**CDS Scientific Council Meeting: Dec 6-7, 2022**
**Virtual Meeting**

**Agenda**

*14h Strasbourg, 13h London, 8h Wash. DC / Boston*

**Tuesday 06 December**
14h - Welcome and introductions
14h10 - CDS Activities 2021-2022. *(M. Allen)*
*14h30 - CDS Information System. (P. Fernique)*
*14h45 - SIMBAD and Biblio. (C. Loup, A. Oberto, S. Lesteven)*
*15h00 - VizieR. (P. Ocvirk, G. Landais)*
*15h15 - Aladin. (T. Boch, C. Bot)*
*15h30 - R&D. (A. Schaaff)*
*15h45 - Demonstration*
16h05 - Break
16h25 - Discussion *(All)*
17h30 - Close

**Wednesday 07 December**
14h - CDS plans and challenges. *(M. Allen)*
14h30 - **Closed session**
17h - Close

**Attendees:**
(All CDS staff in the open sessions)

**Stephen Serjeant** [Chair] (Open University, UK)
**Marica Branchesi** (Gran Sasso Science Institute (GSSI), Italy)
**Eric Peng** (NOIRLab)
**Guido De Marchi** (ESA) - *apologies for Dec 7*
**Michael Sterzik** (ESO) - *apologies for Dec 6*
**Roopesh Ojha** (NASA)
**Denis Veynante** (CNRS, HPC and data)
**Thierry Bret-Dibat** (delegated by Philippe Laudet) (CNES)
**Thierry Forveille** (IPAG)
**Franck Le Petit** (Obervatoire de Paris)
**Chiara Ferrari** (OCA) - *apologies*
**Marian Douspis** (IAS) - *apologies*

**Bruno Bezard** (INSU representative delegated by M. Giard)
**Pierre-Alain Duc** (Dir. Obs. Strasbourg)
TBD (Université de Strasbourg)

# CDS Activity Report 2021-2022

December 6, 2022

Mark Allen - Director CDS

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

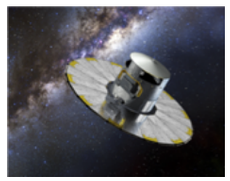
# CDS mission

- **Collect useful data on objects in electronic form**
- **Improve them by critical evaluation and combination**
- **Distribute the results to the international community**
- **Conduct research using the data**

**Science Driven**:

- *Necessary evolutions to meet the scientific reference service needs of the astronomy community*
- *Innovations to meet challenges and ensure sustainability*
  - *Science is changing, technology is changing*

# Contributing to the global astronomy data infrastructure



**Ground and Space Observatories, Instruments and missions**

ESO

esa

cnes

NASA

CDS

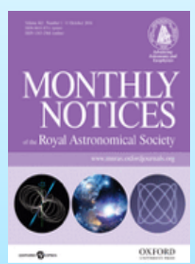
cnrs INSU Observer & comprendre

Université de Strasbourg

Journals

Astronomy & Astrophysics

AAS




+ …

eosc

**Astronomy Data Centres**

astrophysics data system

NED

CADC, MAST, HEASARC, IPAC, + …

**Virtual Observatory**

EURO VO

France

IVOA

ASOV

**H2020 projects:**

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

EUROPLANET 2024 Research Infrastructure

EOSC Future

**Data e-Infrastructures**

RDA RESEARCH DATA ALLIANCE
+ RDA Europe

CORE TRUST SEAL

WORLD DATA SYSTEM

# CDS in 2022

## Service integration and portal

## Objects
Dictionary
Bibliography
Name resolver
TAP

## Catalogues
Photometry service
Associated data service
FTP
TAP

## Sky Atlas
Data integration
Surveys
HiPS/MOC

**+ Aladin Lite** — Embeddable widget

Catalogue
Cross-matching service

## Scientific content curation

**Processing of published literature and reference data**

Acronyms
Identifiers
Citations
Basic Data
Cross-identifications
Catalogues/Tables
Associated data
All-sky surveys, HiPS
Models/Simulations

Vizier and SIMBAD data ingestion pipelines
Tools: DJIN, COSIM, BCS

## Development

**R&D to support development, update and maintenance of CDS science services**

Innovations in astronomical data techniques

Astronomy Informatics

R&D ‘watch’

## R&D programs

New tech testing
Training

## Technical infrastructure operations and maintenance

Servers
Storage
Security
Mirrors

**Science user support** | **Projects** | **Virtual Observatory** | **Astro. community engagement** (Journals / Agencies / Observatories / Data Centres) | **Coordination with Open Science Data Infrastructures**

# 2021-2022

- **Another challenging year that required continued re-planning.**

- **Core work of building CDS content, and operating the services, was maintained at a very high level.**

- **Recruitments of contractors: Documentalist, 2 developer engineers, project engineer, system engineer.**

- **Developments to support internal CDS processes, & developments in response to new scientific needs and also changes at publishers.**

- **HCERES Evaluation** *(but also MESR, CORTECS, CTS, CNRS & CNAP researcher evals).*

# 2021-2022

- **CDS services continue to be heavily used:**
  - ***2.5 million queries/day*** *- see next presentation*
- Contributions: IVOA, EOSC, RDA, IAU, RDG
- Projects — **ESCAPE, EuroPlanet, EOSC-Future, XMM2ATHENA.**
- Community interactions - **ADASS, EAS, SF2A**
- Engagements with large projects via : ESCAPE, CDS participation in Gaia, and new SKA SRC activities.
- Responded to strong demand for CDS involvement in Working/Thematic groups at national level.

# Highlights 1.

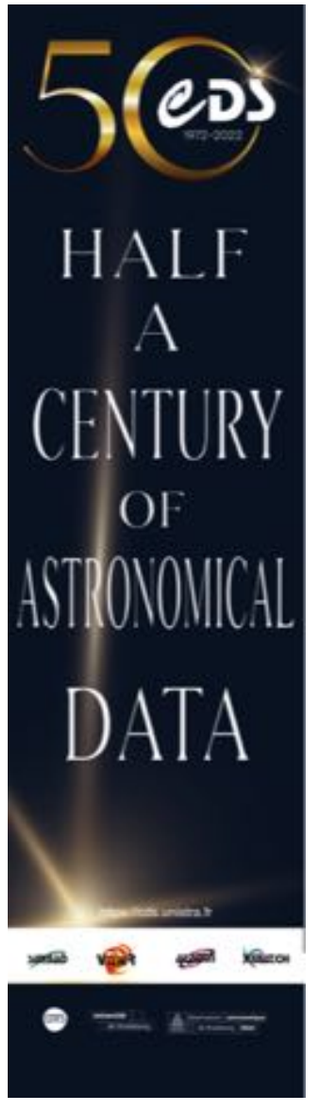
## CDS 50th anniversary

- **April 2022 - Open science and the 50th anniversary of CDS**
  - French launch event with CNRS, CNRS-INSU, UNISTRA, MESR

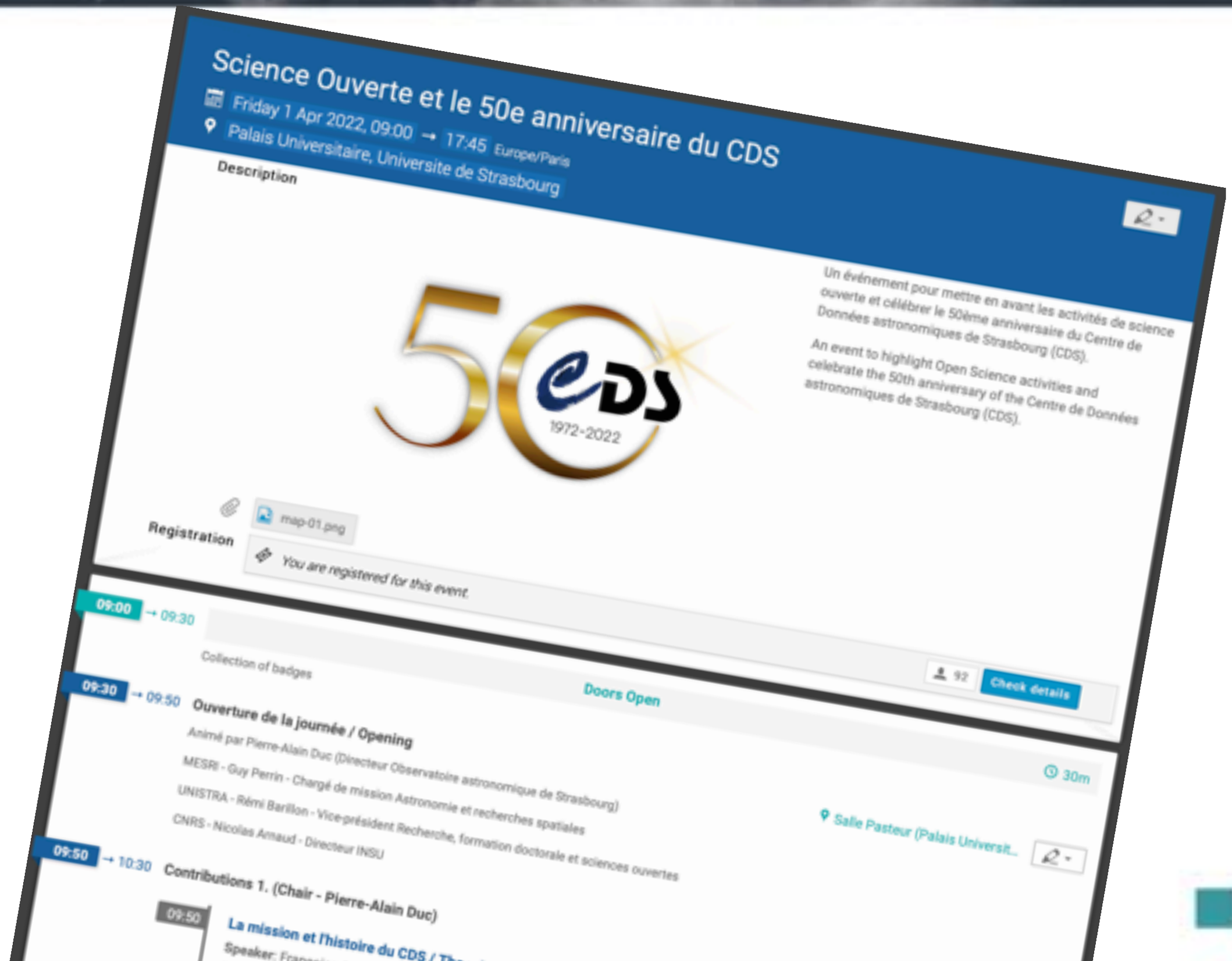
- **Gaia DR3 - 13 June 2022.**
  - 50th anniversary news integrated into into French media events for DR3

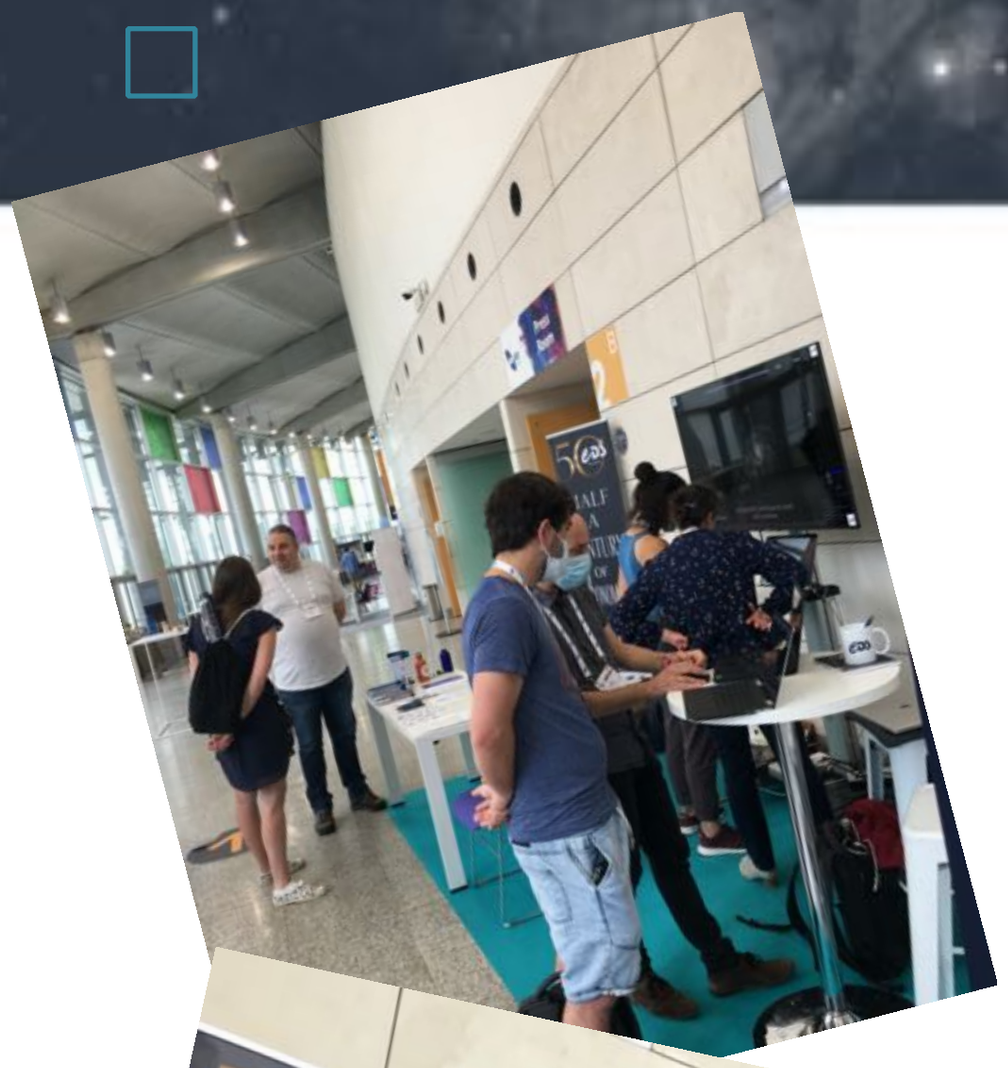
- **SF2A 2022** - Invited presentation - 50 years of CDS, today and future challenges
  - 8 June 2022

- **European Astronomical Society annual meeting**
  - Valencia, Spain, 27 June - 01 July 2022
  - CDS-50 Booth in exhibit hall
  - CDS ‘Lunch Session’




# ‘Open Science and CDS 50th’ French event






eas 2022 Valencia

# 50 years of CDS

## *CDS today and future challenges*

Mark Allen (Director CDS)
and the CDS team

1972-2022

CDS Lunch Session (LS4),
EAS Conference,
1 July 2022

# Highlights 2.

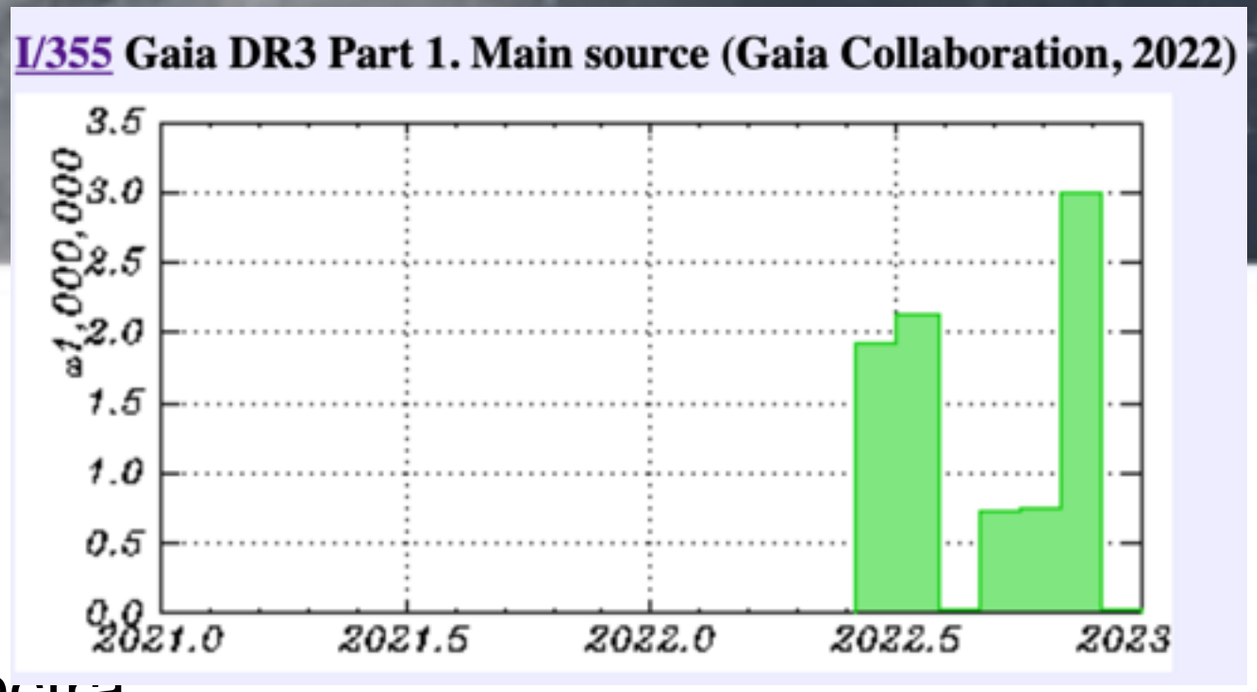
- **CNRS Collective Cristal medal**
  - "*Rewards teams in support functions who have carried out projects with remarkable technical mastery, collective dimension, applications, innovation and reach*".

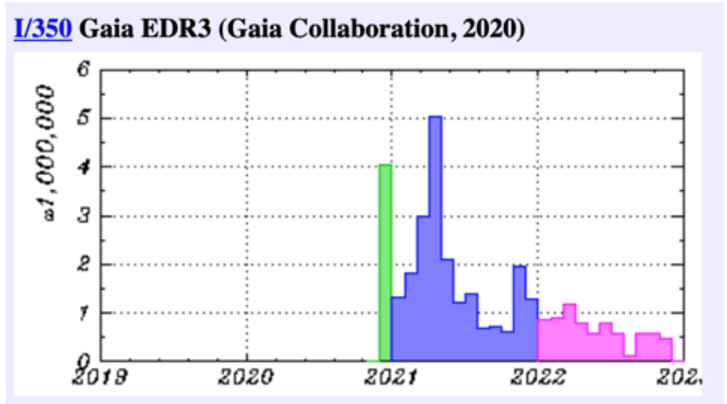


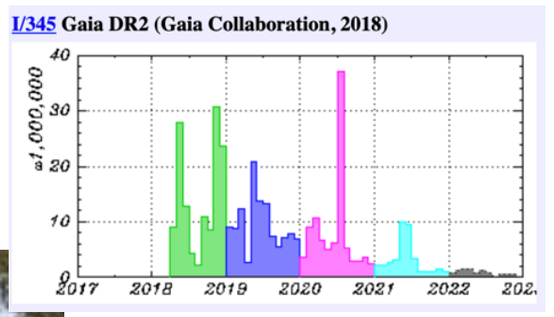
- **Award ceremony 14 Nov 2022**
  - Presented by Director of INSU-CNRS.
  - In presence of ObAS/CDS colleagues, alumni, friends and families.
  - A very welcome celebration for the whole team!

# Highlights 3.

- **Gaia Data Release 3**
- **13 June 2022**
- 1.8 billion records
  - parallaxes, proper motions, spectra
  - 70 individual tables in 6 catalogues
- VizieR, X-Match, Aladin/Aladin Lite, +VO
- 4 million queries in 1st month
- *(Also cross-id of SIMBAD with EDR3)*
- 30m *'galactic panorama'* made with *hip2fits*




partnerships:
ESA/DPAC, CNES



1,8 milliard d'étoiles pour les 50 ans du CDS

Merci à tous les documentalistes, ingénieurs et chercheurs pour leur travail dans la préparation à la mise en ligne des données Gaia DR3 !

http://cds.unistra.fr/gaia

**1.8 billion stars for the 50th anniversary of CDS**

On Monday 13 June 2022, the Strasbourg Astronomical Data Centre (CDS) publishes the Data Release 3 of the ESA Gaia mission in the CDS services. This major data release coincides with the 50th anniversary year of the operations of the CDS (1972-2022), making it an ideal time to celebrate the integration of the 1.8 billion stars from Gaia DR3!

The CDS is a partner data centre for the publication of the Gaia DR3. The data are 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 22 000 catalogues and more than 1000 image surveys. Aladin and Aladin Lite visualisers provide an interactive all-sky view of the entire DR3 using the Hierarchical Progressive Survey (HiPS - https://aladin.cds.unistra.fr/hips/) system.

Gaia DR3 contains a main catalogue of 1 811 709 771 astrophysical sources with their positions, parallaxes, proper motions, radial velocities and many other source properties. In the CDS VizieR catalogue service the Gaia DR3 data are organized into 70 different tables, including tables with visualization plots for the spectra and time series.

The CDS is a data centre for reference astronomy data. It was created in 1972 and is integrated within the Strasbourg Astronomical Observatory (ObAS - https://astro.unistra.fr/en/). CDS is labeled as a Research Infrastructure by the French Ministry of Higher Education and Research. The CDS is also one of the major actors of the development of the international Virtual Observa
and promotion of Open Science in astronomy. In 2022 CDS is celebrating its 50th

# Staff

## Permanent staff:

11 - Researchers
(8 CNAP, 2 CNRS, 1 Emeritus.)
9 - Software engineers
10 - Documentalists

## Contract staff:

3 - Engineers
3 - Documentalists
2 - Postdoc researchers

### Direction and Administration

**M. Allen (Director)**

**P. Fernique (Technical Lead), C. Halter, (**vacant - Admin assistant**)**

### Science

**M. Allen**
**C. Bot**
**L. Cambrésy**
**S. Derriere**
**F. Genova** *[Emeritus]*
**C. Loup**
**G. Monari**
**A. Nebot**
**P. Ocvirk**
**A. Siebert**
**B. Vollmer**

### Software Development and Operations

**T. Boch**
**F. Bonnarel**
**P. Fernique**
**G. Landais**
**S. Lesteven**
**G. Mantelet**
**A. Oberto**
**F-X. Pineau**
**A. Schaaff**
*M. Baumann*
*A. Flint*

### Documentalist

**A. Eisele**
**M. Brouty**
**C. Brunet**
**M. Buga**
**E. Collas**
**M. Neuville**
**E. Perret**
**E. Son**
**P. Vannier**
**P. Vonflie**
*A. Fiallos*
*C. Fix*
*K. Van der Woerd*

### Support (shared with Observatoire de Strasbourg)

**S. Langenbacher, V. Trimbour, C. Saillard, T. Keller,** *M. Misslin*

**Post-doctoral Researchers**
*S. Amodeo*
*C. Lui*

**Project support**
*H. Heinl*
*M. Marchand*

**Ph.D Students**
*L. Correia*
*M. Planaque*

**Software Development Interns**
*10 interns, short term contracts*
*(Apprenticeship - T. Dumortier)*

# Staff - recent changes

**Permanent staff:**

- Documentalist - **F. Marquis** - departed October 2021 *(25 yrs @ CDS)*.
  - CNRS ‘mobility position’ *FSEP* position open in December 2022.

- Adminstrative assistant - **L. Arbousse** - departed April 2021 *(17 yrs @ CDS)*.
  - *replaced by temporary contract in Oct 2021, but left Sept. 2022.*

**Contract staff changes:**

- Engineer (Aladin) : **M. Baumann** - re-joined CDS in March 2022.
- Documentalist - **A. Fiallos** - started May 2022.
- Engineer (VizieR) - **A. Flint** - started May 2022.
- EOSC Future (project engineer) – **M. Marchand** - started Sept 2022.
- Postdoc XMM2ATHENA - **J. Kuuttilia** - started August 2021, left 2022.
  - XMM2ATHENA/CDS postdoc (2yr) being advertised.
- CDS Administrative Project Manager - **C. Holtzinger** - left September 2022.
- Apprenticeship - **T. Dumortier** - started Sept 2022.

# National and European Landscape

**Elements that define high level policies:**

- **French National Roadmap for Research Infrastructures (2021)**
- MESR National Plan for Open Science - 2nd plan released in 2021
- European Cloud Initiative & European Open Science Cloud (EOSC)
- (ESFRI Roadmap)

- CNRS-INSU Prospective
- INSU Astronomy & Astrophysics Prospective
- CNES - French Space Agency prospective
- US Decadal review — 2021
- ASTRONET Science Vision and Infrastructure Roadmap (document planned for end 2022)

# French national roadmap for Research Infrastructures

**Roadmap 'launched' March 2022**

***Emphasises:***

- Open Science aspects.
- Research data in infrastructures.
- Role of IVOA and CDS contribution.
- Role of CDS for data connected to pubs.
- CDS partnerships CNES, ESA, NASA/SAO/ADS, A&A.

***CDS entry:***

- Scientific production and service use.
- Open Science.
- International aspects.
- Society aspects - contracts / training / planetaria

***IR Status***

- Important for CDS support.

# CORTECS network


CDS

## Centre de Données astronomiques de Strasbourg

Domain: Earth and space sciences

The core facility

- Head Manager of the core facility
  **Allen ALLEN**
- Scientifical Manager
  **Mark ALLEN**
- Technical Manager
  **Pierre FERNIQUE**

UMR 7550 - ObAS / Strasbourg Astronomical Observatory

CNRS, UNIVERSITY OF STRASBOURG

Contact the core facility

+33 3 68 85 24 10

Website of the core facility

Central campus : Historical

Established in an unic site

**Collaborations of research**

- ✓ with the academic community
- ✓ to the socio-economic world

CENTRE DE D
OMIQUES DE STRA

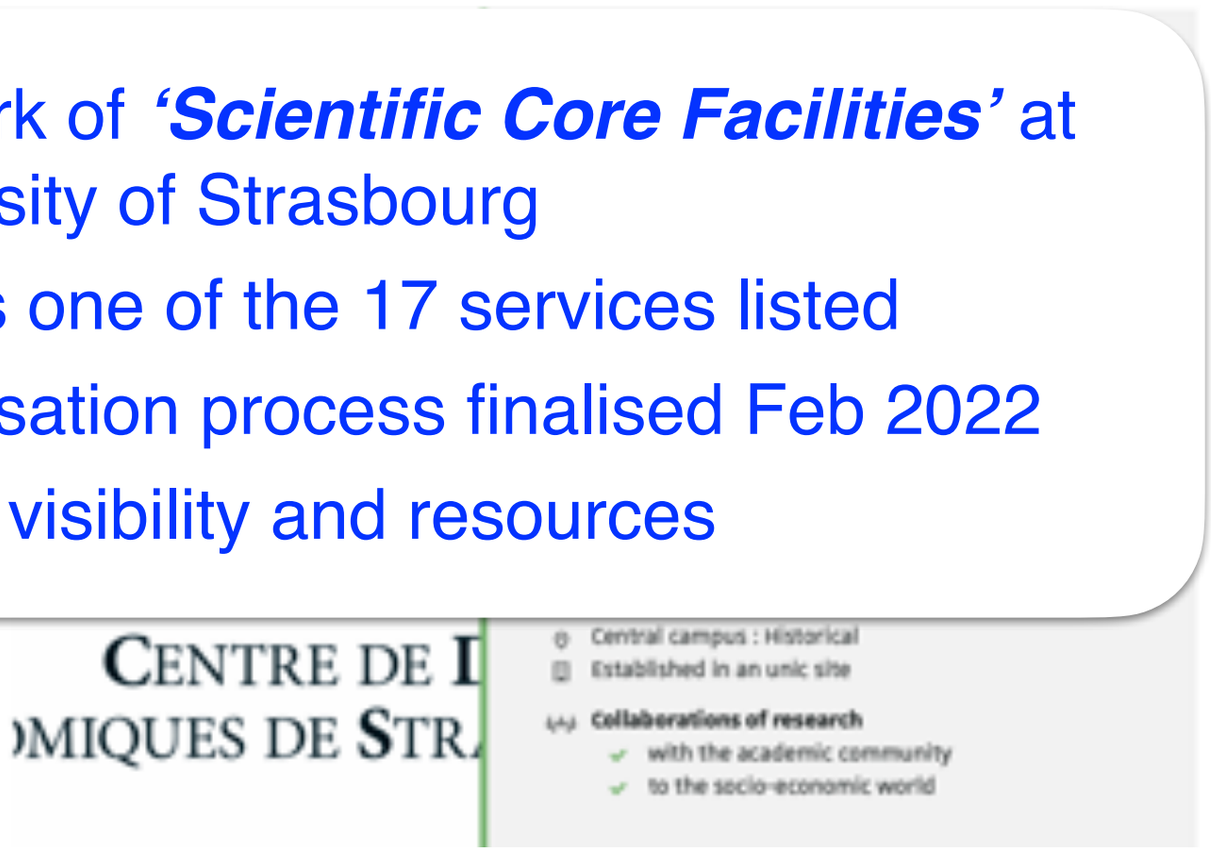
CDS
CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

Certifications and accreditations

CORE TRUST SEAL

# CORTECS network


- Network of ***'Scientific Core Facilities'*** at University of Strasbourg
- CDS is one of the 17 services listed
- Labellisation process finalised Feb 2022
- Brings visibility and resources




Certifications and accreditations

CORE TRUST SEAL

# Recherche Data Gouv - innaugurated July 2022

**New French initiative in the National Plan for Open Science**

— ***CDS is named as one of the Thematic Reference Centres***

- Define the international data description standards in their thematic field;
- Define and disseminate best practices for data collection, documentation, processing and dissemination in their thematic field;
- Support inclusion in and compliance with the international ecosystem.
- Repositories are to be ‘harvested’ by *Recherche Data Gouv;*


https://recherche.data.gouv.fr/en

# HCERES and prospective

- ObAS is being evaluated by HCERES in 2022.
  - Application sent in Summer 2022.
  - Virtual visit of the committee occurred 4-6 October 2022.
  - CDS activities 2016-2021 presented.
- Report expected in next months…

- A ‘Prospective’ document outlining the next 5 years has been submitted to CNRS and University of Strasbourg
  - Based on themes discussed in Science Council meetings 2020/21
  - Feedback expected in early 2023.
  - *More in the ‘Plans and challenges’ presentation tomorrow.*

# Virtual Observatory and Projects

**CDS participation in VO at National, European and International levels:**

- OV-France (&ASOV), Euro-VO

**CDS continues leading role in IVOA**

- Executive board, WG/IG roles
- **Executive Board member for EuroVO** - M. Allen
- **Chair of the Committee for Science Priorities** - A. Nebot
- **Deputy Chair of the Data Access Layer Working Group** - G. Mantelet
- **Chair of the Data Curation and Preservation Interest Group** - G. Landais
- **Deputy Chair of the Radio Astronomy Interest Group** - F. Bonnarel
- **Chair of the Education Interest Group** - H. Heinl
- **Editorial team for the IVOA Newsletter** - S. Amodeo

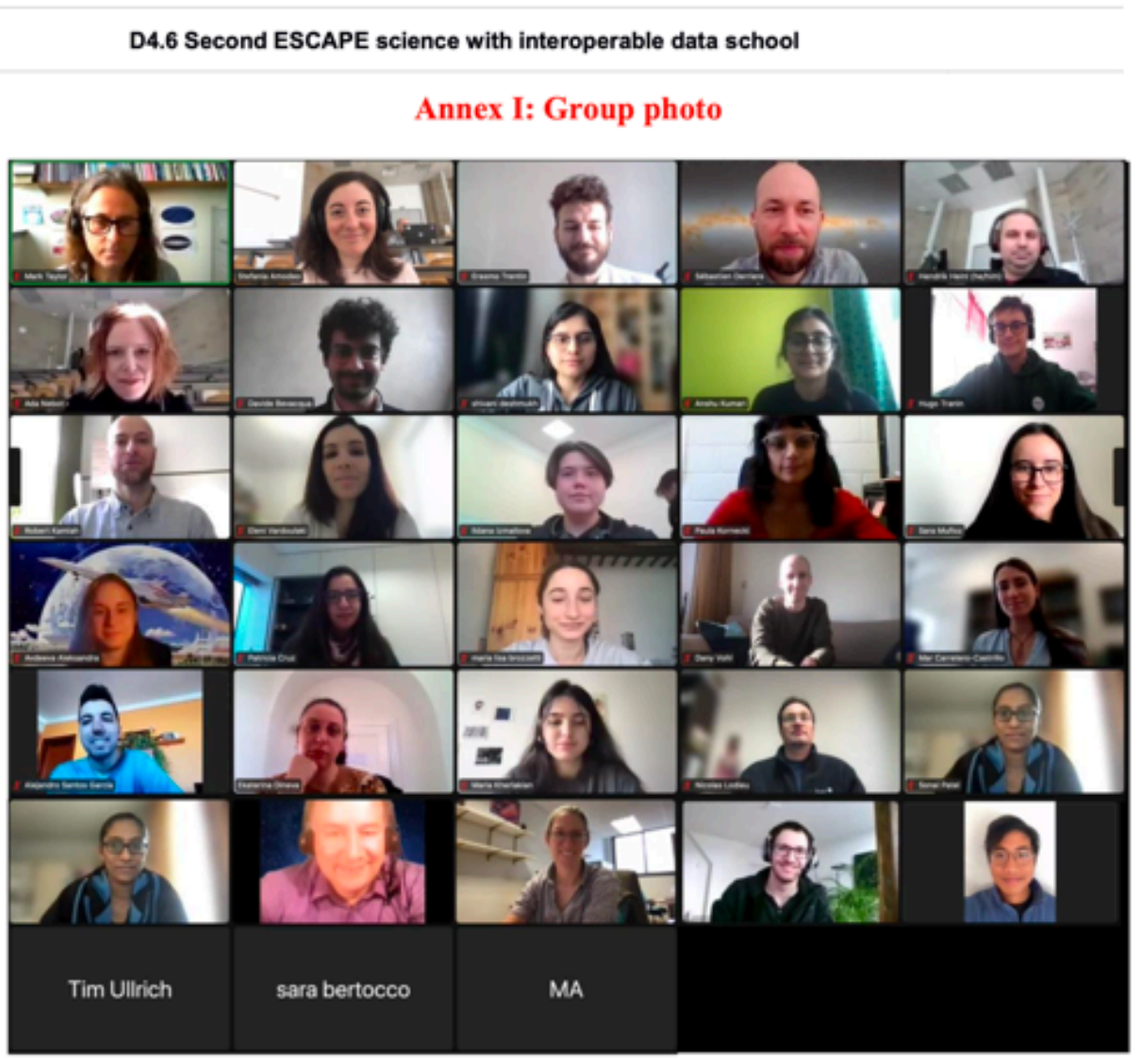
**Important progress for the interoperability of astronomy data and services**

- e.g. Space-Time coverage of data (MOC 2.0 standard) and HiPS system.
- Leadership for integration of Radio Astronomy (+others) into VO framework

# VO training via ESCAPE project

- **2nd Virtual School Feb 2022**
  - Focus on students & postdocs (28)
  - CDS tutorials, + others
  - Support of participant science cases

- **Technology Forum 3, March 2022**
  - **Sharing practical expertise**
  - **‘Hack-a-thon’ mode**



# Projects

- **ESCAPE** - final phase now, conclusion in Jan 2023.

- **Europlanet 2024** Research Infrastructure (EPN-2024-RI)
  - Small but important CDS participation (~20PM).

- **XMM2ATHENA** **(***started 2021)*
  - A. Nebot is the ObAS coordinator. Postdoc to be recruited
  - ~5PM participation of CDS related to X-Matching

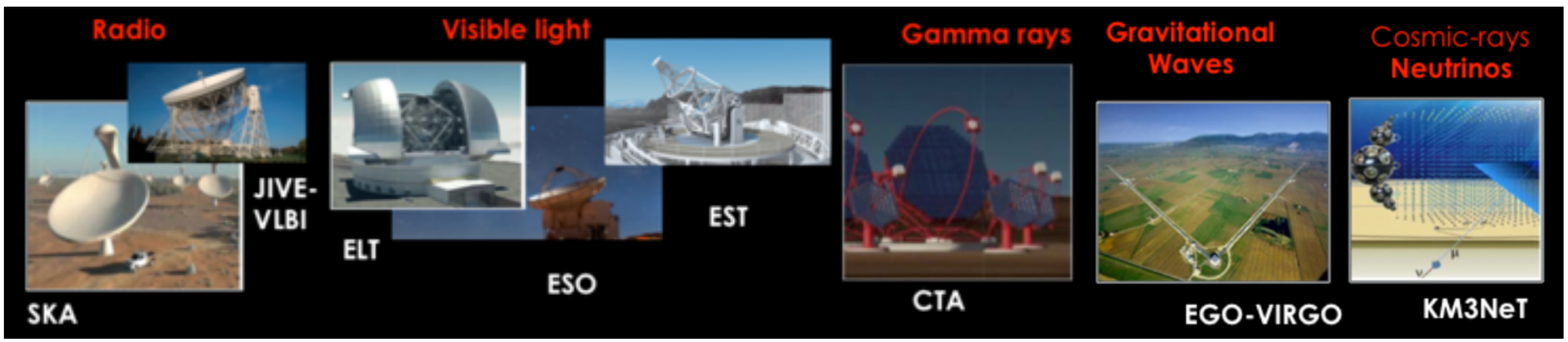
- **EOSC Future** - large (~40 M€) project of science and e-Infrastructures
  - Started April 2021, ends August 2023.
  - CDS is small part for Test Science Cases and training activities. New project engineer (M. Marchand) making fast progress!

# Rec. — Engagements with large projects

**ESCAPE connections to ESFRI and other RIs**

Main engagement with big projects - funded activities related to use of VO (and CDS tools)

Benefits for everyone - e.g. Aladin Lite v3, VO standards, training materials, connections - e.g. SKA




**SKA - Started participation in the SKA Science Regional Centre (SRC) prototyping activities.**

**- Bringing expertise on access to large distributed data and visualisation. Aladin as a SRC visualisation prototype.**

**- F. Bonnarel as main contact - many many meetings to engage in this process**

**Slow progress for Euclid and LSST/Rubin connections in 2022:**

**Euclid mission**

- MOU to be explored for CDS-EC-ESA.
- Aim for publishing reference catalogues in CDS services

**LSST**

- CDS X-Match is part of the FINK broker (for LSST)
- Collaborative discussions ongoing with FINK and other brokers about thier use of CDS services as part of their operations - possible EC project in 2023
- In-kind contributions to be discussed approached (again) - meeting proposed at AAS meeting Seattle.

# Recs. Machine Learning / AI / HPC landscape

- **AI / machine Learning**
  - Received the recommendation (thanks D. Veynante) to start with **Institute for Information Sciences and Technologies (INS2I)**. (https://www.ins2i.cnrs.fr/fr/ins2i)
    - [Olivier Cappé (DAS) and/or Jamal Atif (chargé de mission)].
    - '*They have a complete view of the field and will be able to identify possible partners in relation with CDS needs.*' — *to be followed-up*
  - Experience and partners in ESCAPE on DL/ML for archive products
    - Heidelberg Institute for Theoretical Studies (HITS) [K. Polsterer]
    - Published a MNRAS paper (ESO, CDS), DL prototypes, ESCAPE Deliverable D4.5
    - Experience for DL on archival data - *pros:* new results, *cons:* difficult interpretation
  - Other contacts made in ESCAPE to follow: *[H. Dickinson, Open University]*

- **HPC**
  - Participation in the Action Specifique Numerique. - *A. Schaaff*
  - e.g. Invited CDS presentation at ASNUM 2022 meeting next week (Lyon)
  - Participation in MESR Thematic Group - CoSIN - where issues for Research Infrastructures are discussed, and plan made for future. (*C. Bot, M. Allen, P. Fernique*)

# Staffing needs

- **Documentalist**
- Increasing number of references, and complexity drives need.
- Core data ingestion task.
- Necessary **replacement** of Documentalist who departed in 2021.
- FSEP ‘mobility’ position open now with result known in Jan 2023.

- **Administrative Assistant**
- CDS needs admin staff with expertise on contracts.
- CDS cannot operate without it!
- **Replacement** of Admin Assistant who left in 2020.
- Temporary contractor left Oct 2021.

- **Aladin / visualisation engineer**
  - Unique opportunity to respond to new SKA mission.
  - Build on revolution of HiPS/MOC/Aladin.
  - Critical **replacement** of F. Bonnarel (2024).
  - SKA SRC work would depend on it.

- **VizieR engineer**
- Renewal of Vizier ingestion process underway with contract engineer
- 2 engineers per major CDS service are necessary for operations & development

Note: Expected **retirement** of another engineer in ~2025-2026. (Also **retirement** of another ObAS engineer (~2027) who developed SAADA which is used by CDS)

# Science staff

**CDS Scientists**

- Necessary expertise of active scientists covering wide range of astrophysics research areas.
- Necessary diversity of scientific profiles.
- ***Recruitment being sought via CNRS and CNAP competitions***
- *(CDS supports ~4 candidates/yr)*

**Contribute to all scientific aspects of the CDS mission:**

- Selection and validation of scientific data.
- Providing expert scientific advice and setting the scientific strategy.
- Supervision of the scientific development and evolution  of the CDS services.
- Support the users (hot-line, dissemination, documentation…).

**Priority areas related to CDS service tasks:**

- *Scientific support of SIMBAD*
- *Catalogue, spectral, data cube and multi-dimensional data expertise*
- *Expertise for current and future very large data sets*
- *Data Science and Astro-statistical and expertise (machine/deep learning)*

# Summary

- A challenging, but special year for CDS.
  - Responding to changing conditions.
  - Strong demand for CDS participation in Open Science actions at National and European level.
  - Celebrating CDS 50th!
- High level production — helped by new contractors
- Important progress on core work of operating, maintaining and developing CDS reference services — helped by new contractors
- Started re-engaging with community in live events.
- Final phases of large projects - very demanding but productive.
- Engaged in discussion about the new aspects of the CDS mission - RDG and SKA.

# The CDS Information System

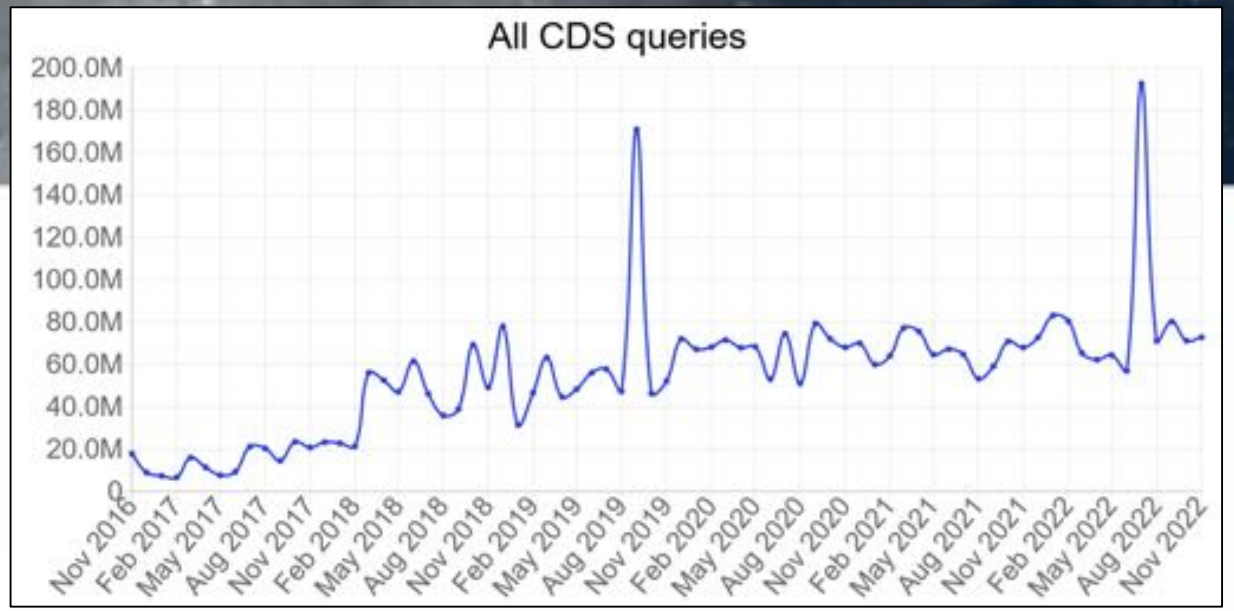
## Overview & stats 2022

CDS Council – 6 & 7 December 2022

Pierre Fernique
on behalf of all the CDS staff

# Global usage 2022

- **2.5 million requests** per day
- **>360K unique IPs** per month




**CDS main metrics *(Oct 2021 - Sep 2022)***

compared to Oct 2020 - Sep 2021

| Main services | **Simbad** *Astro.object DB* | **Vizier** *Catalog service* | **Aladin** *Image service* | **Total** |
|---|---|---|---|---|
| **users/month** | 154.0K +6%<br>Adv.users: 1252 | 45.2K +17% | 360.2K +53% | **> 360.2K** |
| **queries/day** | 344.2K -12% | 431.7K -28% | 1.7M +73% | **2.5M** |
| **load/day** | 7.1GB +32% | no data | 162.3GB +52% | **> 169.3GB** |
| **data volume** | 34.5GB +28% | 79.6TB +51% | 427.3TB +5% | **507.0TB** |
| **data content** | 13.5M obj +7% | 22.8K cats +7% | 1068 HIPS +6% | |
| **reliability** | 99.75% +0.3% | 99.11% -0.5% | 99.93% +0.3% | **> 99.11%** |

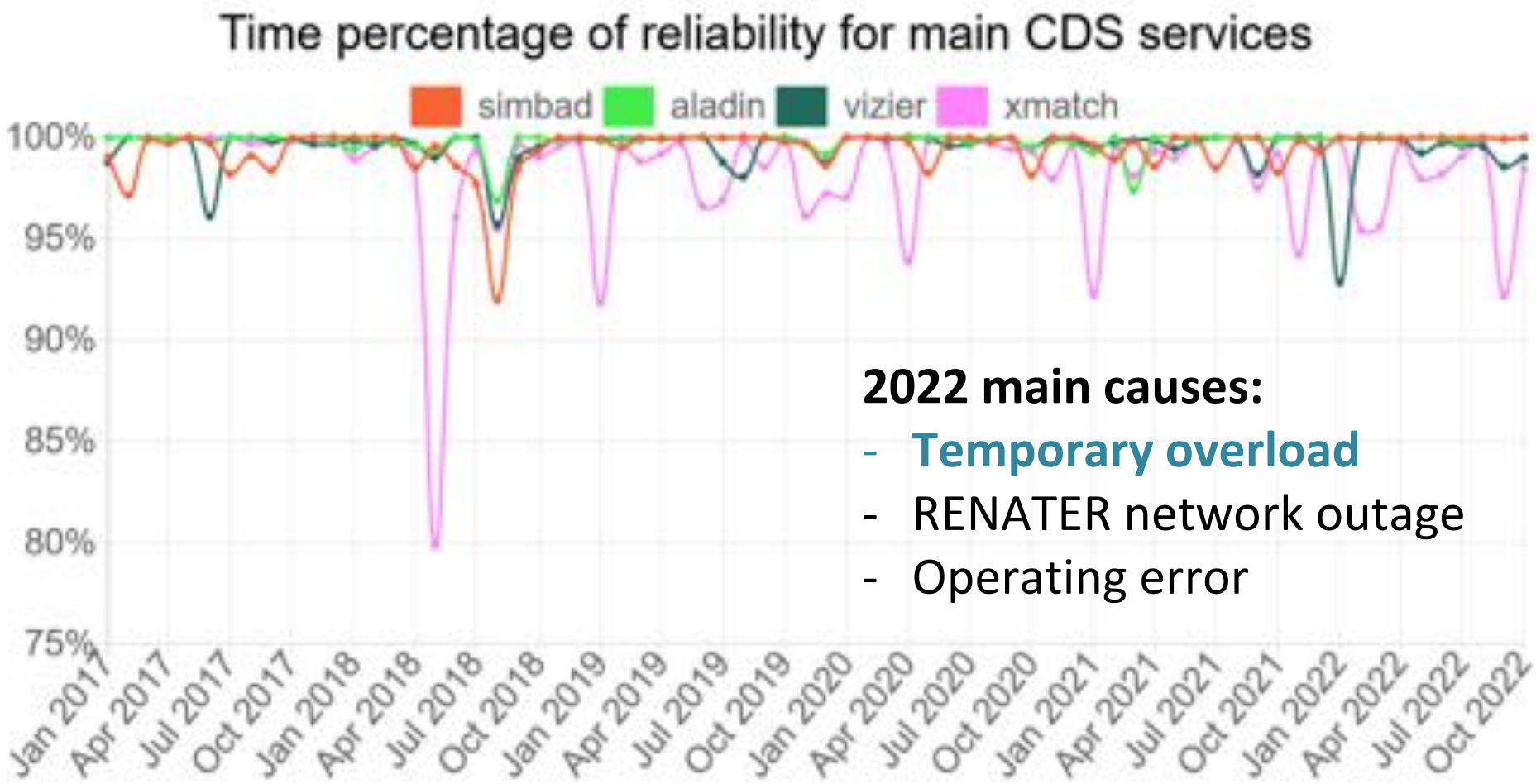
(Bots have been removed of these stats)
**users/month** Based on unique IP (Adv.Users = only based on "advanced scientifical queries") on the main site
**queries/day** Based on effective DB queries (object queries, catalog queries, HiPS tile queries...)
**load/day** Downloaded from CDS servers
**data volume** Volume of service (data + index)
**data content** Number of service "items"
**reliability** Percentage of service availability (mirrors included)

**Queries world repartition** (Simbad 11/2022)

# Constraints and consequences

- **Our goal** : **permanent availability** 24 hours a day, 7 days a week, in a classic academic context (technical staff without on-call duty)
- **Our architecture**: **total duplication : services + data**, on several distinct geographical sites (still in progress)
- **It guarantees**:
  - the **performance** and **continuity** of services
  - the assurance of a **rapid resumption** of service in the event of a major accident at one of the sites (PRA)

# CDS S.I. reliability




**2022 main causes:**

- **Temporary overload**
- RENATER network outage
- Operating error

# CDS service architecture

- 2 **geographically distinct local machine rooms**
- **External mirror sites** (in France and in other countries)
- Server **virtualisation**, via a cluster of VMWare hypervisors, and Docker encapsulations
- Data storage equipment: **2 RAID6 bays synchronised on the 2 local sites** ("CDS Allsky data system")

- **2 local computer rooms**
  => One at the Observatory
  => Replicated elements in Unistra DC
- **7 external sites** (6 VizieR + 1 Simbad mirrors)
- **20 partner HiPS sites** (Aladin tiles)

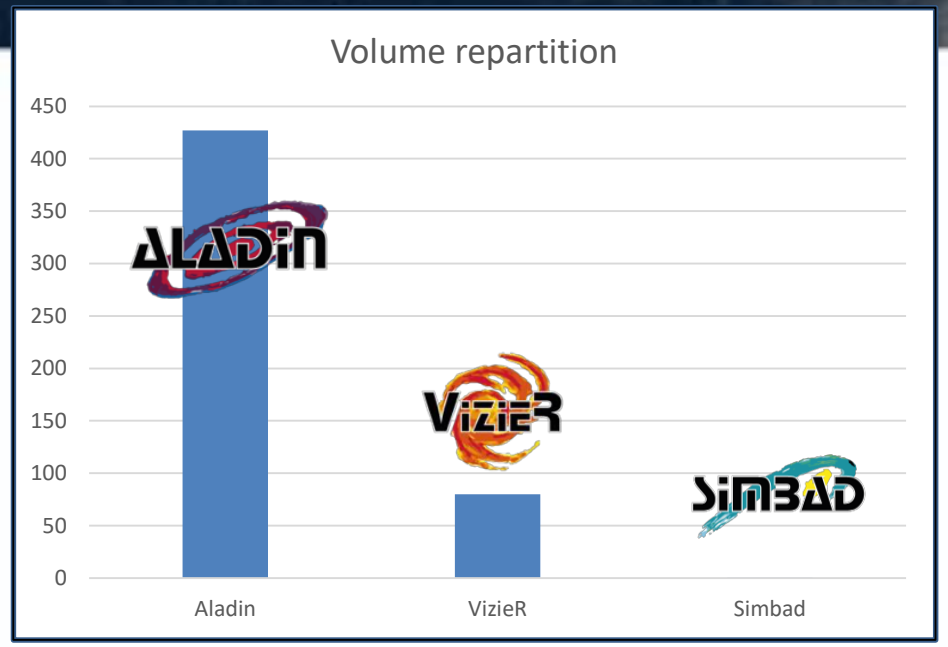
*Challenge in January 2023 with possible 2-hour power cuts.*

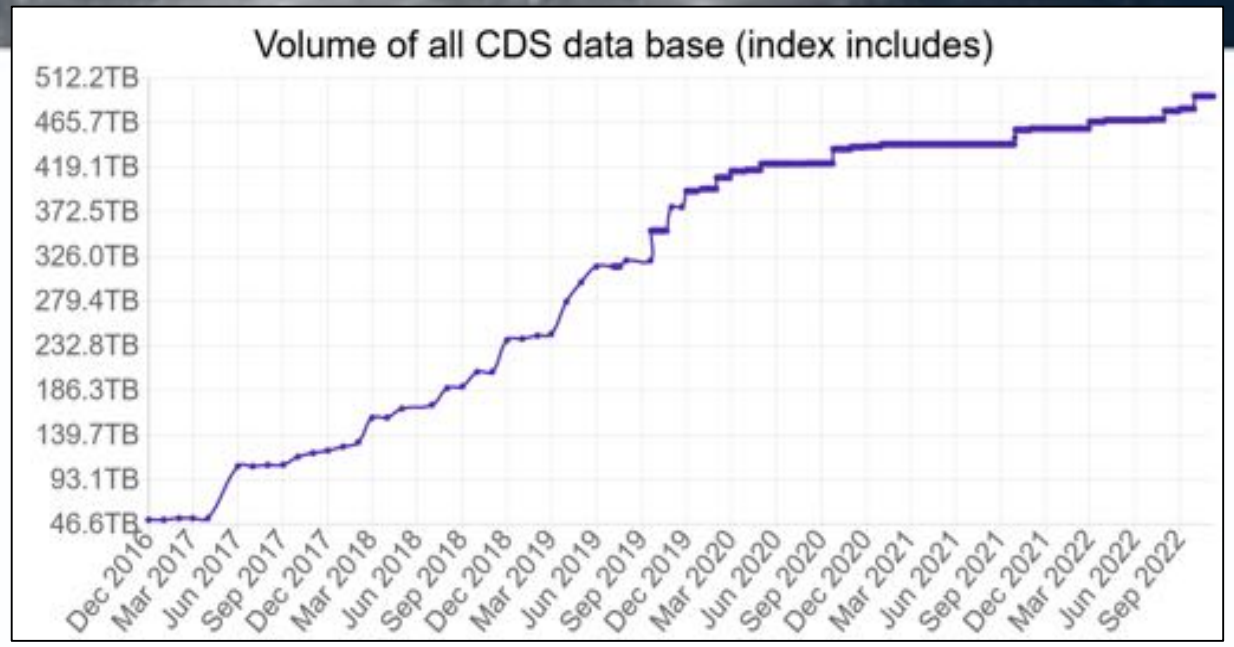
# Local machine rooms evolution

- **Retirement** of CDS installations in the **local server room on the ObAS** site
  - Strong encouragement from the University to retire this room (corresponding to University and national level policies of rationalization/reduction of small server rooms)
  - Loss of the leader of the ObAS logistics team
  - Technical limitations of the server room
- **Our solution: Partnership with the IPHC** (UMR7178) on the CNRS Cronenbourg campus.
  => Technical and functional conditions appear a priori to be adapted to our desired level of operations
- **University Data Centre will become the main local site** for our installations.

*We aim to begin relocation of servers in the first half of 2023*

# Operational storage




- Nov 2022: **520 TB** / 1.6 PB : 90% for Aladin, 9% VizieR
- Evolution relies to the volume of astronomical data published
  (bibliographic data, tables, catalogues and image surveys)

- *Provisional plan : ~5 replicated PB in 2025*

# Operational servers


- CDS services require **low CPU power** (except Xmatch service)
- But **a mistake to undersize the servers** because the availability constraint implies the capacity to absorb "peaks"
- The technical solution: **virtualise the servers** as far as possible on VMWARE hypervisors (presently 3)

- *Provisional plan : Acquisition (2023) of a 4th hypervisor*

# Computer resources for data preparation

- **Computing machines:**
  => **3 big machines** are sufficient for the task (Aladin HiPS, VizieR big catalogs)

- **Storage of original and intermediate data:**
  => In 2022, acquisition of **800TB disk storage space** to facilitate the eventual reprocessing of the original data, (cheap medium, but without a very high guarantee of preservation)

# Software development strategy

- **Open source software** supported by a dynamic community (e.g. Postgres, astropy, etc.)
- **''In-house'' developments** for libraries and tools particularly specific to our activity (e.g. management of spatial indexes, astronomical libraries, bibliography processing, etc.)
- Use of paid software remains exceptional, justified by the absence of an equivalent free solution (e.g. VMWare). These paying solutions are considered as transitory

- At least **two developers per service**.

=> By the end of 2022, this strategy is assured partially with contractors

# Information system

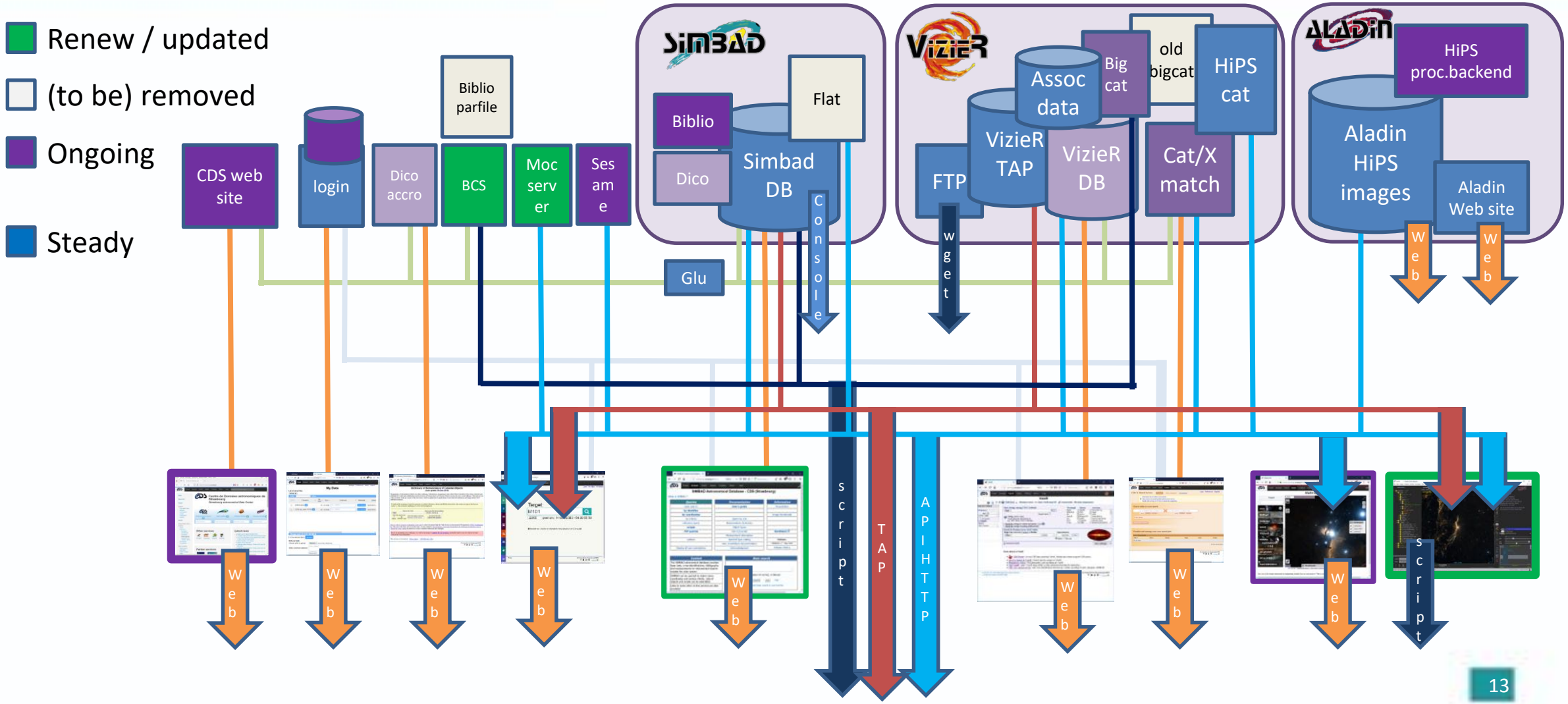
- Nov 2022: **CDS Information service** is based on **24 components**
- The permanent **evolution** of these components **is essential** to map:
  - The current needs (astronomy and data evolution)
  - The CDS human resources

- For each component, we evaluate its sustainability and we plan its renewal, evolution or withdrawal

# Recent developments in the IS

- The deployment of the **new journal article management system**:
  => The "BCS" will replace the old system (parfile based) within a few months.
- The **evolution of the collection registry** (MocServer) in order to take into account both the spatial and temporal coverage.
- The release of the **2 Aladin clients** (Aladin Lite v3 and Aladin Desktop v12)
- The evolution of the **global positional index of VizieR** for a MOC compatible solution (replacing Qbox).
- The redesign of **the CDS website** (prototype phase).

# Evolution detail of IS components



# Questions ?

## The CDS Information System

# SIMBAD : the bibliographic database

A meta-compilation of astronomical objects of interest

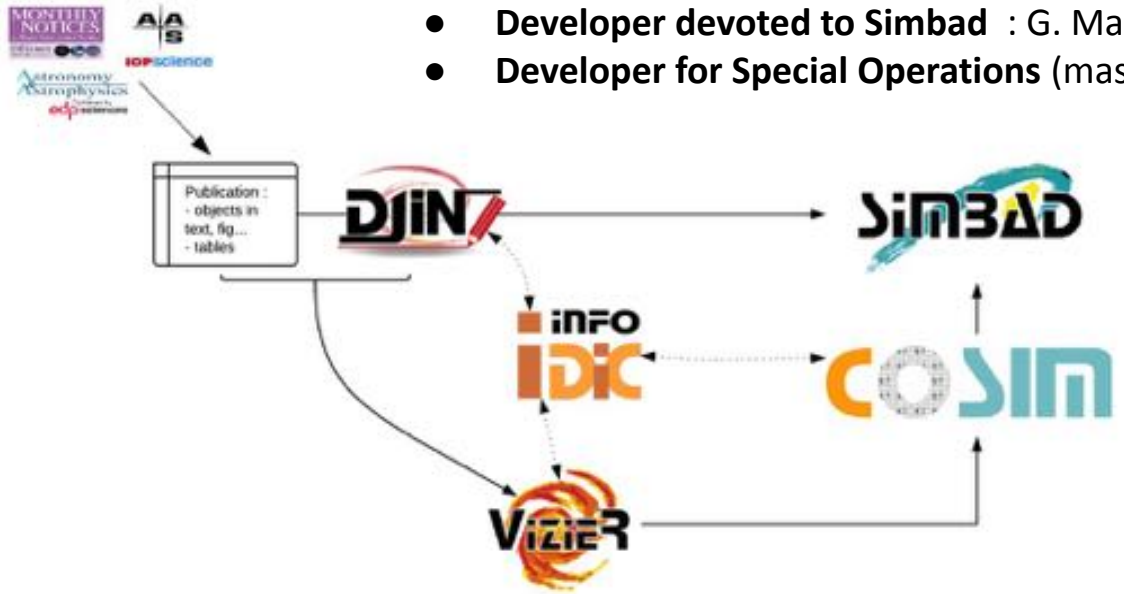
## CDS Council
## December 2022


CDS team

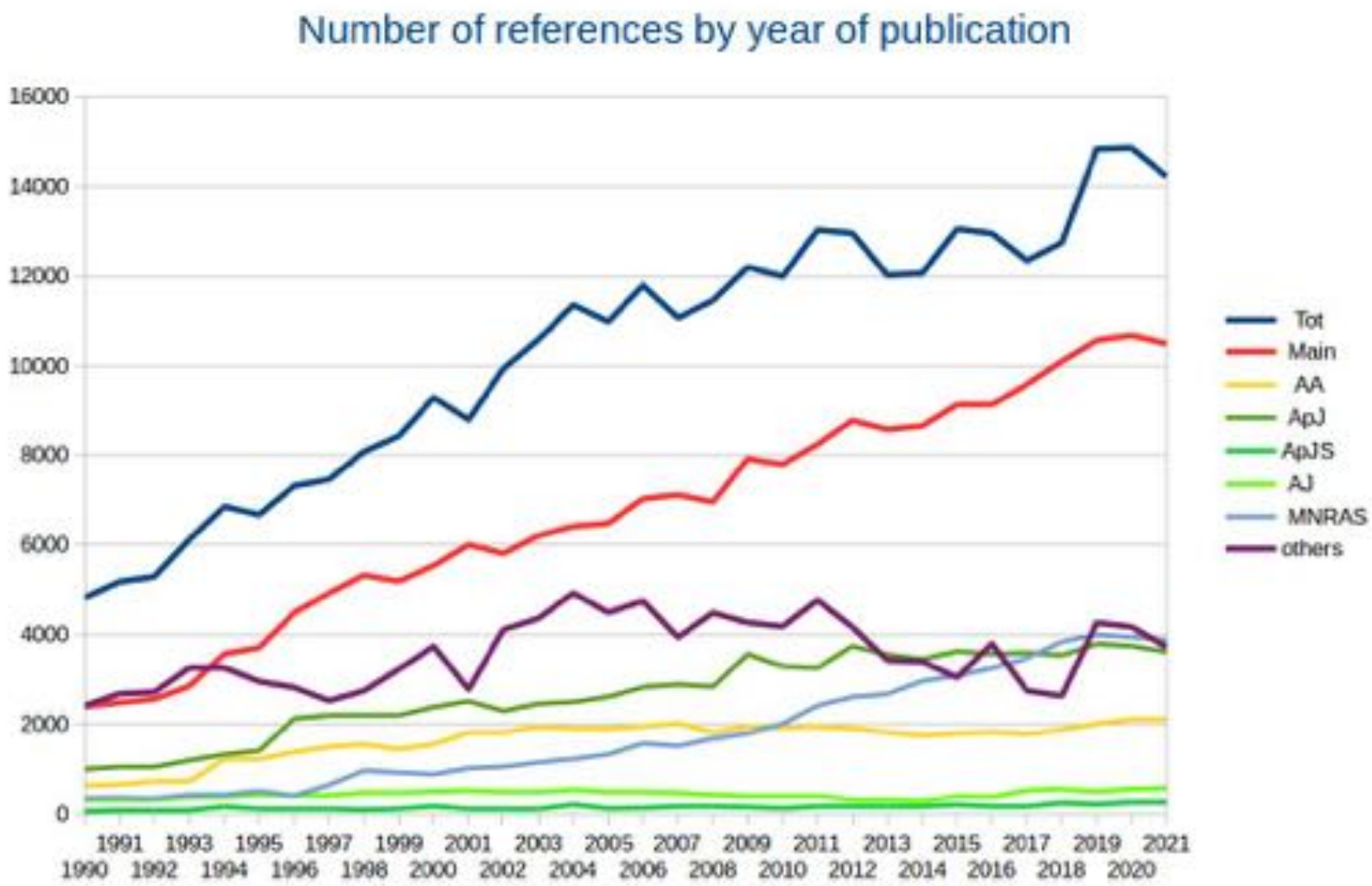
# The Team

- **Bibliography & coordination** : S. Lesteven (engineer)
- **Scientific content** : C. Loup (astronomer)
- **Database & softwares** : A. Oberto (developer)
- **Nomenclature** : B. Vollmer (astronomer)

- **Documentalists** (Data Stewards) :
  - Nomenclature : M. Brouty
  - Ingestion of references via DJIN : A. Eisele, E. Son, M. Neuville, P. Vonflie
  - Ingestion of lists of objects via COSIM : C. Brunet, E. Collas, K. van der Woerd, M. Buga, E. Perret
- **Astronomers** : C. Bot, L. Cambrésy, S. Derrière, F. Genova, G. Monari, A. Nebot, P. Ocvirk, A. Siebert
- **Developer devoted to Simbad** : G. Mantelet
- **Developer for Special Operations** (massive Xid) : T. Boch, F.-X. Pineau



# Workflow : References



# The Content (05/12/2022)

| Overview | N (million) |
|---|---|
| Objects | 13.969 |
| Identifiers | 54.829 |
| References | 0.412 |
| Citations of objects in articles | 31.030 |
| Acronyms | 0.015 |
| **Basic data** | **N (million)** |
| HRV/redshift | 6.500 |
| Proper motions | 7.277 |
| Parallax | 6.572 |
| Spectral type | 0.858 |
| Morphological type | 0.145 |
| Magnitudes (UBVRIGJHKugriz) | 10.660 |

| Collection of measurements | N (million) |
|---|---|
| HRV/redshift | 10.103 |
| Proper motions | 13.954 |
| Parallax | 14.635 |
| Distance | 9.262 |
| Spectral type (spectroscopic only) | 1.222 |
| Teff logg Fe/H (spectroscopic only) | 2.827 |
| Variability | 3.130 |

# Massive Xid with Gaia (E)DR3

- Total number of Gaia DR3 sources : 7,809,560
- Massive Xid beginning of February and in May 2022 : 7.15 million
- Criteria for sources not in Gaia DR2 : like the Xid with Gaia DR2

  No neighbour within 3'' if delta(G) <= 2.5 mag. Dist < 1''

  Criteria for sources that already had Gaia DR2 coordinates :

  Xid done at epoch 2015.5 → stars with PM > 300 mas/yr done separately

  No neighbour within 1'' if delta(G) <= 2.5 mag (as some DR2 have been resolved into 2 components in DR3). Dist < 0.15''.
- 86% of the stars in SIMBAD are identified in Gaia DR3
- Still ingesting tables of objects based on Gaia DR2, a third iteration will be required

# Objects of interest from Gaia DR3

| Variability | Priority | N Period>0 | Status/Comments |
|---|---|---|---|
| Cepheids and RV Tauri | 1 | 14,987 | Done |
| Main sequence oscillators | 1 | 54,476 | Done |
| Long Period Variables | 1 | 392,240 | Done |
| RR Lyrae | 1 | 271,779 | Done |
| Eclipsing Binaries | 1 | 2,190,738 | To be processed (compact companion done) |
| Micro-lensing events | 1 | 363 | Done |
| Rotational Modulation | 2+ | 2,935,262 | |
| **Non-single stars** | **Priority** | **N Period>0** | **Status/Comments** |
| Spectroscopic Binaries | 1 | 220,527 | Being processed (139,200 done) |
| Eclipsing Binaries | 1 | 86,918 | To be processed |
| Astrometric Binaries | 1 | 135,760 | To be processed |
| **Extra-galactic** | **Priority** | **N** | **Status/Comments** |
| Galaxy candidates | 1 | 424,786 | Being processed / Limited to galaxies with a size (RadS>0) |

# Future plans for the content: the distance project

**Purpose : add in SIMBAD objects with a reasonably well known distance**

1) Stars with Plx >= 10, the 100 pc sample

Criterium : e(Plx) < 0.9 to minimize spurious measurements

Includes most very high proper motion stars

Done up to Plx = 30 (11,806 sources)

2) Galaxies from SDSS DR16 with a spectroscopic redshift

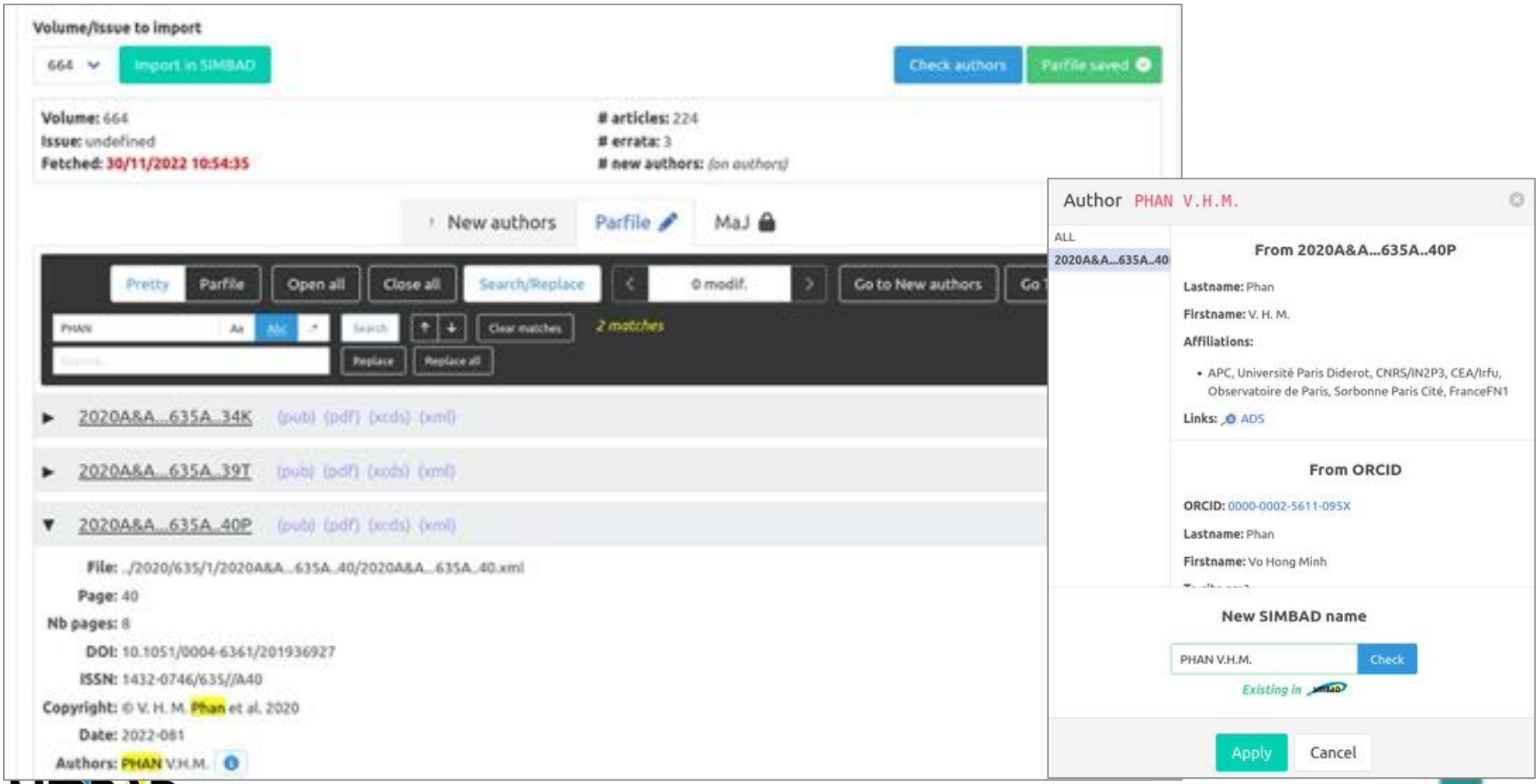
Basic criteria : z > 0.001, no flags

Difficulty : remove false SDSS G that are in fact part of large galaxies

Being studied, working meeting next Thursday

# Bibliographical Center

- Evolutions to improve validation process



# Dictionary

Portal Simbad VizieR Aladin X-Match Other - Help

**Dictionary of Nomenclature of Celestial Objects**

***(Last update: 21-Nov-2022)***

*Result of query:* info cati **RAVE$**

Details on Acronym: **RAVE**

**RAVE** (Radial Velocity Experiment)

**Write:** <<**RAVE** JHHMMSS.s+DDMMSS>>
<<**RAVE** rHHMMSSs+DDMMSS>> (not recognised by Simbad, but found in literature)
<<**RAVE C**HHMMSSs+DDMMSS>>

**N:** 518387

**Object:** * (SIMBAD class: Star)

**Note:** AAO UKST 6dF observations for the Radial Velocity Experiment (RAVE).
DR2: 51829 sources.
DR3: 83072 sources.
DR4: 482194 sources.
DR5: 520701 sources.
DR6: 518387 sources.

**Ref:** =2008AJ....136..421Z

by ZWITTER T., SIEBERT A., MUNARI U., FREEMAN K.C., SIVIERO A., WATSON F.G., FULBRIGHT J.P., WYSE R.F.G., CAMPBELL R., SEABROKE G.M., WILLIAMS M., STEINMETZ M., BIENAYME O., GILMORE G., GREBEL E.K., HELMI A., NAVARRO J.F., ANGUIANO B., BOECHE C., BURTON D., CASS P., DAWE J., FIEGERT K., HARTLEY M., RUSSELL K., VELTZ L., BAILIN J., BINNEY J., BLAND-HAWTHORN J., BROWN A., DEHNEN W., EVANS N.W., RE FIORENTIN P., FIORUCCI M., GERHARD O., GIBSON B., KELZ A., KUIKEN K., MATIJEVIC G., MINCHEV I., PARKER Q.A., PENARRUBIA J., QUILLEN A., READ M.A., REID W., ROESER S., RUCHTI G., SCHOLZ R.-D., SMITH M.C., SORDO R., TOLSTOI E., TOMASELLA L., VIDRIH S., WYLIE-DE BOER E.

Astron. J., 136, 421-451 (2008)

*The radial velocity experiment (RAVE): second data release.*

- Catalogue: <RAVE JHHMMSS.s+DDMMSS> N=51829.

**Ref:** =2011AJ....141..187S

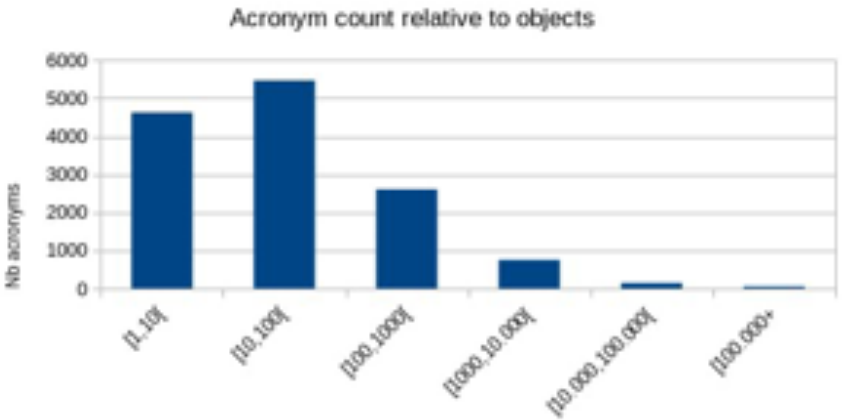
by SIEBERT A. , WILLIAMS M.E.K., SIVIERO A., REID W., BOECHE C., STEINMETZ M., FULBRIGHT J., MUNARI U., ZWITTER T., WATSON F.G., WYSE R.F.G., DE JONG R.S., ENKE H., ANGUIANO B., BURTON D., CASS C.J.P., FIEGERT K., HARTLEY M., RITTER A., RUSSEL K.S., STUPAR M., BIENAYME O., FREEMAN K.C., GILMORE G., GREBEL E.K., HELMI A., NAVARRO J.F., BINNEY J., BLAND-HAWTHORN J., CAMPBELL R., FAMAEY B., GERHARD O., GIBSON B.K., MATIJEVIC G., PARKER Q.A., SEABROKE G.M., SHARMA S., SMITH M.C., WYLIE-DE BOER E.

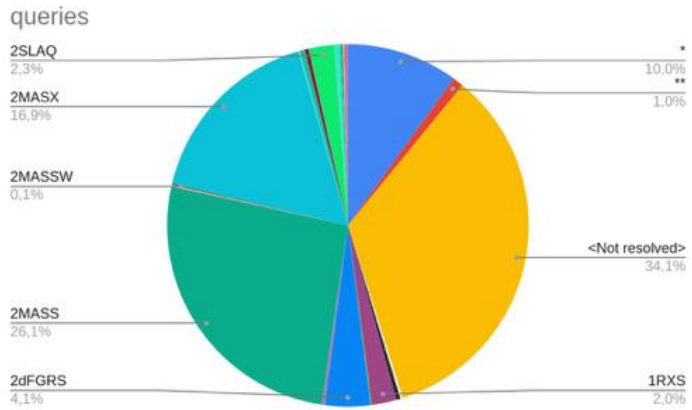
Astron. J., 141, 187 (2011)

*The RAdial velocity experiment (RAVE): third data release.*

- Supersedes the 2nd release (Zwitter et al. 2008AJ....136..421Z, Cat. III/257)
- <RAVE JHHMMSS.s+DDMMSS> N=83072.

A large part of acronyms created are for objects <100

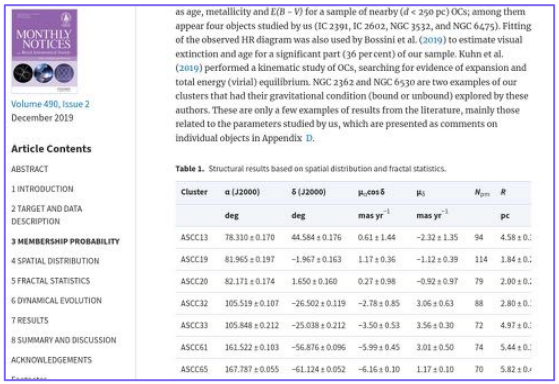



Most acronyms used in SIMBAD web pages queries

# Improving documentalists tools

published tables



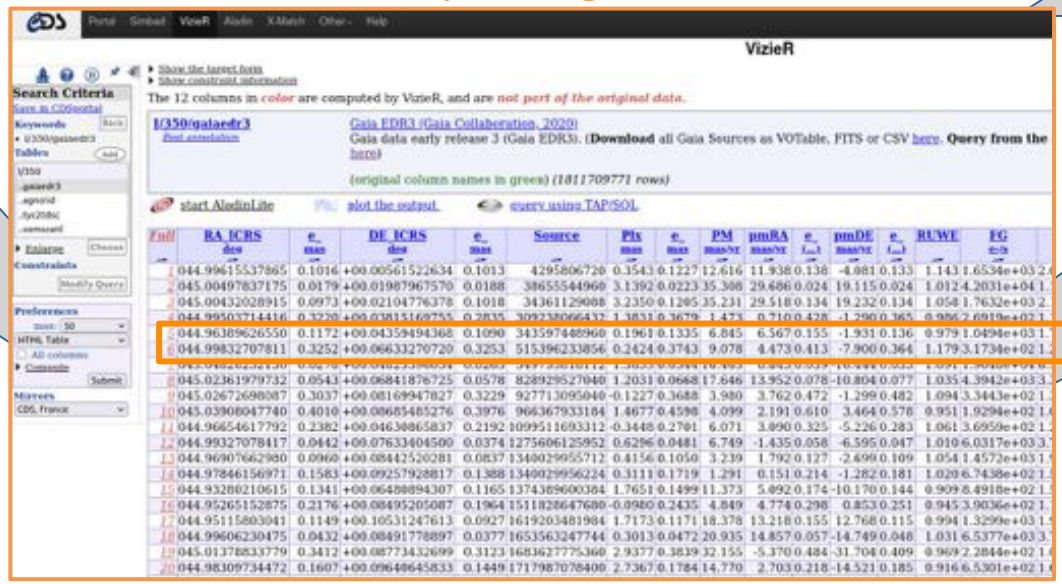

From position / name

GSC4SIM

New catalogs and new input/output interactions


Copy specific data

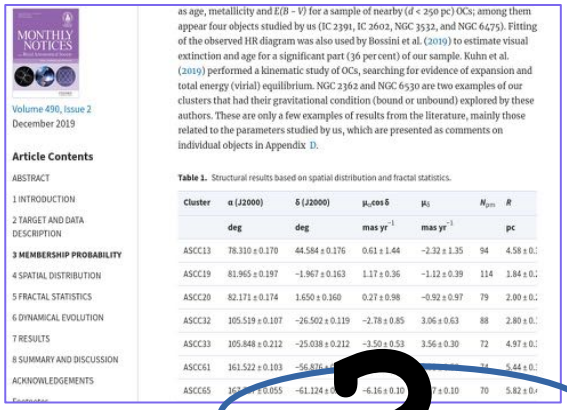
Find corresponding data in Vizier

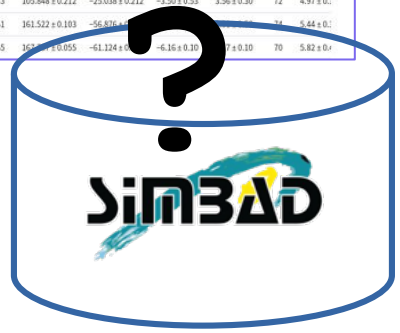


# Improving documentalists tools

**COSIM**

published tables

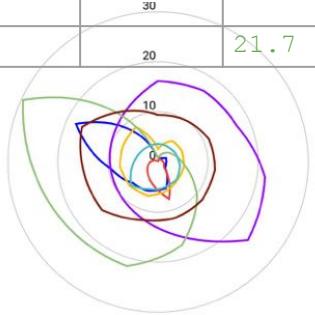



| | |
|---|---|
| UPDATE from identifier | 1221 |
| UPDATE from coordinates | 534 |
| NEW OBJECTS | 100 |
| DISPLAY (to explore more) | 190 |

All distances for possible matches

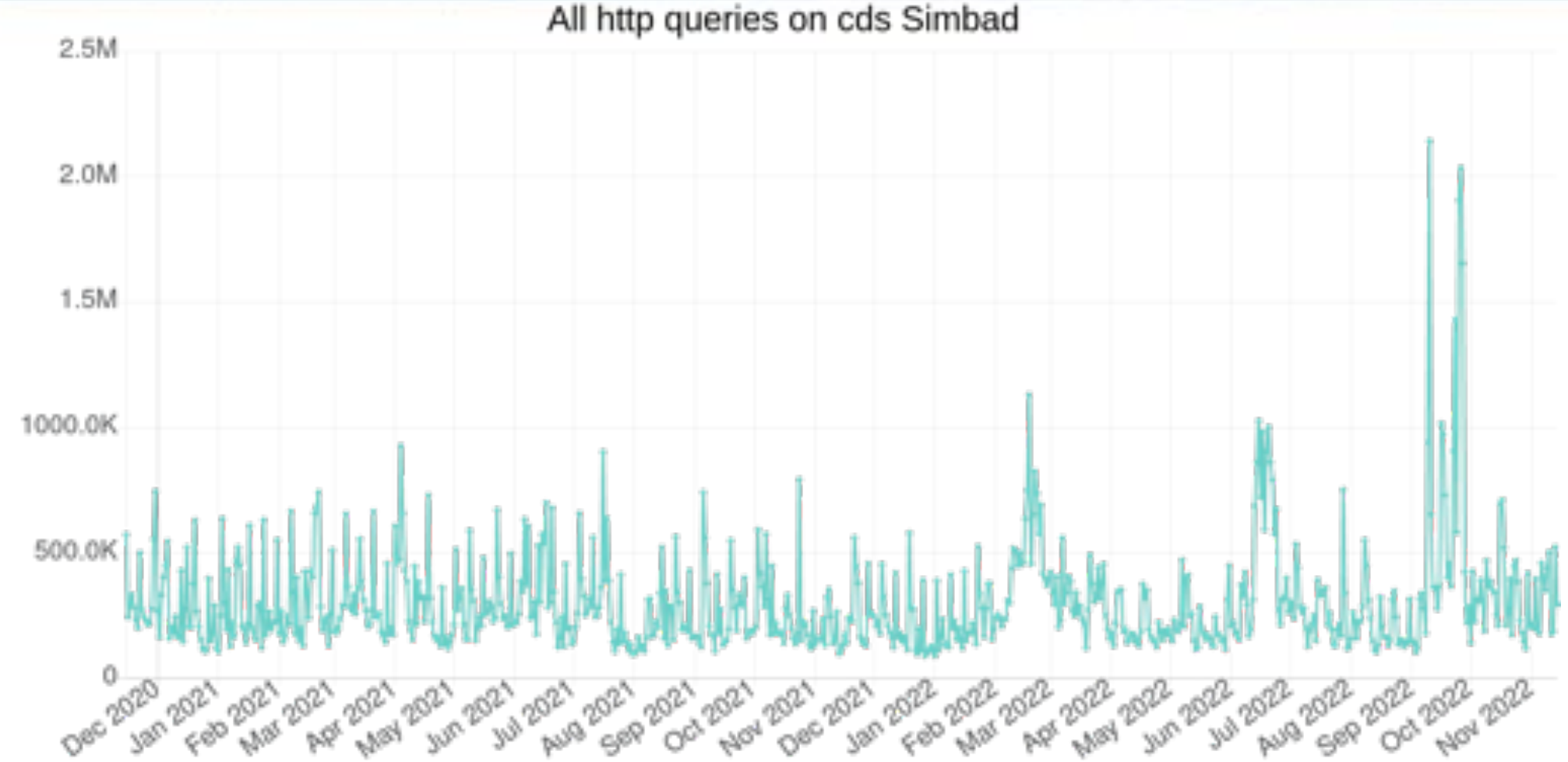
| X/* | Radio/* | denseCore/* | gamma/* | smm/* | PoC/* | BrownDwarf?/* |
|---|---|---|---|---|---|---|
| 0.8 | 0.2 | 2.5 | 16.1 | 0.7 | 3.6 | 9.4 |
| 1 | 0.6 | 51 | 16.9 | 2.2 | 3.6 | 10 |
| 1.7 | 0.8 | 5.2 | 17.4 | 3.6 | 3.9 | 11.1 |
| 1.9 | 1.2 | 5.3 | 21.6 | 5.4 | 4.4 | 11.6 |
| 5.5 | 1.6 | 6 | 23.7 | 10.2 | 5.1 | 11.7 |
| 6.1 | 7.7 | 7 | | 17.3 | 5.2 | 12 |
| 6.5 | | 7.3 | | 21.7 | 7.1 | 14.7 |



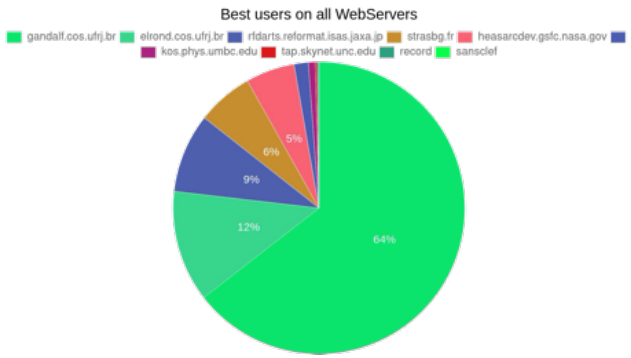

More detailed information: example of distances of possible matches (to explore more) from different types of objects in Simbad with the new table to ingest

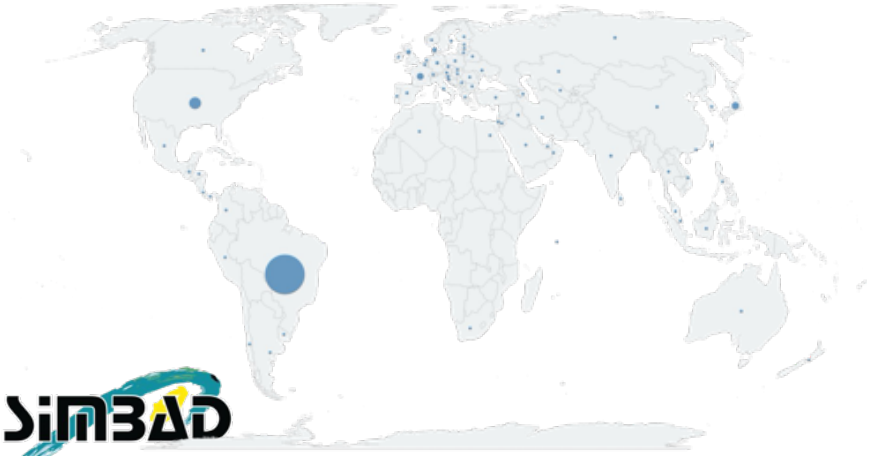
SIMBAD

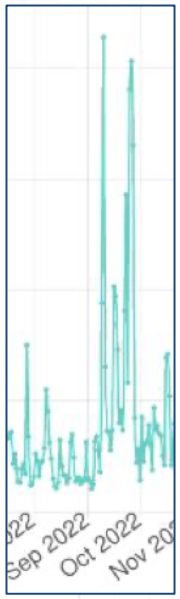
# Usage


All http queries on cds Simbad
2.5M
2.0M
1.5M
1000.0K
500.0K
0
Dec 2020
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Jun 2021
Jul 2021
Aug 2021
Sep 2021
Oct 2021
Nov 2021
Dec 2021
Jan 2022
Feb 2022
Mar 2022
Apr 2022
May 2022
Jun 2022
Jul 2022
Aug 2022
Sep 2022
Oct 2022
Nov 2022

# Usage (Sept. 2022)

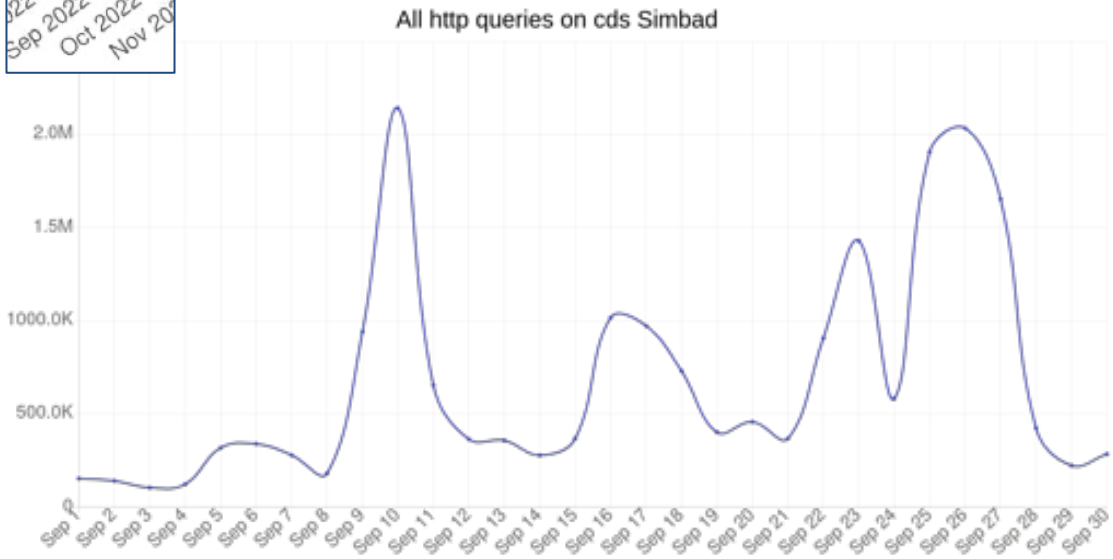






```
format object "
%COO(B)
%PM(B)
%PLX(B)
%RV(B)
%FLUXLIST(G)[%*(B)]"
UCAC4 701-071307

format object "%OTYPE(V)"
GSC 06581-01729
```




# Scientific council 2022

VizieR Staff and contributors:

Astronomers: P.Ocvirk, C. Bot, G. Monari, S.Derriere
Engineers: G.Landais, F.X.Pineau, A.Flint,T.Boch
Documentalists: P.Vannier, E. Perret, C. Fix, A.Fiallos, M. Brouty

Non-CDS: C. Saillard, T. Keller (Strasbourg Observatory)

# Welcome to the VizieR team

## Alicia FLINT

- Ingenieur de Recherche, VizieR developer
- Support VizieR code renewal

## Ana FIALLOS

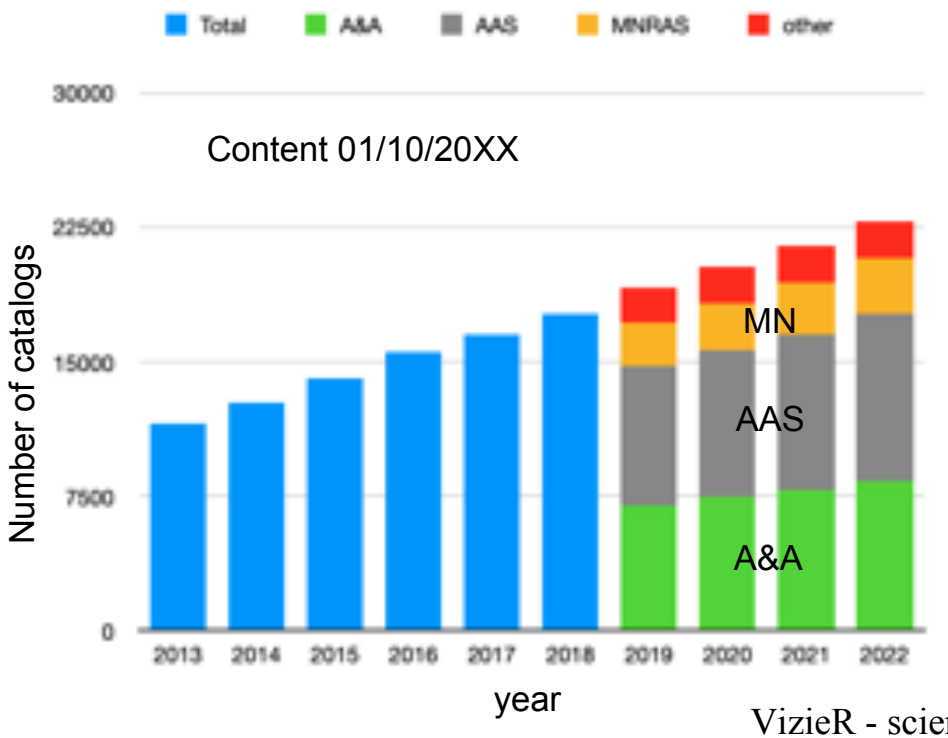
- Ingenieur d'etudes, VizieR documentalist
- Ingestion of catalogues

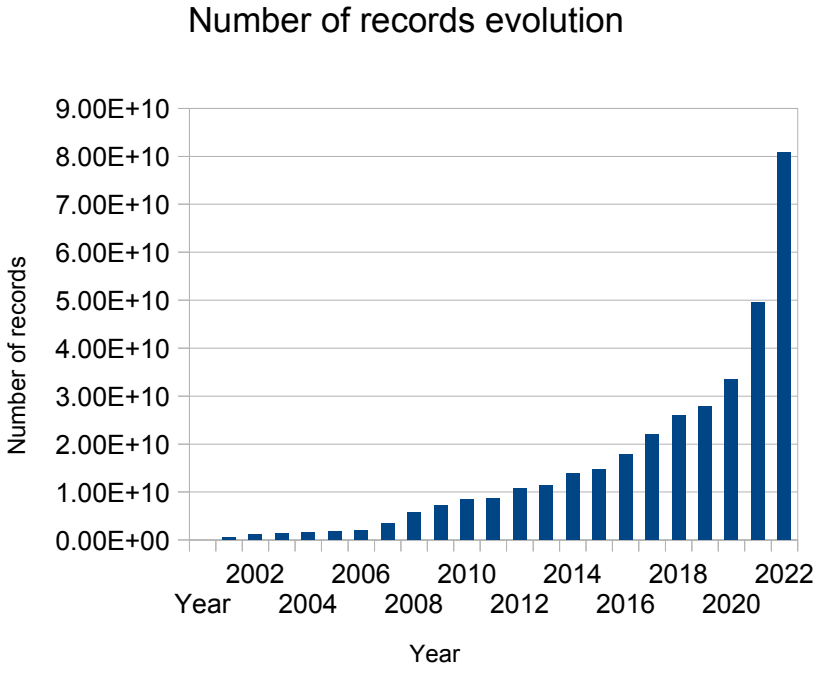
# VizieR content - I

# VizieR content - II

## Ingestion statistitics :

- A steadily increasing volumetry : +1440 in past year
  - Total (05/12/2022): 23 006 catalogues, >52k tables
- Dramatic increase of records : 80 billions records (+62%), 85.5TB (+60%)

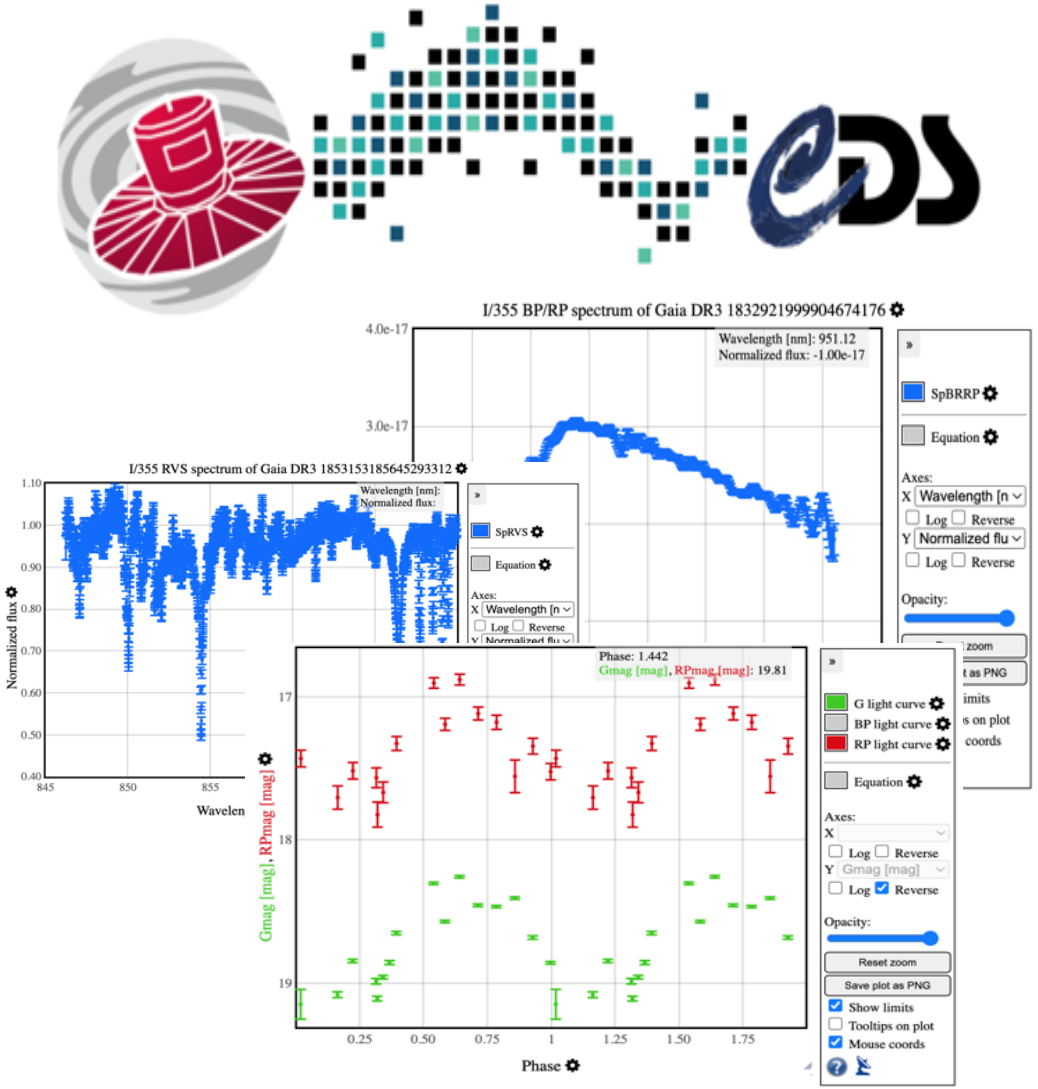


# VizieR content – III : Gaia DR3

## Special operation for Gaia DR3

- 6 catalogs :
  - Main source
  - Extra-galactic
  - Non-single stars
  - Variability
  - Solar system
  - Perf. Verification
- ~70 tables
- including 11 large tables
- Max > 11 billion records
- Tight schedule !



# VizieR content – III : Very large catalogs & co

Very large catalogs ingested in past year:

- SDSS DR16
- Dark Energy Survey DR2
- DESI Legacy Imaging Surveys DR8
- Guide Source Catalogue 2.4.2
- StarHORSE 2021 (long)
- TESS Input catalogue v8.2
- GDR2AP

“Thick” catalogs: > 150 columns

- SDSS-RM (Reverb. Mapping) (up to 472 columns)

In progress/contact made:

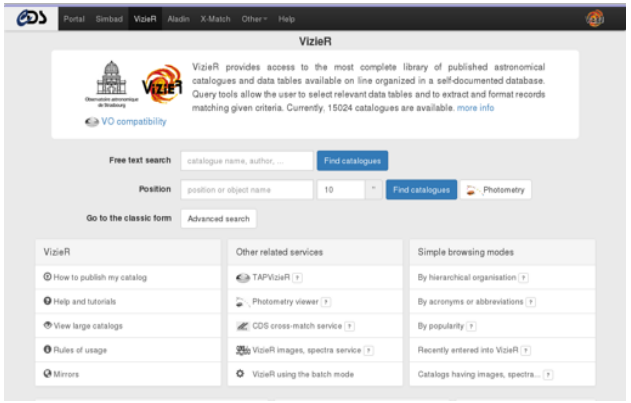
- NOAO Source Catalog
- Pan-STARRS DR2
- EROS

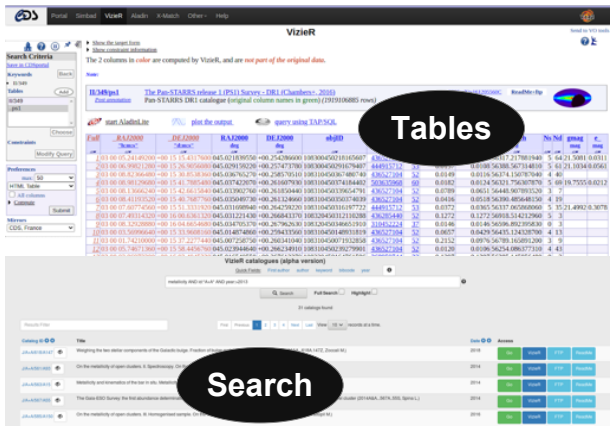
Planned for 2022:

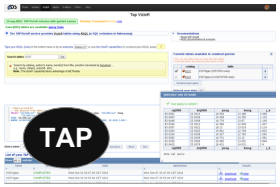
- DES: DESI DR8, DECALS DR9/10
- KIDS DR4
- ESO phase 3: ATLAS-DR4, VPHAS+ DR3.2, VIKING DR4

# Accessing VizieR data

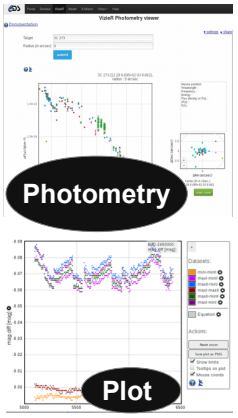
VizieR services and access modes: discovery, table access, visualization, photometry, associated data, astroquery, ...



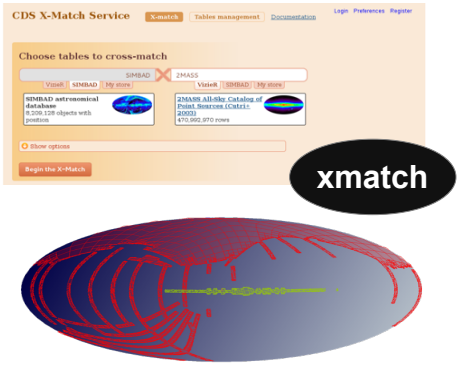



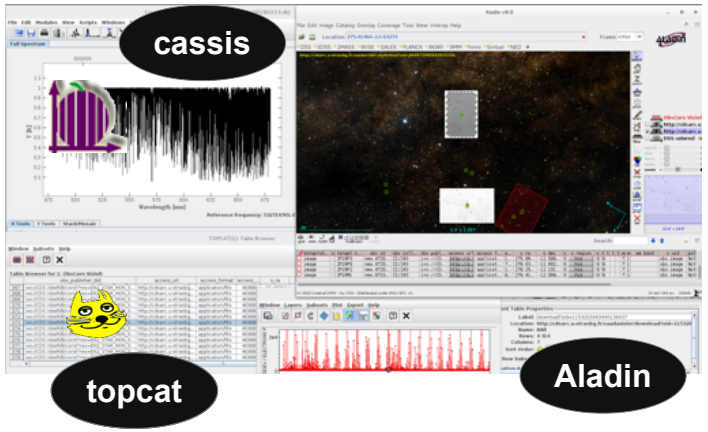





Derived products provided in CDS or data available through software, API..

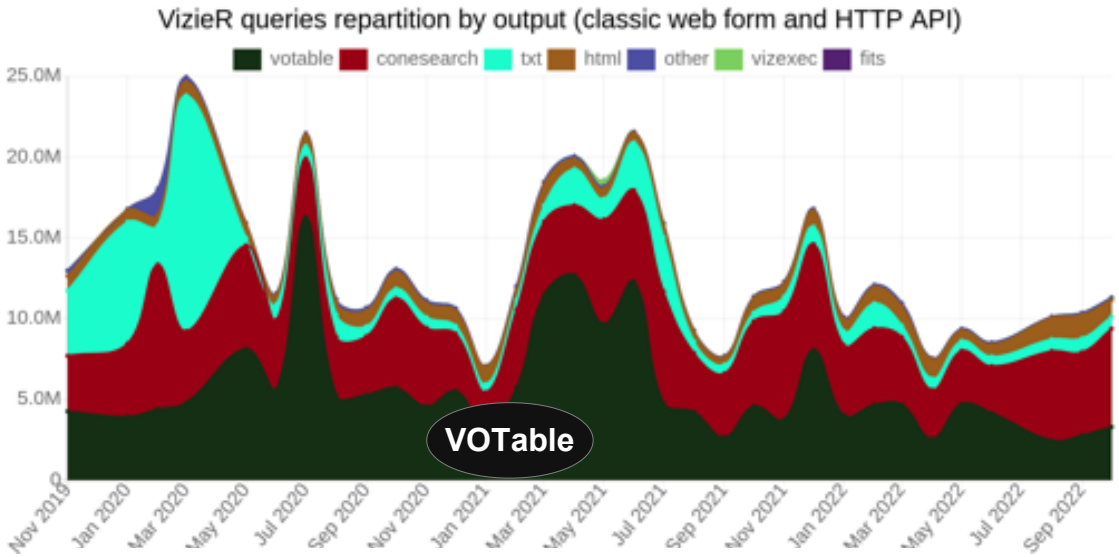




# VizieR usage statistics

VizieR statistics (from the CDS statistics collector)



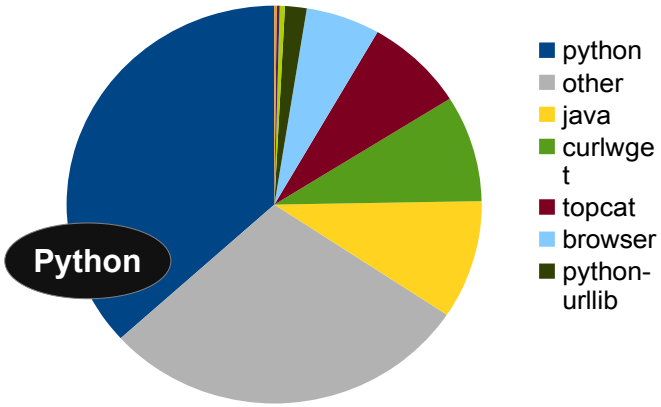

Total : ~430K queries/day

TAP : ~36.1K queries/day

Associated data (spectra/images): ~900 queries/day

Importance of the Virtual Observatory (conesearch+VOTable)~70% of the queries

Queries repartition by user-agent




Importance of Python(37%), TOPcat (8%), curl+wget (9%)

# Recent activity

- VizieR **certification** renewal (Core Trust Seal)
  (G.Landais + P.Fernique + F.Genova + M.Allen)

- **Database** migration for the entire VizieR computer infrastructure (12 databases)
  Postgrtesql9.6 → Postgresql14 (G.Landais)

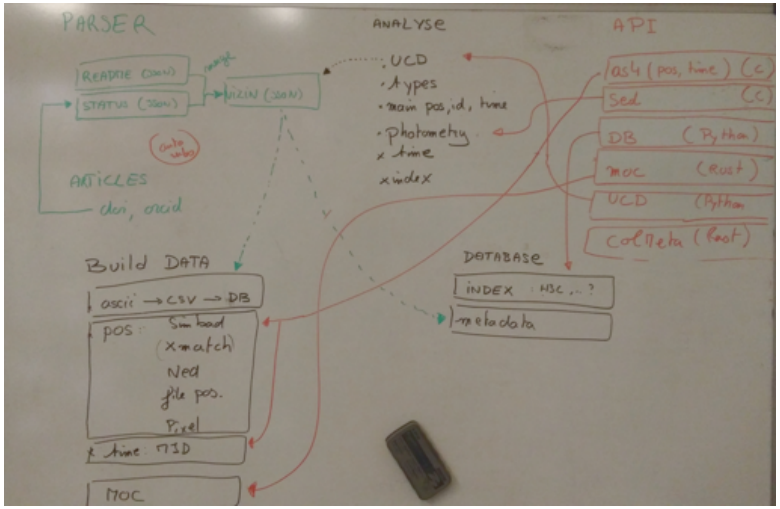
- **Unit** migration to VOUnit for all outputs (G.Landais + S.Derriere)
  unity library (N.Gray)
  example:

| | CDS units | VOunits |
|---|---|---|
| | [Msun/yr] | log(solMass/yr) |
| flux | 10-13ph/cm2/s/MeV | 1e-10ph.cm**-2.s**-1 |

- **Global indexation** for positions using **MOC** (FX. Pineau + A.Vanhulle + G.Landais)
  VizieR moc server: https://cdsarc.cds.unistra.fr/viz-bin/moc

  Improve the VizieR granularity :
  Qbox (order 8: ~12 arcmin)
  → MOC (order 10: ~3arcmin)

- **VizieR ingestion code renewal** :
  a long work just beginning to put the VizieR source code in the long term -
  (G.Landais + Alicia Vanhulle + FX.Pinneau)



# VizieR renewal

**Bilan**

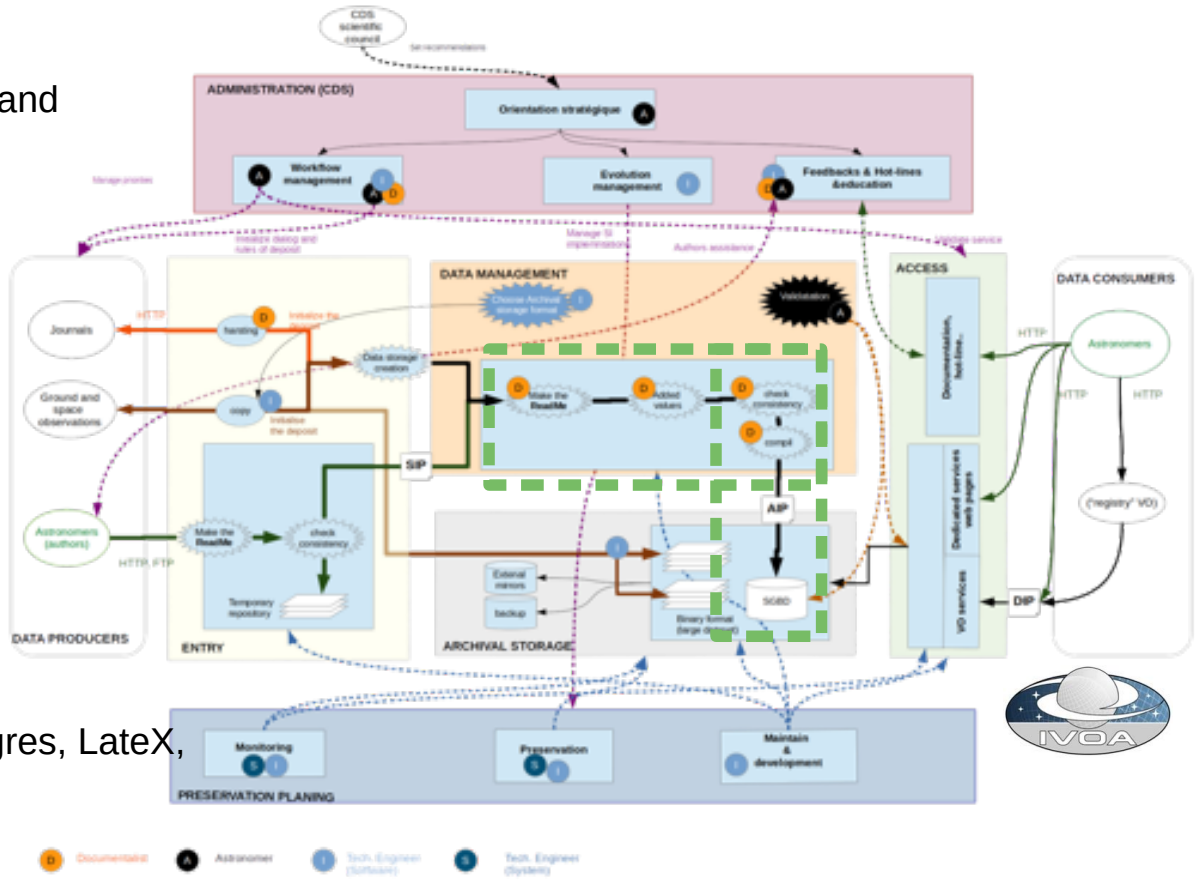
- 30 years  (1992) – success !
- But also: difficulties of maintenance and evolution
- Tech: C, SHELL, AWK, Python, Postgres/Sybase, LateX
- >100K lines codes (ingestion code)

**The new version**

- Conserve the knowledge
- Meet the evolution requirements
- Adapt the technology
- Tech: C/C++, SHELL, Python, Postgres, LateX, ANTLR, ElasticSearch, Rust

Example of updates:

- Global indexation(MOC)
- FAIR Metadata (Licences, Data origin)
- VO compatibility: eg: links  (Datalink)




The VizieR Information System - OAIS schema

# Outlook – 2023 and (a bit) beyond

- **Top Priority = continued support of FAIR data publication from major journals and data producers through variety of access modes**
- => continued support and development of in-house tools to assist catalogue indexation and ingestion
- ~Short term emergency: Associated data workflow
  - SAADA lead dev retiring
  - => requires new strategy:
    - SAADA code assimilation?
    - Rewrite lower level workflow?
    - Opportunity to taylor to VizieR needs
- Med to long term goals:
  - Code renewal (Very high priority, ongoing)
  - Consolidation of VizieR team with new staff

# The Aladin project

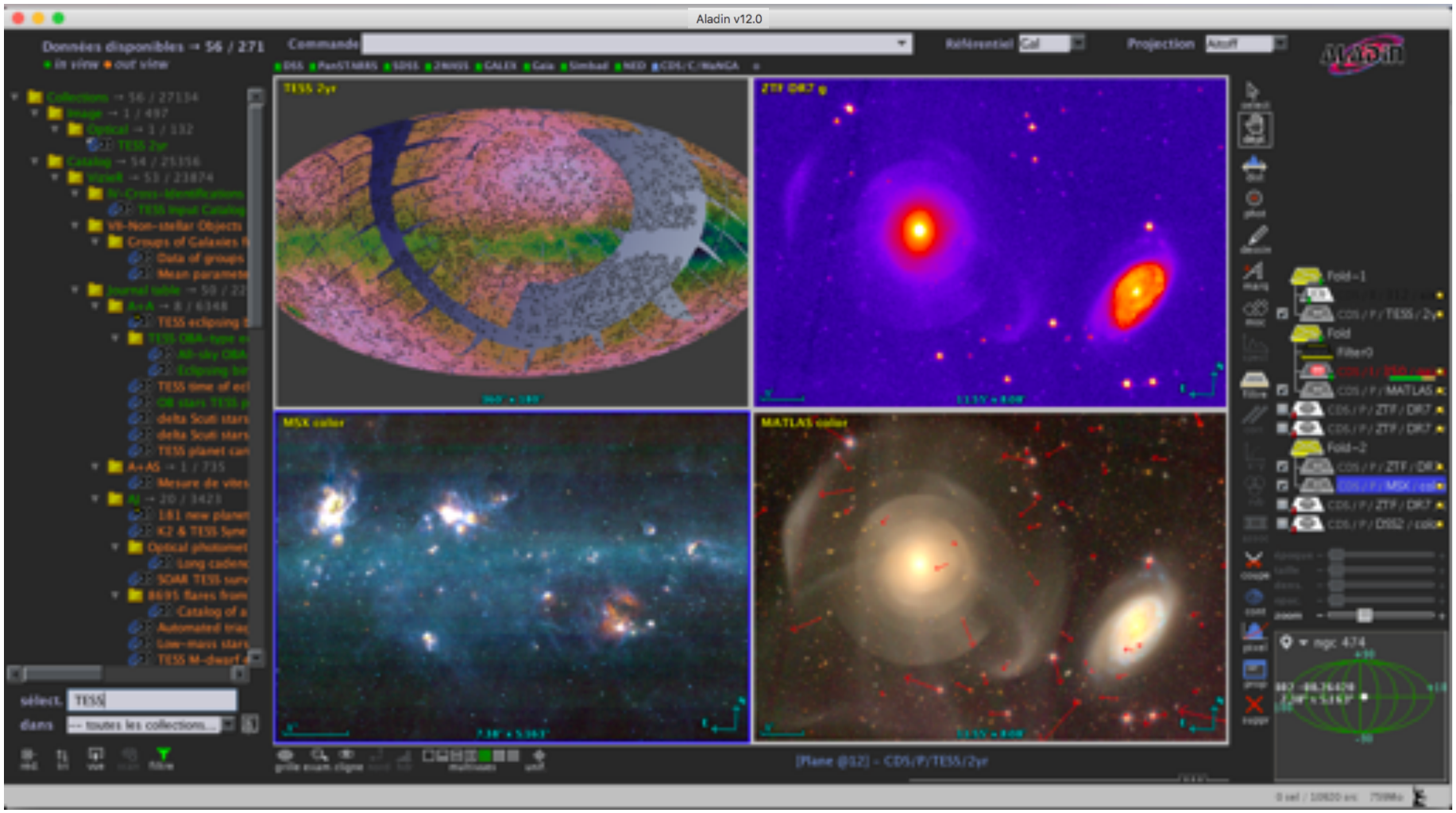
*in 2021 - 2022*

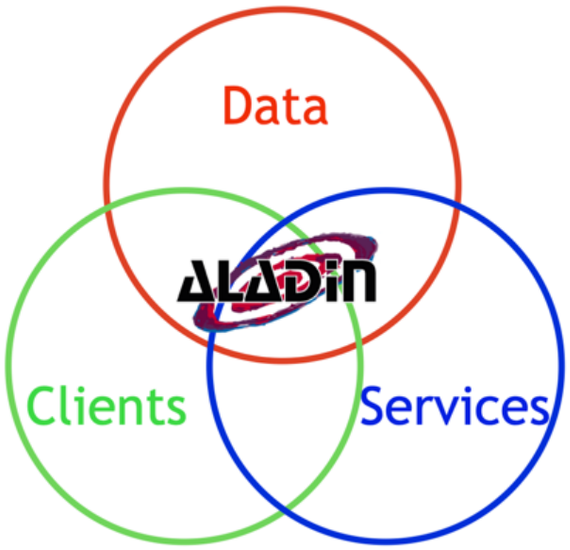
## CDS council - December 6, 2022


CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

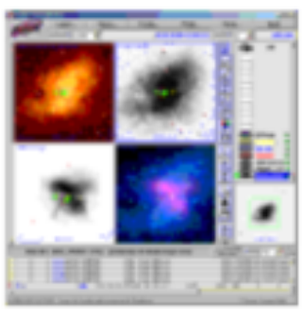
*Thomas Boch*, Caroline Bot, Pierre Fernique, Mark Allen, Matthieu Baumann, François Bonnarel, Mihaela Buga, Sébastien Derriere, Hendrik Heinl, Ada Nebot, François-Xavier Pineau, Manon Marchand, Christophe Saillard, Thomas Keller

# Aladin project




HiPS2FITS
HiPSgen, …



Download
**Aladin Desktop**
on your machine

Preview with
**Aladin Lite**
in your browser

# CDS 50 - Aladin 30



**ALADIN: AN INTERACTIVE DEEP SKY MAPPING FACILITY**

1994IAUS..161..347P

Ph. PAILLOU, F. BONNAREL, F. OCHSENBEIN and M. CRÉZÉ
*Centre de Données Astronomiques de Strasbourg — CDS*
*Observatoire Astronomique de Strasbourg*
*France*

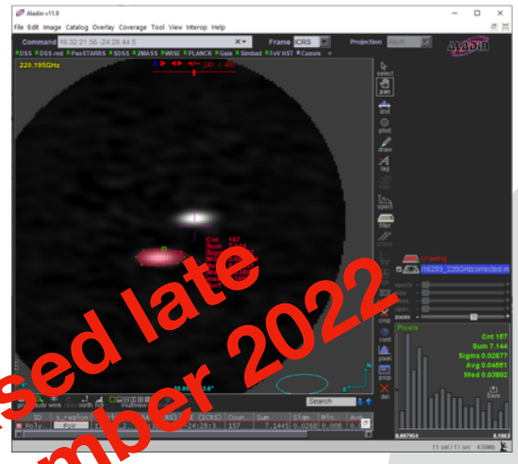
**2. Aladin Project**

CDS began the Aladin project in 1992. The CDS council agreed with the general idea in September 1992. In December 1992, the Astronomy group of INSU/CNRS decided that a priority was the funding needed to start this CDS project. Funding for Aladin's construction was approved in June 1993, after a project review. Aladin will provide access, in addition to a complete sky atlas, to up-to-date data contained in Simbad and catalogue archives. Distribution of this tool to some selected sites is foreseen for the end of 1995.

For the interactive sky atlas, requirements in storage capacity are of the order of one Tbyte for complete sky coverage (1 TB for a complete sky in one colour with a resolution of 1 arcsec and coding on 16 bits; this can be reduced by using data compression). Hardware and software solutions for one Tbyte data archiving exist and are already in use. These solutions rely on 12" WORM optical disks systems and are now used in several places (ESA/ESRIN, STScI, NASA). The access speed to stored information is compatible with the proposed application (less than one minute to access a given sky region image). New technology to archive several tens of Tbytes

# Aladin Desktop v12

## *What's new in release V12 ?*



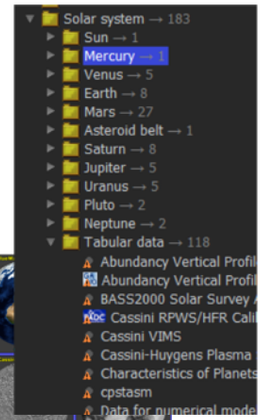
Released late September 2022

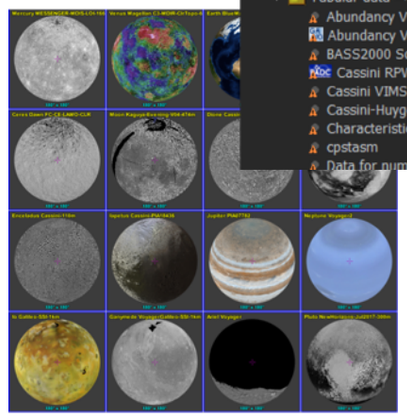
### Photometric tools

- *HiPS support*
- *Pixel extraction by regions*
- *Table of measurements*

### Planetary data support

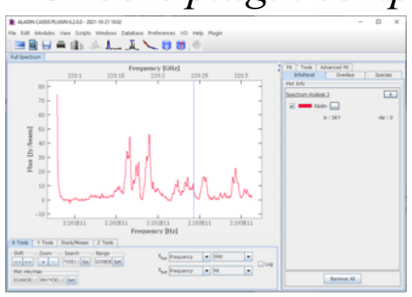
- *HiPS, MOC, catalogs..*



### Spectrum manipulation

- *Spectrum from cubes*
- *CASSIS plugin compatibility(IRAP)*

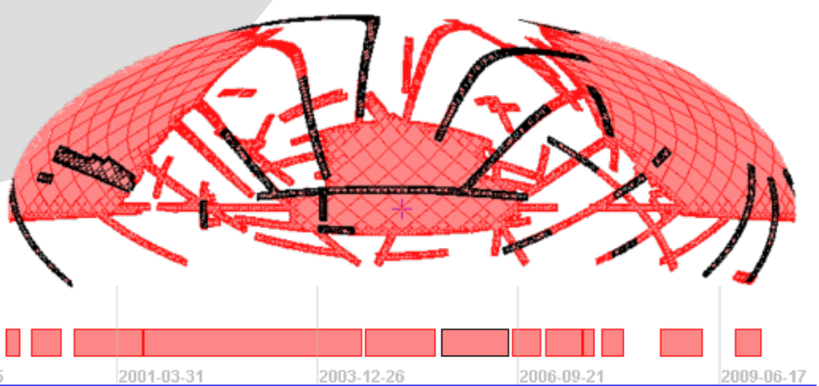


### MOC improvements

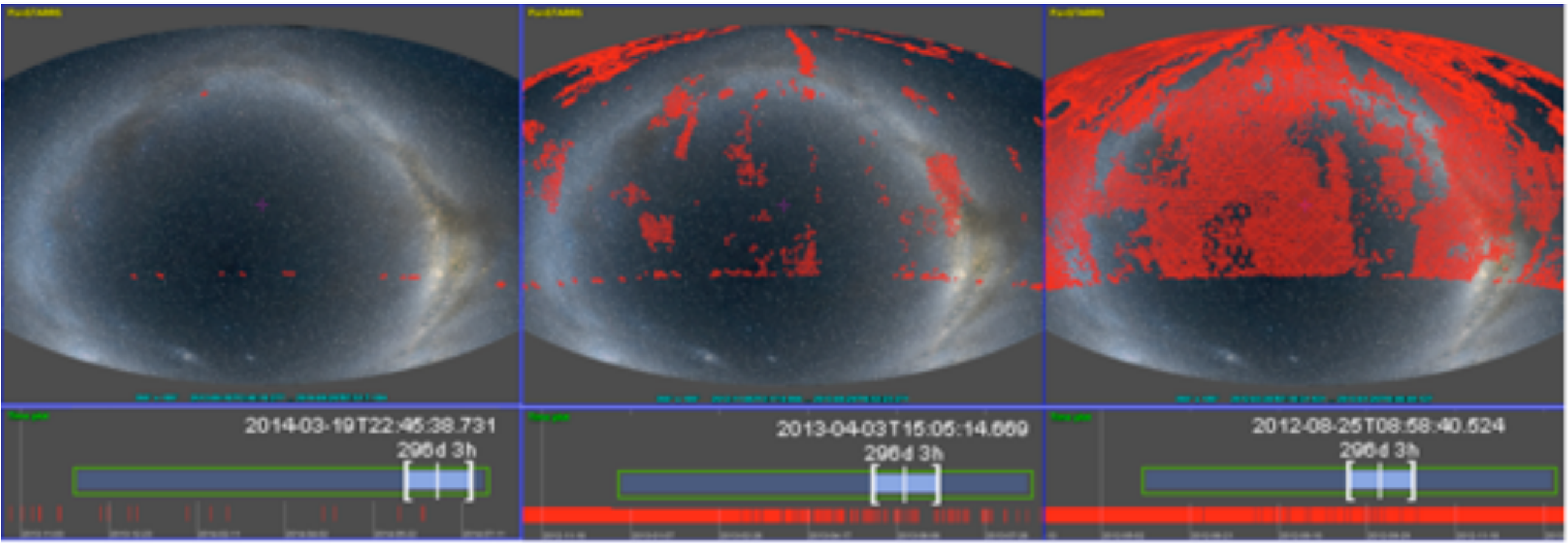
- *IVOA MOC 2.0 full compliance*
- *Resource Tree controlled both by space and time filters*
- *STMOC highlight & selection*

### Cone search update facility

- *Update the contents of a ConeSearch & TAP catalogue plane*

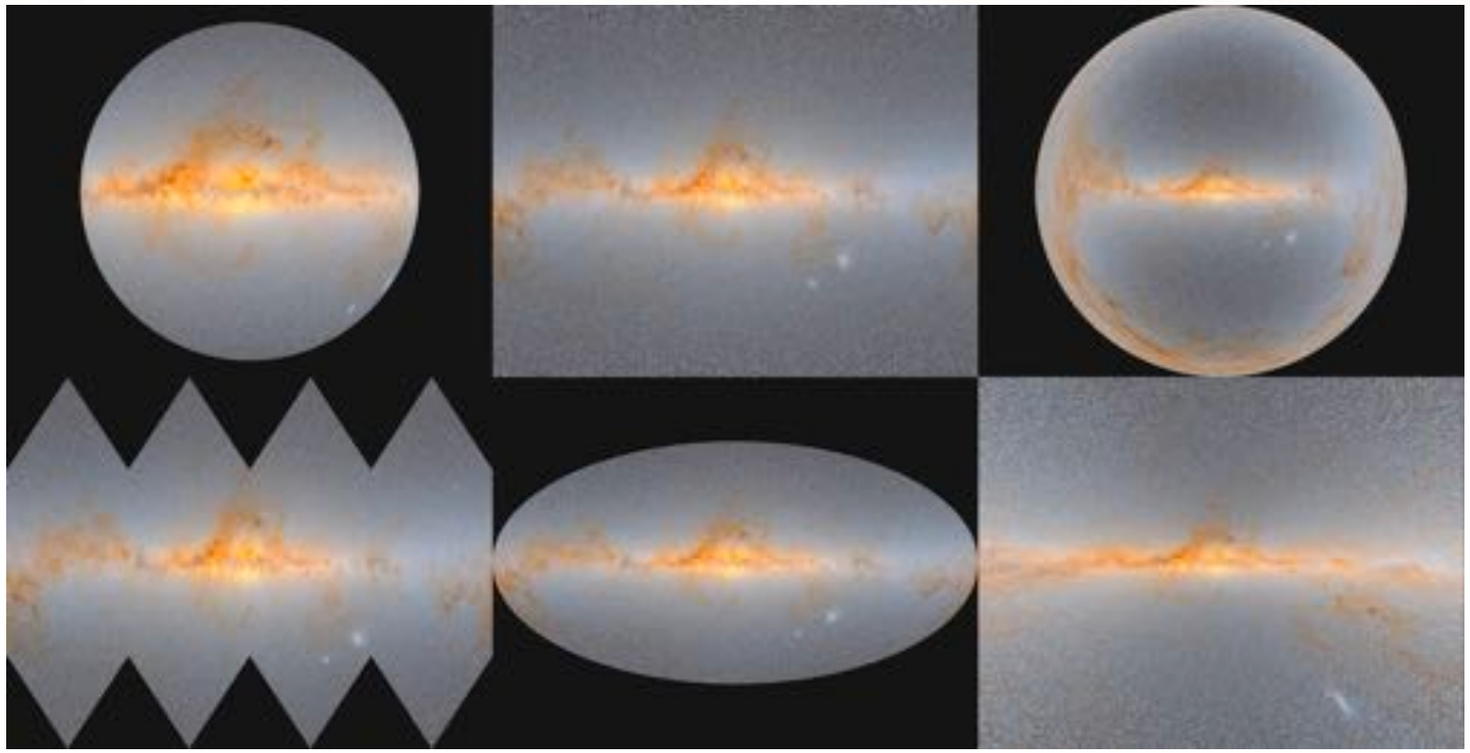


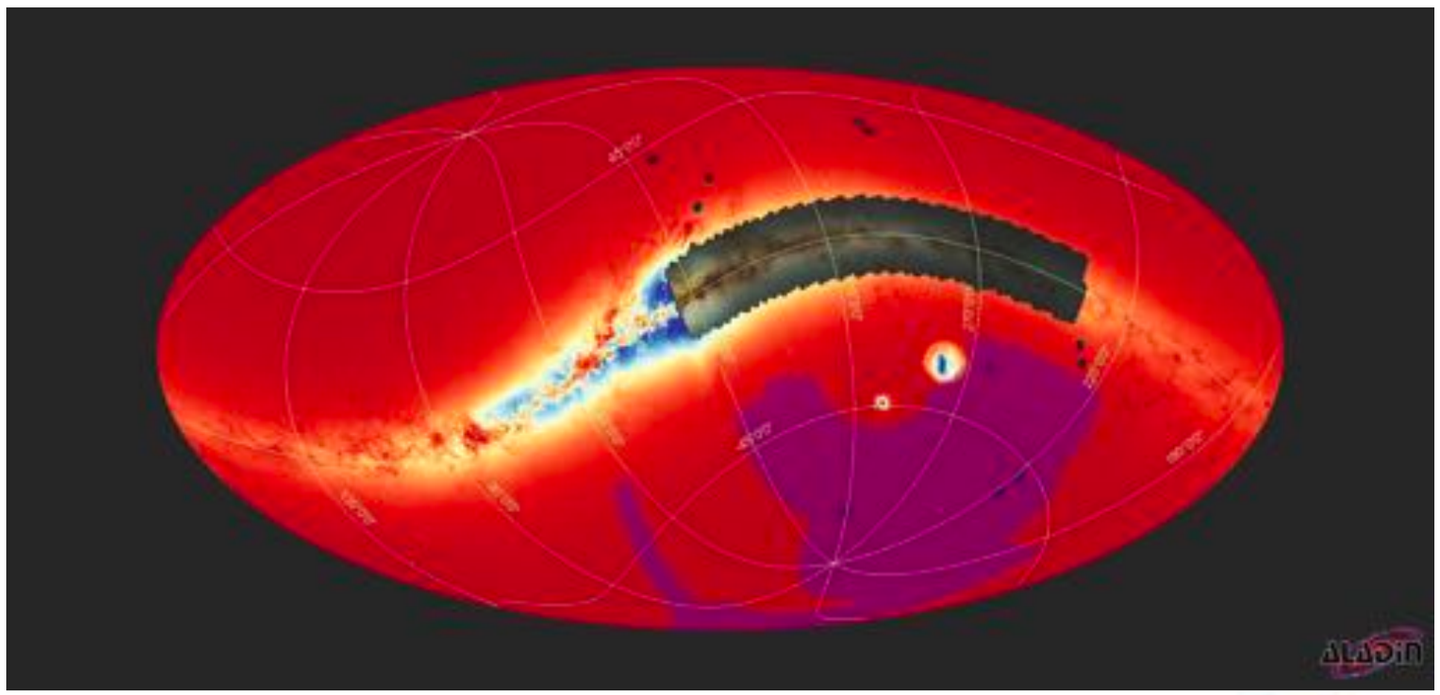
# Space-Time MOC filtering



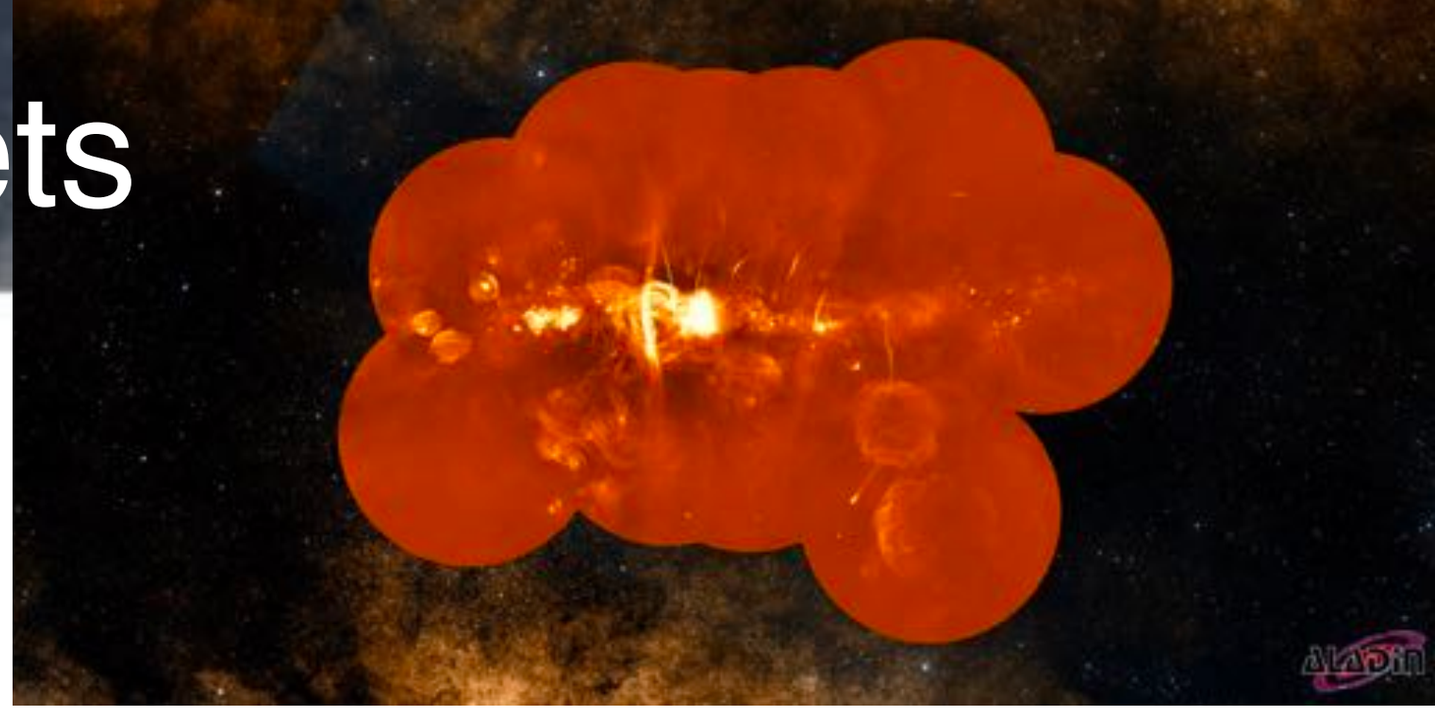
# Aladin Lite v3 in beta

- Beta test release in June 2022
- New features
  - improved GPU rendering (WebGL/Rust)
  - access to FITS tiles
  - new projections
  - coordinates grid
  - access to all image HiPS
  - access to VizieR tables
  - new color maps
- Official release in January 2023

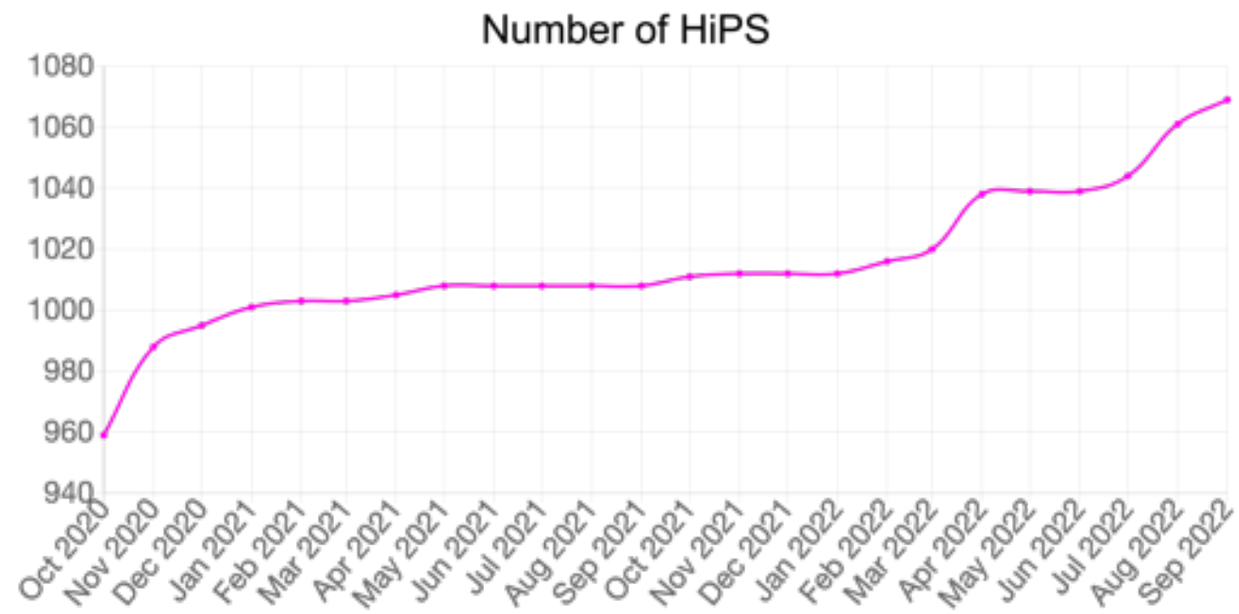


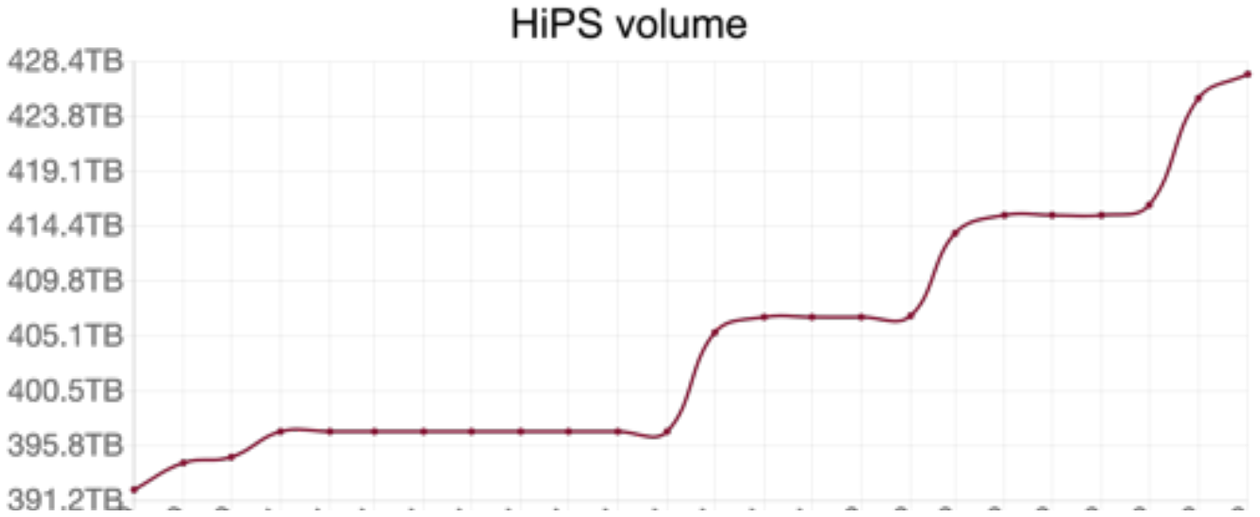


# HiPS datasets

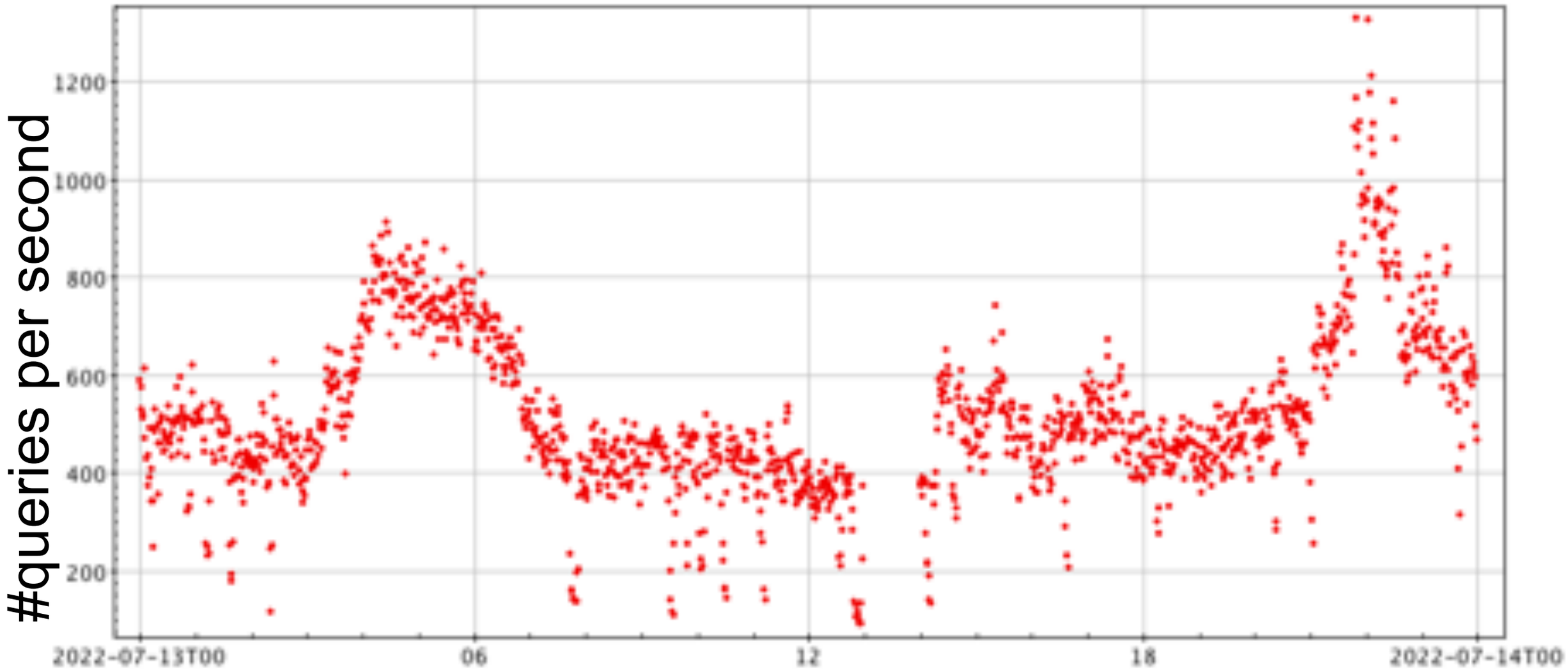


- HiPS creation ongoing
  - MeerKAT
  - JWST first images
  - ESO outreach, HST outreach
  - …
- reprocessing
  - GALEX
  - PanSTARRS
- helping others
  - e.g: LOFAR LoTSS



# HiPS peak of requests



# HiPS access statistics

**360k** users/month

+53%

**1.7M** queries/day

+73%

## Growth driven beyond professional astronomy

# Conclusion and future plans

- 30 years of experience
  - still learning to evolve and stay strong

- Aladin Desktop v12, Aladin Lite v3, HiPS growing
  - continuous interaction with the community
  - e.g: 2 talks and 1 poster at ADASS, 1 VO School

- Involvement in large future projects to guide our R&D
  - time-domain developments, perspective LSST
  - dealing with (large) data cubes
    - CASSIS collaboration —> Aladin Desktop plugin
    - SKA Orange team —> visualisation prototypes

- All this relies on humans as well

https://aladin.cds.unistra.fr/AladinLite/showcase/v3-2022-animation/

# R&D @ CDS

## and various developments

André Schaaff on behalf of the CDS Team

**CDS Scientific Council 2022**

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

IVOA

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

# Foreword

- Technological evolutions are fast & various (interactions, visualization, mobility, components, Big & Open Data, Clouds, etc.).

- The R&D activity is well identifed since 2 decades, structured and involving several persons of the staff with the help of (10 to 12 per year) interns and short contracts.

- By the way, a continous training through the presentation of the results.

# Internship programme

- After a slowdown in 2021, we had again 12 interns in 2022 who worked with us on a few topics, R&D and various developments

- An intern, Thibault Dumortier, hired as apprentice for one year from September 2022.

- IT Job Market still tight and hiring is not easy.

# Internship programme, remark

- This presentation is not exhaustive, other developments and experiments were also done without being explicitly explained here even if interns were involved.
- Example: 2 interns, Baptiste Robert (IUT Charlemagne, Nancy) and Thibault Dumortier (IUT Schuman, Strasbourg), worked with François-Xavier Pineau, cf. the X-Match service presentation. They worked with the Rust language.

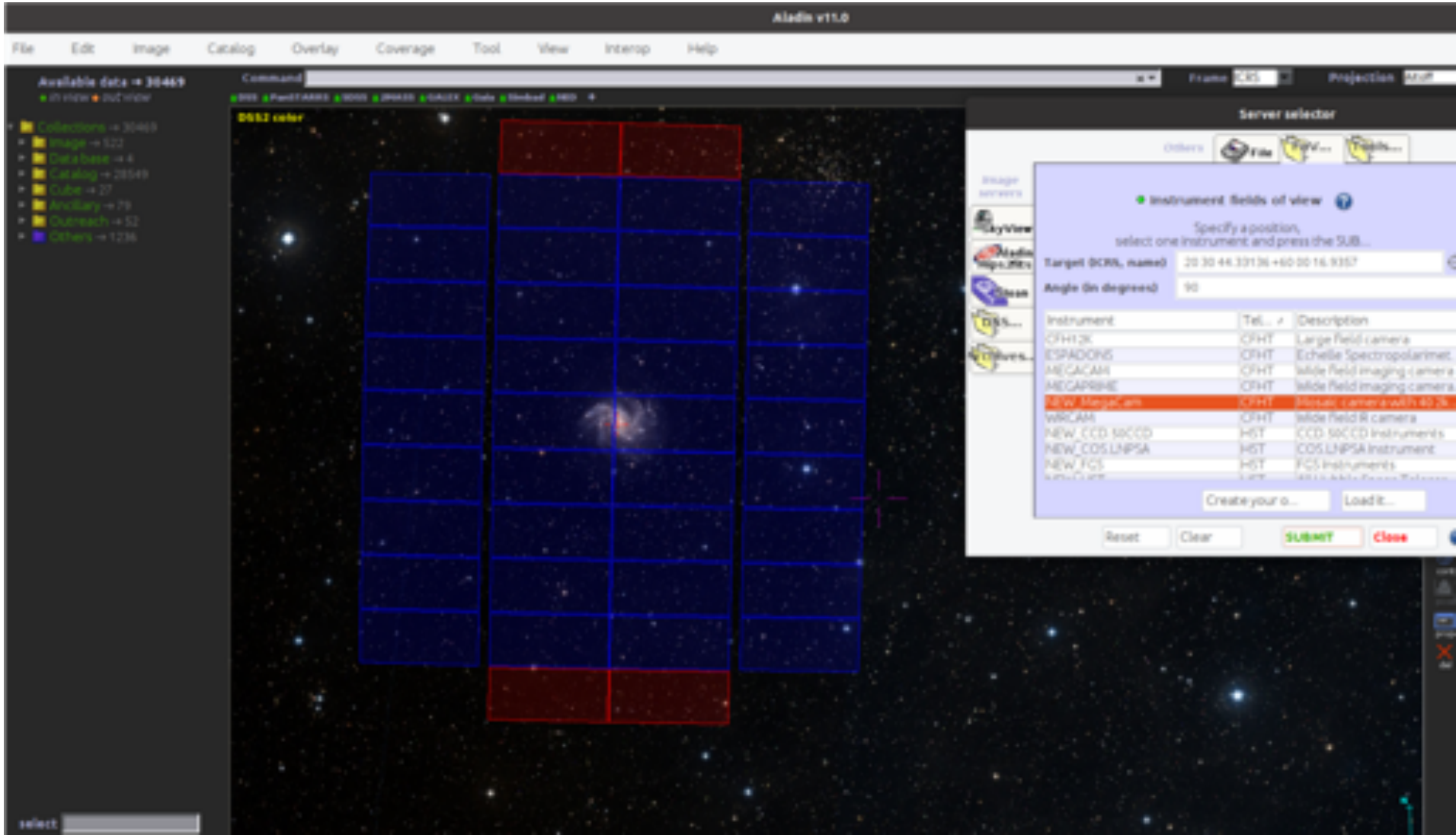
- Gilles Landais worked without the help of an intern on a LaTex parser (based on ANTLR) in the frame of VizieR to validate configuration files.

# Updating representation of instruments'Field of View with recent IVOA standards

- Aladin Desktop is regularly used for preparing and checking observations.
- It allows the overlay of a large set of instrumental Fields of View represented in VOTable XML.
- Such FoV can also be created by drawing with the help of a graphical editor
- Update with recent IVOA standards needed for better interoperability

**F. Bonnarel, M. Louys , T. Boch, L. Michel**
**Intern: Clément Nogueira (IUT Schuman, Strasbourg)**

# Selecting instruments footprints in Aladin Desktop

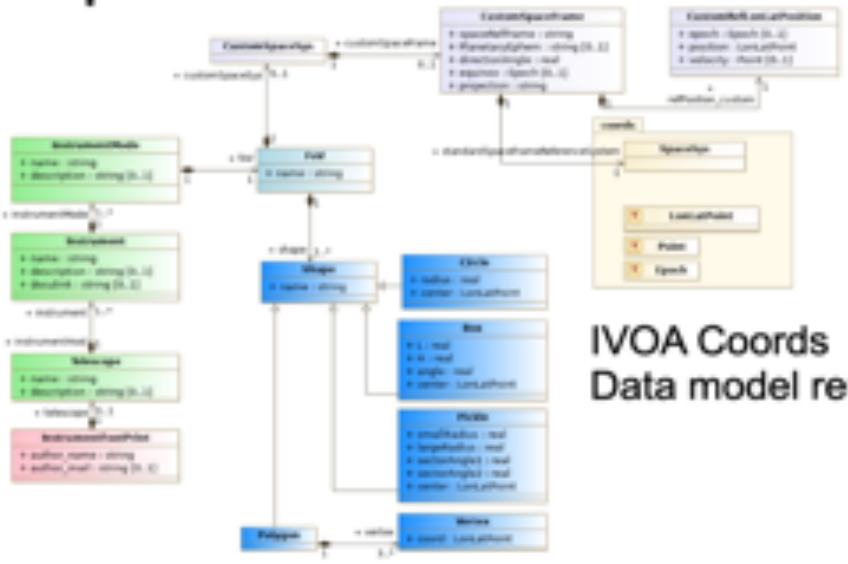


# Foot print Data model Update

Annotation syntax

International Virtual Observatory Alliance

Model Instances in Votables
Version 1.0
IVOA Proposed Recommendation 2022-09-02
Working group
DM
This version
http://www.ivoa.net/documents/mivot/20220902



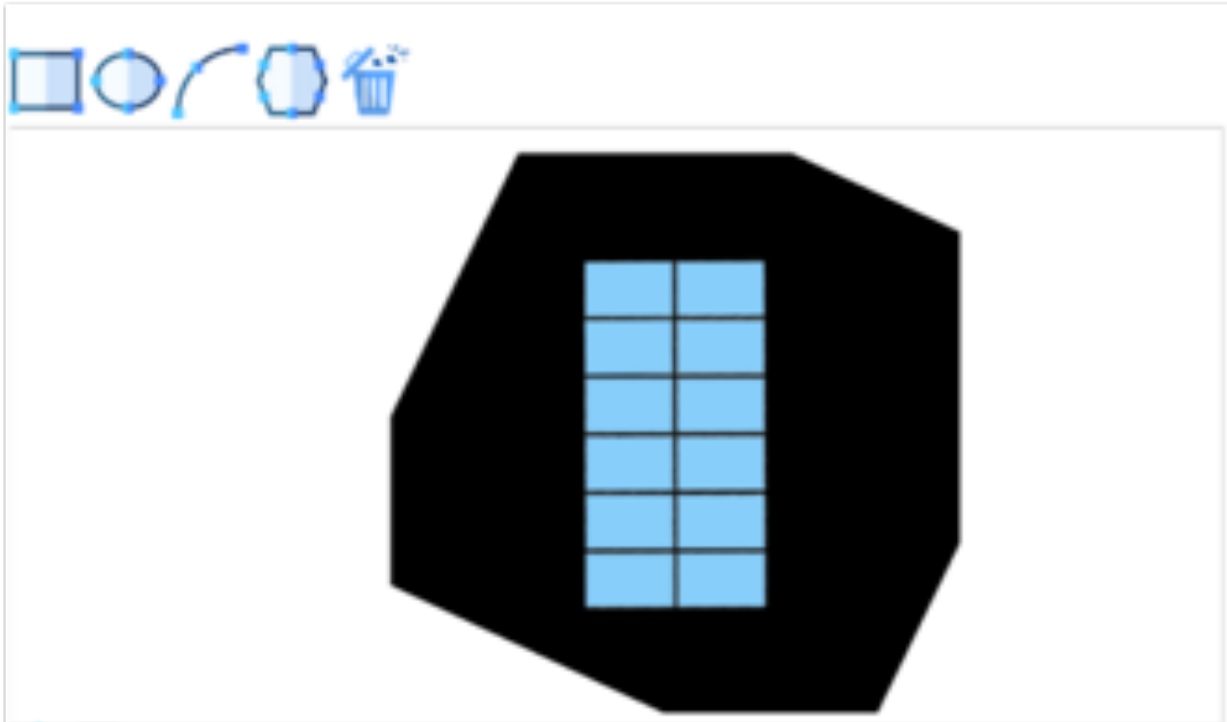
IVOA Coords Data model reuse

XML representation

```
<RESOURCE type="meta">
  <VODML>
    <REPORT status="OK">Hand-made mapping: ad hoc model so far</REPORT>
    <MODEL name="instfov" url="https://www.ivoa.net/xml/instfov.xml"/>
    <MODEL name="coords" url="https://www.ivoa.net/xml/Coords-v1.xsd"/>
    <GLOBALS>
      <INSTANCE dmid="_footprint" dmtype="instfov:InstrumentalFootPrint">
        <ATTRIBUTE dmrole="instfov:InstrumentalFootPrint.author_name"
                   dmtype="ivoa:string" value="Laurent MICHEL"/>
        <ATTRIBUTE dmrole="instfov:InstrumentalFootPrint.author_email"
                   dmtype="ivoa:string"
                   value="laurent.michel@astro.unistra.fr"/>
        <INSTANCE dmrole="instfov:InstrumentalFootPrint.telescope"
                  dmtype="instfov:Telescope">
          <ATTRIBUTE dmrole="instfov:Telescope.name" dmtype="ivoa:string"
                     value="MockedTelescope"/>
          <ATTRIBUTE dmrole="instfov:Telescope.description"
                     dmtype="ivoa:string" value="mocked telescope"/>
          <COLLECTION dmrole="instfov:Telescope.instrument">
            <INSTANCE dmtype="instfov:Instrument">
              <ATTRIBUTE dmrole="instfov:Instrument.name"
                         dmtype="ivoa:string" value="MockCamera"/>
              <ATTRIBUTE dmrole="instfov:Instrument.description"
                         dmtype="ivoa:string" value="mocked camera"/>
```

https://saada.unistra.fr/fov/

# Drawing and recording instruments footprints



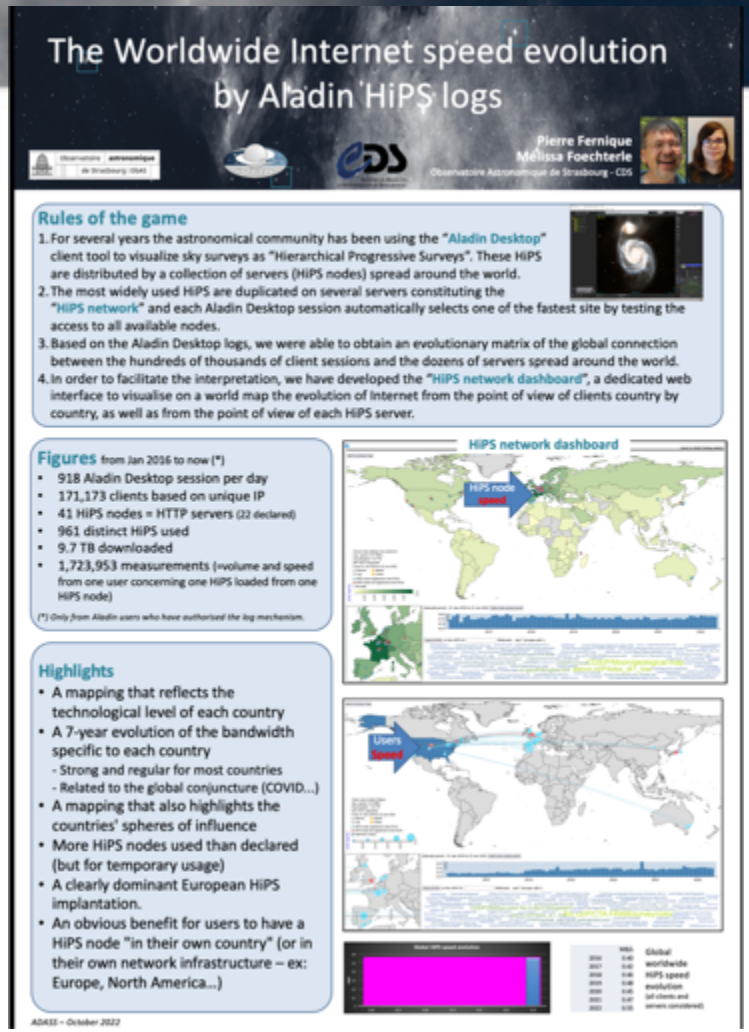

**Java script editor** : allows to generate instruments FOV by drawing them from scratch or on top of a template image

# The Worldwide Internet speed evolution by Aladin HiPS logs

- For several years the astronomical community has been using the Aladin Desktop client tool to visualize progressive sky surveys (HiPS) which are distributed through a collection of servers (HiPS nodes) spread around the world.
- The most used HiPS are duplicated on several servers and each Aladin Desktop session automatically selects one fast site.
- The Aladin Desktop logs were exploited to obtain an evolving matrix of the global connection between the hundreds of thousands of client sessions and the dozens of servers around the world.

**Highlights**

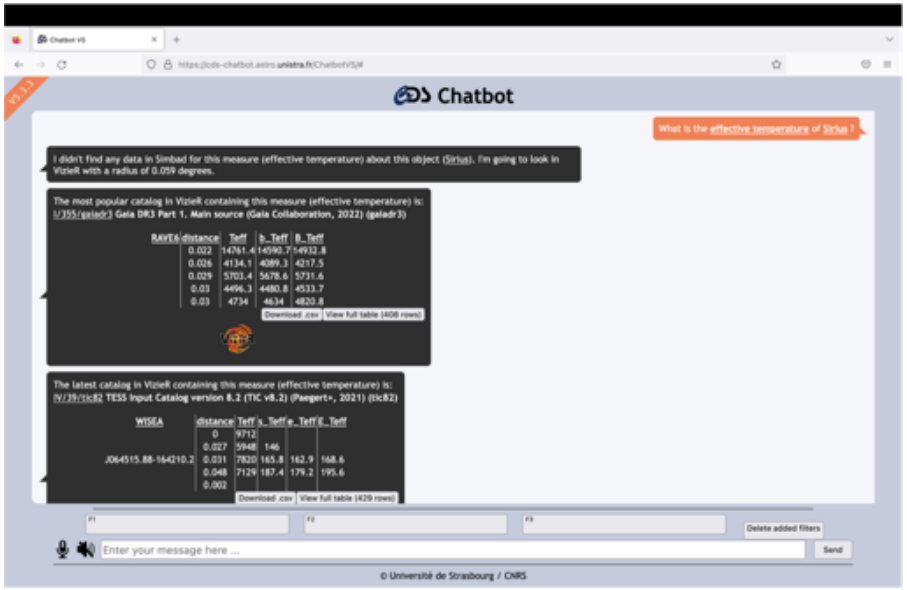
- A mapping that reflects the technological level of each country (A 7-year evolution of their bandwidth), strong and regular for most countries.
- A clearly dominant European HiPS implantation.
- An obvious benefit for users to have a HiPS node "in their own country" (or in their own network infrastructure – ex: Europe, North America…)
- Also related to the global conjuncture (COVID, ...)




**P. Fernique Intern: Mélissa Foechterlé (IUT Schuman, Strasbourg)**

# Chatting with the services

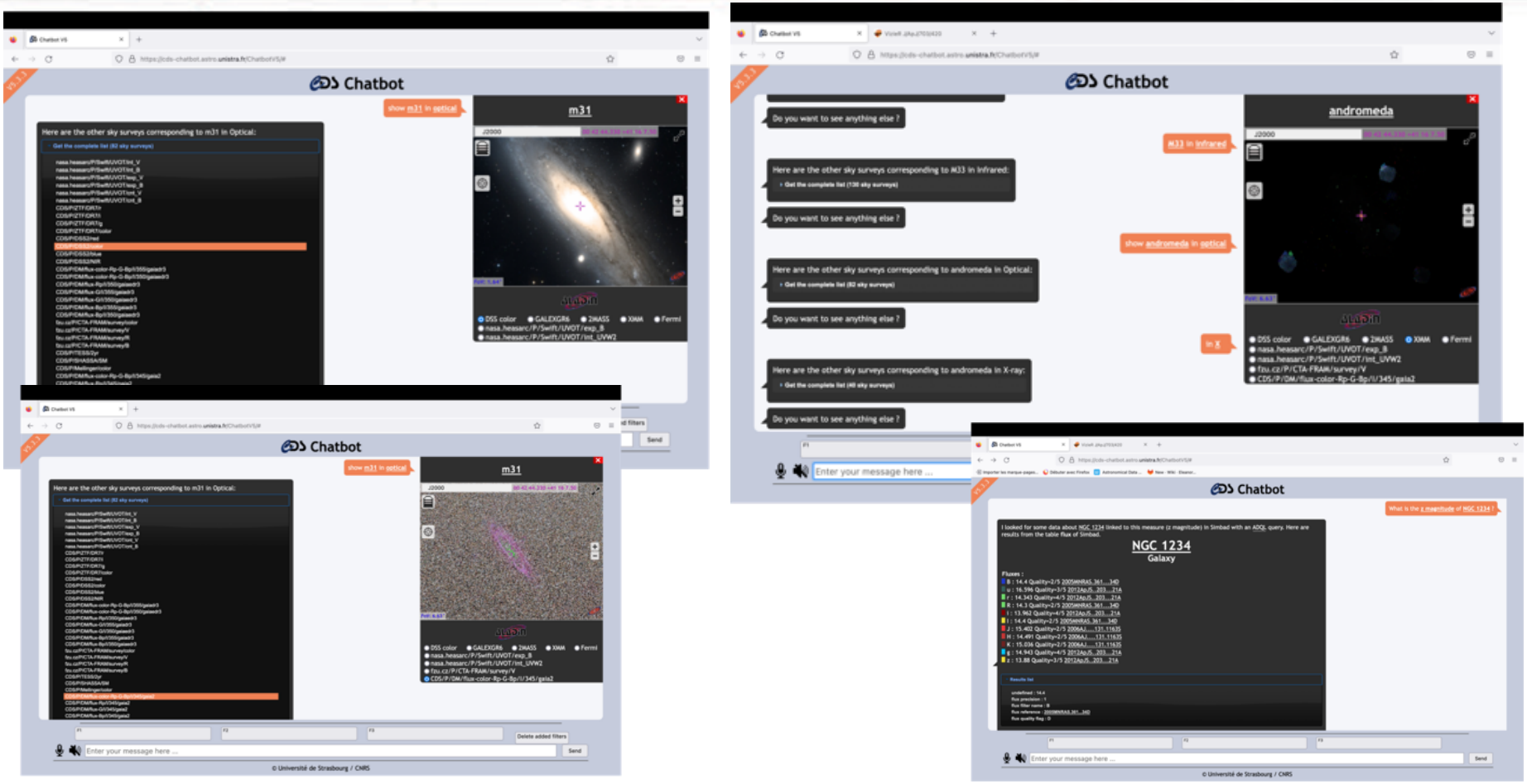
- A long-term work started in 2017, based on an Open source platform (RASA).
- Natural Language query translation to understandable queries by the CDS services.
- Several presentations at IVOA, ADASS, Astroinformatics.
- Time is coming to propose it on the new CDS Lab webpage for a larger testbed




**It benefits from authors in Simbad, missions and wavelengths in VizieR, DJIN to recognize identifiers in a text, UCDs, ADQL / TAP, the Sesame name resolver, Aladin Lite, => in-house and VO effort**


**A. Schaaff, T. Boch, S. Derriere. M. Brouty**
**Intern: Eléanore Renaud (UTBM)**

# Chatting with the services (2)

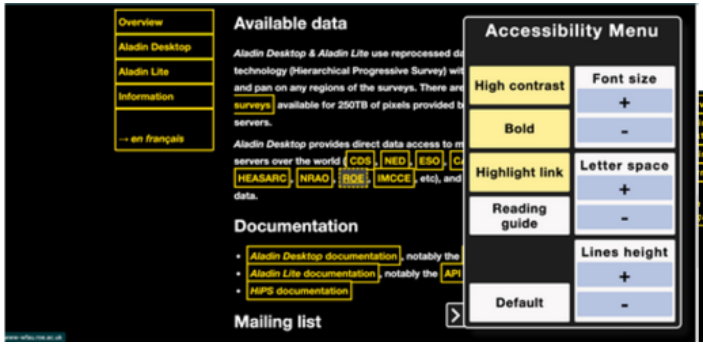



It is dedicated to CDS services but due to its IVOA protocols usage, it will be possible to query outside services in the future

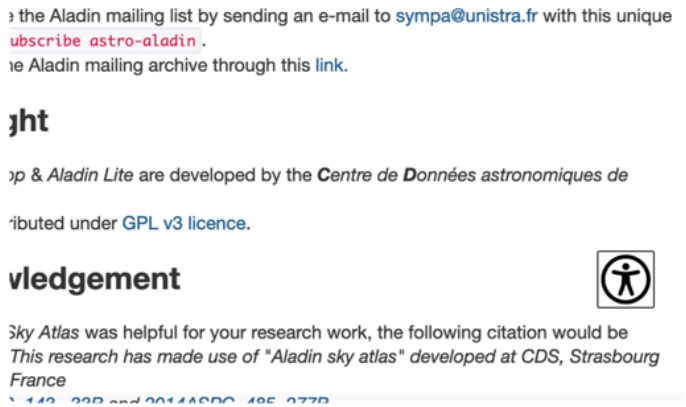
# Inclusive CDS

- This work was following a SARA 2020 topic “Getting everyone connected: software and hardware interfaces that make a difference for professionals and students with disabilities in on-line work; making **online accessible to** everyone around the World.”


Exemple for view accuracy disability








**A. Schaaff, G. Mantelet, T. Boch**
**Intern: Paul Devaux (IUT Charlemagne )**

# Various works @ SSC XMM

- Linked to VO, a development of the Web client for TAP service with a smart resource locator
- This tool is able to manage connections with multiple TAP nodes at the same time, it can also easily browse join tables

**L. Michel (SSC XMM)**
**Intern: Cyril Obrecht (UTBM)**

- Linked to VO too, a TAP service delivering XMM data
  - This service is able to map on the fly searched data on the MANGO VO data model (draft).
  - This capability, unique at the time of writing, has been used as a reference implementation for the mapping syntax standard.

**L. Michel (SSC XMM), M. Louys, F. Bonnarel**
**Intern: Ihsane Errami (IUT Schuman, Strasbourg)**

- In the theme of Code to data, a service able to run XMM-Newton analysis tasks on lists of photons detected inside a sky region drown on the screen.

**L. Michel, P. Maggi, G. Vasilopoulos (SSC XMM)**
**Intern: Alexandre Viala (UTBM)**

# Future investigations

- A continuous R&D effort to provide an added value to the data access & presentation mechanisms.
- A Lab part on CDS Website in 2023.
- Continuous efforts to make the data and services accessible to Everyone : Inclusive CDS

# Conclusion

- A coverage with various spin-offs:
  - improving the services, updating the staff skills
  - new (but not enough by itself !) technologies.

- A team work at the Observatory level: contracts, workstations, presentation of the services and the professions to integrate quickly the newcomers, etc.
- Also a contribution to IT student Education

# CDS Plans and challenges

December 7, 2022

Mark Allen - Director CDS

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

# CDS Strategy

- **Constantly evolving** — based on scientific needs
- **Main strategy** for core services is well established:
  - Pursue the CDS mission at the highest possible level
  - Science-driven Data Centre for “reference” data
  - Spirit of Open Science, and application of FAIR principles

- **Themes**

**i) Reinforcement of core mission** — trusted reference data centre
**ii) Enabling science with the CDS services** —supporting community use of CDS services, and development of the CDS science team
**iii) Engagement with the astronomy community**
**iv) Adaptation and innovation** — responding to science needs and increasing volume
**v) Building on success of CDS by maintaining specialised staff profiles & teamwork**

# Service plans ~5 yr timescale

- **SIMBAD**
  - Exploit new hierarchy of object types in query services.
  - Cross-ids with successive releases of Gaia.
  - Strategies and tools to systematically cross-id large surveys.
  - Work toward SEDs based on cross-ids.
- **VizieR**
  - Towards a merge of normal and large catalogue pipelines.
  - Necessary re-writing of legacy code.
- **X-Match**
  - Next generation cross-matching service.
  - Continue to work on the role of the X-Match service for event brokers.
- **Aladin**
  - Exploitation of WebGL enabled Aladin Lite — next innovations.
  - Scientific services based on HiPS.
  - Use of Aladin in notebooks - ipyaladin.
  - Work on the long-term evolution of Aladin Desktop and Aladin Lite.

# Plans 2022-2023

**Continuation of core work of building CDS content and operating services.**

**Re-engage with close partners** (after ~3 yrs without in-person meetings)

- **SAO/ADS** — use of CDS generated object-reference mapping.
  — indexing of VizieR catalogues.
  — common interests in supporting Open Science.
- **A&A** — processing of A&A papers for SIMBAD and VizieR.
- **ESO** — use of CDS systems in ESO operations.
  — VO evolution and next steps after ESCAPE.
- **ESA** — use of CDS systems in e.g. ESASky and ESA Datalabs
- **CNES** — CDS systems to support use of space mission data.

# Plans 2022-2023

- **CDS Service Developments.** *(Mostly from yesterday's presentations)*
  - SIMBAD distance project — add objects to SIMBAD with well known distances: Stars and Galaxies.
  - Dictionary renewal continues.
  - Continued improvements of Documentalist tools.
    - Manage with changes in publisher systems.
    - Respond to documentalist needs - e.g. tracking progress.
  - VizieR renewal - see next slides.
  - Aladin Lite v3 - support use by astronomers and projects.
  - Progress on the new X-Match prototype.
- **R&D for CDS services in new platforms and 'Virtual Research Environments'** - discuss with ESA, SKA, etc. , EC projects?

# Outlook – 2023 and (a bit) beyond

- **Top Priority = continued support of FAIR data publication from major journals and data producers through variety of access modes**
- => continued support and development of in-house tools to assist catalogue indexation and ingestion
- ~Short term emergency: Associated data workflow
  - SAADA lead dev retiring
  - => requires new strategy:
    - SAADA code assimilation?
    - Rewrite lower level workflow?
    - Opportunity to taylor to VizieR needs
- Med to long term goals:
  - Code renewal (Very high priority, ongoing)
  - Consolidation of VizieR team with new staff

FNSO proposal

# VizieR renewal

**Bilan**

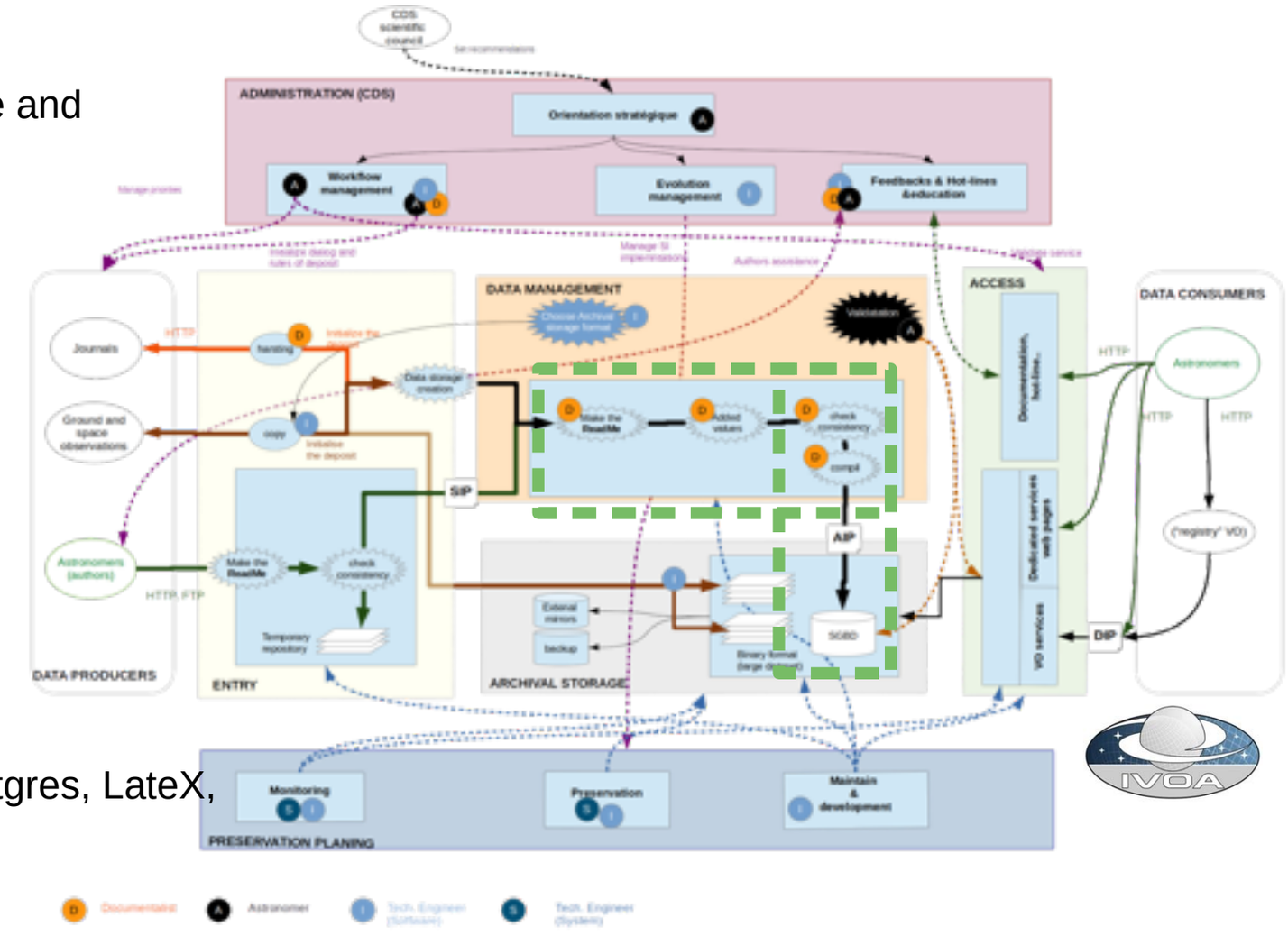
- 30 years  (1992) – success !
- But also: difficulties of maintenance and evolution
- Tech: C, SHELL, AWK, Python, Postgres/Sybase, LateX
- >100K lines codes (ingestion code)

**The new version**

- Conserve the knowledge
- Meet the evolution requirements
- Adapt the technology
- Tech: C/C++, SHELL, Python, Postgres, LateX, ANTLR, ElasticSearch, Rust

Example of updates:

- Global indexation(MOC)
- FAIR Metadata (Licences, Data origin)
- VO compatibility: eg: links  (Datalink)




The VizieR Information System - OAIS schema

# Important data sets in 2023

**Maintain approach of being as complete as possible with main astronomy journals.**

**What is the "reference data" from future large missions/ surveys that should be made available at CDS ?**


| Telescope/instrument | Number of records | ETA | manageable at CDS? (tech-wise) | Band/depth |
|---|---|---|---|---|
| Gaia all epochs | 10^9x100epochs~1e11 | 2022-2024 | no | |
| Gaia flat | 1.7x10^9 | available in VizieR | yes | |
| Pan-STARRS DR1-2 all epochs | 2.10^9x10epochs~2^10 | done | ~yes? | |
| Pan-STARRS DR1-2 flat | 2.10^9 | DR1 in VizieR | yes | optical wide |
| ESO Phase III cats | 1-2x10^9 | ongoing | yes | optical/NIR wide |
| LSST/Rubin all epochs | 4^10x[~250epochs] = 10^13 | 2025-2030 | no | optical wide |
| LSST flat | 40x10^9 | 2025-2030 | yes | optical wide |
| EUCLID | ~10^10 galaxies | 2028 and beyond | yes | optical/NIR wide |
| JWST | deep extragal surveys, narrow FoV, high res | 2022 and beyond | yes | NIR deep |
| Nancy Grace RST | a few 10^9 stars, 4x10^8 galaxies | launch 2027 | yes | orange / NIR deep/medium wide |
| ELT | ~narrow FoV, very high res (0.005") | 2027 | yes | optical/NIR deep |
| Giant Magellan Telescope | deep extragal surveys, narrow FoV, very high res (0.01") | 2029 | yes | optical/NIR deep |
| TMT | deep extragal surveys, narrow FoV, very high res (0.015") | 2027 | yes | NUV/optical/mid-IR |
| 21cm / SKA | up to 10^9 gals | 2030-ish? | yes | radio wide deep |
| GW: LIGO/PTA/LISA | ? | operating and upcoming | yes | grav. wave |
| CTAO | ? | 2024 | yes | gamma |
| KM3NET | several 10^5 - 10^6 | | yes | neutrino |
| ATHENA | ? | | yes | X-ray |
| SVOM | ? | | yes | gamma |

**2023:**

- DES: DESI DR8, DECALS DR9
- NOAO Source Catalogue
- KIDS DR4
- ESO catalogues
- +++

# Plans 2022-2023

## CDS infrastructure

- Get through the winter period — with possible electricity cuts.
- Push ahead to have operational duplication of CDS services.
  - Involves installation of new firewalls and a hyperviser.
- Finalise plans for moving servers into IPHC.
- Continue to move servers as planned into University Data Centre.
  - Depends on readiness of the Data Centre.
- Plan for major upgrades: All-Sky-Data ~ 5PB in ~2025.

# Local machine rooms evolution

- **Retirement** of CDS installations in the **local server room on the ObAS** site
  - Strong encouragement from the University to retire this room (corresponding to University and national level policies of rationalization/reduction of small server rooms)
  - Loss of the leader of the ObAS logistics team
  - Technical limitations of the server room
- **Our solution:** **Partnership with the IPHC** (UMR7178) on the CNRS Cronenbourg campus.
  => Technical and functional conditions appear a priori to be adapted to our desired level of operations
- **University Data Centre will become the main local site** for our installations.

*We aim to begin relocation of servers in the first half of 2023*

# Evolution detail of IS components

- Renew / updated
- (to be) removed
- Ongoing
- Steady

CDS web site

old

HiPS

ALADIN

HiPS proc.backend


web

# Recherche Data Gouv - inaugurated July 2022

**New French initiative in the National Plan for Open Science**

— ***CDS is named as one of the Thematic Reference Centres***

- Define the international data description standards in their thematic field;
- Define and disseminate best practices for data collection, documentation, processing and dissemination in their thematic field;
- Support inclusion in and compliance with the international ecosystem.
- Repositories are to be ‘harvested’ by *Recherche Data Gouv;*


https://recherche.data.gouv.fr/en

# Projects

- **ESCAPE** - final phase now, conclusion in Jan 2023.
  - ESCAPE follow-on projects:
  - Proposals for EC calls - consortia forming now.
  - Expect to make use of ***Cascading Grants*** of a big cluster project.

- **Europlanet 2024** Research Infrastructure (EPN-2024-RI)
  - Small but important CDS participation (~20PM).

- **XMM2ATHENA (***started 2021)*
  - A. Nebot is the ObAS coordinator. Postdoc to be recruited
  - ~5PM participation of CDS related to X-Matching

- **EOSC Future** - large (~40 M€) project of science and e-Infrastructures
  - Started April 2021, ends August 2023.
  - CDS is small part for Test Science Cases and training activities. New project engineer (M. Marchand) making fast progress!

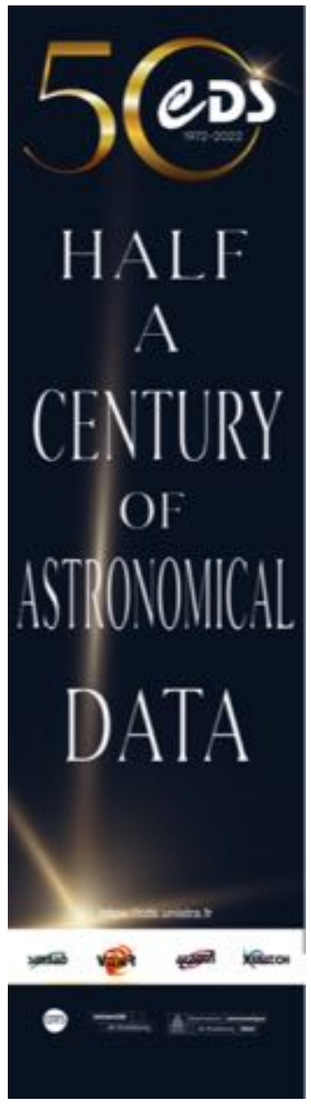
# Events for 2023

## CDS 50th anniversary … continuing into 2023 :-)

- **American Astronomical Society (AAS) meeting**
  - 8-12 January, Seattle.
  - CDS-50 theme for exhibit booth (4 CDS staff).
  - CDS interaction with US community.
  - Aladin demonstrations - Aladin Lite v3, Aladin Desk. V12, HiPS network.
  - Gaia@CDS services (in particular with Berkeley prize)
  - Re-connections: AAS publishing, SAO/ADS, NED, NASA, Rubin/LSST.

- **SF2A 2022** - Hosted by Observatoire astronomique de Strasbourg.
  - 20-23 June 2023.
  - Proposal in preparation for an Open Science session.

- **European Astronomical Society annual meeting**
  - Krakow, Poland, 10-14 July 2023.
  - CDS Booth in exhibit hall
  - Proposed Special session ( decisions before end 2022)
    - Science with the Virtual Observatory: status, success cases, and the future
  - EOSC training at the booth (?)




# Events for 2023.

- **IVOA**
  - Interoperability meeting, Bologna, Italy, 8-12 May 2023
  - Large participation important - after 6 virtual meetings…
  - Planning for the future of VO activities:
    - (SKAO now a IVOA member, Rubin/LSST engaged - ‘VO first approach’).
    - European approach - defining what next after ESCAPE.

- **ADASS 2023** - Tucson, USA, 5-9 November 2023
  - Plus IVOA interoperability meeting afterwards.
  - CDS participation (with economic/environmental considerations).

- **Others:**
  - ASOV, RDA, EOSC, Science visits, Science conferences, Data Infrastructure conferences…
  - *Looking ahead - a Focus Session has been proposed by IVOA for the IAU GA, Cape Town, August 2024. Envisage having associated hands-on workshop.*

# Pursue the staffing replacements - discussion with INSU in January 2023.

- **Documentalist**
- Increasing number of references, and complexity drives need.
- Necessary **replacement** of Documentalist who departed in 2021.
- FSEP ‘mobility’ position open now with result known in Jan 2023.

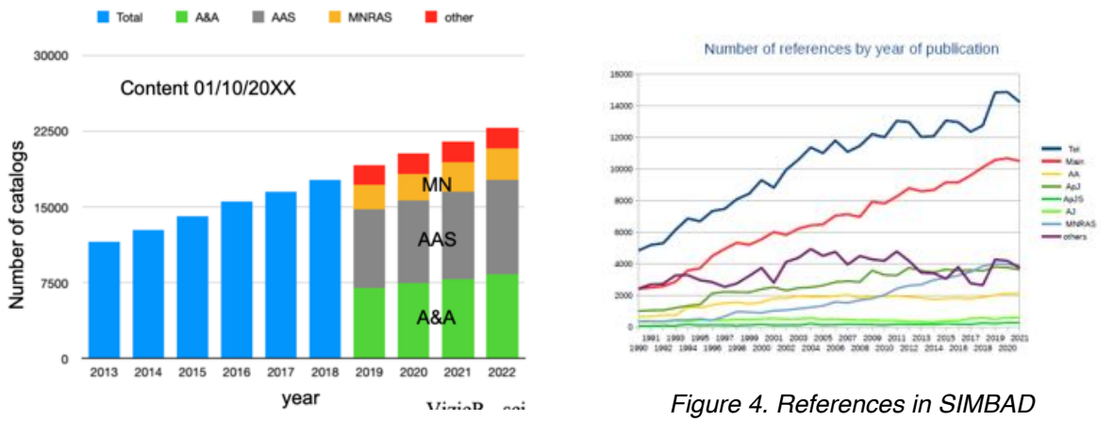



*Figure 4. References in SIMBAD 1990-2021*

- **Administrative Assistant**
- CDS needs admin staff with expertise on contracts.
- CDS cannot operate without it!
- **Replacement** of Admin Assistant who left in 2020.
- Temporary contractor left Oct 2021.

- **Aladin / visualisation engineer**
  - Unique opportunity to respond to new SKA mission.
  - Build on revolution of HiPS/MOC/Aladin.
  - Critical **replacement** of F. Bonnarel (2024).
  - SKA SRC work would depend on it.

- **VizieR engineer**
  - Renewal of Vizier ingestion process underway with contract engineer
  - 2 engineers per major CDS service are necessary for operations & development

Note: Expected **retirement** of another engineer in ~2025-2026. (Also **retirement** of another ObAS engineer (~2027) who developed SAADA which is used by CDS)

# Science staff

**CDS Scientists**

- Necessary expertise of active scientists covering wide range of astrophysics research areas.
- Necessary diversity of scientific profiles.
- ***Recruitment being sought via CNRS and CNAP competitions***
- *(4 candidates being supported for 2023)*

**Contribute to all scientific aspects of the CDS mission:**

- Selection and validation of scientific data.
- Providing expert scientific advice and setting the scientific strategy.
- Supervision of the scientific development and evolution  of the CDS services.
- Support the users (hot-line, dissemination, documentation…).

**Priority areas related to CDS service tasks:**

- *Scientific support of SIMBAD*
- *Catalogue, spectral, data cube and multi-dimensional data expertise*
- *Expertise for current and future very large data sets*
- *Data Science and Astro-statistical and expertise (machine/deep learning)*

# CDS services in context of education and public — e.g. new planetarium in 2023


# Summary

- **An exciting new year coming up after the challenges of 2022:**
  - CDS operations in a more certain environment (let's hope).
  - Re-connecting with partners in-person.
  - New proposals: EC, FNSO. Also CNES APR.
  - Pursue engagements with projects - in particular SKA.
  - CDS to establish functional activities connected with Recherche Data Gouv.
  - Defining the next phases for VO interoperability (and connections to EOSC and other Open Science infrastructure).

- **Challenges:**
  - Staffing replacements.
  - Strengthen the CDS scientific team — and how we describe it.
  - Positioning of CDS services in context of large projects and science platforms.
  - All-Sky-Data renewal — find a funding mechanism.
  - Ensure smooth transitions for changes of infrastructure - server rooms.

# Open Science in Astronomy - workshop in SF2A

The themes of the workshop will be :

- The use of Open Science tools and services for astronomy research
- Current and future reference data sets in astronomy
- Plans and challenges for data services in astronomy
- Innovations to support scientific discovery with big a rich astronomy data sets
- The FAIR principles in astronomy and beyond.
  - FAIR – Findable, Accessible, Interoperable and Re-usable