

# Les documentalistes et la FAIRisation des données scientifiques : un travail d'équipe inter-métiers

"Interopérabilité et pérennisation des données de la recherche: comment FAIR en pratique ? Retours d'expériences"

Paris, le 27 novembre 2018



Soizick Lesteven  
pour l'équipe du CDS

# □ CDS : ses objectifs

- 46 ans au service de la communauté scientifique
- Depuis 1972, les objectifs du CDS sont :
  - Collecter l'information « utile » sur les objets astronomiques au format électronique
  - Améliorer cette information par évaluations critiques et comparaisons
  - Distribuer cette information à la communauté astronomique internationale
  - Conduire des recherches en utilisant ces données



# □ CDS : ses services



- La base de données de référence mondiale pour l'identification, la nomenclature et la bibliographie des objets astronomiques



- La collection des catalogues astronomiques et des tables publiées dans les journaux de la discipline mais aussi les données associées



- L' Atlas interactif du ciel, la collection d'images de référence

→ découverte, visualisation et manipulation de données des services de l'OV y compris CDS



- Le dictionnaire de nomenclature des objets célestes

→ un million de requêtes par jour sur l'ensemble des services

# □ CDS : son contenu (octobre 2018)



- ~ 9.6 millions d'objets,
- ~ 32 millions d'identificateurs,
- ~ 300.000 références bibliographiques



- 17.673 catalogues (38.000 tables)
- ~ 27 milliards de sources



- 575 relevés d'images de référence du ciel
- ~ (203 TB pixels)



- ~ 24.000 acronymes

# □ CDS : son organisation

- Le CDS est une infrastructure de recherche de la feuille de route nationale, qui dépend du CNRS et de l'Université de Strasbourg, incluse dans l'Observatoire Astronomique de Strasbourg.
- Le CDS est une « équipe intégrée » de 34 personnes :
  - Des **chercheurs**, qui connaissent l'astronomie, son évolution et les besoins des utilisateurs
  - Des **documentalistes** qui ont en charge la création du contenu des bases de données
  - Des **informaticiens** qui créent les systèmes, développent les outils et les interfaces utilisateurs et participent activement à l'OV
- Un travail sur le long terme
  - Les différents services se sont construits au fil du temps en fonction des besoins de la communauté astronomique et de l'évolution de la science et des techniques
  - Evolution du travail et des compétences
  - Anticipation sur les besoins et sur les évaluations des techniques disponibles à chaque étape

# □ CDS : son contexte

Participations à différents projets/missions



Participations à l'OV



Collaborations avec les centres de données



CADC, MAST, HEASARC, IPAC, + ...

Gestion des données



Collaborations avec les journaux



# □ CDS : ses documentalistes

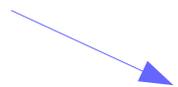
- Ingestion des données :
  - A partir des articles publiés dans les journaux de notre discipline
    - Recherche des objets astronomiques avec leurs données pour Simbad
    - Intégration des tables et autres données associées pour Vizier
    - Sélection de tables pour ajouter les données dans Simbad
  - Maintien du dictionnaire de nomenclature de l'UAI
  - Description des images du ciel pour Aladin
  - Avec le soutien des astronomes de l'équipe pour le choix des données à ingérer, l'expertise, la validation

**Table 2**  
SDSS Data for CSS RRab Stars

ID	$u_0$	$g_0$	$r_0$	$i_0$	$z_0$	$V_{\text{gsr}}$ ( $\text{km s}^{-1}$ )	[Fe/H]
CSS_J004056.6-020802	18.19	17.03	16.76	16.69	16.63	...	...
CSS_J003621.0-015958	17.55	16.45	16.15	16.08	16.07	...	...
CSS_J004212.4-004251	18.22	17.16	16.91	16.82	16.81	123.7	-2.19
CSS_J004424.5-002743	18.97	17.88	17.59	17.51	17.49	-78.6	-1.47
CSS_J005150.9-024858	17.05	15.96	15.74	15.69	15.68	...	...
CSS_J005328.6-004321	19.03	17.84	17.59	17.57	17.55	-134.1	-1.71
CSS_J004923.6-001800	18.98	17.76	17.51	17.48	17.44	-89.5	-1.20
CSS_J005338.1-000303	17.76	16.73	16.44	16.42	16.38	...	-1.92
CSS_J010533.7-002344	19.30	18.20	17.96	17.91	17.88	...	-1.82
CSS_J011742.0-020819	17.11	16.02	15.82	15.74	15.75	...	...
CSS_J011723.6-020434	18.50	17.38	17.12	17.02	16.99	...	...
CSS_J011046.4-020214	18.71	17.61	17.41	17.41	17.35	...	...
CSS_J012924.9-024121	19.23	18.13	17.91	17.82	17.85	...	...
CSS_J012159.8-014415	19.21	18.07	17.80	17.71	17.71	...	...
CSS_J012206.6-011023	17.65	16.53	16.40	16.33	16.36	...	-2.32

**Notes.** Column 1: gives the CSS ID; Column 2–6: give the extinction-corrected SDSS magnitudes of the RRab; Column 7: gives the velocity in the galactic standard of rest, based on SDSS spectra; Column 8: gives the metallicity based on SDSS spectra.

(This table is available in its entirety in a machine-readable form in the [online journal](#). A portion is shown here for guidance regarding its form and content.)

 7.760 lignes

Summary | **ReadMe** | VizieR | Browse | FTP | Tar

J/ApJ/763/32 Galactic halo RRab stars from CSS (Drake+, 2013)

Probing the outer Galactic halo with RR Lyrae from the Catalina Surveys. Drake A.J., Catelan M., Djorgovski S.G., Torrealba G., Graham M.J., Belokurov V., Koposov S.E., Mahabal A., Prieto J.L., Donalek C., Williams R., Larson S., Christensen E., Beshore E. <Astrophys. J., 763, 32 (2013)> =2013ApJ...763...32D

**ADC Keywords:** Stars, variable ; Stars, distances ; Photometry, SDSS ; Surveys

**Keywords:** galaxies: stellar content; Galaxy: formation; Galaxy: stellar content; Galaxy: structure; stars: variables: RR Lyrae

**Abstract:**

**Description:**  
The Catalina Sky Survey began in 2004 and uses three telescopes to cover the sky between declination 6°-75° and +65° in order to discover near-Earth objects (NEOs) and potential hazardous asteroids (PHAs). Each of the survey telescopes is run as separate sub-surveys. These consist of the Catalina Schmidt Survey (CSS) and the Mount Lemmon Survey (MLS) in Tucson Arizona, and the Siding Spring Survey (SSS) in Siding Spring, Australia.

**File Summary:**

FileName	Lencl	Records	Explanations
ReadMe	80	.	This file
<a href="#">table1.dat</a>	118	12379	Parameters of RRab stars
<a href="#">table2.dat</a>	63	7760	SDSS data for CSS RRab stars

**See also:**

- [B/Vsx](#) : AAVSO International Variable Star Index VSX (Watson+, 2006-2014)
- [II/294](#) : The SDSS Photometric Catalog, Release 7 (Adelman-McCarthy+, 2009)
- [J/AJ/144/9](#) : Catalina Surveys periodic variable stars (Drake+, 2014)
- [J/AJ/144/114](#) : Radial velocities of 6 field RR Lyrae (Sesar+, 2012)
- [J/AJ/144/39](#) : Galactic RRab from the GEOS RR Lyr Survey (Le Borgne+, 2012)

[0.1/72.1] Heliocentric distance ([pos.distance](#))

[0.006/6.6] V band extinction ([Note 2](#)) ([phys.absorption;em.opt.V](#))

Ephemeris MJD ( $\eta$ ) ([time.epoch](#))

**Byte-by-byte Description of file:** [table2.dat](#)

Bytes	Format	Units	Label	Explanations
1- 4	A4	---	---	[CSS_] CSS identifier;
5- 20	A16	---	ID	<[DCD2013] CSS J+HMSS.s+DDMSS> in Simbad
22- 26	F5.2	mag	u0mag	[10.1/25] Extinction corrected SDSS u band magnitude
28- 32	F5.2	mag	g0mag	[10.3/25.1] Extinction corrected SDSS g band magnitude
34- 38	F5.2	mag	r0mag	[11/24.7] Extinction corrected SDSS r band magnitude
40- 44	F5.2	mag	i0mag	[11/24.4] Extinction corrected SDSS i band magnitude
46- 50	F5.2	mag	z0mag	[11.5/22.8] Extinction corrected SDSS z band magnitude
52- 57	F6.1	kn/s	Vgsr	[-359/398]? Galactic standard of rest velocity ( <a href="#">1</a> )
59- 63	F5.2	[Sun]	[Fe/H]	[-3/-0.7]? Metallicity ( <a href="#">1</a> )

**Note (1)** : Based on SDSS spectra. The galactic standard of rest assumes the solar peculiar motion of (U,V,W) = (9, 12+220, 7)km/s

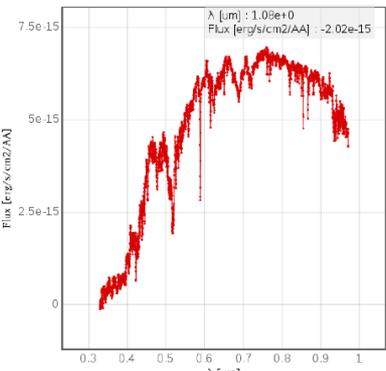
**History**  
From electronic version of the journal

**Reference:**  
Drake et al. Paper II. [2013ApJ...765...154D](#) Cat. [J/ApJ/765/154](#)

(End) Greg Schwarz [AAS], Emmanuelle Perret [CDS] 09-Oct-2014

**UCD** (Ultra Compact Disk) →

**plots** →



Mise à jour des données + un gros travail d'expertise pour les documentalistes !  
 Pour chaque catalogue : Description, standardisation, , vérification, metadonnées et données ajoutées (UCD, KW, filters, coordonnées, liens, plots, ...)

# □ CDS : ses documentalistes

- Ingestion des données mais surtout
  - Valorisation des données
    - Identifier, sélectionner, vérifier, homogénéiser, comparer, analyser, corriger ....
    - décrire, cross-identifications, ajouts des métadonnées, des données associées
- ⇒ Apporter une forte valeur ajoutée aux données

# □ CDS : ses documentalistes

Consulter les autres Bases de Données

Poser la question aux astronomes

Ecrire aux auteurs

En discuter avec les autres documentalistes de l'équipe



Participer aux réunions de travail

Rechercher dans d'autres articles

Proposer des améliorations des outils logiciels

Compter sur son expérience qui s'enrichit continuellement

# □ CDS : FAIR

- ⇒ Fort impact sur la recherche
  - F** : Données fiables, bien décrites, correctement répertoriées
  - A** : Données accessibles par tous via les services du CDS
  - I** : Données « OV compatibles » donc interopérables avec d'autres données et d'autres logiciels et via des pipelines
  - R** : Données facilement réutilisables
- Cela est possible grâce à la complémentarité des profils de « l'équipe intégrée » de documentalistes, astronomes et informaticiens.