

SIMBAD : the bibliographic database

A meta-compilation of astronomical objects of interest

**CDS Council
November 2023**



CDS team



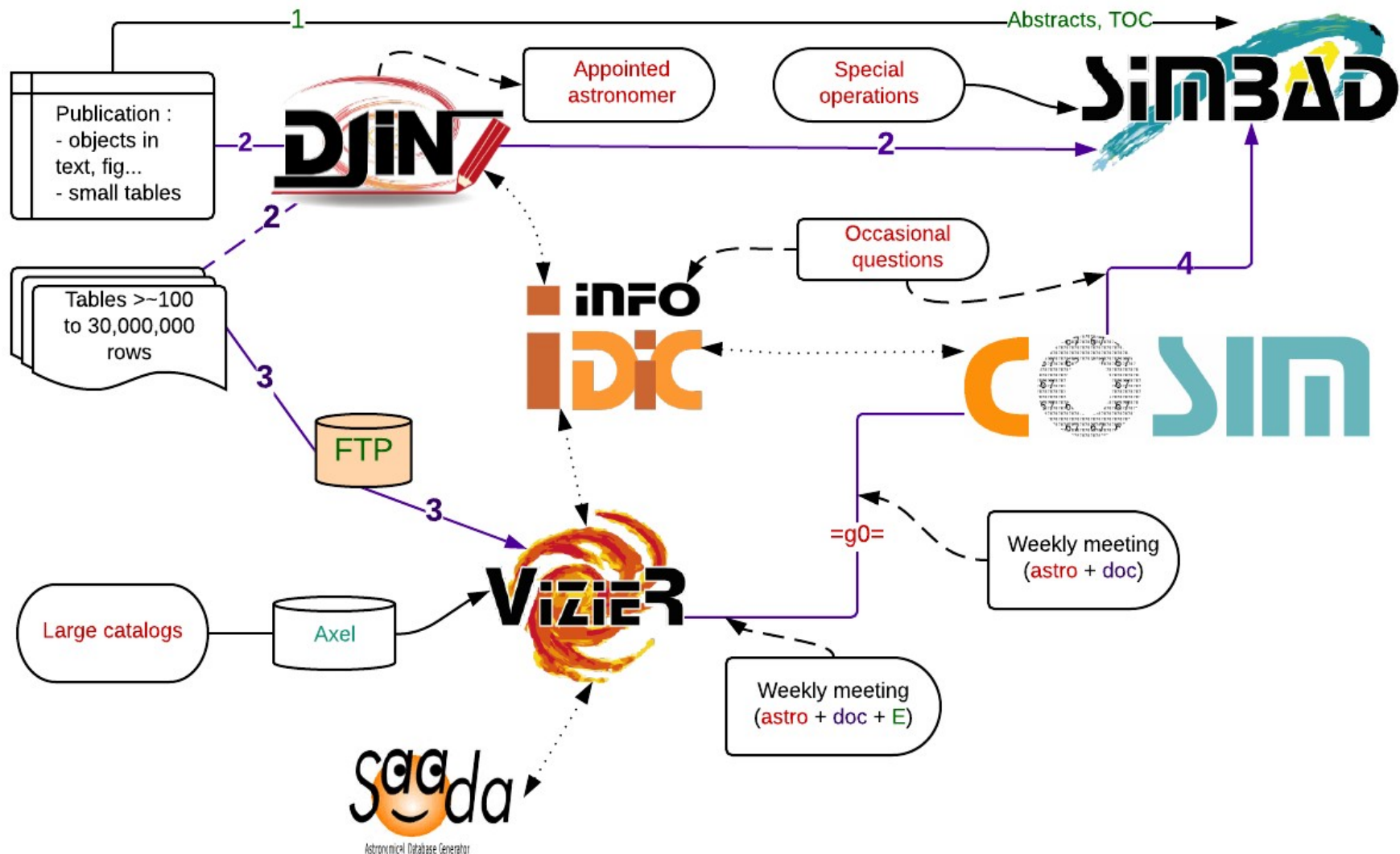


The Team

- **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 for Simbad** : G. Mantelet
- **Developer for Special Operations** (massive Xid) : T. Boch, F.-X. Pineau

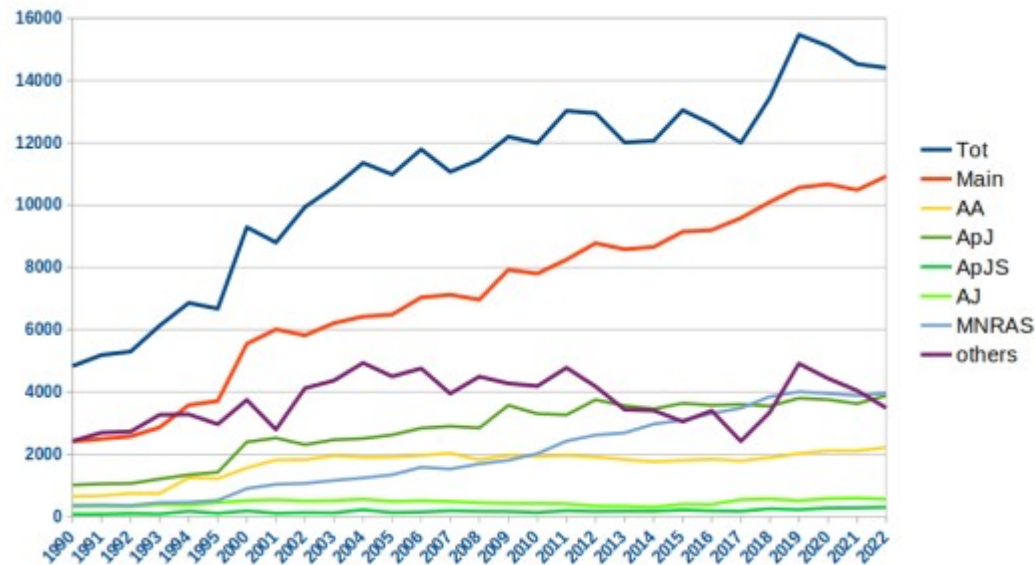
Workflow





Bibliography

Number of references per year of publication since 1991



Main journals: 10921 articles published in 2022 (10485 in 2021) : stable
Other journals: 3487 articles published in 2022 (4043 in 2021) : decreasing

Workload for the DJIN team : 15 to 25 articles per day



Tables of objects : prioritisation

Apraisal in bi-weekly meetings with astronomers & documentalists :

From 01/01 to 31/10/2023 : 1354 references evaluated, 66% with priority 1 for ingestion, 9% with priority 2

Criteria for priority 1

- Well characterized (nature, evolutionary stage) or rare objects (e.g. Blazars or Hot Subdwarfs)
- Spectroscopy of Stars & Galaxies, but also Molecular Clouds or Dense Cores
- Secure members of sets of Stars / Galaxies (mostly clusters, but also Streams, Moving Group, Group of G)
- Validated cross-identifications
- Known objects with additional data or just to link the reference

Criteria for priority 2 : mostly unsecure candidates

Criteria for non ingestion

- Photometric catalogues of uncharacterized sources
- X-match catalogues
- Poor astrometry or lack of coordinates

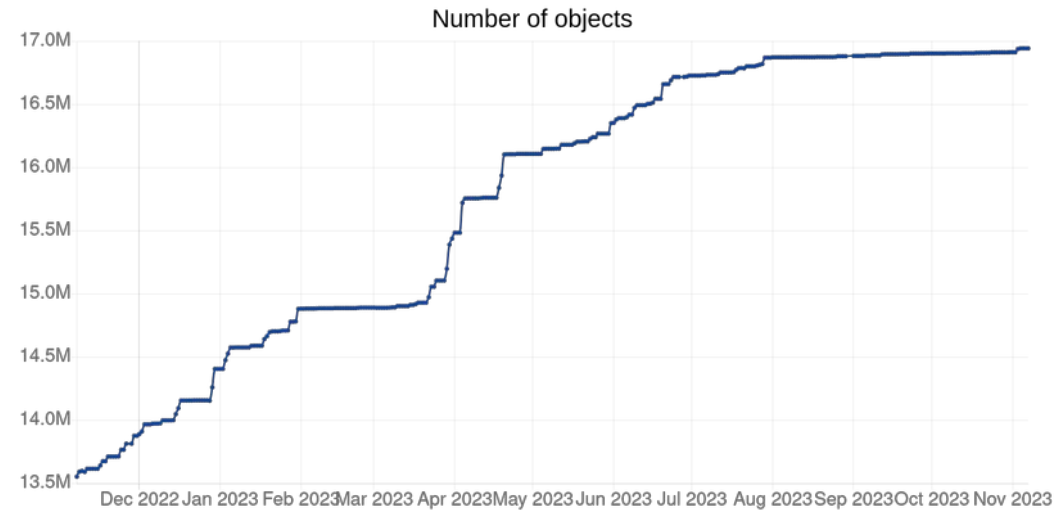
- Workload for the COSIM team : 700 to 1000 tables per year, from ~10 to more than 1 million objects
- Done in 2023 until end of October : 900, compared to 766 in 2022
- The improvement of the astrometry with Gaia and the revised relations between object types have speed up the cross-identifications process with the content of SIMBAD




Statistics on the Content

From 31/10/2022 to 31/10/2023

Overview	N (million)	+ in 1 yr
Objects	16.946	2.977
Identifiers	61.474	6.645
References	0.427	0.015
Citations of objects in articles	38.134	7.104
Basic data		
HRV/redshift	7.636	1.136
Proper motions	9.065	1.788
Parallax	7.713	1.141
Spectral type	0.946	0.088
Morphological type	0.147	0.003
Collection of measurements		
HRV/redshift	12.290	2.187
Proper motions	15.819	1.865
Parallax	15.884	1.249
Distance	9.543	0.281
Spectral type (spectroscopic only)	2.019	0.797
Teff logg Fe/H (spectroscopic only)	4.120	1.293
Variability	3.283	0.153



The spectacular growth of nearly 3 million objects last year reflects an important - and expected - change in astronomy surveys which increasingly include spectroscopically characterized sources that make them suitable for ingestion in SIMBAD.



Large catalogues ingested in SIMBAD

Spectroscopic surveys

- SDSS catalogues of QSOs (DR7,9,10,12,14,16) : 750k objects
- RAVE DR2 to DR6 : 451k objects
- GALAH DR2 & DR3 : 589k objects
- APOGEE-2 DR16 : 366k objects (DR17 still to be done)

Memberships (from many references)

- 999k stars (candidates) members of clusters, associations, streams, or moving group
- 1336k galaxies (candidates) members of groups or clusters

Variable Stars

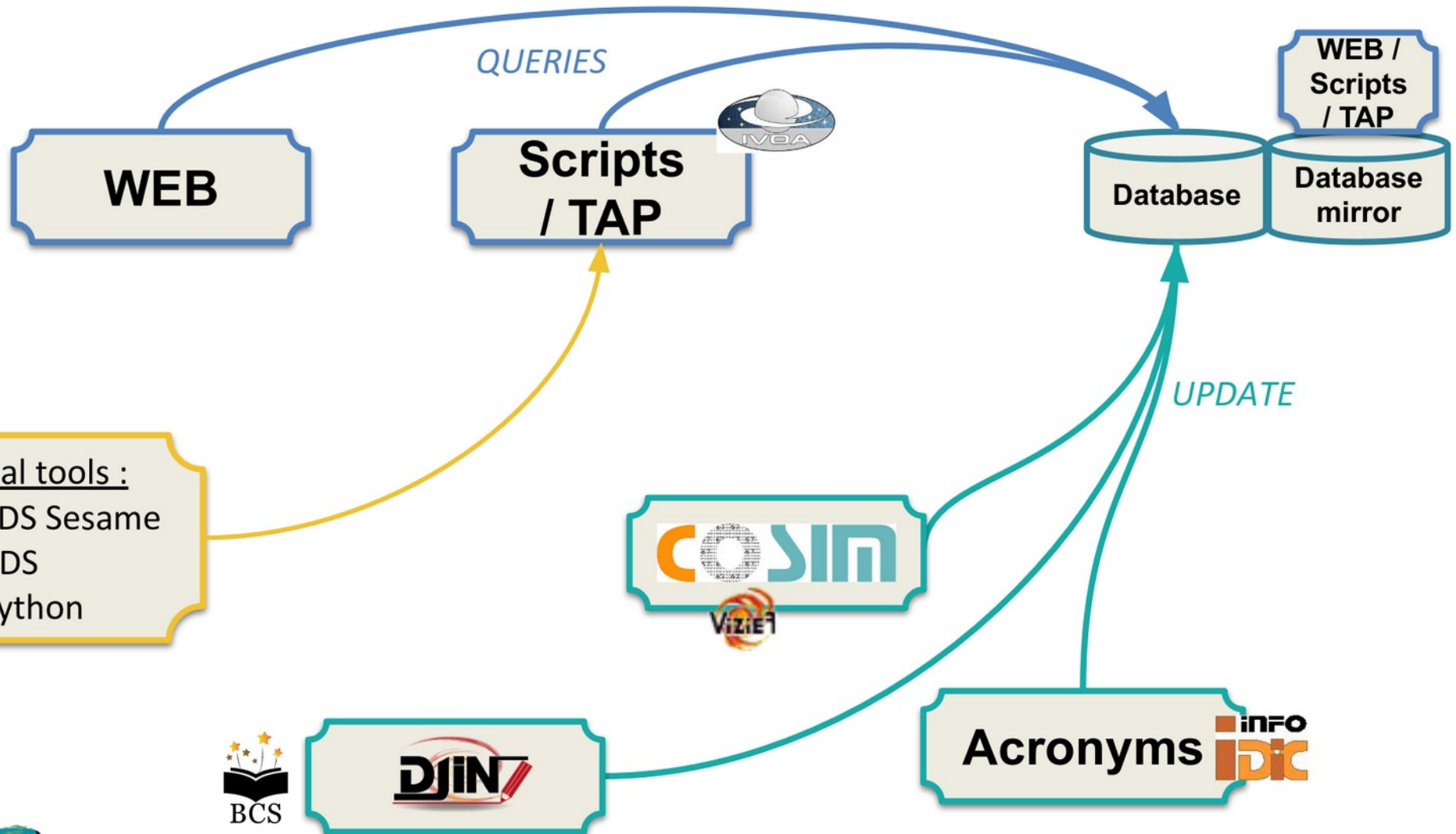
- OGLE Variables (several articles) : 756k objects
- Gaia DR3 Long Period Variables : 1721k objects
- Gaia DR3 RR Lyrae : 272k objects
- (Gaia DR3 Eclipsing Binaries still to be done)

Historical

- Hipparcos : 118k objects
- Tycho2 : 2559k objects
- HYPERLEDA : 956k objects



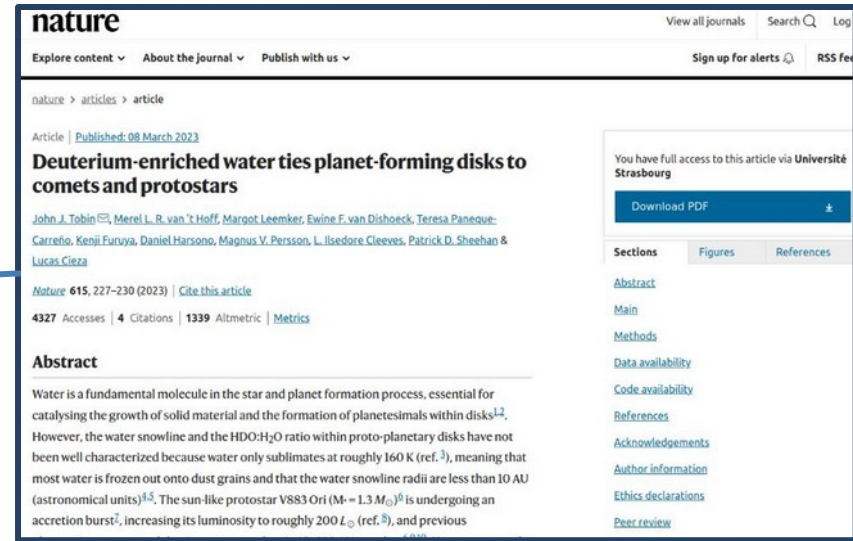
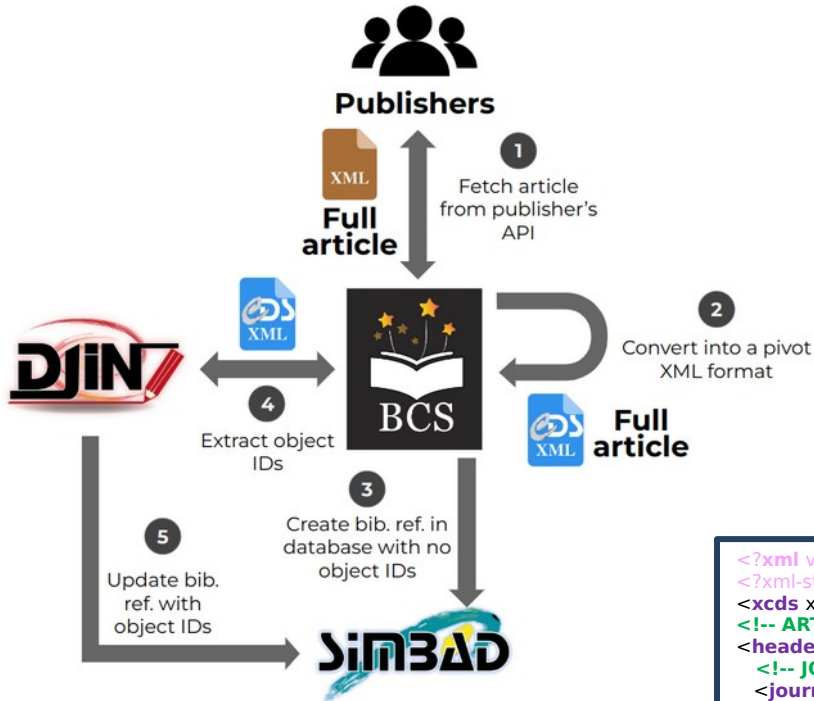
SIMBAD software





Bibliographical Center

Journal
A&A
AJ
ApJ
ApJL
ApJS
ATel
MNRAS
MNRASL
NatAs
Natur
PASP
PASJ
PASJS
RAA
RNAAS
Sci



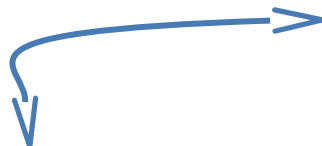
```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="../../../../XSL/HTML/xcds.xsl"?>
<xcds xmlns:fn="http://www.w3.org/2005/xpath-functions" xmlns:xsi="http://www.w3.org/2001/XMLSchema"
<!-- ARTICLE HEADER -->
<header>
<!-- JOURNAL -->
<journal id="Natur" issn="1476-4687">Nature</journal>
<!-- PUBLICATION -->
<volume>615</volume>
<issue>7951</issue>
<publisher>Nature Publishing Group</publisher>
<page type="first">227</page>
<page type="last">230</page>
<page type="count">4</page>
<pub_date>
<month>03</month>
<year>2023</year>
</pub_date>
<copyright>2023 The Author(s), under exclusive licence to Springer Nature Limited</copyright>
<!-- TITLE -->
<title>Deuterium-enriched water ties planet-forming disks to comets and protostars</title>
<!-- IDS -->
<idno type="bibcode">2023Natur.615..227T</idno>
<idno type="DOI">10.1038/s41586-022-05676-z</idno>
<!-- AUTHORS -->
<authors>
<author><firstname>John J.</firstname> <lastname>Tobin</lastname> <affiliation ref="AFF0"
/><idno type="ORCID">0000-0002-6195-0152</idno> <email>jtobin@nrao.edu</email></author>
  
```

New: Transformation from HTML web article to CDS-XML format to easily process (and **homogeneously**)



The dictionary is a registry of all acronyms and their formats. It serves the users to correctly write an identifier, and it is a key stone for SIMBAD to check that all identifiers are correct.



Dictionary of Nomenclature of Celestial Objects (Last update: 02-Nov-2023)

Result of query: info cati TYC

Obj_Type	Acronym	(Explanation)	Format
E Star	ACT	(Astrographic Catalogue Tycho)	TYC FFFF NNNNN N
E Star	TRC	(Tycho Reference Catalogue)	TYC FFFF NNNNN N
E Star	TYC	(Tycho mission)	TYC FFFF-NNNNN-N
+ Star	Tycho	(Tycho SNR)	[RCM2004] Tycho AN
E Star	Tycho	(Tycho mission)	TYC FFFF NNNNN N

Acronyms flagged by + are *completely incorporated in Simbad*
Acronyms flagged by E are *available in electronic form at CDS*

```

cat = TYC
description = Tycho mission
create = 06.11.06
update = 06.11.06
version = 1.0 23-Oct-1998: Genova
version = 1.1 05-Mar-2009: bvollmer;
version = 1.2 05-Mar-2009: bvollmer;
type = Star
pgm = space()
  uint(binaire, maximum=9999, minimum=0, justification=d)
  text(facultatif, texte="-")
  uint(binaire, maximum=65535, minimum=0)
  text(facultatif, texte="-")
  uint(binaire, maximum=9, minimum=0)
format = FFFF-NNNNN-N
exemple = /TYC 1-13-1
exemple = /TYC 9537-380-1
exemple = /TYC 99-2884-2
exemple = /TYC 9038-2799-3
exemple = /TYC 1327 606 4/TYC 1327-606-4
    
```

Acronyms numbers:

12 638 acronyms like [ABC0000]
(among 15 376) for SIMBAD usage

Eg:

[JE82] 296
[VV2000] J042315.8-012033

Dictionary of Nomenclature of Celestial Objects (Last update: 02-Nov-2023)

Result of query: info cati TYC

Obj_Type	Acronym	(Explanation)	Format
E Star	ACT	(Astrographic Catalogue Tycho)	TYC FFFF NNNNN N
E Star	TRC	(Tycho Reference Catalogue)	TYC FFFF NNNNN N
E Star	TYC	(Tycho mission)	TYC FFFF-NNNNN-N
+ Star	Tycho	(Tycho SNR)	[RCM2004] Tycho AN
E Star	Tycho	(Tycho mission)	TYC FFFF NNNNN N

Acronyms flagged by + are completely incorporated in Simbad
Acronyms flagged by E are available in electronic form at CDS

2 text files →

```

cat = TYC
description = Tycho mission
create = 06.11.06
update = 06.11.06
version = 1.0 23-Oct-1998: Genova
version = 1.1 05-Mar-2009: bvollmer;
version = 1.2 05-Mar-2009: bvollmer;
type = Star
pgm = space()
  uint(binaire, maximum=9999, minimum=0, justification=d)
  text(facultatif, texte="-")
  uint(binaire, maximum=65535, minimum=0)
  text(facultatif, texte="-")
  uint(binaire, maximum=9, minimum=0)
format = FFFF-NNNNN-N
exemple = /TYC 1-13-1
exemple = /TYC 9537-380-1
exemple = /TYC 99-2884-2
exemple = /TYC 9038-2799-3
exemple = /TYC 1327 606 4/TYC 1327-606-4
    
```

acronym	
PK	acroid: SERIAL
	acronym: VARCHAR
	created: DATE
	modified: DATE
	pub_status: SMALLINT
	simbad_acro: VARCHAR
	explanation: VARCHAR
	deprecated: BOOLEAN
	type: VARCHAR
	created_from_lortet: BOOLEAN

synonym_of	
PK,FK1	acroid: INTEGER
PK,FK2	parent: INTEGER

format	
PK,FK3	acroid: INTEGER
PK	format: VARCHAR
	position: SMALLINT

format_example	
PK,FK4	acroid: INTEGER
PK	example: VARCHAR
	position: SMALLINT

acroref	
PK,FK5	acroid: INTEGER
PK,FK6	refid: VARCHAR
	position: SMALLINT

version	
PK,FK7	acroid: INTEGER
PK	version: VARCHAR
	date: DATE
	author: VARCHAR
	message: VARCHAR
	position: SMALLINT

reference	
PK	refid: VARCHAR
	lortet: BOOLEAN
	deprecated: BOOLEAN
	first_author: VARCHAR

GSC4SIM



published tables

MONTHLY NOTICES
Volume 490, Issue 2
December 2019

Article Contents

ABSTRACT

1 INTRODUCTION

2 TARGET AND DATA DESCRIPTION

3 MEMBERSHIP PROBABILITY

4 SPATIAL DISTRIBUTION

5 FRACTAL STATISTICS

6 DYNAMICAL EVOLUTION

7 RESULTS

8 SUMMARY AND DISCUSSION

ACKNOWLEDGEMENTS

as age, metallicity and E(B - V) for a sample of nearby ($d < 250$ pc) OCs; among them appear four objects studied by us (IC 2391, IC 2602, NGC 3532, and NGC 6475). Fitting the observed data with the models of Kuhn et al. (2010) to estimate visual extinction, we performed a kinematic study of OCs, searching for evidence of expansion and total energy (virial) equilibrium. NGC 2362 and NGC 6530 are two examples of our clusters that had their gravitational condition (bound or unbound) explored by these authors. These are only a few examples of results from the literature, mainly those related to the parameters studied by us, which are presented as comments on individual objects in Appendix D.

Table 1. Structural results based on spatial distribution and fractal statistics.

Cluster	α (J2000)	δ (J2000)	$\mu_{\text{core}} S$	β	N_{mem}	R
	deg	deg	mas yr ⁻¹	mas yr ⁻¹	pc	
ASCC13	78.310 ± 0.170	44.584 ± 0.176	0.61 ± 1.44	-2.32 ± 1.35	94	4.58 ± 0.1
ASCC19	81.965 ± 0.197	-1.967 ± 0.163	1.17 ± 0.36	-1.12 ± 0.39	114	1.84 ± 0.1
ASCC20	82.171 ± 0.174	1.650 ± 0.160	0.27 ± 0.98	-0.92 ± 0.97	79	2.00 ± 0.1
ASCC32	105.519 ± 0.107	-36.502 ± 0.119	-2.78 ± 0.85	3.06 ± 0.63	88	2.80 ± 0.1
ASCC33	105.848 ± 0.212	-25.038 ± 0.212	-3.50 ± 0.53	3.56 ± 0.30	72	4.97 ± 0.1
ASCC61	161.522 ± 0.103	-56.876 ± 0.096	-5.99 ± 0.45	3.01 ± 0.50	74	5.44 ± 0.1
ASCC65	167.787 ± 0.055	-61.124 ± 0.052	-6.16 ± 0.10	1.17 ± 0.10	70	5.82 ± 0.1

GSC4SIM

Specific query for large catalog to add in SIMBAD



Copy specific data

From position / name

Find corresponding data in Vizier

VizieR

Search Criteria

Keywords: I/350/galaxedr3

Tables: I/350, galaxedr3, agncrid, lyc2t6dc, comsancal

Constraints: (Modify Query)

Preferences: max: 50, HTML Table, All columns, Compute

Mirrors: CDS, France

The 12 columns in color are computed by VizieR, and are *not part of the original data*.

I/350/galaxedr3
Gaia EDR3 (Gaia Collaboration, 2020)
Gaia data early release 3 (Gaia EDR3). (Download all Gaia Sources as VOTable, FITS or here)

(original column names in green) (1811709771 rows)

Full	RA ICRS	e_{mas}	DE ICRS	e_{mas}	Source	Pix	e_{mas}	PM	pmRA	$e_{\text{mas/yr}}$	pmDE
	deg		deg					mas/yr	mas/yr		mas/yr
1	044.99615537865	0.1016	+0.00561522634	0.1013	4295806720	0.3543	0.1227	12.616	11.938	0.138	-4.081
2	045.00497837175	0.0179	+0.01987967570	0.0188	3865544960	3.1392	0.0223	35.308	29.686	0.024	19.115
3	045.00432028915	0.0973	+0.02104776378	0.1018	34361129088	3.2350	0.1205	35.231	29.518	0.134	19.232
4	044.99503714416	0.3220	+0.03815169755	0.2835	309238066432	1.3831	0.3679	1.473	0.710	0.428	-1.290
5	044.96389626550	0.1172	+0.04359494368	0.1090	343597448960	0.1961	0.1335	6.845	6.567	0.155	-1.931
6	044.99832707811	0.3252	+0.06633270720	0.3253	515396233856	0.2424	0.3743	9.078	4.473	0.413	-7.900
7	045.04828232130	0.0278	+0.04825396034	0.0265	549755818112	1.5835	0.0344	16.465	0.843	0.039	-16.444
8	045.02361979732	0.0543	+0.06681676725	0.0576	626929527040	1.2031	0.0668	17.646	13.952	0.078	-10.804
9	045.02672698087	0.3037	+0.08169947827	0.3229	927713095040	-0.1227	0.3688	3.980	3.762	0.472	-1.299
10	045.02698947346	0.4010	+0.08695489276	0.3926	965639933184	1.6270	0.4508	1.890	3.161	0.510	-3.450
11	044.96654617792	0.2382	+0.04630865837	0.2192	1099511693312	-0.3448	0.2701	6.071	3.090	0.325	-5.226
12	044.99327078417	0.0442	+0.07633404500	0.0374	1275606125952	0.6296	0.0481	6.749	-1.435	0.058	-6.595
13	044.96907662980	0.0960	+0.08442520281	0.0837	1340029955712	0.4156	0.1050	3.239	1.792	0.127	-2.699
14	044.97846156971	0.1583	+0.09257928817	0.1388	1340029956224	0.3111	0.1719	1.291	0.151	0.214	-1.282
15	044.93280210615	0.1341	+0.06480894307	0.1165	1374389600384	1.7651	0.1499	11.373	5.092	0.174	-10.170
16	044.95265152875	0.2176	+0.08495205087	0.1964	1511828647680	-0.0980	0.2435	4.849	4.774	0.298	0.853
17	044.95115803041	0.1149	+0.10531247613	0.0927	1619203481984	1.7173	0.1171	18.378	13.218	0.155	12.768
18	044.99606230475	0.0432	+0.08491778897	0.0377	1653563247744	0.3013	0.0472	20.935	14.857	0.057	-14.749
19	045.01378833779	0.3412	+0.08773432699	0.3123	1683627775360	2.9377	0.3839	32.155	-5.370	0.484	-31.704
20	044.98309734472	0.1607	+0.09640645833	0.1449	1717987078400	2.7367	0.1784	14.770	2.703	0.218	-14.521

IN PROGRESS ...

```
# MATCHES :
# 444 #--- 2mass : 1 matches
# 82 #--- gaia3 : 0 matches
# 39 #--- gaia3 : 2 matches
# 1 #--- gaia3 : 3 matches
# 80 #--- gaia3 : (accepted < 1") 1 matches
# 242 #--- gaia3 : ( too far > 1") 1 matches
```



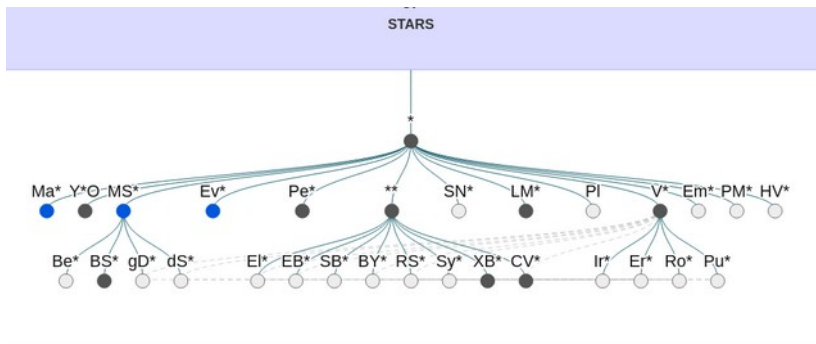


SIMBAD API

New API to query SIMBAD by cone: allow filter by any criteria (distance/number of papers)

```
http[s]://.../simbad/sim-cone?RA=..&DEC=..&SR=0.1&RESPONSEFORMAT=json&MAXREC=5&ORDER_BY=biblist
```

Categorisation of object types, are now available programmatically



```
SELECT
  otypedef.path,description,COUNT(*)
FROM ident,basic
JOIN otypedef ON basic.otype=otypedef.otype
JOIN h_link ON (oid=child AND parent=ident.oidref)
WHERE id = 'Hyades Moving Group' AND (membership >=90)
AND is_candidate = 0 AND otypedef.otype='V*..'
GROUP BY otypedef.path,description
```

path	description	COUNT_ALL
** > ** > BY**	"BY Dra Variable"	66
** > ** > EB**	"Eclipsing Binary"	2
** > ** > RS**	"RS CVn Variable"	3
** > MS* > dS**	"delta Sct Variable"	1
** > V**	"Variable Star"	4
** > V* > Er**	"Eruptive Variable"	13
** > V* > Ro**	"Rotating Variable"	1

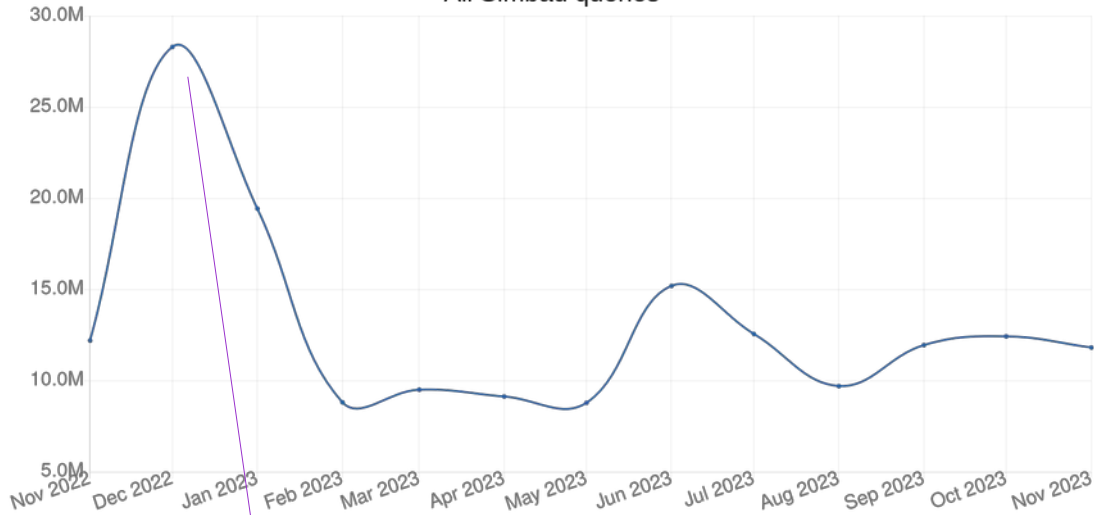
+ more in progress
contributing
Python libraries





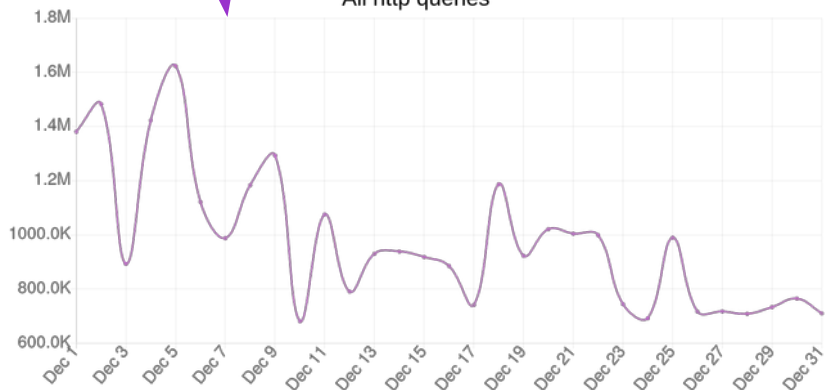
Usage

All Simbad queries

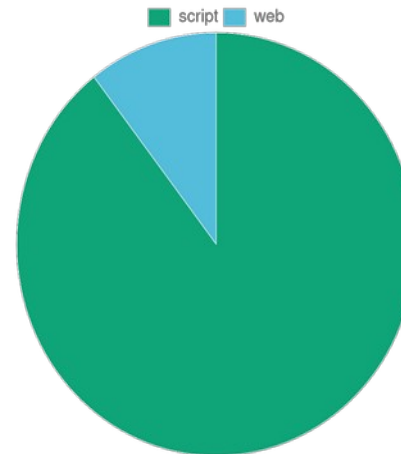


Queries per day : from 283K to 1.6M

All http queries



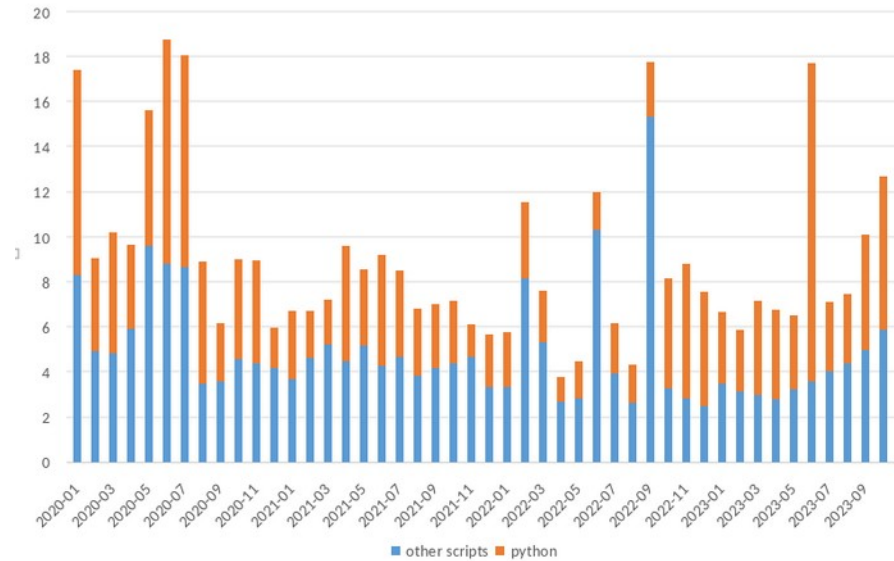
Web vs script on cds Simbad



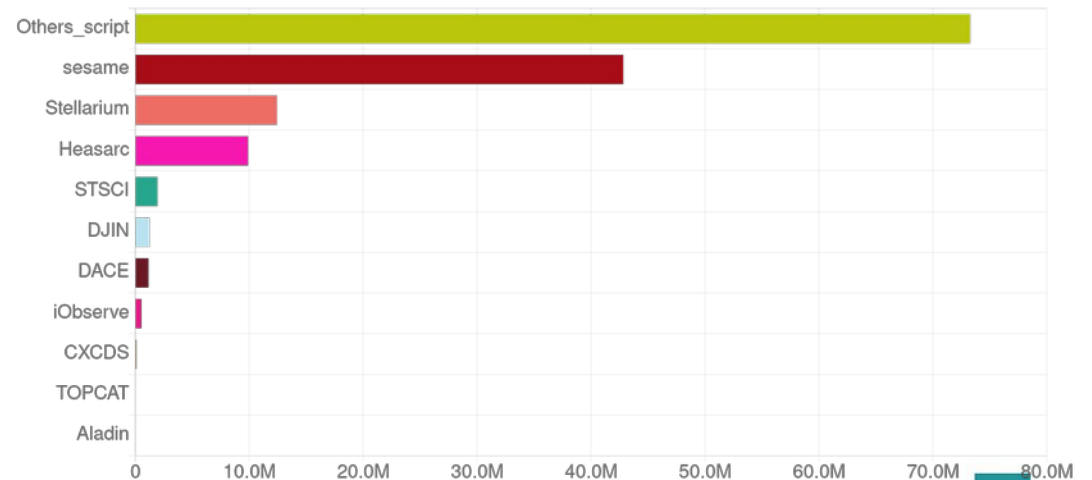


Scripts usage

Python in script queries over the past years



Top repartition of script agent on queries all WebServers



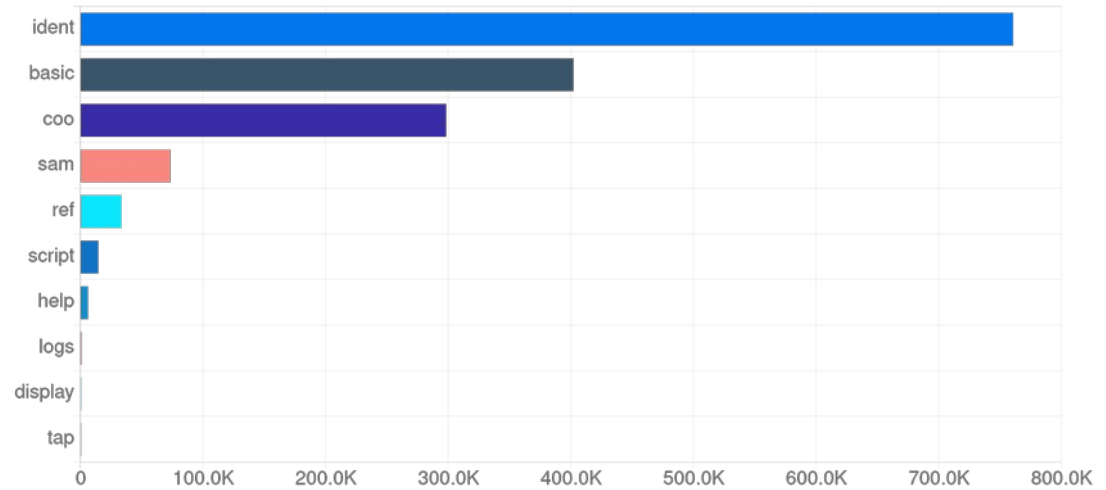
Queries (by scripts) grouped by the provenance of users



WEB usage

Distinct usage of web pages in SIMBAD

Top repartition of web pages forms queried on cds Simbad



Provenance of users on the Web

Top repartition of Referrer on queries

