



HAL
open science

Impact of Astronomy Evolution on the Documentalists' Activities at CDS

Evelyne Son, Soizick Lesteven, Fabienne Marquis, Françoise Genova, Mark Allen

► **To cite this version:**

Evelyne Son, Soizick Lesteven, Fabienne Marquis, Françoise Genova, Mark Allen. Impact of Astronomy Evolution on the Documentalists' Activities at CDS. Multidimensional Astronomy Librarianship: From Open Science to the Preservation of Astronomical Heritage, Royal Astronomical Society, Jun 2021, Virtual conference, United Kingdom. 10.3847/25c2cfef.eacea881 . hal-03372191

HAL Id: hal-03372191

<https://hal.science/hal-03372191>

Submitted on 9 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Impact of Astronomy Evolution on the Documentalists' Activities at CDS

2 EVELYNE SON ¹, SOIZICK LESTEVEN ¹, FABIENNE MARQUIS ¹, FRANÇOISE GENOVA ¹ AND MARK G. ALLEN ¹

3 ¹*Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France*

4 ABSTRACT

5 The CDS is a science driven data centre for reference astronomy data. The objectives have been
6 defined since its creation in 1972: collecting, improving and distributing data for the international
7 astronomical community. However, science and techniques have significantly evolved over the past 50
8 years, resulting in a deluge of data in terms of both quantity and complexity. The CDS has been
9 able to constantly adapt to these evolutions to offer the community appropriate services and quality
10 data with high added value. In this presentation we will focus mainly on the impact of these changes
11 on the the skills of the CDS data stewards, the so-called documentalists, in terms of specialisation,
12 adaptation to technologies and understanding of data taking into account the evolution of astronomy.
13 All this while ensuring a high level of quality and quick availability of the data.

14 *Keywords:* Astronomy databases (83) — Astronomy web services (1856) — History of astronomy
15 (1868) — Astronomy data analysis (1858) — Interdisciplinary astronomy (804)

16 1. THE CDS, A SCIENCE DRIVEN DEVELOPMENT

17 The CDS has evolved a lot since its creation in 1972 (50 years ago), following both astronomy and technology
18 evolutions. From the beginning the CDS goals have remained the same: collect useful data on objects in electronic
19 form; improve them by critical evaluation and combination; distribute the results to the international community and
20 conduct research using these data.

21 One of the most important aims is to give the astronomers the necessary resources to conduct their research. That
22 means taking into account the research evolution, and make the database content constantly adapt and evolve. It also
23 means taking technological evolutions into account, and making the database systems and user interfaces evolve. We
24 also need to take into account the continuously increasing volume of data, by adapting tools and work practices to
25 cope with this deluge of data. The continuous increase in data complexity also requires greater knowledge and skills.

26 The CDS has been able to develop more and more services over the years, thanks to the evolution of techniques and
27 science, and the availability of data: different versions of SIMBAD¹ the database dedicated to identification, data and
28 bibliography associated to the objects of interest (Wenger et al. (2000), Wenger et al. (2006)); VizieR² the database
29 dedicated to tables and other data attached to publications and catalogues (Ochsenbein et al. 2000); the Dictionary
30 of Nomenclature of Celestial Objects³ defining the origin, formats, and definition of the acronyms used in link with
31 the IAU⁴(Lortet et al. 1994); Aladin⁵ the atlas of the sky and image database (Bonnarel et al. 2000); the CDS Portal⁶
32 giving access to all the CDS services through a unique page (Boch & Derriere 2010); and the X-Match Service⁷ allowing
33 one to make fast cross-matches of very large tables and catalogues (Pineau et al. 2020; Boch et al. 2012).

Corresponding author: Evelyne Son
evelyne.son@astro.unistra.fr

¹ <http://simbad.u-strasbg.fr/simbad/>

² <http://vizier.u-strasbg.fr/viz-bin/VizieR>

³ <http://cds.u-strasbg.fr/cgi-bin/Dic-Simbad>

⁴ <https://www.iau.org/>

⁵ <https://aladin.u-strasbg.fr/aladin.gml>

⁶ <http://cdsportal.u-strasbg.fr/>

⁷ <http://cdsxmatch.u-strasbg.fr/#tab=xmatch&>

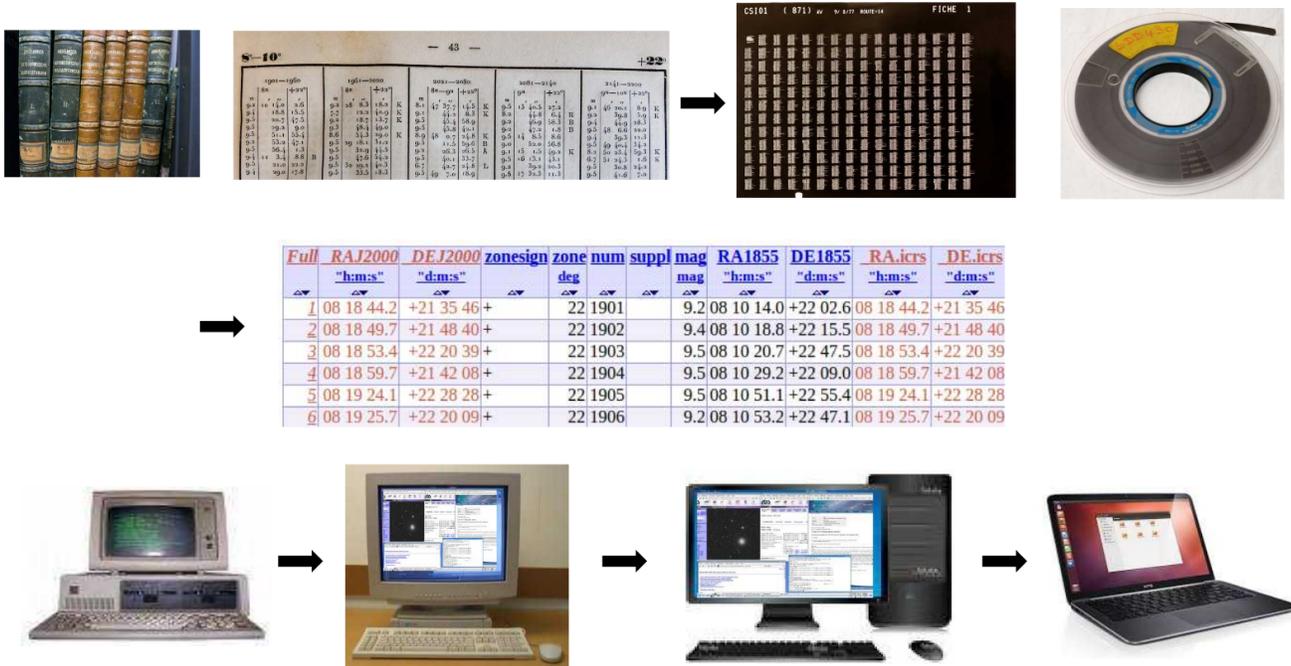


Figure 1. The Bonner Durchmusterung (BD) catalog through technological evolution at CDS.

2. EVOLUTION OF SCIENCE AND TECHNIQUES

To have a good illustration of how the science and technology have evolved in the past 50 years at CDS, we can take the example of the BD (Bonner Durchmusterung) catalogue⁸ first published in 1852. At the CDS, as illustrated in Figure 1, we first had the paper version of this catalogue, stored in books. The catalogue was reproduced on microfiches in the 70's (Ochsenbein et al. 1981), saved on magnetic tape in the 80's. To consult this catalogue, magnetic tapes had to be requested and sent through postal services. It could take several weeks between the request and the consultation of data. The computers also evolved a lot between the 80s – 90s and today, from core computers to working stations. Data storage has grown significantly, as have sharing capacities, with the advent of the internet. The first use of the internet in France (at that time its precursor arpanet) was a demo of SIMBAD made by the NASA partners at the LISA I conference in Washington (also an IAU meeting)⁹. The BD catalogue was entered as an electronic table in Vizier at the end of the 90s. It is now available from a laptop, almost instantaneously from everywhere, and can be found and queried through the Virtual Observatory. As one can imagine, typing microfiches and creation of online tables with interactive data are not the same, and the work practices of data stewards has evolved a lot during these decades!

As for astronomy and scientific development, a good illustration of the huge advances in the field is to consider the emergence and development of ground and space observatories. These brought new wavelengths in SIMBAD, through observations made in UV, X or gamma-ray bands. Also IRAS brought new data in the infrared, and Gaia added the magnitude G in optical band in SIMBAD. These surveys brought a large number of objects, and new data in SIMBAD, like the redshifts added with IRAS (Bidelman 1988), proper motions and parallaxes added with USNO (Ochsenbein 1984). The level of astrometric precision grew spectacularly in 30 years, between the Hipparcos and the Gaia missions, from milli-arcseconds to tens of microarcseconds! Also these missions brought more accuracy in proper motions, parallaxes etc.

⁸ <https://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=I/122>

⁹ <https://interstices.info/au-coeur-de-la-premiere-connexion-francaise-a-larpanet/> - text in French. The event is described in sections "Un réseau astronomique" (an astronomical network) and "Une liaison transatlantique" (a transatlantic liaison).

<i>PG 0849+580</i>			
Type: V*			
Coord 1950.0 = 08 49 52.9	+58 00 02	mb, mv = 15.3	V4
Coord 2000.0 = 08 53 44.5	+57 48 39	sp type = CV	
gal = 159.00 +38.83			
./.			
PG 0849+580	V* BZ UMa	SV* SVS 1719	
WD 0849+58			
Measurements:			
V*: 1			
References: 4			
1989A&AS...78..145 1988ApJ...334..422 1988MNRAS.233..513 1988IAUC.4691.....			

Figure 2. The star PG 0849+580 in SIMBAD in 1990. Few data were associated to the object: 4 identifications and 4 bibliographical references, coordinates, 1 object type, 1 magnitude in V band.

3. EVOLUTION OF THE SERVICES

3.1. Evolutions in SIMBAD

The primary requirement of the database SIMBAD was to give the cross-identifications for objects and keep the link with bibliography. After 1972, there was the CSI (Catalogue of Stellar Index) and the BSI (Bibliographical Star Index). SIMBAD was born from the fusion of the two (Egret 1983). We had only stars in the database, and these were from just a few catalogues. The data were restricted to: object types, object names, coordinates, magnitudes in B and V bands, spectral types, proper motions and some other measurements (as illustrated in Figure 2). There were few publications taken into account (dating back to 1950, and only those with stars). There were few people at CDS, but many collaborators from other institutes participated to construct SIMBAD¹⁰.

Today in SIMBAD we find much more data: object types, coordinates, magnitudes, radial velocities and/or redshifts, parallaxes, proper motions, spectral and morphological types, hierarchies between objects, occurrences and positions of the objects in articles. There are more fields, and better quality and precision (see Figure 3). All the CDS services can interact together (SIMBAD, Vizier and Aladin, see Wenger et al. (1996)) and with the Virtual Observatory tools (see Genova et al. (2001), Genova (2007)).

A good example of how documentalists have to adapt is to consider the evolution of object types in SIMBAD. At the beginning there were only stars in the CSI. Galaxies were added in 1983 (Dubois et al. 1983). Then there were only 4 object types: stars, galaxies, stars in galaxies and unknown. In the 90s there were already 100 object types, and today we have more than 200 object types in SIMBAD¹¹. The new object types classification provides hierarchy between the different types and a complex tree structure (Figure 4). To deal with all the object types in SIMBAD today requires a rather good knowledge of astronomy!

3.2. Evolutions in Vizier

As for Vizier, the documentalists need to deal with various data. Catalogues are larger and larger, which means dealing with more and more rows and columns in tables (today we can have as many as 2 billion rows in one table). There are more data, which are more and more precise. The documentalists have to describe each column, each data, which format and units are used, to standardize them, to add notes, Read Me, Unified Content Descriptors, links. . . They have to deal with various content (time series, surveys) and different formats, and to homogenize the data to create standardized tables and other output formats. We also add data associated to publications in Vizier, like spectra, light curves, images. . . which have to be curated through downloads, plots and so on.

A good example of working mode adaptation are the UCDs (for Unified Content Descriptors) used for Vizier tables since 1998. The UCDs give a semantic description of columns contents. This is a list of standard words and combinations to express semantic meaning. This allows one to find, select and compare data through the different

¹⁰ Among them: the Astronomisches Rechen Institut of Heidelberg (for astrometry and stars' positions), Geneva Observatory and the Institut d'Astronomie de l'Université de Lausanne (photometric data), Marseilles Observatory (radial velocities), La Plata Observatory (spectral classification), Paris-Meudon Observatory (bibliography), Institut d'Astrophysique de Paris (IAP, extragalactic), Besançon Observatory (double stars). Non-exhaustive list (F. Ochsenbein, private communication).

¹¹ <http://simbad.u-strasbg.fr/simbad/sim-display?data=otypes>

V* BZ UMa -- Cataclysmic Variable Star

Other object types: **DN*** (), **X** (RBS,RX,...), * (SBSS,USNO,...), **V*** (V*,AAVSO), **CV*** (2019MNRAS), **V*?** (SV*), **IR** (2MASS), **UV** (PG)

ICRS coord. (ep=j2000) : 08 53 44.1734051661 +57 48 40.595271757 (Optical) [0.0393 0.0391 90] A 2018yCat.1345....0G

FK4 coord. (ep=B1950 eq=1950) : 08 49 52.3021477380 +58 00 04.121281981 [0.0393 0.0391 90]

Gal coord. (ep=j2000) : 159.0025580313121 +38.8332773744221 [0.0393 0.0391 90]

Proper motions mas/yr : 26.411 -16.016 [0.082 0.090 90] A 2018yCat.1345....0G

Radial velocity / Redshift / cz : V(km/s) 18 [0] / z(-) 0.00006 [0.00000] / cz 18.0 [0.0] (Opt) C 2012ApJS...203...21A

Parallax (mas): 6.5572 [0.0638] A 2018yCat.1345....0G

Spectral type: M5.5:Ve D 1994MNRAS.271..323R

Fluxes (13) :

B	15.95	[0.37]	E	2001yCat.1271....0S
V	17.58	[0.28]	E	2008AJ...136..735L
R	14.98	[0.40]	E	2001yCat.1271....0S
G	16.2056	[0.0166]	C	2018yCat.1345....0G
I	15.38	[0.38]	E	2008AJ...136..735L
J	14.824	[0.043]	C	2003yCat.2246....0C
H	14.440	[0.062]	C	2003yCat.2246....0C
K	14.085	[0.056]	C	2003yCat.2246....0C
u (AB)	15.974	[0.005]	B	2009yCat.2294....0A
g (AB)	16.377	[0.003]	B	2009yCat.2294....0A
r (AB)	16.050	[0.004]	B	2009yCat.2294....0A
i (AB)	16.074	[0.004]	B	2009yCat.2294....0A
z (AB)	15.880	[0.007]	B	2009yCat.2294....0A

SIMBAD with radius 2 arcmin



VizieR photometry viewer within radius arcsec

Identifiers (16) :

An access of full data is available using the icon Vizier near the identifier of the catalogue

SV* SVS 1719	RBS 728	SDSS J085344.00+574841.0	V* BZ UMa
LINEAR 5482517	RX J0853.7+5748	SDSS J085344.17+574840.6	[ZEH2003] RX J0853.7+5748 1
2MASS J08534416+5748406	IRXS J085343.5+574846	TIC 316276691	AAVSO 0846+58
PG 0849+580	SBSS 0849+580	USNO-B1.0 1478-00239954	Gaia DR2 1037411284055821568

Plots and Images



Figure 3. The star PG 0849+580 in SIMBAD today. Only the first page is displayed here, with the measurements and data associated to the object: 16 identifiers, links to have access to images and plots of this object. Today in SIMBAD there are 139 bibliographical references linked to the object (on the increase) and more extra-data available.

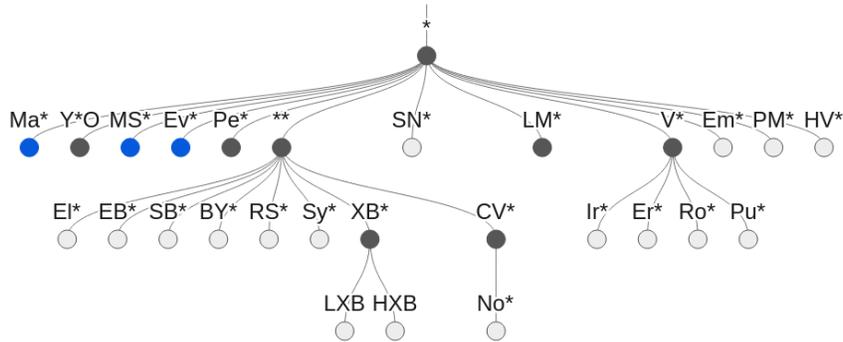


Figure 4. An excerpt of the new object types classification for SIMBAD.

87 tables. This is also a tool to manage data. They are used to validate catalogue descriptions, and also used since 2004
88 as a standard for VOTables in the Virtual Observatory (Derriere et al. 2005).

89 4. A HIGH IMPACT ON SCIENCE

90 Standards have been used for many years in astronomy to describe bibliography and data. A few examples are
91 Bibcodes (Schmitz et al. 1995), UCDs, README in Vizier, and the FITS format. These standards existed long before
92 the Virtual Observatory era. In a first step they were created as tools to manage or exchange data, and they were largely
93 used internationally. As interoperability requires standards, some of them became Virtual Observatory standards, and
94 thus participate in the FAIRisation of astronomical data. Standards are a good example of the impact that data

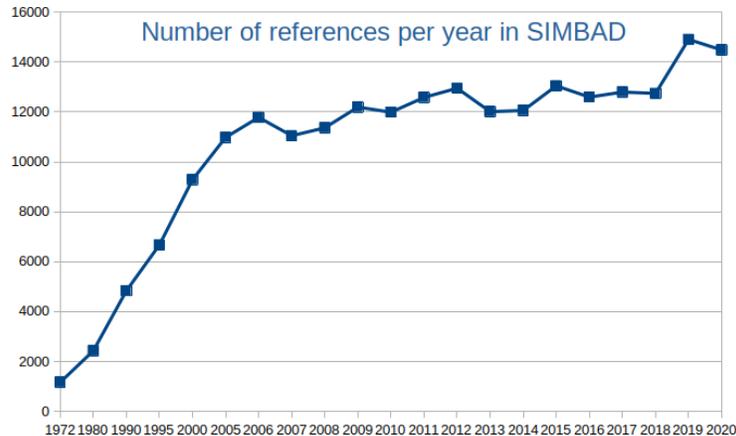


Figure 5. Increase in the number of bibliographical references per year in SIMBAD.

95 stewards' practices may have on science. Indeed, if science has an impact on documentalists activities, the work done
 96 by documentalists has also definitely an impact on science. Since the beginning of the CDS, the work regularly done
 97 by the documentalists allows to check the data, homogenise them, compare them and detect errors in the publications.
 98 These good practices ensure the accuracy and the quality of the data, and also provide a high level of added value to
 99 the data. They also enable data to be Findable, Accessible, Interoperable, and Reusable.

100 5. IMPACT ON DOCUMENTALISTS' ACTIVITIES

101 Today, the process from data reception to completion of data curation at the CDS is quite complicated and involves
 102 several services, which means several kinds of documentalists, with different specialisations. The CDS is a cooperative
 103 team of around 38 people, equally divided into astronomers, documentalists and computer scientists (Perret et al.
 104 2015). A team work was developed over the years to adapt to the evolution of both the quality and quantity of data.
 105 During the process there are ongoing interactions with the IT engineers to improve and/or develop specific tools, and
 106 ongoing interactions with the astronomers to ensure data and metadata quality and select the more relevant data.
 107 Documentalists share their expertise in-house, with common documentation, the sharing of good and new practices,
 108 regular meetings, trainings and seminars. They constantly adapt their practices to maintain the workflow and quality
 109 of the services.

110 To summarise the impact of astronomy evolution during the past decades on the documentalists' activities at CDS,
 111 one first needs to mention the significant increase in the number of published articles and data (Figure 5 and Woelfel
 112 et al. (2007)). Along with this constraint, data stewards have to get a deep understanding of the different aspects
 113 of their work, that means: data understanding, identification, selection and verification of data, homogenisations,
 114 descriptions and corrections of data. Their expertise has to evolve constantly by gaining knowledge of new data and
 115 new science topics, adaptation to new tools and new formats, new ways of performing work and new team organisation.
 116 This means permanent learning, adaptation and knowledge development. Today we have more tools to deal with data,
 117 but also more data to curate, and despite the extraordinary developments of technology and tools, the data stewards
 118 expertise remains essential and mandatory, in particular for some parts of the work that remain non-automatable, for
 119 which more knowledge and analytical skills are required from data stewards.

120 6. CONCLUSION

121 Science and technology have evolved together during the past decades, producing data and permitting to store,
 122 curate and share it. At the core of this process, the work of data stewards is essential, it involves curating data,
 123 producing added value, using and spreading standards, as well as producing FAIR data. This enables Open Science,
 124 and the availability of data has been revolutionising scientific working methods. The data curated in this way are
 125 useful for science and have an high impact on its development. Thus the daily work of documentalists evolves with
 126 science, but is also entirely part of its evolution.

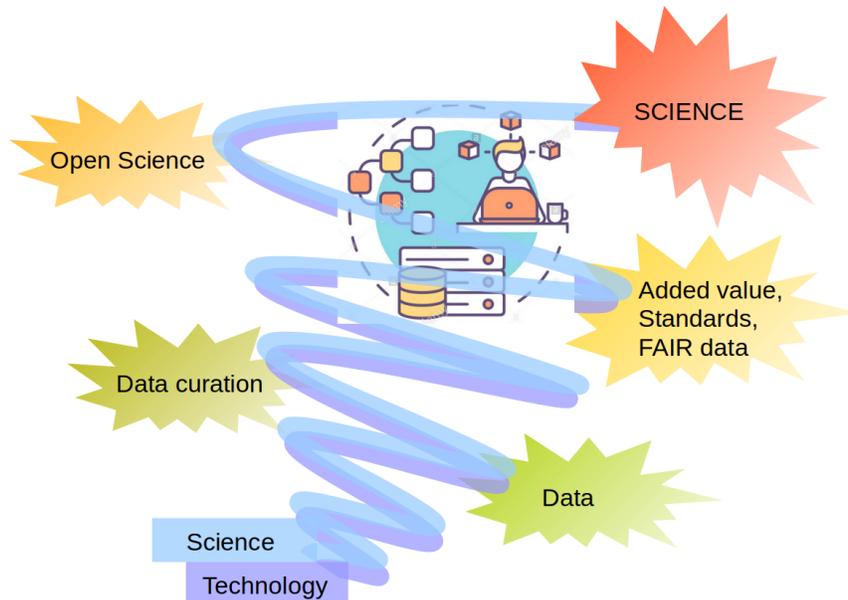


Figure 6. The evolution of astronomy and technologies have a deep impact on data stewards practices, but these practices also have an impact on science.

We want to thank here some retired people of the CDS, who participated to this paper by giving us information about the beginning and history of the CDS: François Ochsenbein, Marc Wenger and Pascal Dubois. We also would like to remind that the CDS as we know it today is the result of 50 years of continuous team work and collaborations, so we have a thought for all our collaborators around the world and throughout the time.

REFERENCES

- Bidelman, W. P. 1988, *Bulletin d'Information du Centre de Donnees Stellaires*, 35, 52
- Boch, T., & Derriere, S. 2010, in *Astronomical Society of the Pacific Conference Series*, Vol. 434, *Astronomical Data Analysis Software and Systems XIX*, ed. Y. Mizumoto, K. I. Morita, & M. Ohishi, 221
- Boch, T., Pineau, F., & Derriere, S. 2012, in *Astronomical Society of the Pacific Conference Series*, Vol. 461, *Astronomical Data Analysis Software and Systems XXI*, ed. P. Ballester, D. Egret, & N. P. F. Lorente, 291
- Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, 143, 33, doi: [10.1051/aas:2000331](https://doi.org/10.1051/aas:2000331)
- Derriere, S., Preite Martinez, A., Williams, R., et al. 2005, *An IVOA Standard for Unified Content Descriptors Version 1.10*, IVOA Recommendation 19 August 2005, doi: [10.5479/ADS/bib/2005ivoa.spec.0819D](https://doi.org/10.5479/ADS/bib/2005ivoa.spec.0819D)
- Dubois, P., Ochsenbein, F., & Paturel, G. 1983, *Bulletin d'Information du Centre de Donnees Stellaires*, 24, 125
- Egret, D. 1983, *Bulletin d'Information du Centre de Donnees Stellaires*, 24, 109
- Genova, F. 2007, in *JENAM-2007, "Our Non-Stable Universe"*, 77–77
- Genova, F., Bonnarel, F., Dubois, P., et al. 2001, in *Astronomical Society of the Pacific Conference Series*, Vol. 225, *Virtual Observatories of the Future*, ed. R. J. Brunner, S. G. Djorgovski, & A. S. Szalay, 176
- Lortet, M. C., Borde, S., & Ochsenbein, F. 1994, *A&AS*, 107, 193
- Ochsenbein, F. 1984, *Bulletin d'Information du Centre de Donnees Stellaires*, 27, 203
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23, doi: [10.1051/aas:2000169](https://doi.org/10.1051/aas:2000169)
- Ochsenbein, F., Bischoff, M., & Egret, D. 1981, *A&AS*, 43, 259

- 165 Perret, E., Boch, T., Bonnarel, F., et al. 2015, in
166 Astronomical Society of the Pacific Conference Series,
167 Vol. 492, Open Science at the Frontiers of Librarianship,
168 ed. A. Holl, S. Lesteven, D. Dietrich, & A. Gasperini, 13
- 169 Pineau, F.-X., Boch, T., Derrière, S., & Schaaff, A. 2020, in
170 Astronomical Society of the Pacific Conference Series,
171 Vol. 522, Astronomical Data Analysis Software and
172 Systems XXVII, ed. P. Ballester, J. Ibsen, M. Solar, &
173 K. Shortridge, 125
- 174 Schmitz, M., Helou, G., Lague, C., et al. 1995, *Vistas in*
175 *Astronomy*, 39, 272
- 176 Wenger, M., Ochsenbein, F., Bonnarel, F., Lesteven, S., &
177 Oberto, A. 2006, in Astronomical Society of the Pacific
178 Conference Series, Vol. 351, Astronomical Data Analysis
179 Software and Systems XV, ed. C. Gabriel, C. Arviset,
180 D. Ponz, & S. Enrique, 662
- 181 Wenger, M., Fernique, P., Genova, F., et al. 1996, in
182 American Astronomical Society Meeting Abstracts, Vol.
183 189, American Astronomical Society Meeting Abstracts,
184 06.02
- 185 Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*,
186 143, 9, doi: [10.1051/aas:2000332](https://doi.org/10.1051/aas:2000332)
- 187 Woelfel, F., Genova, F., Bonnarel, F., et al. 2007, in
188 Astronomical Society of the Pacific Conference Series,
189 Vol. 377, Library and Information Services in Astronomy
190 V, ed. S. Ricketts, C. Birdie, & E. Isaksson, 43