

Les UCDs - métadonnées pour l'astronomie

Sébastien Derriere

CDS, Observatoire astronomique de Strasbourg

derriere@astro.u-strasbg.fr

Contexte: données et métadonnées en astronomie

Les astronomes doivent faire face à de grands volumes de données hétérogènes:

- articles, publications
- catalogues (données tabulaires)
- spectroscopie
- images, relevés du ciel

Les données sont produites par de nombreux projets: comment les comparer?

Catalogues astronomiques (tables):

214.386166	-57.767818	16.926	15.777	99.999	0.09	0.19	9.99	17.067	15.508	99.999
214.535889	-57.767764	16.458	15.562	99.999	0.07	0.17	9.99	16.496	15.457	99.999
214.401036	-57.767685	14.974	14.391	99.999	0.04	0.11	9.99	15.021	14.549	99.999
214.569711	-57.767623	17.971	15.777	99.999	0.18	0.19	9.99	17.394	15.553	99.999
214.349915	-57.767576	16.975	99.999	99.999	0.10	9.99	9.99	16.840	99.999	99.999
214.550993	-57.767487	16.801	15.716	99.999	0.09	0.18	9.99	16.605	15.682	99.999
214.557370	-57.767406	99.999	16.525	13.594	9.99	0.27	0.22	99.999	15.544	12.905
214.404212	-57.767370	15.848	14.973	99.999	0.05	0.13	9.99	15.654	15.197	99.999
214.296113	-57.767262	15.161	13.266	99.999	0.04	0.08	9.99	15.055	13.271	99.999
214.238914	-57.767254	15.363	14.061	99.999	0.04	0.10	9.99	14.916	14.106	99.999
214.286765	-57.767228	15.694	13.984	99.999	0.05	0.09	9.99	15.784	14.019	99.999
214.595510	-57.767131	17.716	16.170	99.999	0.15	0.23	9.99	17.274	16.150	99.999
214.466317	-57.767040	15.975	13.680	12.353	0.06	0.09	0.12	15.998	13.686	12.836
214.503014	-57.767008	17.436	99.999	99.999	0.13	9.99	9.99	17.548	99.999	99.999
214.483010	-57.766971	99.999	16.015	99.999	9.99	0.21	9.99	99.999	16.370	99.999
214.470701	-57.766933	16.031	13.917	12.708	0.06	0.09	0.14	16.025	13.909	13.395
214.382021	-57.766657	18.085	99.999	99.999	0.19	9.99	9.99	19.044	99.999	99.999
214.261358	-57.766521	17.167	15.984	99.999	0.11	0.20	9.99	17.209	16.248	99.999
214.611797	-57.766361	17.103	15.149	99.999	0.10	0.14	9.99	16.741	14.812	99.999
214.261771	-57.766321	17.388	15.712	99.999	0.13	0.18	9.99	16.713	15.023	99.999
214.531274	-57.766314	16.179	14.323	13.036	0.06	0.11	0.16	16.099	14.255	13.102
214.561229	-57.766304	17.249	15.955	99.999	0.11	0.20	9.99	17.213	17.242	99.999
214.251734	-57.766279	16.455	14.660	13.244	0.07	0.12	0.18	16.426	14.683	12.860
214.491574	-57.766278	17.009	14.856	13.533	0.10	0.13	0.22	16.628	14.731	13.337
214.591738	-57.766200	17.174	99.999	99.999	0.11	9.99	9.99	16.903	99.999	99.999
214.311793	-57.766161	15.516	14.837	99.999	0.05	0.12	9.99	15.496	14.697	99.999
214.341976	-57.766060	99.999	15.804	99.999	9.99	0.19	9.99	99.999	15.705	99.999
214.271051	-57.766037	16.967	15.725	99.999	0.10	0.18	9.99	17.014	16.130	99.999
214.581047	-57.765921	15.941	13.637	12.239	0.06	0.09	0.12	15.858	13.620	12.295
214.421628	-57.765900	99.999	16.214	13.928	9.99	0.23	0.27	99.999	16.593	13.955
214.411201	-57.765870	15.641	14.798	99.999	0.05	0.12	9.99	15.604	14.415	99.999
214.511290	-57.765842	99.999	16.320	99.999	9.99	0.24	9.99	99.999	18.163	99.999
214.461607	-57.765814	17.931	15.866	99.999	0.17	0.19	9.99	17.819	15.194	99.999
214.551082	-57.765781	15.537	12.979	11.402	0.05	0.08	0.09	15.540	12.948	11.385
214.471125	-57.765746	17.131	99.999	99.999	0.20	9.99	9.99	18.831	99.999	99.999

RA

DEC

magI magJ magK



Tables astronomiques

Une description avec des noms de colonnes ne suffit pas!

Le service **VizieR** du centre de Données astronomiques de Strasbourg (**CDS**) donne accès à 4000 catalogues différents.

Chaque catalogue peut contenir plusieurs tables. Une table peut contenir quelques centaines de colonnes - La plus grande contient un milliard d'objets.

Les origines des tables sont diverses :

- publication électronique de tables des journaux
- numérisation d'anciennes données (plaques photo)
- observations récentes avec détecteurs numériques

Description des tables

Dans VizieR, une description standardisée est fournie pour chaque catalogue (fichier **ReadMe**), mais:

- Origines, formats et descriptions hétérogènes.
- Chaque auteur peut nommer les colonnes à sa guise (120 noms pour une magnitude V dans VizieR) et utiliser n'importe quelle unité.
- A priori, **aucune comparaison automatisée n'est disponible** entre différentes tables.

Comment rechercher des catalogues d'après leur contenu ?

« Trouver les catalogues contenant une mesure de mouvement propre, ou une magnitude Johnson B... »

Unified Content Descriptors

Les **UCDs** (Unified Content Descriptors) ont été créés pour fournir une **description sémantique** du contenu des colonnes.

Origine Bottom-Up : exploration manuelle de centaines de catalogues dans VizieR (P. Ortiz)

A chaque nouveau concept, un nouvel UCD était créé.
UCDs: chaîne de caractères avec une définition:

PHOT_EXTINCTION_ISM Interstellar extinction.
POS_GAL_LAT Galactic latitude

Unified Content Descriptors

~1500 UCDs suffisent pour décrire le contenu de 100,000 colonnes dans VizieR.

UCDs construits dans une structure hiérarchique à 4 niveaux

La structure n'est pas fondamentale, c'était juste:

- un moyen simple de regrouper des éléments suffisamment proches au niveau du sens
- un moyen de spécifier le contexte (rendre explicite de l'information implicite)

Les UCDs ne sont pas un modèle de données, ni une ontologie!

Utilisation et outils

Des outils ont été développés pour démontrer comment les UCDs peuvent être utilisés pour sélectionner des catalogues, comparer des données, etc...

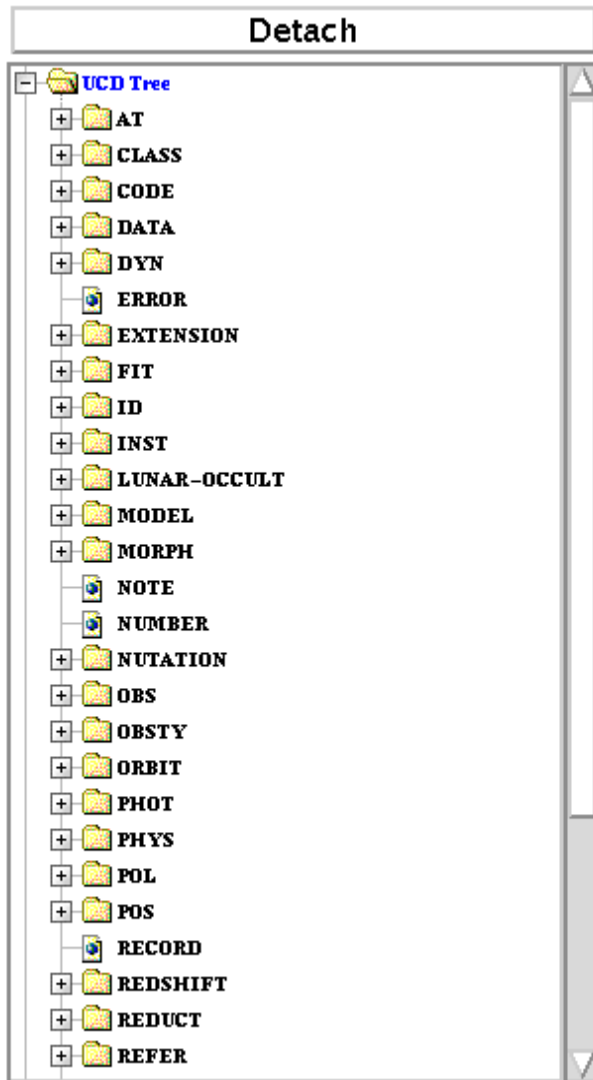
<http://vizier.u-strasbg.fr/UCD/>

Les outils incluent un browser pour explorer l'arbre, et la recherche d'UCDs à partir de texte.

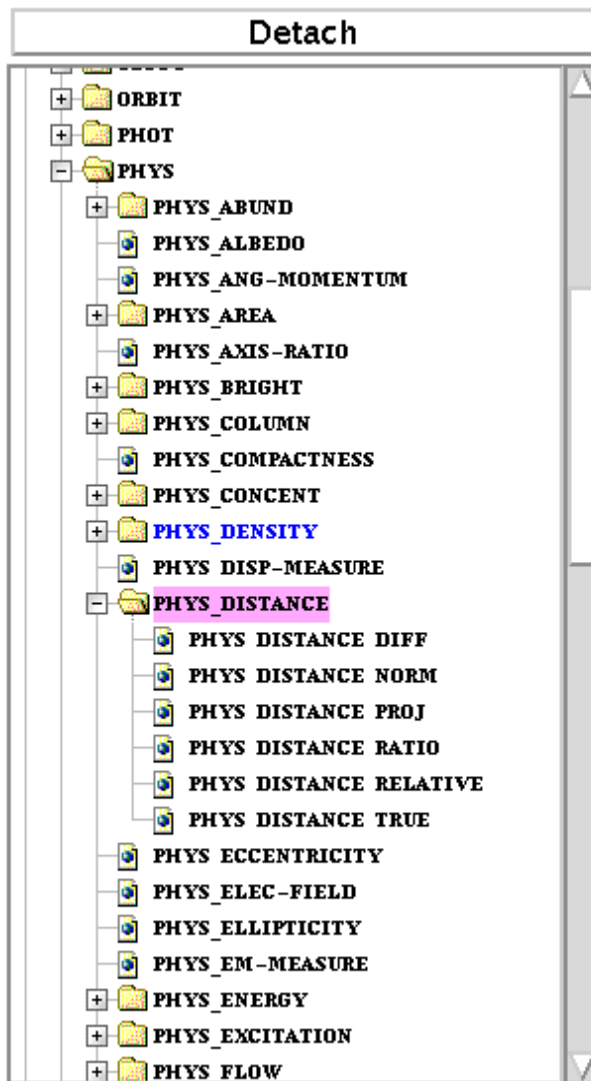
UCD browser

UCD Browser

The left frame displays the UCD tree. Clicking on a leaf will display some informations about it in the right frame.



UCD browser



UCD information

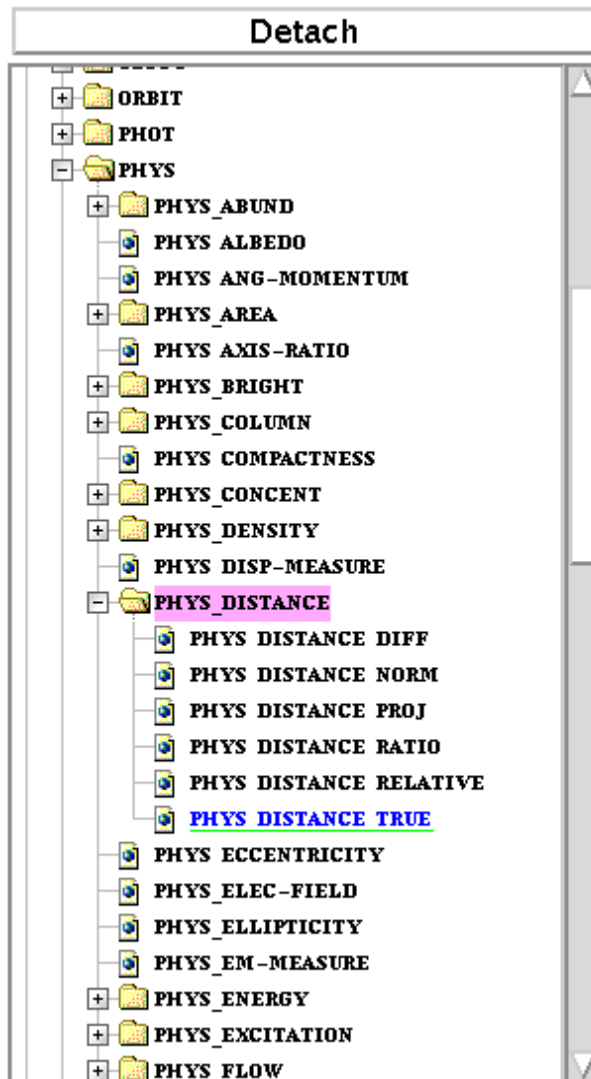


[AVO](#) · [ESO](#) · [ST-ECF](#) · [AstroGrid](#) · [CDS](#) · [Terapix](#) · [Jodrell Bank](#)

UCD **PHYS_DISTANCE** represents: **Distance and related quantities**

PHYS_DISTANCE is node #945 at level 1 in the UCD tree (roots are level 0).
There are 6 elements under this node.

UCD browser



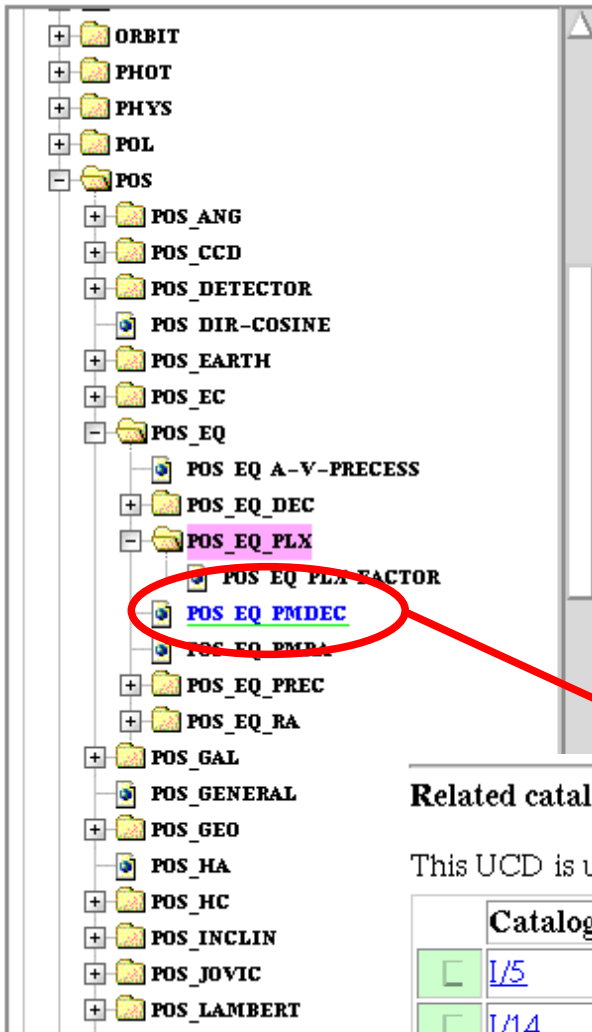
UCD **PHYS_DISTANCE_TRUE** represents: **Distance (true, linear distance)**

PHYS_DISTANCE_TRUE is leaf #951 at level 2 in the UCD tree (roots are level 0).

Statistics for this UCD:

Column names and units associated to UCD: **PHYS_DISTANCE_TRUE**
(there are 127 different column names and 11 different units).

Frequency:	column name	Frequency:	unit
137	Dist	192	kpc
18	R	141	pc
18	r	65	Mpc
17	D	13	AU
12	dist	2	100pc
12	d	2	10+6km
9	Z	2	---
8	z	1	km
7	Dhel	1	km/s
6	Dist2	1	10-4
6	Dgal	1	al
5	zDist		
4	Y		
4	xDist		
4	yDist		
4	CODIST		
4	HDist		
3	Rg		
3	Distance		
3	Rdist		
3	Xdist		
3	Ydist		



Trouver des catalogues intéressants à partir des UCDs

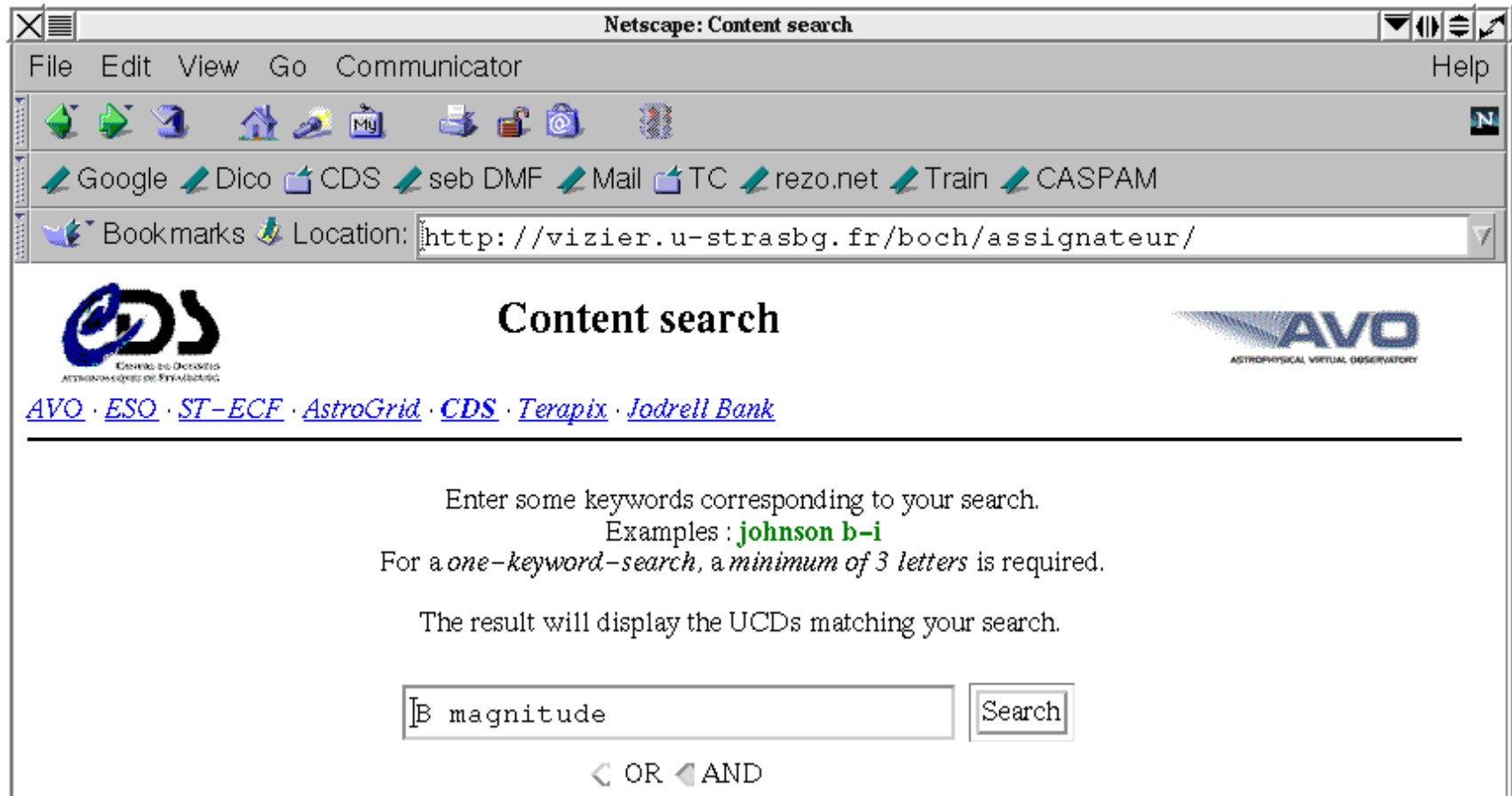
Related catalogues in VizieR:

This UCD is used in 209 columns, in 134 different catalogues (187 tables) of VizieR.

	Catalogue	Title	Bibcode
<input type="checkbox"/>	I/5	Proper Motions in Cape Zone Catalogue -40/-52 (Spencer Jones H.+ 1936)	
<input type="checkbox"/>	I/14	Proper Motions of 1160 Late-Type Stars (Fogh Olsen, 1970)	1970A&AS....2....69O
<input type="checkbox"/>	I/40	WASHINGTON 20 Catalog (Morgan, 1933)	
<input type="checkbox"/>	I/61B	AGK3 Catalogue (Dieckvoss, Heckmann 1975)	1975QB6..A15.....D
<input type="checkbox"/>	I/62C	Perth 70: Positions of 24900 Stars (Hog+ 1976)	1976AAHam...9....1H
<input type="checkbox"/>	I/68A	Positions and Proper Motions in alpha Per cluster (Fresneau, 1980)	1980BICDS..18...81F

Recherche à partir de texte

Ex: magnitude B







The screenshot shows a Netscape browser window titled "Netscape: Content search". The address bar contains the URL "http://vizier.u-strasbg.fr/boch/assignateur/". The page content includes the CDS logo (Centre de Données Astronomiques de Strasbourg) and the AVO logo (Astrophysical Virtual Observatory). Below the logos, there is a navigation menu with links to AVO, ESO, ST-ECF, AstroGrid, CDS, Terapix, and Jodrell Bank. The main content area contains instructions for searching: "Enter some keywords corresponding to your search. Examples : johnson b-i. For a one-keyword-search, a minimum of 3 letters is required. The result will display the UCDs matching your search." Below this text is a search input field containing "B magnitude" and a "Search" button. At the bottom of the search area, there are radio buttons for "OR" and "AND".

Results for the search **B MAGNITUDE** :

- [UCD Leaves](#)
- [Related categories](#)

Suggested UCD leaves

UCD	Description	Relevance	Associated catalogues (limited to 5)
PHOT_MAG_B	Blue B magnitude	 66 %	I/121 I/125A I/209A I/219 I/222
PHOT_JHN_B	Johnson magnitude B (JHN)	 50 %	I/197A I/207 II/7A II/117 II/207
PHOT_STR_B	Stroemgren Magnitude b (STR)	 50 %	II/170 II/188 II/200 J/A+A/294/135 J/A+A/304/415
PHOT_PHG_B	Photographic blue magnitude B (includes the O magnitude of POSS)	 50 %	I/80 I/129 I/199 I/200

Allows contents-based catalogue selection

Comparaisons automatiques

Si deux colonnes sont décrites par le même UCD, elles peuvent être comparées !

UCD + unités = **Conversions de colonnes**

[AVO](#) · [ESO](#) · [ST-ECF](#) · [AstroGrid](#) · [CDS](#) · [Terapix](#) · [Jodrell Bank](#)

I/146/ppm1 Positions and Proper Motions – North (Roeser+, 1988)
Catalogue PPM-North

_r	recno	PPM	DM	Mag	Sp	RAJ2000	DEJ2000	pmRA	pmDE	Npos	e_RAs	e_DEd	e_pmRA	e_pmDE	EpRA-1900
arcmin				mag		"h:m:s"	"d:m:s"	s/yr	arcsec/yr		10mas	10mas	mas/yr	mas/yr	yr
0.0930	164887	164887	+04 3559	10.6	F8	17 57 24.373	+04 36 09.20	-0.0014	0.032	4	10	10	4.7	4.8	25.67

I/239/tyc_main The Hipparcos and Tycho Catalogues (ESA 1997)
The main part of Tycho Catalogue

_r	recno	TYC	Proxy	RAhms	DEdms	Vmag	r_Vmag	RA(ICRS)	DE(ICRS)	AstroRef	Plx	pmRA	pmDE	e_RAdeg	e_DE
arcmin						mag		deg	deg		mas	mas/yr	mas/yr	mas	m
0.1046	35715	425 1844 1		17 57 24.42	+04 36 09.0	10.43		269.35174824	4.60249678		27.50	41.60	37.50	32.20	2
7.3676	35741	425 2502 1		17 57 48.97	+04 40 05.8	9.54		269.45402305	4.66828815	X					

I/239/tyc_main converted columns :

recno	pmDE
35715	0.0375
35741	

The following conversions have not been performed :

Column name	From	To	Reason
VTmax	mag	mag	Useless

Dans le contexte du VO

Un Observatoire Virtuel pour l'astronomie:

- Plusieurs projets (AVO, NVO, AstroGrid, ...)
- Fournir aux astronomes des outils pour trouver, combiner et exploiter toutes les données disponibles.
- International Virtual Observatory Alliance (IVOA)

Il faut une bonne **INTEROPERABILITÉ** entre les différents services.

Première étape: **VOTable 1.0**

Standard XML pour l'échange de données tabulaires

Transporte les métadonnées avec les données: **avec les UCD**

votable.xml

```
<RESOURCE ID="yCat_1239" name="I/239">
  <DESCRIPTION>The Hipparcos and Tycho Catalogues (ESA 1997)</DESCRIPTION>
  <COOSYS ID="J2000_1991.250" system="eq_FK5" equinox="J2000" epoch="1991.250"/>
  <TABLE ID="I_239_hip_main" name="I/239/hip_main">
    <DESCRIPTION>The Hipparcos Main Catalogue</DESCRIPTION>
    <!-- RowName: ${HIP} -->
    <!-- MatchedRows: 50 (upper limit) -->

    <!-- Now comes the definition of each field -->
    <FIELD name="_RAJ2000" ucd="POS_EQ_RA_MAIN" ref="J2000" datatype="char" arraysize="12" unit="&quot;h:m:s&quot;">
      <DESCRIPTION>Right ascension (FK5) Equinox=J2000.0 Epoch=J2000, proper motions taken into account (computed by VizierR, not part of the original data)</DESCRIPTION>
    </FIELD>
    <FIELD name="_DEJ2000" ucd="POS_EQ_DEC_MAIN" ref="J2000" datatype="char" arraysize="12" unit="&quot;d:m:s&quot;">
      <DESCRIPTION>Declination (FK5) Equinox=J2000.0 Epoch=J2000, proper motions taken into account (computed by VizierR, not part of the original data)</DESCRIPTION>
    </FIELD>
    <FIELD name="HIP" ucd="ID_MAIN" datatype="int" width="6">
      <DESCRIPTION>Identification (HIP number) (H1)</DESCRIPTION>
      <LINK href="http://vizir.u-strasbg.fr/local/cgi-bin/VizieR-5?-info=XML&amp;-out.add=,&amp;-source=I/239/hip_main&amp;recno=${recno}"/>
    </FIELD>
    <FIELD name="RAhms" ucd="POS_EQ_RA_MAIN" datatype="char" arraysize="11*">
      <DESCRIPTION>Right ascension in h m s, ICRS (J1991.25) (H3)</DESCRIPTION>
    </FIELD>
    <FIELD name="DEdms" ucd="POS_EQ_DEC_MAIN" datatype="char" arraysize="11*">
      <DESCRIPTION>Declination in deg ' ", ICRS (J1991.25) (H4)</DESCRIPTION>
    </FIELD>
    <FIELD name="Vmag" ucd="PHOT_JHN_V" datatype="float" width="5" precision="2" unit="mag">
      <DESCRIPTION>? Magnitude in Johnson V (H5)</DESCRIPTION>
    </FIELD>
    <FIELD name="RA(ICRS)" ucd="POS_EQ_RA_MAIN" ref="J2000_1991.250" datatype="double" width="12" precision="8" unit="deg">
      <DESCRIPTION>?* alpha, degrees (ICRS, Epoch=J1991.25) (H8)</DESCRIPTION>
    </FIELD>
    <FIELD name="DE(ICRS)" ucd="POS_EQ_DEC_MAIN" ref="J2000_1991.250" datatype="double" width="12" precision="8" unit="deg">
      <DESCRIPTION>?* delta, degrees (ICRS, Epoch=J1991.25) (H9)</DESCRIPTION>
    </FIELD>
    <FIELD name="Plx" ucd="POS_PARLX_TRIG" datatype="float" width="7" precision="2" unit="mas">
      <DESCRIPTION>? Trigonometric parallax (H11)</DESCRIPTION>
    </FIELD>
  </TABLE>
</RESOURCE>
```

Dans le contexte du VO

Les UCDs offrent une bonne couverture du domaine sémantique car ils ont été créés à partir d'une base très importante de ressources.

Ils sont déjà utilisés dans certains services du VO:

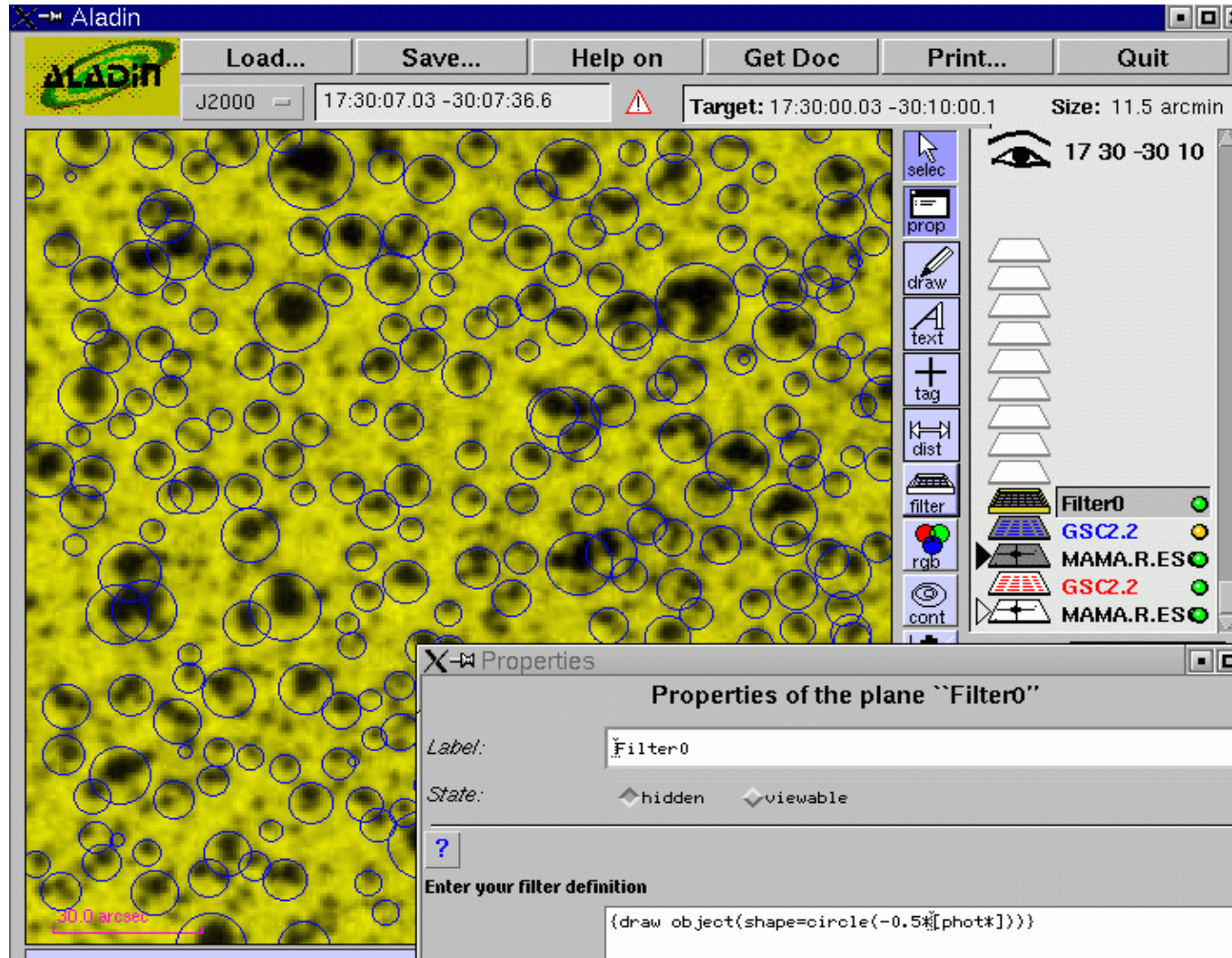
Cone search (initiative NVO): trouver les sources autour d'une position donnée, dans un catalogue.

- › ucd= "ID_MAIN"
- › ucd= "POS_EQ_RA_MAIN"
- › ucd= "POS_EQ_DEC_MAIN"

Renvoie un résultat en VOTable (éventuellement avec UCDs).

Dans le contexte du VO

Aladin (atlas interactif): contrôler l'apparence des sources d'un catalogue.



Perspectives pour les UCDs dans le VO

Les UCDs offrent un moyen standard de décrire le contenu sémantique. Ils doivent être améliorés pour être plus flexibles, et couvrir de nouveaux domaines (travail en cours!).

Changements prévus:

- Définitions d'une liste : **core UCDs**, validés par l'IVOA.
- **Namespaces** temporaires pour manipuler localement des UCDs
- Utilisation d'**atomes (specificateurs)**:

ivoa:**PHOT_JHN_V**,**ERROR**

Assignment des UCDs

Les descriptions des éléments comprennent:

- un nom
- une unité
- une courte description

Catalog	---	Catalogue (H=Hipparcos) (H0)	
HIP	---	Identifieur (HIP number) (H1)	
Proxy	---	Proximity flag (H2)	Note on Proxy: this
flag			
RAhms	---	Right ascension in h m s, ICRS (J1991.25) (H3)	
DEdms	---	Declination in deg ' ", ICRS (J1991.25) (H4)	
Vmag mag	?	Magnitude in Johnson V (H5)	
VarFlag	---	? Coarse variability flag (H6)	Note on VarFl

Assignment des UCDs

UCD	Dataset	Type	Name	Unit	Description
<input checked="" type="radio"/> ID_MAIN <input type="radio"/> PHYS_ABUND_[FE/H] <input type="radio"/> PHOT_JHN_H <input type="radio"/> ID_CATALOG <input type="radio"/> PHOT_FLUX_HALPHA <input type="radio"/> <input type="text"/>	hip_main.dat	A	Catalog	---	Catalogue (H=Hipparcos) (H0)
<input checked="" type="radio"/> ID_NUMBER <input type="radio"/> <input type="text"/>	hip_main.dat	I	HIP	---	Identifier (HIP number) (H1)
<input checked="" type="radio"/> REMARKS <input type="radio"/> CODE_MISC <input type="radio"/> <input type="text"/>	hip_main.dat	A	Proxy	---	Note on Proxy: this flag provides a coarse indication of the presence of nearby objects within 10arcsec of the given entry. If non-blank, it indicates that 'H' there is one or more distinct Hipparcos Catalogue entries, or distinct components of system from h_dm_com.dat 'T' there is one or more Proximity flag (H2)
<input checked="" type="radio"/> POS_EQ_RA_MAIN <input type="radio"/> <input type="text"/>	hip_main.dat	A	RAhms	---	Right ascension in h m s, ICRS (J1991.25) (H3)
<input checked="" type="radio"/> POS_EQ_DEC_MAIN <input type="radio"/> <input type="text"/>	hip_main.dat	A	DEdms	---	Declination in deg ' ", ICRS (J1991.25) (H4)
<input checked="" type="radio"/> PHOT_JHN_V <input type="radio"/> <input type="text"/>	hip_main.dat	F	Vmag	mag	? Magnitude in Johnson V (H5)
<input checked="" type="radio"/> CODE_VARIAB <input type="radio"/> <input type="text"/>	hip_main.dat	I	VarFlag	---	Note on VarFlag: the values are 1: < 0.06mag ; 2: 0.06-0.6mag ; 3: >0.6mag ? Coarse variability flag (H6)
<input checked="" type="radio"/> REFER_CODE <input type="radio"/> <input type="text"/>	hip_main.dat	A,@c	r_Vmag	---	Note on r_Vmag: the source is G = ground-based, H=HIP, T=Tycho Source of magnitude (H7)

<http://vizier.u-strasbg.fr/UCD/assign/>

Perspectives pour les UCDs dans le VO

Les UCDs pourraient être utilisés:

- **Dans les registries** (descriptions des ressources et services);
- **Dans les Data Models** (attachés aux attributs des modèles OO) --- déjà fait pour le modèle IDHA;
- **Dans VOQL**, langage de requête du VO (basé sur XML? SQL?).

Vers une ontologie:

UCDs comme point de départ pour construire une ontologie de l'astronomie... MDA!

Extension à de nouveaux domaines !