

# SIMBAD improvements



---

CDS Council  
October 2018



C. Loup,  
A. Oberto,  
S. Lesteven

# Bibliographical Team

Soizick Lesteven / Cécile Loup

## References

Year, journal, volume, pages,  
title, authors, abstract, keyword,  
bibcode, copyright, DOI

From publishers to SIMBAD

**Soizick Lesteven**  
**Magali Neuville**



## Dictionary of Nomenclature

Dictionary update  
(analysis, creation and update  
of acronyms)  
Update links to NED database

**Marianne Brouty**  
**Fabienne Woelfel**

## HiPS

Aladin update  
Creation and description  
of HiPS

**Mihaela Buga**



## SIMBAD update

**Extraction of information from articles**  
Objects, identifiers, fundamental data,  
measurements, references, .....

**Aline Eisele** : A&A, PASJ, PASP, ...  
**Magali Neuville** : ApJ, ApJS, Sci, Natur, ...  
**Evelyne Son** : AJ, MNRAS  
**Philippe Vonflie** : MNRAS

## Tables

### VizieR

Standardization,  
description of tables, and  
data valorization.  
VizieR update

**Marianne Brouty** : ApJ  
**Emmanuelle Perret**: ApJS, ApJ  
**Tiphaine Pouvreau** : ApJ  
**Patricia Vannier**: A&A, MNRAS,...



=g=

## SIMBAD update

Via electronic tables  
(selections, extractions, scripts,  
cross-identifications)

**Catherine Brunet**  
**Mihaela Buga**  
**Esther Collas**  
**Emmanuelle Perret**  
**Fabienne Woelfel**

## Scientific Expertise

SIMBAD : **C. Loup, C. Bot, L. Cambrésy, A. Nebot, P. Ocvirk, A. Siebert, B. Vollmer**

Dictionary : **B. Vollmer**

VizieR : **P. Ocvirk, C. Bot, S. Derrière**

## IT Development

SIMBAD : **A. Oberto, T. Boch, S. Lesteven, G. Mantelet**

Dictionary : **A. Oberto**

VizieR : **G. Landais, T. Boch, F.-X. Pineau**

Aladin : **P. Fernique, T. Boch**

# □ Gaia DR2 in SIMBAD

Expected by many users even just after the release !

Available for users end of June as announced

**A successful team work. Operation anticipated more than 10 years ago**

## Criteria :

- Only objects with sub-arcsecond astrometry : 6.5 millions
- Objects with insufficient astrometry completely ignored
- No neighbour in SIMBAD within 3'', = resolution of 2MASS (310,000 discarded)
- No neighbour in Gaia DR2 within 3'' and  $\Delta G < 3$  mag (270,000 discarded)
- Distance between SIMBAD and Gaia DR2 positions  $< 1''$

## Results : 4.5 millions objects have been upgraded

- 4.5 million coordinates, Gaia DR2 identifiers, G magnitudes
- 3.8 million parallaxes ( $>0$ ) and proper motions
- 2.0 million radial velocities

	<b>Stars</b>	<b>Galaxies</b>
Total	5.0	3.3
Astrom. $< 1''$	4.6	2.2
<b>Gaia DR2</b>	<b>4.0</b>	<b>0.5</b>



# □ Gaia DR2 in SIMBAD : next

## Objects unrecovered in Gaia DR2 :

- Close neighbours : mostly crowded regions (MCs, Galactic Bulge, ...)
- Too faint for Gaia (mostly galaxies, but also Cataclysmic, brown dwarfs, etc...)
- And of course objects with insufficient astrometry

## How to improve ?

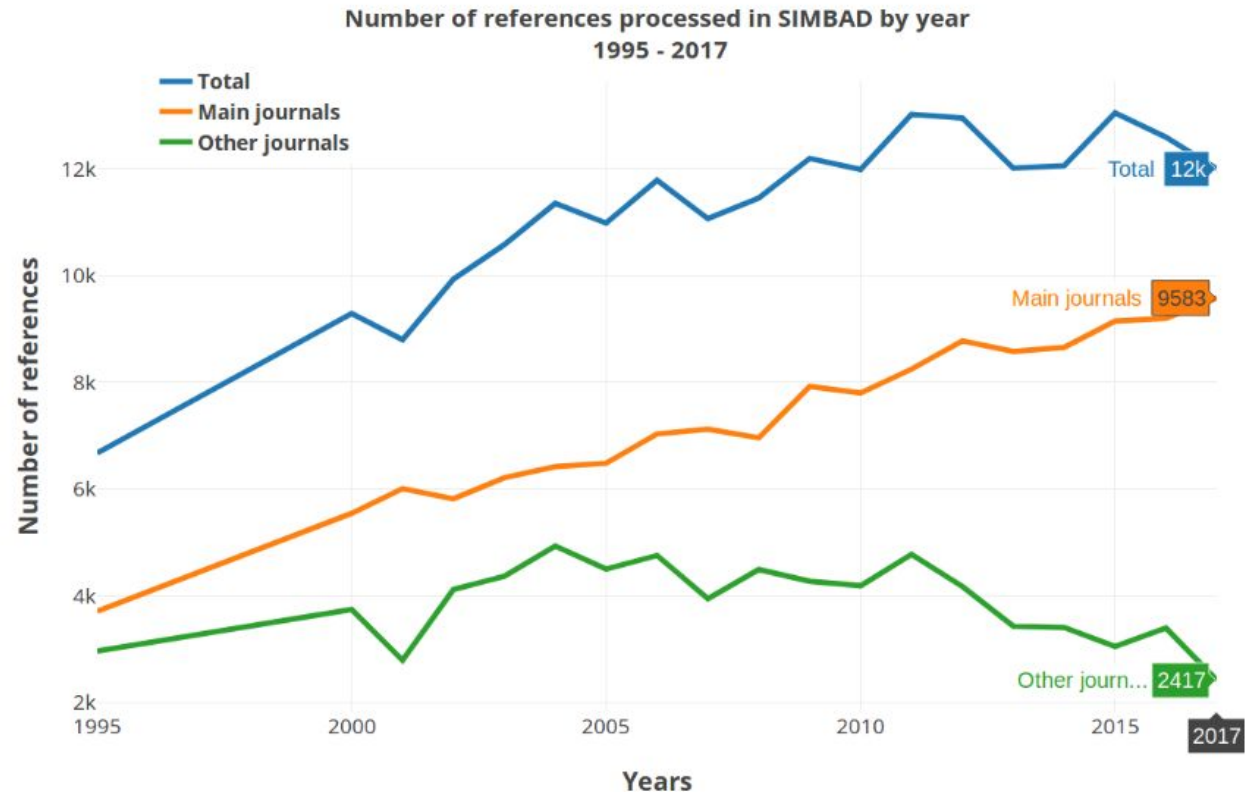
- Release the neighbour criteria to 1.5'' for objects without 2MASS identification (especially OGLE, SDSS, etc...)
- Special operations for objects with astrometric accuracy of 1-10 '' (e.g. CRTS) : time consuming

## High proper motion stars ( $PM > 50$ mas/yr) : 325,000

- 85% recovered in Gaia DR2 with the massive Xid
- However only 50% for those with  $PM > 500$  mas/yr → all done
- Dedicated operation → only 7% missing now
- Next step : include all new HPM stars from Gaia DR2

# □ Bibliography

- 12,000 references processed in SIMBAD in 2017
- The quantity of articles in the main journals are still growing (MNRAS).
- All efforts are done to keep the quality of the process (expertise work for the documentalists)
- To be fully processed, an article can be analysed up to 4 times (DJIN, VizieR, COSIM, Dictionary)

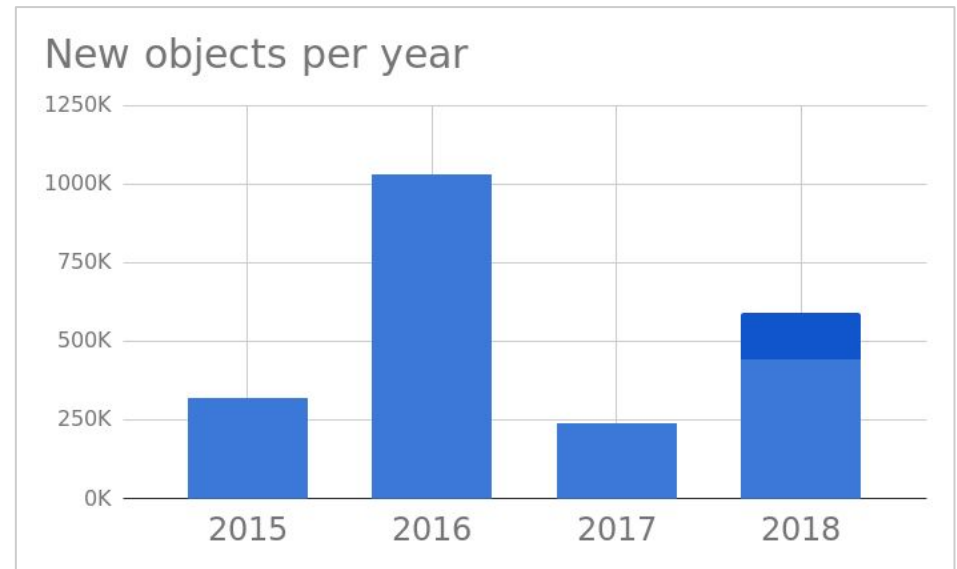
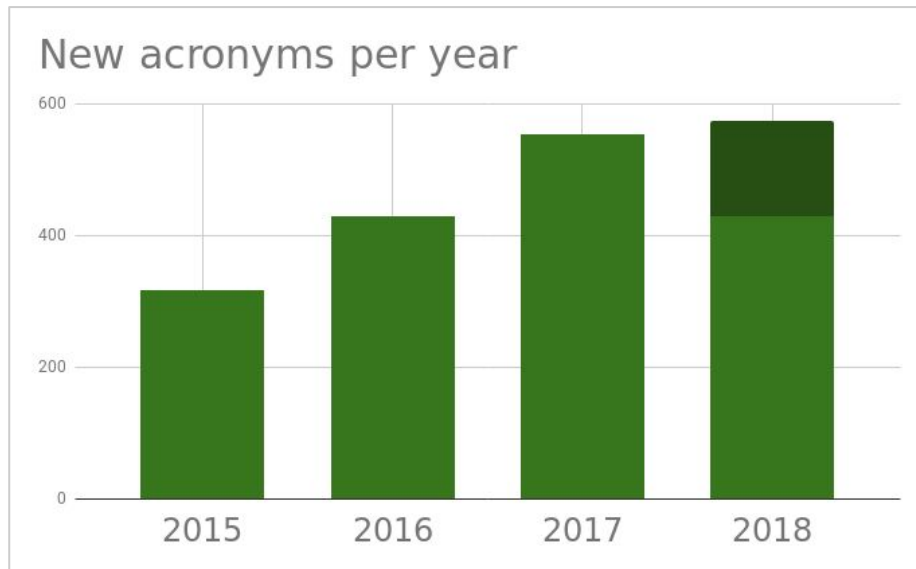


# □ Bibliography

- The new «bibliographical data entry» tools are now operational.
- The full bibliographical data are now in SIMBAD, which avoids the duplication of data. They are accessible and correctable in real time by the full team.

# □ SIMBAD content growth

- +5.7 million identifiers (Gaia) (triple in 10 years)
- Acronyms: 13800 => +50/month
- Objects: 9.6M => +28,600 /month (double in 10 years)



# □ Objects in Tables

The number of tables is growing, and the number of objects in tables is growing too.

**Tables with priority 1** are even growing more :

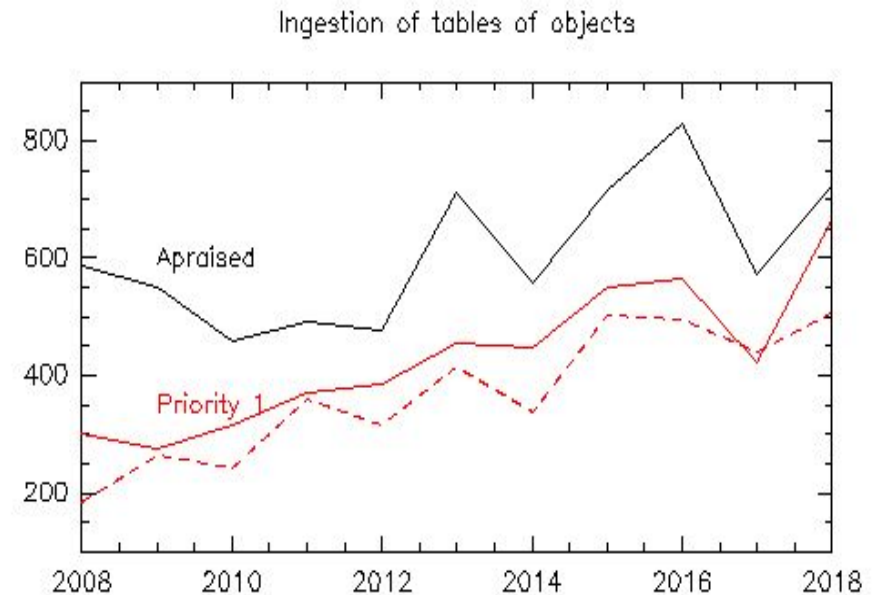
- More articles with spectroscopy
- More articles with membership

**Backlog for appraisal :**

- About 500 in 2017
- About 800 tables, finishing 2016
- More from publications
- More from acronyms & Vizier
- → pre-appraisal by B. Vollmer & C. Loup

**Ingestion :**

- Backlog negligible, but will increase with the pre-appraisal
- Improvement of COSIM efficiency since the Gaia Xid, thanks to a better astrometry





# □ Some numbers of objects

In the last years the number of objects has amazingly increased. Even for some classes of “rare” objects the numbers are getting impressive :

A few examples :

**AGNs & QSOs** : **402,000**, soon close to 600,000 (+ QSOs SDSS DR12)

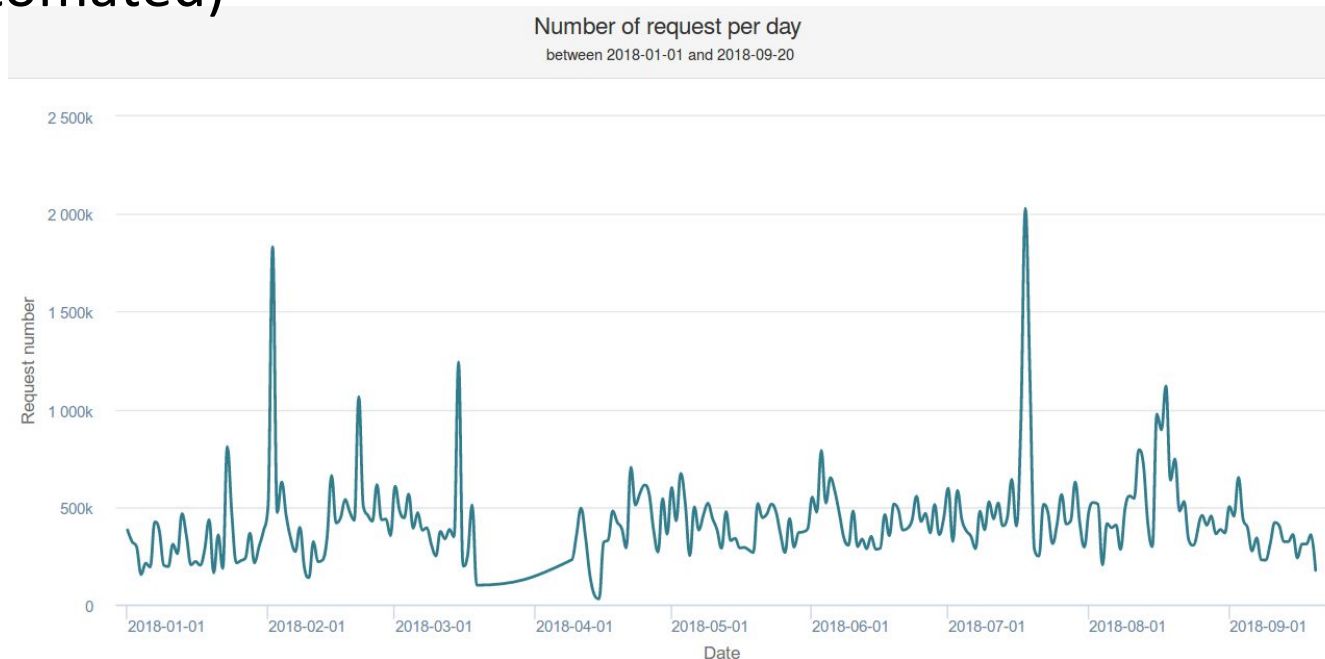
**Variable stars** : **692,000** (89,000 in the GCVS)

**Eclipsing binaries** : **105,000**, soon close to 500,000 (+ OGLE Bulge)

**White dwarfs** : **39,000** (14,300 in the WD), compared to 5,000 15 years ago

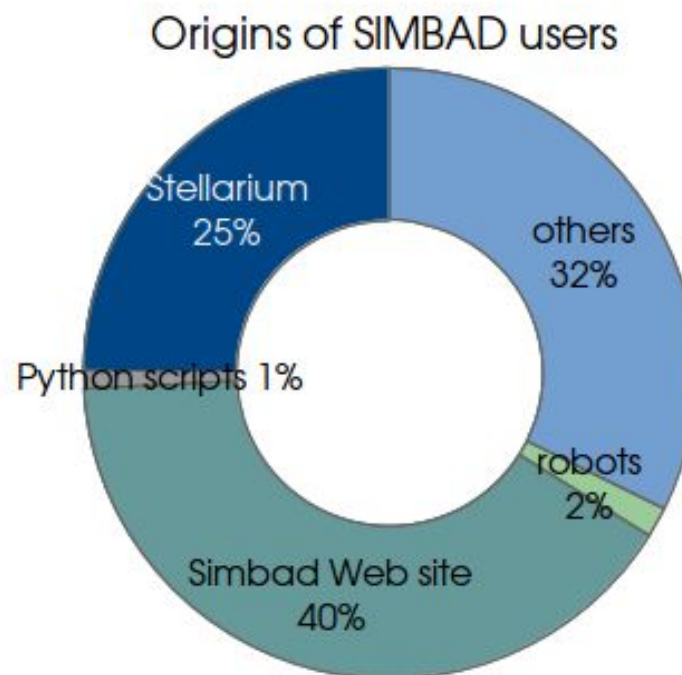
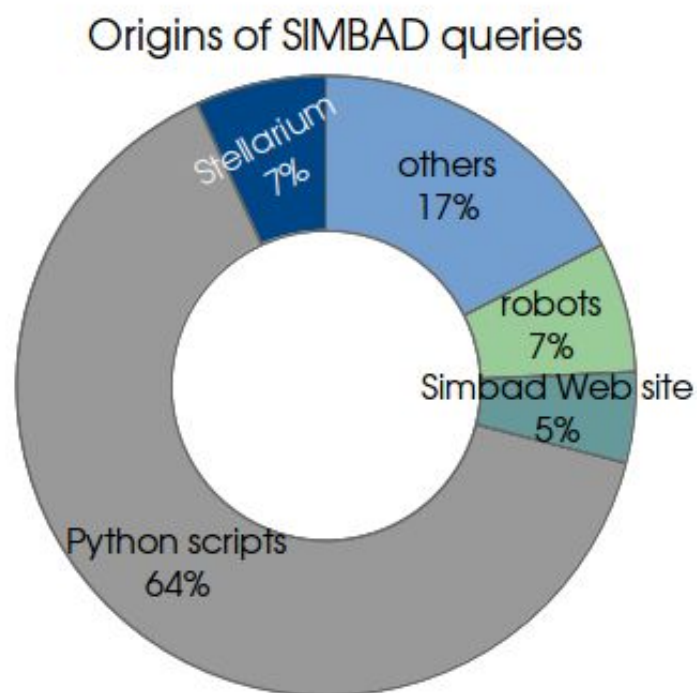
# □ SIMBAD Usage (12 last months)

- $\pm 500,000$  queries and 5GB per day
- $\pm 160,000$  monthly unique visitors
- 2.3% queries are blacklisted (overload)
- 6% of visitors are human users  
(versus scripts/automated)



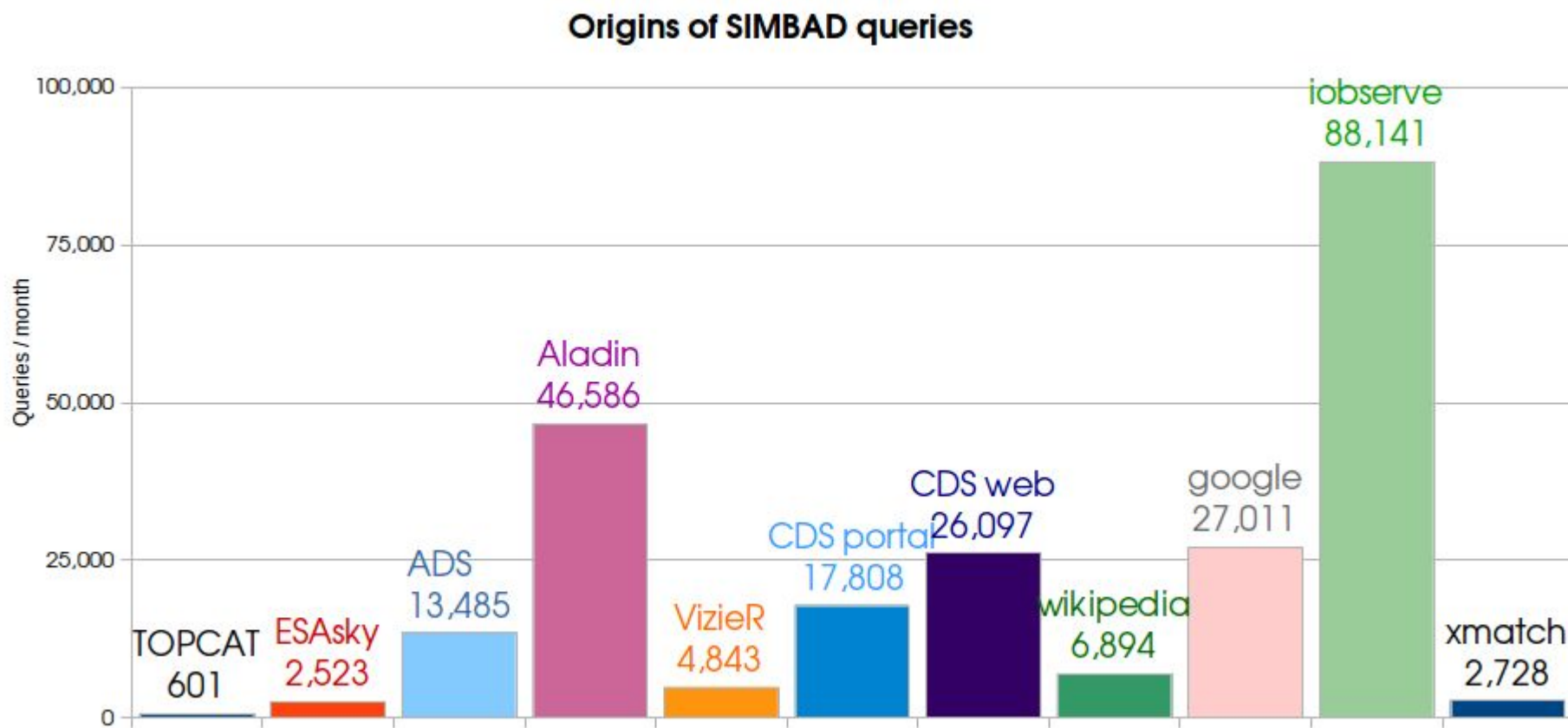
# □ SIMBAD Usage (3 last months)

- A large activity is coming from 1% of users through python scripts.
- Most of people are **SIMBAD users** or **Stellarium's**.



# □ SIMBAD Usage (3 last months)

- Other queries are coming from: (note that Xmatch/TOPCAT are mainly for long list)





Document Settings

BIBCODE/DOI  
< 2016A&A...596A..29M >

☰ 🔍 ✓ 📏 ↔

MENTIONS (20)

- > ? 2MASS J19560294-3207
- > ? 2MASS J20100002-2807
- > ? HIP 102141B (1)
- > ? HIP 11437A (3)
- > ? HIP 11437B (1)
- > ? J05335981-0221325 (1)
- > ? J06131330-2742054 (1)
- > ? J19102820-2319486 (1)
- > ? J19233820-4606316 (1)
- > ? J20013718-3313139 (1)
- > ? J20055640-3216591 (1)
- > ? J20333759-2556521 (1)
- > ? J21100535-1919573 (1)
- > ? NGC 2264 (3)
- > ? TWA 22 (1)
- > ... TX Psa (1)
- > ? TYC 1186- 0706-1 (1)
- > ? TYC 1208- 0468-1 (1)
- > ? TYC 6878 0195 1 (2)
- > ? TYC 9486- 927-1 (1)

**Fig. 2.** Top panel: distribution of Li EW versus  $V - K_s$  color in three color ranges:  $0.5 < V - K_s < 3.4$  mag (19 stars),  $3.4 < V - K_s < 5.9$  mag (21 stars), and  $5.4 < V - K_s < 5.9$  mag (6 stars) of  $\beta$  Pictoris members with known rotation period and with overplotted (s) fits. Middle panel: residuals from the fit with horizontal dotted lines indicating the amplitude of dispersion. Bottom panel: EW residuals versus rotation period in the three color ranges. Solid lines are linear fits,  $r$  is the Pearson linear correlation significance level.

members (see the legend in Fig. 1). Detailed information for each member is given in Paper I and the criterion to distinguish close from wide components is given in Messina et al. (Paper III)

TYC 1186-0706-1	00	23	34.66	+20	14
GJ 2006A	00	27	50.23	-32	33
GJ 2006B	00	27	50.35	-32	33
2MASS J01112542+1526214A	01	11	25.42	+15	26
2MASS J01351393-0712517	01	35	13.93	-07	12
TYC 1208-0468-1	01	37	39.42	+18	35
HIP 10679	02	17	24.74	+28	44
HIP 10680	02	17	25.28	+28	44
HIP 11437A	02	27	29.25	+30	58
HIP 11437B	02	27	28.05	+30	58
HIP 12545	02	41	25.90	+05	59
GJ 3305	04	37	37.30	-02	29
2MASS J04435686+3723033	04	43	56.87	+37	23
HIP 23200	04	59	34.83	+01	47
HIP 23309	05	00	47.10	-57	15
HIP 23418A	05	01	58.80	+09	59
BD-211074A	05	06	49.90	-21	35
BD-211074B	05	06	49.90	-21	35

nary whose tidal locking between the two components may have altered the internal mixing and, consequently, enhanced the Li depletion rate with respect to single stars. The two other stars are TYC 6878 0195 1 and HIP 11437A, both components of wide binaries that are expected to have evolved, rotationally, as single stars. We only note that HIP 11437A hosts a debris disc.

#### 4. The rotation-Li depletion correlation

We first analyse the possible correlation between rotation and Li EW to investigate what role the rotation plays in producing the observed Li EW dispersion among stars with similar masses.

This information will be used for a more accurate LBD to infer the association age.

We analyse the following three sub-group stars with  $0.5 \leq V - K_s < 3.4$  mag totaling stars with  $3.4 \leq V - K_s \leq 4.5$  mag totaling very red stars with  $5.4 < V - K_s < 5.9$  mag that correspond to the onset of the Li de  $V - K_s > 5.9$  mag have not undergone any L blue stars sample, we observe a trend of increasing color (decreasing mass), we observe the opposite trend; in the red stars we observe a trend of increasing Li EW with increasing color (decreasing mass), as shown in the top panel of Fig. 2, linear fits purposes, i.e., approximating the mass dependence of the correlation with the rotation period. In the plot the residuals with respect to the linear fit (HIP 11437A), represented by a square and a red circle, were excluded from the fit indicating stars that are more Li rich than the negative residuals indicate stars that are more negative the average. In the case of blue stars, we find a plitude of the dispersion  $\sigma = 90$  mÅ of the fit, whereas in the case of red stars the plitude of the dispersion is about twice as large,  $\sigma = 125$  m, represent peak-to-peak amplitudes of dispersion on the rotation period, we computed

VALIDATED MENTIONS (20)

- > ✓ 2MASS J01112542
- > ✓ 2MASS J01351393
- > ✓ 2MASS J04435686
- > ✓ 2MASS J05082725
- > ✓ 2MASS J05241914
- > ✓ 2MASS J16430128
- > ✓ 2MASS J18151564
- > ✓ 2MASS J18202275
- > ✓ 2MASS J20434114
- > ✓ 2MASS J21103147
- > ✓ BD-21 1074A (1)
- > ✓ BD-21 1074B (1)
- > ✓ \* bet Pic (12)
- > ✓ GJ 2006 A (1)
- > ✓ GJ 2006 B (1)
- > ✓ GJ 3305 (1)
- > ✓ GSC 08350-01924
- > ✓ HD 173167 (1)
- > ✓ HD 191089 (1)
- > ✓ HIP 102141 (1)
- > ✓ HIP 102409 (1)
- > ✓ HIP 103311 (1)
- > ✓ HIP 10679 (1)
- > ✓ HIP 10680 (1)
- > ✓ HIP 112312 (1)
- > ✓ HIP 11437 (1)
- > ✓ HIP 12545 (1)

SELECTION DETAILS

MENTION (1)

State: Validated ✓

Position: d - Data/Table

Raw id: 2MASS J20434114-2433534

Source: Auto 1

SIMBAD OBJECT

Simbad Identifier: 2MASS J20434114-243353

QUICK ADD Occurrences

Positions: 0

Add command(s)

EM\* SthA 182 (\*\*) 24 refs

coo: 20 43 41.142 -24 33 53.19

pm: +62 -60

rvz: v:-6.1

plx: +35.1

sp: M4.1V+M3.7V

Identifiers:

- EM\* SthA 182
- 2MASS J20434114-2433534
- UCAC3 131-444990

SIMBAD SCRIPT

```

18
19 o 2MASS J05241914-1601153, +d, +2MASS
20 valid
21
22 o 2MASS J16430128-1754274, +d, +2MASS
23 valid
24
25 o 2MASS J18151564-4927472, +d, +2MASS
26 valid
27
28 o 2MASS J18202275-1011131, +d, +2MASS
29 valid
30
31 o 2MASS J20434114-2433534, +d, +2MASS
32 valid
33
34 o 2MASS J21103147-2710578, +d, +2MASS
35 valid
36
37 o BD-21 1074A, +d, +BD -211074A, =1
38 valid
39
40 o BD-21 1074B, +d, +BD -211074B, =1
41 valid
42
43 o GJ 2006 A, +d, +GJ 2006A, =1
44 valid
45
46 o GJ 2006 B, +d, +GJ 2006B, =1
47 valid
48
49 o GJ 3305, +d, +GJ3305, =1
50 valid
51
52 o GSC 08350-01924, +d, +GSC 08350- 01
53 valid
54
55 o HD 173167, +d, +HD 173167, =1
56 valid
57
58 o HD 191089, +d, +HD 191089, =1
59 valid
60
61 o HIP 560, +d, +HIP 560, =1
62 valid
63
64 o HIP 10679, +d, +HIP 10679, =1
65 valid
66
67

```

Simulate Execute

# □ DJIN2

- Not enough efficient to be used in production
  - Learned some hard lessons from this development (technology choices that didn't work well)
  - Needs to be fully adapted to documentalists' process
  - **New revision** of the project and restart some parts



# COSIM

- New graphical interface in development to improve speed and interactivity

The screenshot displays the COSIM software interface. On the left, there are several control panels for search parameters and visualization options. The main window shows a grayscale astronomical image of a star field with a red circle highlighting a specific region. A data table is visible at the bottom of the interface.

**Search Parameters:**

- Fichier d'entrée: parfile-tests, 1850CDT..1784..227M
- Le rayon de recherche: .SEARCH 60, .SEARCH COO, .SEARCH NOCOO
- Le type d'objet: .OT +1 EB\*, .OT -1 \*IC, .OT +/- n, .OT +/- n .TRUST OT .IGNORE OT
- Les coordonnées: .SIGMA COO, .COO (bornes), .TRUST COO, .IGNORE COO, .REPLACE EQUAL COO, .REPLACE NO COO
- Le score ACRO: .ACRO +/- n, .ACRO (bornes), .TRUST ACRO

**Visualization Options:**

- Options: Visualisation
- Base image layer: DSS2 Red (F+R), Color map: nativ, Reverse
- Overlay layers: overlay\_1, catalog, VizierR:1850CDT..1784..227M, overlay, Simbad
- Tools: Export view as PNG

**Data Table:**

MAIN_ID	OTYPE	RA	DEC	COO_ERR_MAJA	COO_ERR_MINA	COO_ERR_ANGLE	PNRA	PMDEC	B	V	R	J	H	K	SP_TYPE	GALDIM_MAJA
CL* NGC 2422 PMS 389	*inCL	07 37 16.83432	-14 31 29.7658	0.0274	0.0262	90	-7.13	1.012	16.04	15.1	13.107	12.615	12.478			

# □ Server infrastructure (with G. Mantelet)

- Many SIMBAD clients, many access protocols: Web, Python, Stellarium, Aladin, VizieR → need access **uniformisation** using SQL-like (TAP) queries:
  - More stable
  - Normalised by the VO
  - More and more data are downloaded (in 2018 : 4 million rows)



# □ Server infrastructure (with G. Mantelet)

- Work in progress to extract **re-usable modules** from the core of SIMBAD
- New “docker” system for database system ready:
  - Easy install
  - Reproducible
  - Available for recovery plan