPS.
- Depending on the progress of Aladin Lite, support the transition from Aladin Lite to a more heavy client (possibly within 3 to 5 years).
- Renewal of the All-sky data storage to be planned for ~2025-26.
- Development of new services centered on HiPS (à la hips2fits): e.g. SEDs, HiPS cubes analysis, R&D on analysis tools using the hierarchical aspect of HiPS (e.g. filtering)
- Follow the evolutions of data centres. Under consideration are the virtualization of the alasky/alaskybis servers. Links with Data Lakes can also be a topic to follow.
- The long term evolution of Aladin Desktop will be considered: for example to identify which niche should be targeted compared to Aladin Lite. Other considerations include the evolution of the user/graphical interface. Which language? And at which cost? This will require R&D.

4.7 R&D

Since the last Scientific Council 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 7 interns and 1 short-term contract. Engineers from the ObAS GALHECOS team and from the Observatory IT Support were also involved in some actions. It is also important to mention that in several cases, the expertise of astronomers and documentalists is also very helpful, and at the end it is a team work.

As reported in last year's report, in 2020, 12 very motivated students did their internships under difficult COVID-19 conditions (working from home with remote supervision). This was difficult but however we maintained at least the same level than during the previous years. In 2021 we hired less interns than usual (7) as the perspectives of working under the continuing COVID-19 conditions was not enthusiastic. In 2022 we hope to return to the same level as usual.

We continued through two internships the work around the query of the CDS services in Natural Language. The objectives were both to extend its functionalities and to improve the user experience. The improvement of its voice interaction capability opens a new field of investigation for people with disabilities. An ongoing action is to evaluate the use of the Unified Astronomy Thesaurus (UAT) vocabulary. Tests with a growing panel of users is planned to prepare the opening on the CDS website through a CDS Lab section.

We started in September 2021 a new study called "Inclusive CDS" (poster at ADASS 2021). Its aim is to give an access to the services and data to all, including people with disabilities.

An intern worked on the continuation of the study about the integration of large tables in VizieR.

Linked to the IVOA, an intern (hired also for a short contract) continued the development of a CDS own VOSpace (access and sharing of astronomical data files) framework.

An intern worked also on the annotation of TAP responses on the fly with IVOA data models, it was presented during a TAP BoF at ADASS 2021.

An intern also worked on the Adaptation of TAPHandle to the complex data structures.

5. Projects

5.1 Virtual Observatory and Open Science projects

IVOA

CDS continues to play a leading role in the development of the Virtual Observatory. The full list of contributions to the IVOA is being compiled[9] in the document “CDS Participation in IVOA”, which includes a list of the standards developed in 2020-2021, and the CDS contributions to the IVOA interoperability meetings November 2020 and May 2021. We continue to acknowledge the long term collaboration of M. Louys (ICUBE) and L. Michel (ObAS) with CDS on the development of the VO. Their contributions to IVOA are included in the document.

CDS has made significant contributions to the leadership of the IVOA, in particular via various working group chair positions. Ada Nebot finished her term leading the Time Domain Interest / group and take on a big challenge of the Committee for Science Priorities. Gregory Mantelet is the new deputy Chair of the DAL Working Group, and Hendrik Heinl is now the Chair of the Education Interest Group. These responsibilities are listed below:

- **Executive Board member for EuroVO** - M. Allen
- **Chair of the Committee for Science Priorities** (since May 2021) - A. Nebot
- **Deputy Chair of the Data Access Layer Working Group** (since May 2021) - G. Mantelet
- **Chair of the Time Domain Interest Group** (May 2017 - May 2021) - A. Nebot
- **Chair of the Data Curation and Preservation Interest Group** - A. Schaaff
- **Deputy Chair of the Radio Astronomy Interest Group** - F. Bonnarel
- **Chair of the Education Interest Group** - H. Heinl *(CDS-ESCAPE Support Engineer)*
- **Editorial team for the IVOA Newsletter** - S. Amodeo *(CDS-ESCAPE Science Support Postdoc)*

Papers about the IVOA meetings have been published in the ADASS proceedings (Allen et al. 2020[10]) on behalf of the IVOA Exec and Technical Coordination Group. Details of the European contributions are also described in the ESCAPE project Milestone reports (see below).

ESCAPE

The **ESCAPE** (**E**uropean **S**cience **C**luster of **A**stronomy & **P**article physics **E**SFRI research infrastructures) project started in February 2019. CDS leads one of the six work packages (WP4, ‘CEVO’) on *Connecting ESFRI projects to EOSC through VO framework.* An overview of this project was provided in the 2020 report.

In April 2021 M. Baumann finished his contract on this project (having made very important contributions to mocpy and Aladin Lite — see Aladin and X-Match reports). H. Heinl continues as The CDS-ESCAPE support engineer providing technical support to the running of the Work Package, and also contributing his VO expertise to training events. An ESCAPE postdoc, Stefania Amodeo started in March 2021, providing much needed support and new ideas for the training materials and communications.

[9] Document to be available prior to the council meeting.

[10] https://ui.adsabs.harvard.edu/abs/2019arXiv191201481A/abstract

The main WP4 activities in 2020-21 have been:

- **D4.4** - Intermediate Analysis Report on VO data and service integration into EOSC - *Delivered November 2020*
- **4th WP4 Milestone** - Progress and Priorities at IVOA - *reported December 2020.*
- **D4.3 - 1st Science with interoperable data school, On-line, 8-12+19 February 2021**
- **WP4 2nd Technology Forum, On-line 13-15 April 2021**
- ESCAPE Progress meeting and General Assembly, On-line, 28-29 September 2021
- **5th WP4 Milestone** - - Progress and Priorities at IVOA - *reported June 2021*

Many other ESCAPE meetings of have been organised F. Bonnarel focusing on VO tools and standards for radio astronomy and high energy astronomy. All the activities are listed on the WP4 wiki pages[11]. The work was presented at the EAS and an article is being prepared fora Springer special volume on "Data-Intensive Radio Astronomy".

The project successfully passed a review of the first reporting period on 27 November 2020. The project activated 'Article 51' of the grant agreement in response to the COVID-19 pandemic, and a number of deliverables and milestones have been re-scheduled to take into account delays, and the project now has a 6-month no-cost extension so that the end date is now January 2023.

EOSC - European Open Science Cloud

The ESCAPE project is being done in the context of building the EOSC and brings CDS into contact with various projects connected with EOSC. The FAIRsFAIR project[12] is one of the important projects for EOSC that is *"Fostering Fair Data Practices in Europe"* and *"aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle".* M. Allen often represents ESCAPE in FAIRsFAIR events, and is one of the FAIRsFAIR European Group of FAIR Champions (EGFC). F. Genova is on the High Level Advisory Committee for this project.

F. Genova served as the vice-chair of the FAIR Working Group of the EOSC Executive Board[13], which published a series of reports on different aspects of FAIR in EOSC, including "Six Recommendations for Implementation of FAIR Practices"[14]. The working Group produced its final reports at the beginning of 2021. Astronomy is amply cited in several of the reports, and provides one of the Case Studies of the EOSC Strategic and Innovation Agenda[15] (SRIA) V1.0, which was published in February 2021.

The EOSC is now formalised as an Association which is a legal entity (as a Belgium AISBL - International non-profit association). CNRS and UNISTRA are both members of the EOSC Association. A number of Advisory Groups and Task Forces have been set up to guide the implementation of EOSC. With the support of CNRS, M. Allen was selected as a coordinator for the Task Force on *Use Engagement and Adoption*, and the initial activities of the Task Force have been set up in April-October 2021. M. Allen also participate in the CNRS level coordination of involvement in EOSC.

---

[11] https://wiki.escape2020.de/index.php/WP4_-_CEVO

[12] https://www.fairsfair.eu

[13] https://www.eoscsecretariat.eu/working-groups/fair-working-group

[14] https://op.europa.eu/en/publication-detail/-/publication/4630fa57-1348-11eb-9a54-01aa75ed71a1/language-en/format-PDF/source-166584930

[15] https://www.eosc.eu/sria

European Project - EOSC-Future

CDS is part of a the EOSC-Future project. Following a long negotiation period this project began on 01 April 2021. The proposal was led by the Technopolis Consulting Group Belgium and involves 34 partners comprising e-Infrastructure and research community partners over many scientific domains. From the proposal: "*EOSC Future integrates, consolidates, and connects e-infrastructures, research communities, and initiatives in EOSC to develop the EOSC-Core and EOSC-Exchange. EOSC Future will expand the EOSC ecosystem, integrate existing but disparate initiatives, and hand over key project outputs to the EOSC Association. EOSC Future will unlock the potential of European research through a vision of Open Science for Society."* This project involves 40 M€ of funding, with CDS being a small partner (20 PM) for involvement in test science cases and training activities to enable community use of open science resources. Activities have begun, in particular in coordination with ESCAPE, and increased CDS effort will be applied to this project in 2022.

EuroPlanet Project / EuroPlanet Research Infrastructure 2020-2024

CDS is participating in the **Europlanet 2024 Research Infrastructure (EPN-2024-RI)** which started in February 2020. This is a large project (~10 M€) with 56 partners in both industrial and academic sectors. CDS has a minor participation (~3 PM, 20k€) to work on the VESPA activities for interoperability of planetary science data. The CDS effort in this project will be spread over the duration of the project as needed, and has so far concentrated on the planetary catalogues in VizieR (as mentioned in the VizieR report in section 4.3)

5.2 The Research Data Alliance

The CDS continued to be active in the Research Data Alliance (RDA). As anounced last year, CNRS continued to host RDA France after the end of the RDA Europe 4.0 project, which funded it as one of its National Nodes from March 2018 to September 2020. The support to the RDA and its French chapter RDA France expressed in the First National Open Science Plan (PNSO - Plan National pour la Science Ouverte, published in July 2018) was renewed in the Second National Plan published in July 2021. RDA France and the RDA have been supported by the National Fund for Open Science since 2019. Françoise Genova continues to co-lead RDA France with Laurence El Khouri (Direction des Donnéees Ouvertes de la Recherche, CNRS). The RDA started the Regional Advisory Board (RAB), which currently gathers all the regions and countries which support the RDA financially, in June 2021. Françoise Genova is one of the two co-chairs of the Board, which are invited members of the RDA Council and lead the Regional Board, which gathers all the "regions" interested in the RDA with different levels of interest.

The RDA is supported by the PNSO because it is the international forum for discussion and elaboration of good practices for the data aspects of Open Science, and also because it is a powerful tool for the acculturation of the national community to these aspects of Open Science. More than 1000 people with very different profiles are in touch with RDA France. It is also important that the astronomical community stays informed of the RDA work and participates when relevant, to share our knowledge and to make sure that the RDA recommendations are acceptable for us. The IVOA is for instance taken as one of the examples of "Global Open Research Commons" (GORC) tackled by RDA GORC Interest Group[16] and a Working Group[17], which started with in mind generic frameworks such as the EOSC but should also include thematic ones.

[16] https://www.rd-alliance.org/groups/global-open-research-commons-ig

[17] https://www.rd-alliance.org/groups/gorc-international-model-wg

5.3 XMM2ATHENA

CDS is a small but important part of the ObAS participation in a successful proposal in the Horizon 2020 call on SPACE-30-SCI-2020: Scientific data exploitation. The ObAS participation was initially organised by Christian Motch from the GALHECOS, and since his retirement Ada Nebot has taken the ObAS coordinator role. The project XMM2ATHENA is designed to enhance and improve the data exploitation of the current European Space Agency mission XMM-Newton and notably the EPIC (European Photon Imaging Cameras) as well as the Optical Monitor (OM) and put into place the necessary software to exploit the future European Space Agency mission Athena. The consortium is made by eight institutes based in France (IRAP, CEA Saclay, ObAS), Spain (Universidad de Catabria), UK (Leicester, MSSL), Germany (AIP, MPE) and Greece (NOA). The CNRS/ObAS leads WP2 (Multi-wavelength/messenger counterparts) which makes full use of the CDS and its specialised knowledge in this area, as well as as members of the XMM-SSC and will advance algorithms designed in the framework of the former ARCHES project that they led. A. Nebot contributes to the project as part of her independent scientific research. The CDS involvement (~3 PM over 30 months) will be related to catalogue cross-matching (F.-X Pineau) with expected benefits for the main CDS X-Match efforts. XMM2ATHENA[18] started on 01 April 2021. In August 2021 J. Kuuttila joined the team with a two year postdoctoral position, to study the young stellar and binary contributions to the X-ray emission of our Galaxy.

[18] http://xmm-ssc.irap.omp.eu/xmm2athena/

6. Status of the 2020 Recommendations from the CDS Scientific Council

"We recommend CDS continue their strategic thinking on the level of involvement and commitment to large proprietary survey projects, on a case by case basis, with a particular view to the creating and/or supplying the added value products in which CDS excels (such as, but not limited to, cross-matching)." This recommendation is being implemented and **rolls over to the next review year.**

Response:
In the past year we have engaged with the Euclid mission Consortium. A presentation was made to the Euclid Consortium Board in June 2021 to express the CDS interest in hosting the reference data sets from Euclid once they become public, and the need to have early access (with a model similar to the current one with Gaia). The EC Board encouraged CDS to propose a number of individuals as members of the Consortium and to explore a MOU between EC, ESA and CDS. Initial applications for the consortium have been made, and we seek dialogue with ESA.

We have also continued to collaborate with the French partners involved in the FINK LSST event broker, and we have recently participated in discussions for the coordination of the various brokers and their use of the CDS services as part of their future operations.

"*We recommend that CDS continues its engagement in the development of the Virtual Observatory and its promotion at international and national level."* This recommendation is being implemented (even in the face of the obstacles from the global health crisis) and **rolls over to the next review year.**

Response:
CDS has been very active in the development and leadership of the VO in the past period (as detailed in Section 5) with newly approved roles. Major contributions have been made such as leading the MOC 2.0 standard, and highlighting the 'space-time' indexing of astronomy data in the community to raise awareness of these new capabilities. The integration of CDS in VO is also reinforced by the recent upgrade of our publishing registry, as described in the highlights.

"We recommend that CDS continues to aid the professional development of documentalists and, particularly for the fixed-term contract documentalists, mentors them into new roles." We thank the CDS director for this response, and we recognise that group working conditions have been challenging during the pandemic. Mentoring and training of staff on fixed-term appointments remains important at all times, so ***we will roll over the recommendation.***

Response:
As indicated in 2020, all staff, including contractors, have access to and are encouraged to make use of the training courses made available via UNISTRA and CNRS. This has continued in 2020-21 as many courses have been made available in a virtual mode. CDS staff and contractors have participated in professional development courses in the areas of python programming, well-being at work, English courses, statistics courses, and to the great benefit of ObAS some courses in emergency health procedures in the work-place.

"We recommend that CDS continue to be mindful of the changing user base and their preferred communication channels, including the growing use by members of the science-inclined public." We are pleased to see the new dashboard monitoring of CDS services. We are satisfied that the communities' needs are being addressed through their preferred communication channels, but as this is an ongoing and changing requirement, ***we roll this recommendation over to the coming year.***

Response:
CDS continues to make good use of the CDS Dashboard as a global solution. In 2020-21 many events were done in virtual mode, and have reached some new audiences because virtual events are generally more accessible (with low or zero registration fees, and no travel costs). The EAS meeting, and also AAS meeting had many participants, as did ADASS and IVOA and we note that this has enabled more people to attend.

"We recommend that CDS consider interactions with the Zooniverse and other citizen science platforms to consider whether on a case by case basis there could be sources of authoritative data sets prior to publication in peer reviewed journals, or even without such publication." We recognise that there has not been capacity to implement this recommendation during the difficult working conditions of the past year, **so this item rolls over to the next review year.**

Response:
Some interaction with Zooniverse was pursued, with potential use of Aladin Lite for their platform, but this did not proceed due to technical mis-match of technologies. Emphasis has however been placed on interaction with planetaria, and the developers of planetarium software (e.g. Digistar) and especially with the Strasbourg planetarium which is in construction. Information will be presented during the Council meeting, following a workshop with the planetarium in 24-25 November 2021.

"We recommend that the appointment of a replacement for Prof. Whitelock is mindful of the Council's demographic balance in all senses, including but not limited to gender, geography and expertise." We understand that this is still in progress, so **this recommendation rolls over to the following year.** We look forward to developments.

Response:
This is a topic for the CDS authorities CNRS-INSU and UNISTRA. A suggestion from the GW community has been made.

"The CDS YouTube video tutorials are excellent and we suggest that these are also hosted from the CDS website for the benefit of those who find it difficult to access YouTube." We note that **this item is rolling over to the next review year** and will be implemented in due course.

Response:
As outlined in section 3.1 we have started to host video material on the UNISTRA service.

New recommendations from 2020

1. *The urgent staffing priority remains the need for more documentalists, who have more workload than capacity, and for succession plans for retiring permanent staff.*

Response:
We thank the Scientific Council for this recommendation and recognition of the urgent staffing priority for documentalists. The situation has become more urgent in 2021 with the unexpected departure of a permanent documentalist. A presentation of the FTE and overall needs for recruitment will be made to the Council during the meeting.

2. We recommend CDS continues its strategic planning for the next HCERES exercise, considering in particular the boundary of what is made FAIR at CDS in the context of forthcoming missions and observatories.

Response:
The status of an assessment of future large data sets will be presented during the meeting. Templates for the HCERES exercise have recently been obtained, and our responses will be prepared in late 2021 and early 2022.

3. The Council recommends that effort be put into updating the Aladin manual, especially the English version.

Response:
We are very pleased to indicate that this is done as reported in Section 4.5.

4. We recommend CDS develops plans for collaborative ventures for deploying machine learning technologies.

Response:
The CDS has collaboration with partners (HITS, ESO) with a specific task within the ESCAPE project for Deep Learning applied to archival data. A recent paper on this topic has recently been published[19]. We welcome the offer made by D. Veynante for contacts within CNRS on this topic and we will follow-up on this when in-person interactions are possible, and we seeks to gain knowledge and contacts for ML/DL/AI.

5. We recommend continuing the training and organisation of schools to teach how to use the VO to access and manipulate data, and push forward CDS tools.

Response:
A VO School in the context of the ESCAPE project was held in February 2021 and another is planned for February 2022. New tutorials are in preparation.

6. We recommend CDS continues to play a leading role at IVOA and at French level to promote VO, and help the French CERs (Centres d'Expertise Régionaux) develop their expertise.

Response:
The CDS activities in IVOA are detailed in section 5.1. The CDS continues to have strong collaborations with other French institutes on VO topics via the ASOV and other direct interactions. For example in the past period there has been much activity with Observatoire de Paris (EuroPlanet, ESCAPE), and also with colleagues who develop the CASSIS tool. CDS is also active in various French collaborative networks for computing and big data. We also note that A. Nebot serves on the ASOV Council. There was no meeting of the CER in 2021, but there were many INSU led meetings that gathered data related actors.

7. The Council notes and encourages the relationships forged with the Strasbourg university data center to host part of its computing and storage resources. We recommend CDS continues to follow developments in the French and European research landscapes in terms of digital infrastructures (e.g. CNRS plans to develop a combined HPC / HPDA offer for national research infrastructures, EOSC, EuroHPC) and to take advantage, where appropriate, of the opportunities offered.

Response:
The first moves of CDS servers into the UNISTRA data centre have been done successfully in 2021. Discussions have been held on the INSU Exascale initiative, and we are involved in EOSC. More help to to keep well informed and connected on these topics is always welcome!

[19] https://ui.adsabs.harvard.edu/abs/2021MNRAS.501.6026S/abstract