SIMBAD improvements

CDS Council
October 2018

C. Loup,
A. Oberto,
S. Lesteven

Centre de Données
Astronomiques de Strasbourg

Bibliographical Team

Soizick Lesteven / Cécile Loup

References

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

From publishers to SIMBAD

Soizick Lesteven
Magali Neuville

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
PhilippeVonflie : 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

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

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**

April 2017

Gaia DR2 in SIMBAD

Expected by many users even immediatly after the release
Available for users end of June as announced
A successful team work, 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 deltaG < 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

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

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 ?

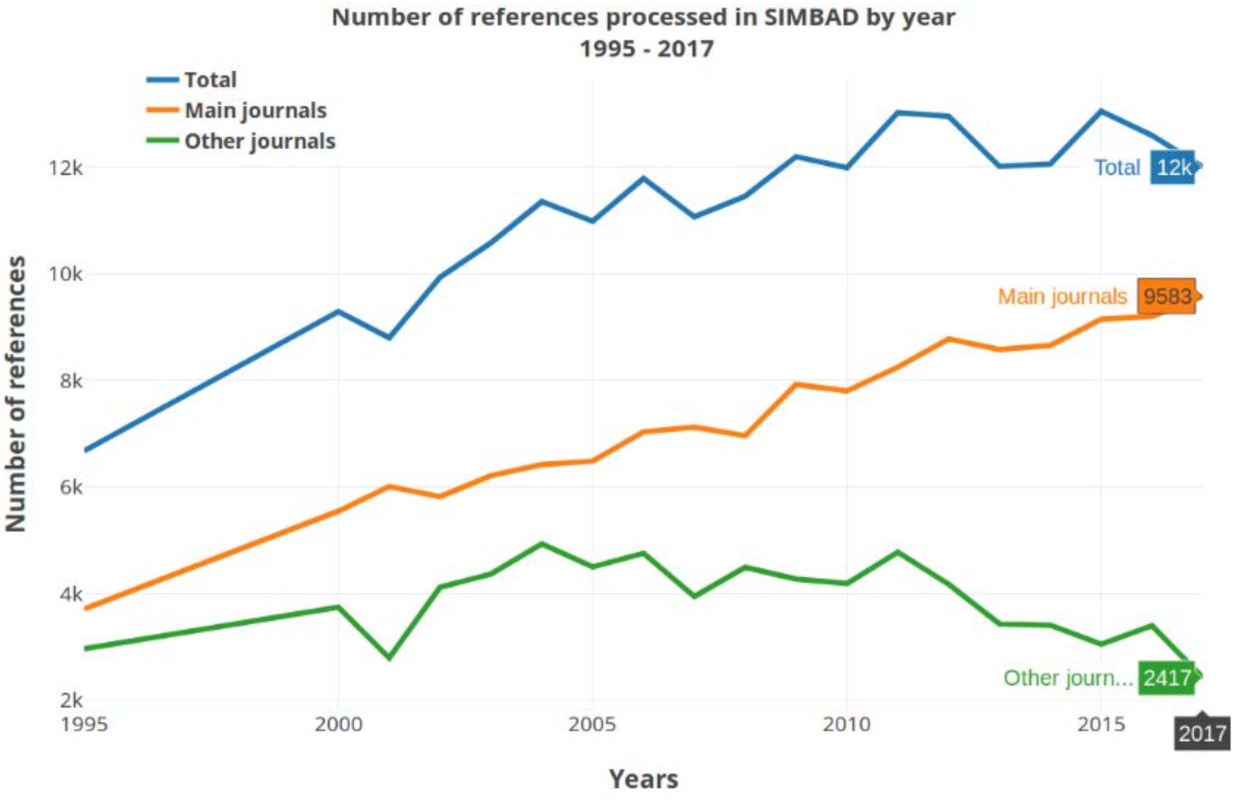
- Release the neighbour criteria to 1.5'' for objects without 2MASS identification (especially OGLE, SDSS, etc…)
- Special operations for objects with astrometric accuracy of 1-10 '' (e.g. CRTS) : time consuming

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
- Dedicated operation → only 7% missing now; 1400 new common PM/plx pairs
- 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)

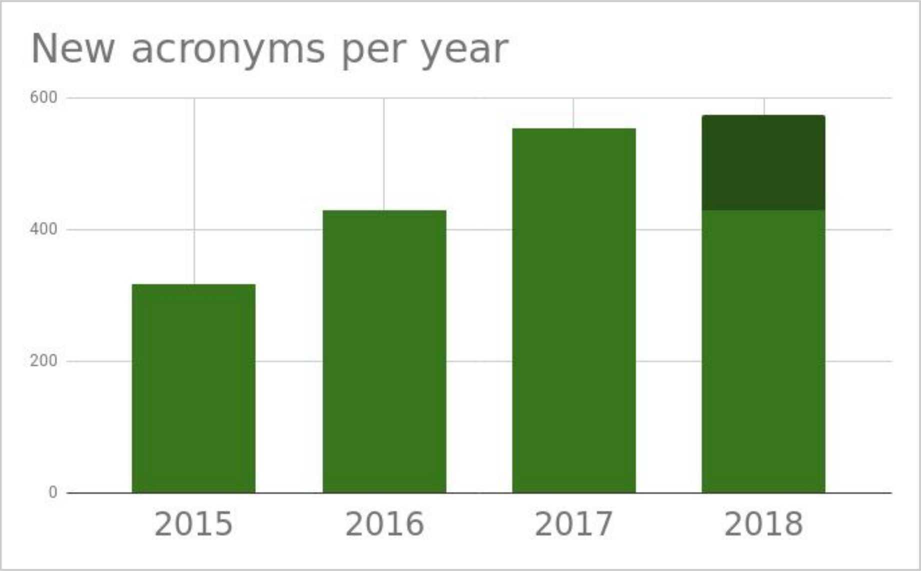


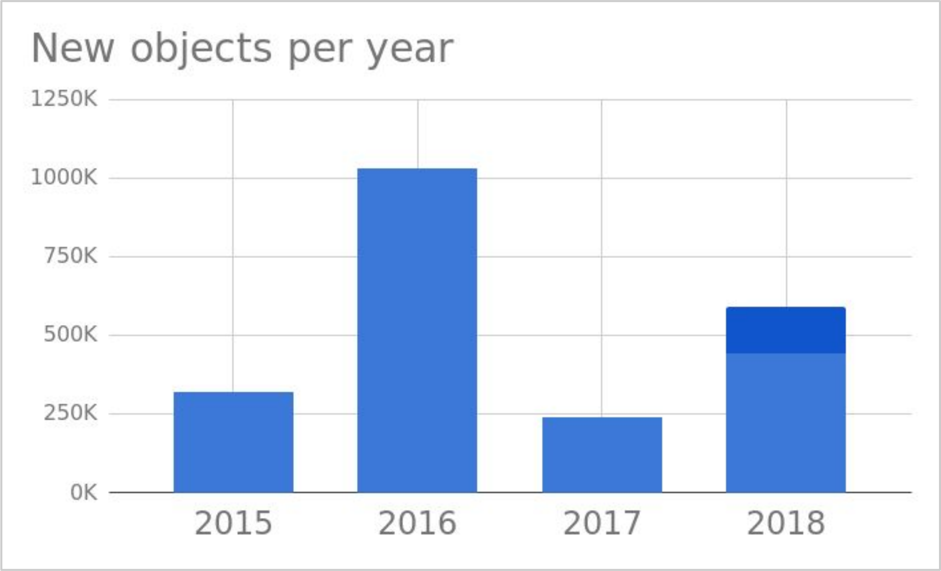
Bibliography

- 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)**
- +4% since last year:
  - Acronyms: 13800 => +50/month
  - Objects: 9.6M => +28,600 /month **(double in 10 years)**





Objects in Tables

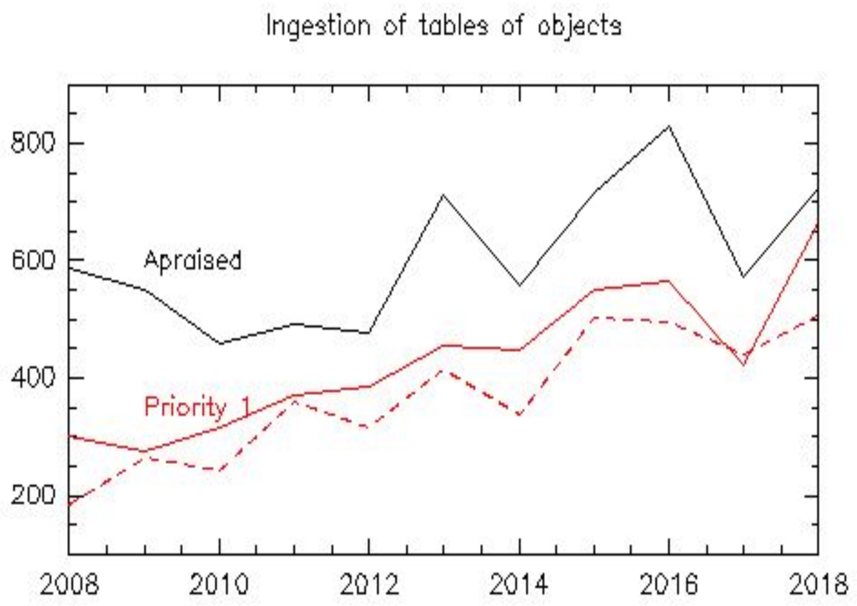
The number of tables is growing, and the number of objects in tables is growing too.

Tables with priority 1 are even growing more :

- More articles with spectroscopy
- More articles with membership

Backlog for apraisal :

- About 400 in 2017
- About 800 tables, finishing 2016
- More from publications
- More from acronyms & VizieR
- → pre-apraisal by B. Vollmer & C. Loup



Ingestion :

- Backlog negligible, but will increase with the pre-apraisal
- Improvement of COSIM efficiency since 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 huge :

A few examples :

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

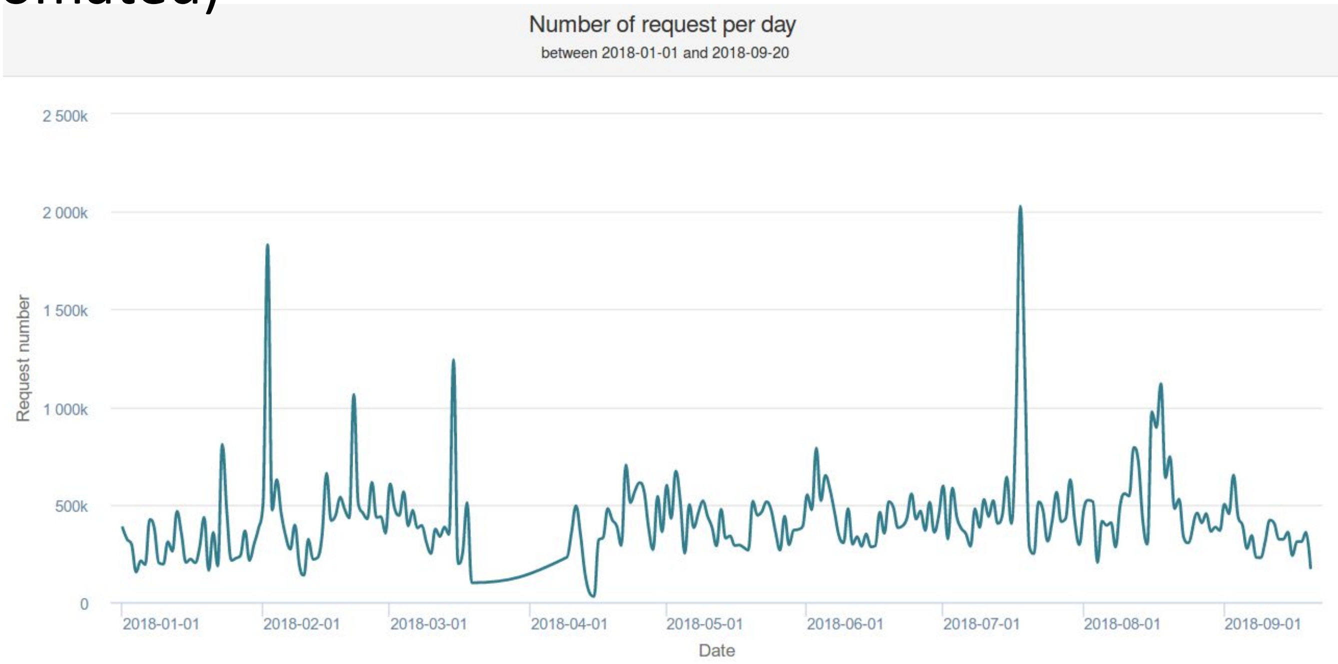
Variable stars : 692,000 (89,000 in the GCVS)

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

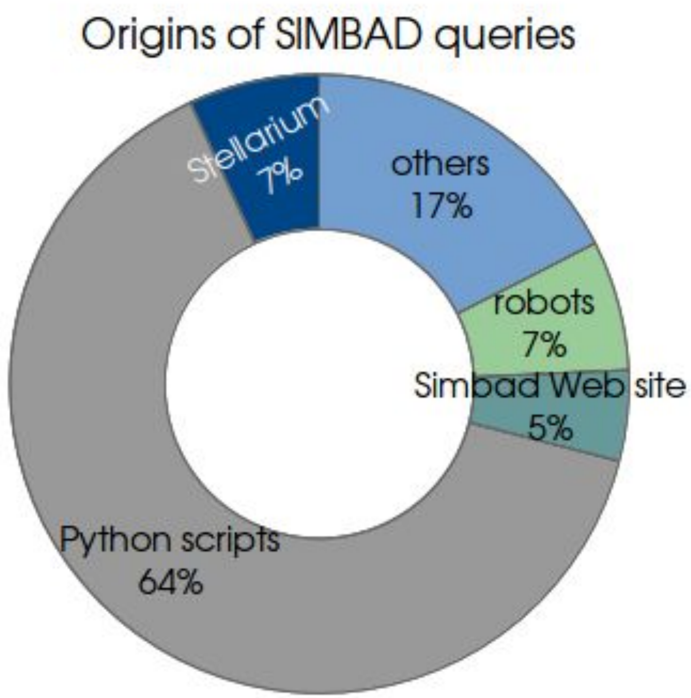
SIMBAD Usage (12 last months)

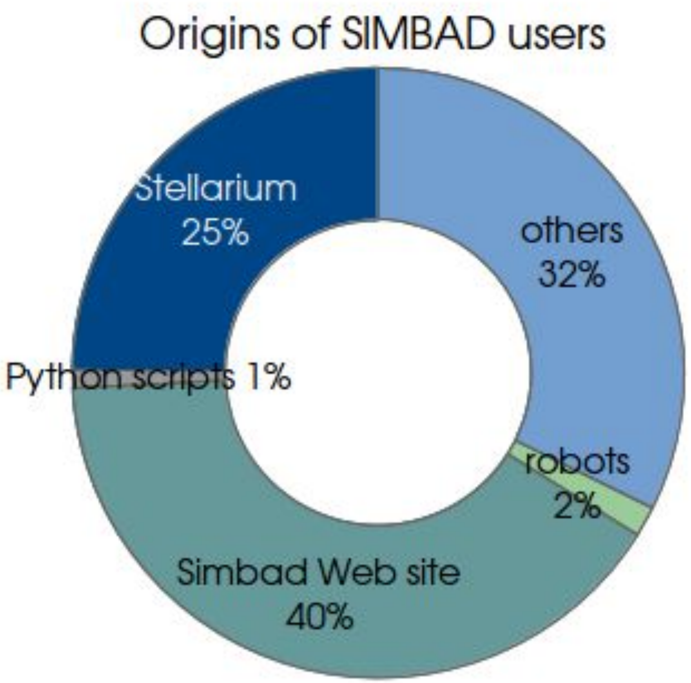
- ± 500,000 queries and 5GB per day
- ± 160,000 monthly unique visitors
- 2.3% queries are blacklisted (overload)
- 6% of visitors are human users
  (versus scripts/automated)



SIMBAD Usage (3 last months)

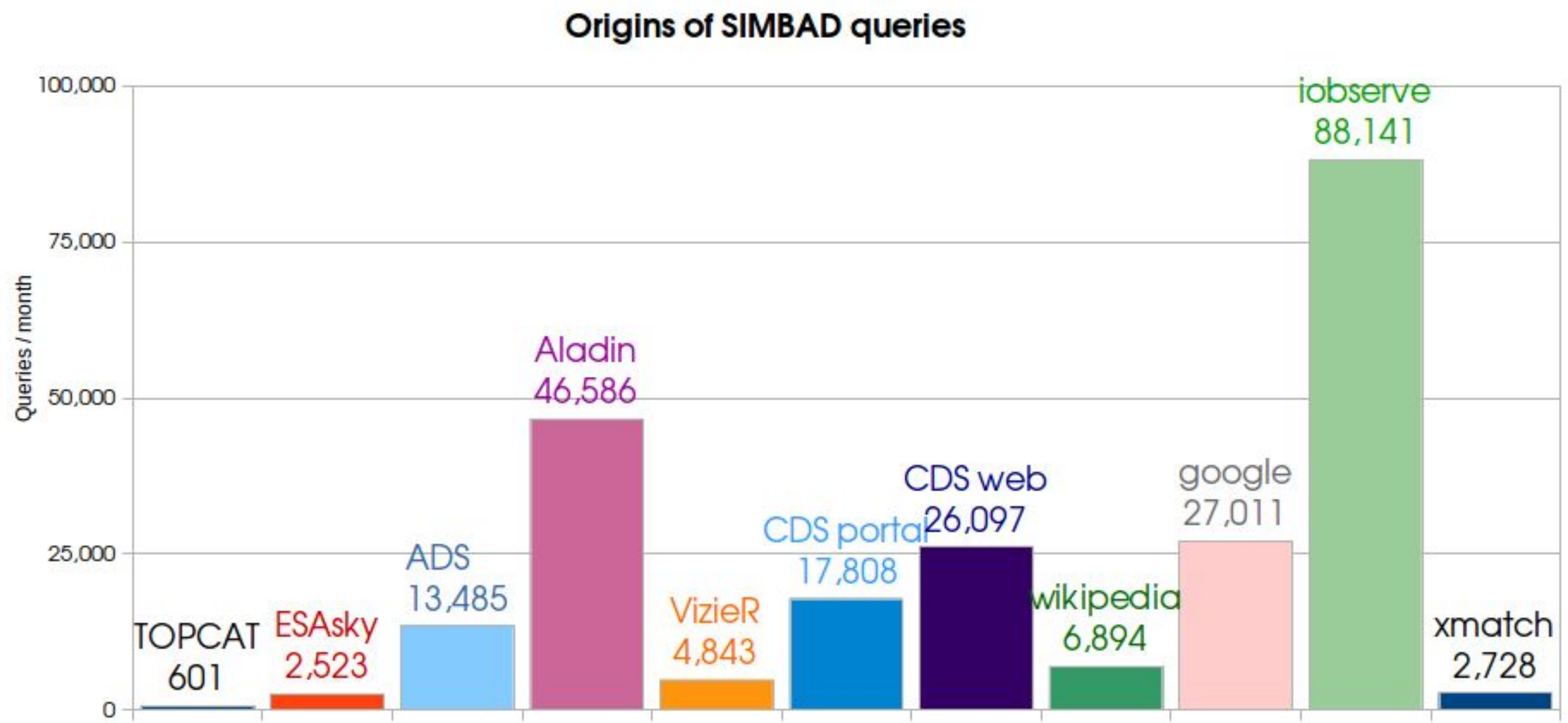
- A large activity is coming from 1% of users through python scripts.
- Most of people are SIMBAD users or Stellarium's.



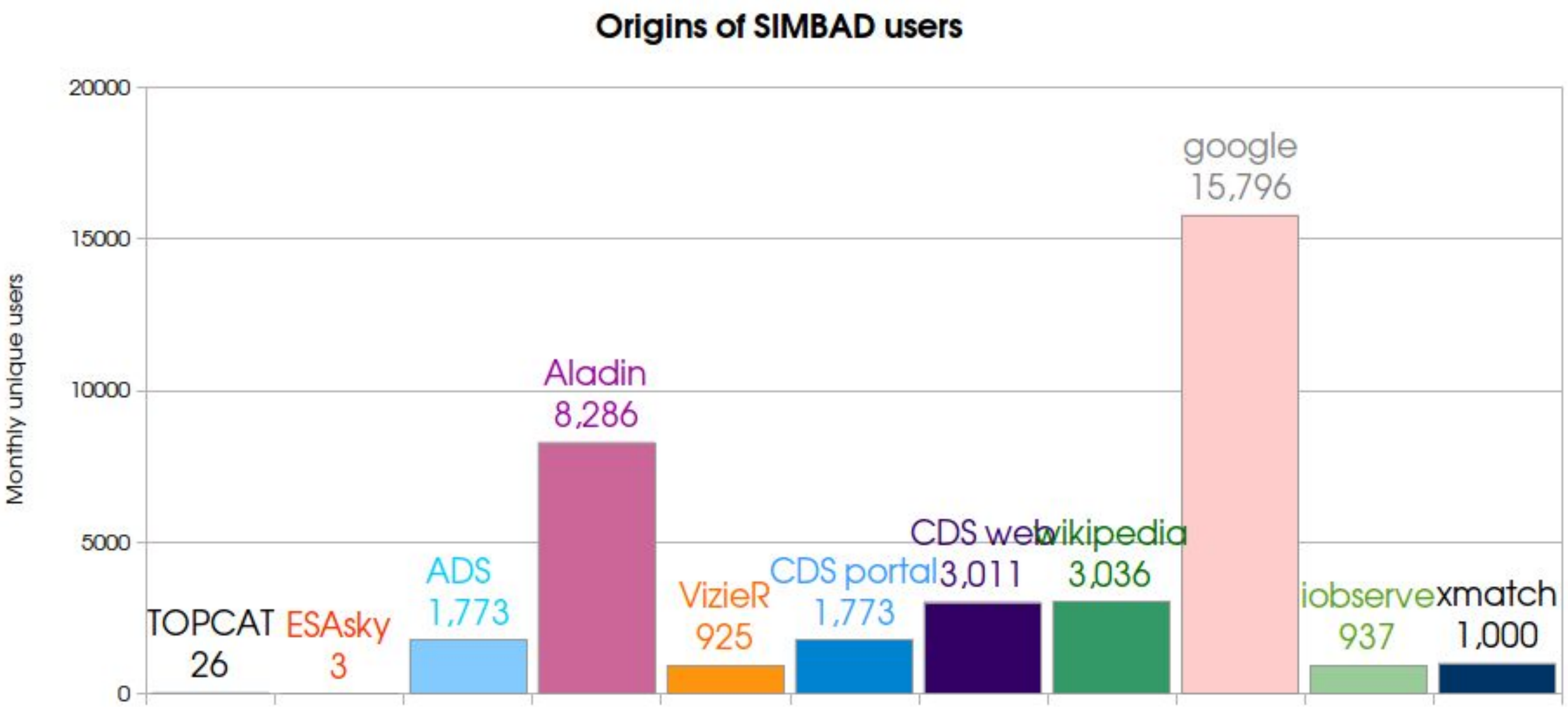


SIMBAD Usage (3 last months)

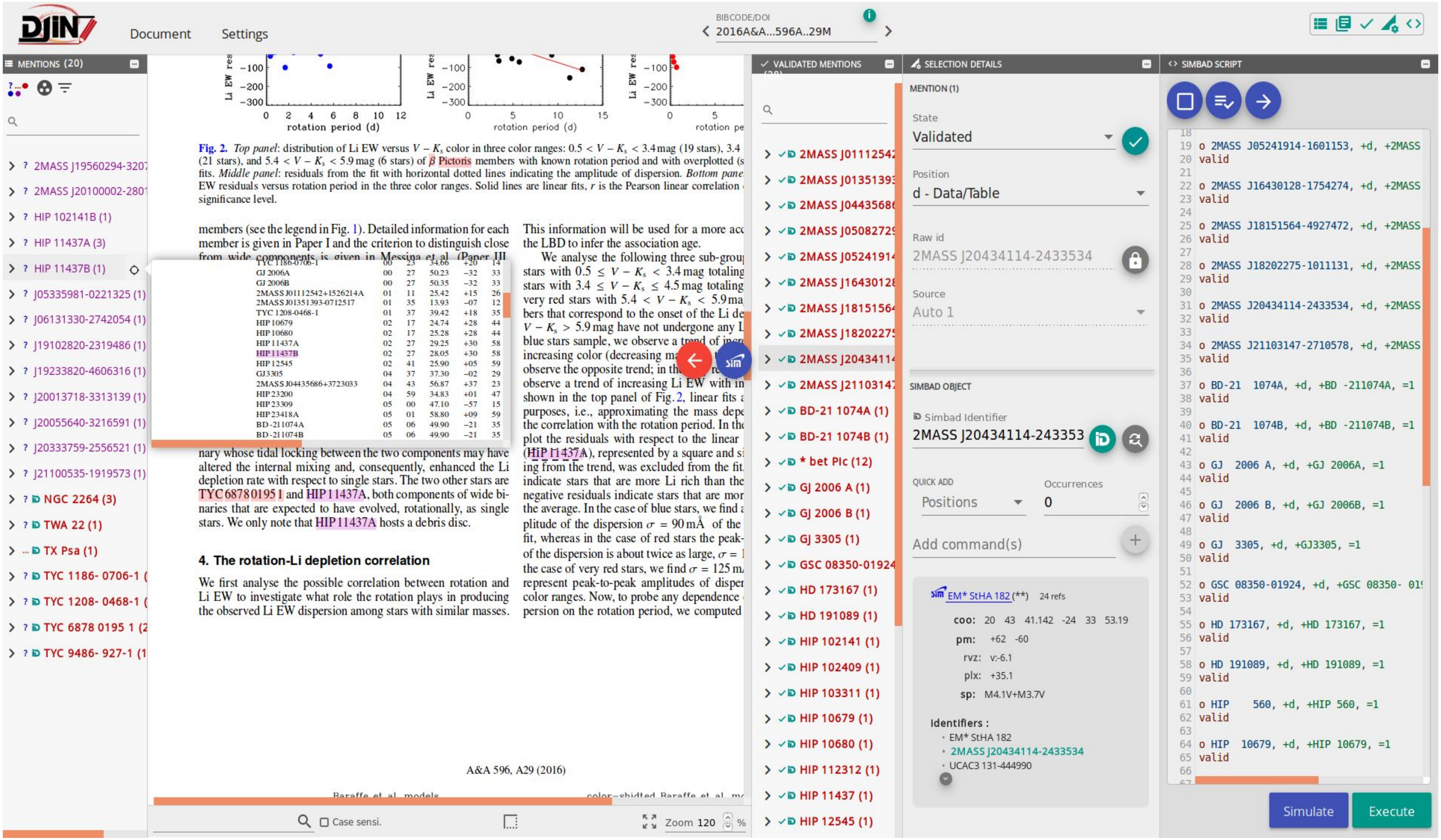
- Other queries are coming from: (note that Xmatch/TOPCAT are mainly for long list)



SIMBAD Users



DJIN2



Document
Settings
BIBCODE/DOI
2016A&A...596A..29M
MENTIONS (20)
VALIDATED MENTIONS
SELECTION DETAILS
SIMBAD SCRIPT
MENTION (1)
State
Validated
Position
d - Data/Table
Raw id
2MASS J20434114-2433534
Source
Auto 1
SIMBAD OBJECT
Simbad Identifier
QUICK ADD
Positions
Occurrences
Add command(s)
Simulate
Execute

DJIN2

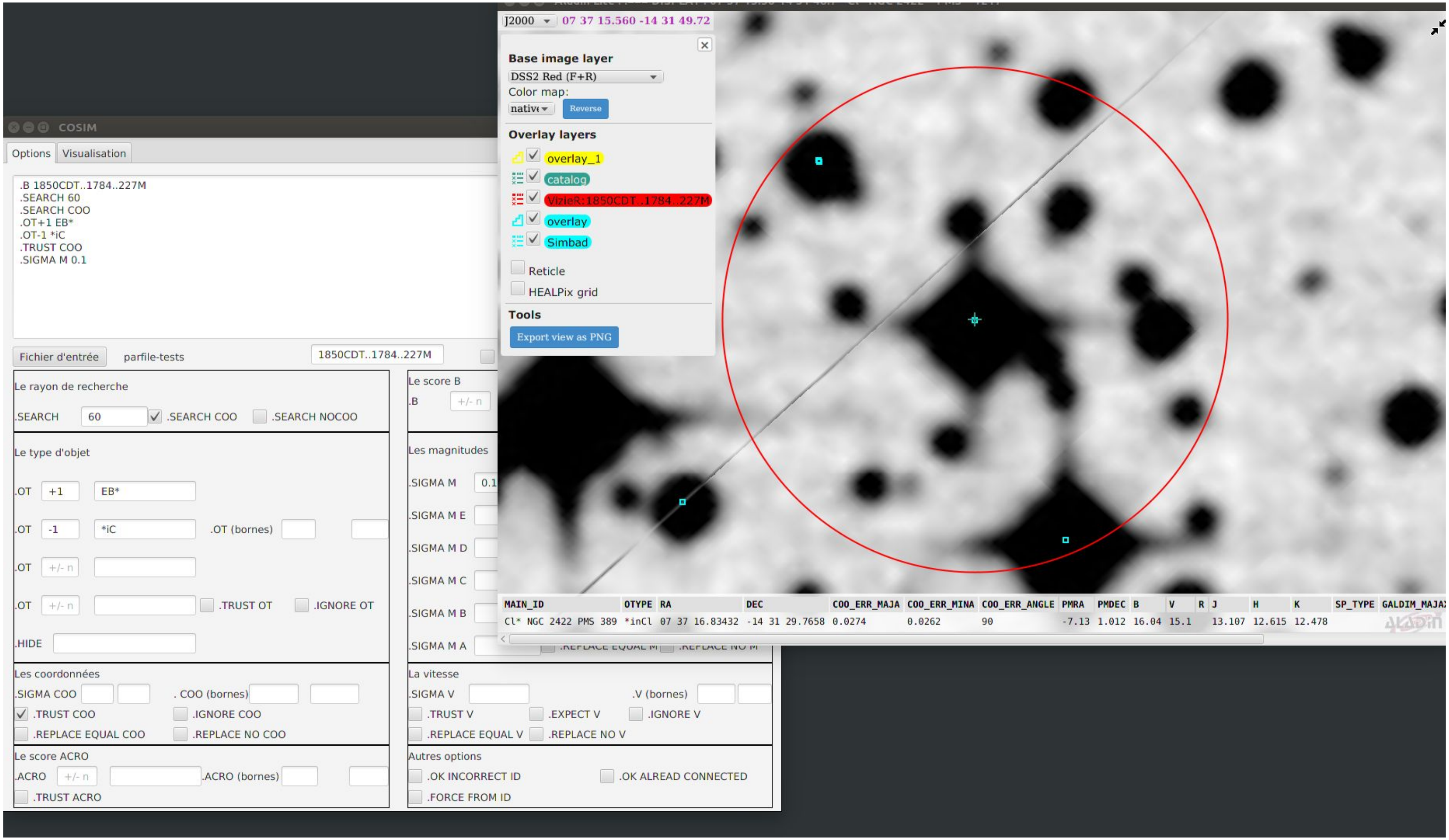
- Not enough efficient to be used in production

  → new feature in development: add known SIMBAD nomenclature to improve results

  → To be done: more tests and improvements to get responsive behaviour under many browsers

COSIM

- New graphical interface in development to improve speed and interactivity



Server infrastructure (with G. Mantelet)

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

SIMBAD Usage in 2018