

Les UCD

Unified Content Descriptors

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
214.263858	-57.766521	17.167	15.984	99.999	0.11	0.20	9.99	17.209	16.248	99.999
214.617997	-57.766361	17.103	15.149	99.999	0.10	0.14	9.99	16.741	14.812	99.999
214.267771	-57.766221	17.888	15.712	99.999	0.13	0.18	9.99	16.713	15.023	99.999
214.533074	-57.766314	16.179	14.823	13.036	0.06	0.11	0.16	16.099	14.255	13.102
214.565329	-57.766304	17.249	15.955	99.999	0.11	0.20	9.99	17.213	17.242	99.999
214.256734	-57.766279	16.455	14.860	13.244	0.07	0.12	0.18	16.426	14.683	12.860
214.495374	-57.766278	17.009	14.856	13.533	0.10	0.13	0.22	16.628	14.731	13.337
214.597738	-57.766200	17.174	99.999	99.999	0.11	9.99	9.99	16.903	99.999	99.999
214.316993	-57.766161	15.816	14.837	99.999	0.05	0.12	9.99	15.496	14.697	99.999
214.347076	-57.766060	99.999	15.804	99.999	9.99	0.19	9.99	99.999	15.705	99.999
214.270551	-57.766037	16.967	15.125	99.999	0.10	0.18	9.99	17.014	16.130	99.999
214.588947	-57.765921	15.941	13.837	12.239	0.06	0.09	0.12	15.858	13.620	12.295
214.426528	-57.765900	99.999	16.814	13.928	9.99	0.23	0.27	99.999	16.593	13.955
214.411001	-57.765870	15.541	14.798	99.999	0.05	0.12	9.99	15.604	14.415	99.999
214.517990	-57.765842	99.999	16.820	99.999	9.99	0.24	9.99	99.999	18.163	99.999
214.465007	-57.765814	17.431	15.866	99.999	0.17	0.19	9.99	17.819	15.194	99.999
214.557982	-57.765781	15.137	12.797	11.802	0.05	0.08	0.09	15.540	12.948	11.385
214.474225	-57.765746	18.131	99.999	99.999	0.20	9.99	9.99	18.831	99.999	99.999

RA DEC mag1 mag2 mag3

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 **UCD (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. Ex d'UCD1:

PHOT_EXTINCTION_ISM Interstellar extinction.
POS_GAL_LAT Galactic latitude

UCD1

~1500 **UCD1** suffisent pour décrire le contenu de 100,000 **colonnes** dans Vizier.

UCD1 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 UCD1 ne sont pas un modèle de données, ni une ontologie!

Outils pour les UCD1

Des outils ont été développés pour démontrer comment les UCD1 peuvent être utilisés pour sélectionner des catalogues, comparer des données, etc...

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

Les outils incluent un browser pour explorer l'arbre, et la recherche d'UCD1 à partir de texte.

Trouver des catalogues intéressants à partir des UCD1

Related catalogues in VizieR:

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

Catalogue	Title	Bibcode
L15	Proper Motions in Cape Zone Catalogue -40/-52 (Spencer Jones H.+ 1936)	
L14	Proper Motions of 1160 Late-Type Stars (Fogh Olsen, 1970)	1970RAAS....2...620
L40	WASHINGTON 20 Catalog (Morgan, 1933)	
L61B	AGK3 Catalogue (Dieckvoos, Heckmann 1975)	1975OB6..A1S.....D
L62C	Perth 70: Positions of 24900 Stars (Hog+ 1976)	1976RAHAm...3...1H
L68A	Positions and Proper Motions in alpha Per cluster (Fresneau, 1980)	1980BICDS...18...81P

S. Derrière - Tutoriel OVFrance 11-13/10/2004 9

Assignment des UCD1

UCD	Dataset	Type	Name	Unit	Description
ID_MAIN					
PHYS_ABUND_IFE/HI					
PHOT_JHN_H	hip_main.dat	A	Catalog	---	Catalogue (H=Hipparcos) (H0)
ID_CATALOG					
PHOT_FLUX_HALPHA					
ID_NUMBER	hip_main.dat	I	HIP	---	Identifier (HIP number) (H1)
REMARKS					
CODE_MISC	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)
POS_EQ_RA_MAIN	hip_main.dat	A	RAhms	---	Right ascension in h m s. ICRS (J1991.25) (H3)
POS_EQ_DEC_MAIN	hip_main.dat	A	DEdms	---	Declination in deg ". ICRS (J1991.25) (H4)
PHOT_JHN_V	hip_main.dat	F	Vmag	mag	? Magnitude in Johnson V (H5)
CODE_VARIAB	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)
REFER_CODE	hip_main.dat	A	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/old/assign/>

S. Derrière - Tutoriel OVFrance 11-13/10/2004 10

Comparaisons automatiques

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

UCD + unités = Conversions de colonnes

S. Derrière - Tutoriel OVFrance 11-13/10/2004 11

Conversion results

AVO · ESQ · ST · ECF · AstroGrid · CDS · Terapix · Jodrell Bank

1/146/ppm1 Positions and Proper Motions - North (Roemer+, 1983)
Catalogue PPM-North

r	reco	PPM	DM	Mag Sp	RAJ2000	DEJ2000	pmRA	pmDE	Npos	e_RAs	e_DEd	pmRA	pmDE	EpRA-1900
arcmin			mag	"h.m.s"	"d.m.s"	s/yr	arcsec/yr	10mas	10mas	mas/yr	mas/yr	mas	mas	yr
0.0920	164887	164887	+04 3559	10.6 F8	17 57 24.273	+04 36 09.20	-0.00	0.032	4	10	10	4.7	4.8	25.67

1/239/tyc_main The Hipparcos and Tycho Catalogues (ESA 1997)
The main part of Tycho Catalogue

r	reco	TYC	Proxy	RAhms	DEdms	Vmag r	Vmag	RA(ICRS)	DE(ICRS)	AstroRef	Ptx	pmRA	pmDE	e_RAdeg	e_DE
arcmin				mag	mag	mag	mag	deg	deg	mas	mas/r	mas/yr	mas	mas	m
0.1046	35715	1844		17 57 24.42	+04 36 09.0	10.43		269.35174924	4.60249678		27.50	41.80	37.50	32.20	2
7.3676	35741	2502		17 57 48.97	+04 40 05.8	9.54		269.45442305	4.66828815	X					

1/239/tyc_main converted columns :

reco	pmDE
35715	0.0375
35741	

The following conversions have not been performed :

Column name	From	To	Reason
V_Tmag	mag	mag	Useless

S. Derrière - Tutoriel OVFrance 11-13/10/2004 12

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="Cat_1239" name="1/239">
  <DESCRIPTION>The Hipparcos and Tycho Catalogues (ESA 1997)/</DESCRIPTION>
  <UCDSYS ID="J2000_1991_250" unit="eq_FIS" equinox="J2000" epoch="1991.250"/>
  <TABLE ID="1_239_hip_main" name="1/239/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 in hours, J2000.0 Epoch=J2000, proper motions taken into account (computed by VizierR, not part of the original data)/</DESCRIPTION>
    <FIELD name="DECJ2000" ucd="POS_EQ_DEC_MAIN" ref="J2000" datatype="char" arraysize="12" unit="&quot;d:m:s&quot;"/>
    <DESCRIPTION>Declination in degrees, J2000.0 Epoch=J2000, proper motions taken into account (computed by VizierR, not part of the original data)/</DESCRIPTION>
    <FIELD name="ID_MAIN" ucd="ID_MAIN" datatype="int" width="6"/>
    <DESCRIPTION>Internal catalog number (ID)/</DESCRIPTION>
    <LINK href="http://vizier.u-strasbg.fr/local/cgi-bin/VizieR-5?-info=WL&out_addr=&source=1/239/hip_main&skip_screens=(reco)"/>
    <FIELD name="RPMs" ucd="POS_EQ_RA_MAIN" datatype="char" arraysize="11x"/>
    <DESCRIPTION>Right ascension in hours, ICRS (J1991.25) (H8)/</DESCRIPTION>
    <FIELD name="Decrs" ucd="POS_EQ_DEC_MAIN" datatype="char" arraysize="11x"/>
    <DESCRIPTION>Declination in deg. " ", ICRS (J1991.25) (H4)/</DESCRIPTION>
    <FIELD name="Vmag" ucd="PHOT_JHN_V" datatype="float" width="5" precision="2" unit="mag"/>
    <DESCRIPTION>? magnitude in Johnson V (H8)/</DESCRIPTION>
    <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 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) (H8)/</DESCRIPTION>
    <FIELD name="Plx" ucd="POS_PARX_TRIG" datatype="float" width="7" precision="2" unit="mas"/>
    <DESCRIPTION>? trigonometric parallax (H11)/</DESCRIPTION>
  </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:

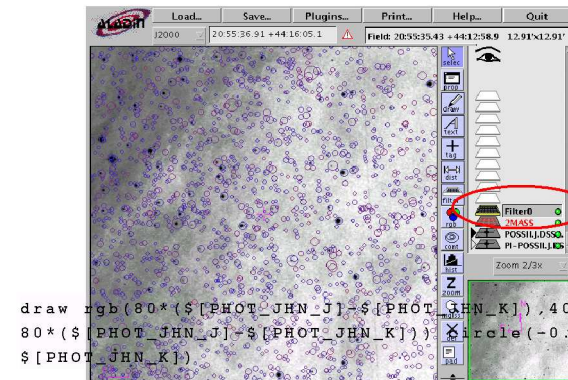
Cone search (initiative NVO): trouver les sources autour d'une position donnée, dans un catalogue.

- > 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.



TP!

Perspectives pour les UCDs dans le VO

Les UCDs offrent un moyen standard de décrire le contenu sémantique.

Utilisation:

- **Dans les registries** (descriptions des ressources et services);
- **Dans les Data Models** (attachés aux attributs des modèles OO) --- déjà fait pour le modèle IDHA;

Vers une ontologie? projet MDA

Extension à de nouveaux domaines !

Les nouveaux UCD (UCD1+)

Description sémantique **simple**, entre

- la complexité du langage naturel
- la rigidité d'un modèle de données

Chaque mot du vocabulaire a une définition associée.

Changement de syntaxe: `__ -> . ; . .`

Changement de philosophie:

Plusieurs **mots** peuvent composer un seul **UCD**.

Un mot peut contenir plusieurs **atomes**.

ex: `PHOT_JHN_B -> phot.mag;em.opt.B`

(UCD1+ ne sont pas sensibles à la casse)

Les nouveaux UCD (UCD1+)

L'esprit reste le même qu'avec les UCD1:

- décrire des quantités (*qu'est-ce que c'est?*)
- avec un niveau de détail **raisonnable**
- pour pouvoir faire des comparaisons entre des ensembles hétérogènes (interopérabilité)

Les nouveaux UCD (UCD1+)

Démarche:

- On part de ce qui existe (ce que les astronomes manipulent, mesurent, cataloguent)
- On tente d'homogénéiser des quantités très variées

On a environ 500 mots dans les UCD1+ (contre 1500 UCD1)

Les nouveaux UCD (UCD1+)

On peut ré-utiliser un mot dans plusieurs UCD:

- phot.mag;em.opt.R
- phot.mag;em.opt.B

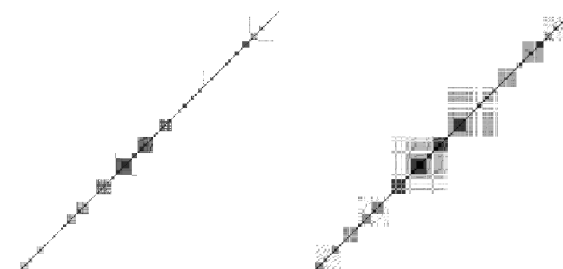
Plus de souplesse et de précision que les UCD1:

- ERROR
- stat.error;phot.mag;em.opt.R

Le **premier mot** porte l'essentiel du sens !

Matching function:
comparaison 'floue' entre les UCD1+

On peut calculer une ressemblance entre deux UCD qui n'est pas forcément 0 ou 1.



On ne crée pas en général de mots pour chaque instance d'un concept:

- pas de src.NGC1234 mais src existe
- pas de phys.at.elem.Fe, phys.at.elem.Cu, mais un terme phys.atmol.element

Cela dit, on a créé des termes pour les différentes parties du spectre électromagnétique:

- em.opt.R
- em.IR.K
- em.UV.100-200nm

UCD1+ vs Ontologies

Ontologie:

Philo: Syn. de Métaphysique générale
Investigation sur le sens de l'être distingué des étants
Discours issu de la logique mathématique et de la linguistique, qui traite des termes utilisés pour désigner les êtres constitutifs de la réalité.

En pratique, une ontologie :

- permet la description formelle d'un domaine, compréhensible par les spécialistes du domaine et les programmes;
- rend explicite ce qui est habituellement implicite;
- résulte d'une approche consensuelle;
- facilite l'interopérabilité entre les outils.

Ontologie:

Définition de:

- **concepts**, comparables aux classes OO (ex: étoile)
- un concept possède des **propriétés** (slots)
 - > distance
 - > type spectral
 - > température
- **facets**, qui sont des contraintes
 - > étoile n'est pas galaxie
- **instances**, qui sont les réalisations (ex: Vega)

Une propriété peut être commune à différents concepts. ex: température d'un instrument

UCD1+ vs Ontologies

Les UCD2 sont formés à partir d'une liste de mots. Ceux-ci peuvent être réutilisés.

ex: `phys.temperature;src` et
`phys.temperature;instr`

Certains concepts ont été introduits dans le vocabulaire. Ils ne peuvent cependant pas être utilisés comme mot primaire.

Certains concepts n'ont pas été introduits (étoile, galaxie)

Les UCD1+ ne sont pas une ontologie, mais ils pourront aider à appliquer les techniques du Web Sémantique à l'astronomie...

Maintenance des UCD1+

- <http://vizier.u-strasbg.fr/UCD/>
 - documents, outils...
- "UCD board" chargé de maintenir la liste des mots

Concept	Description	Status
phys.temperature;src	Temperature of the source	active
phys.temperature;instr	Temperature of the instrument	active
phys.temperature;obj	Temperature of the object	active
phys.temperature;obs	Temperature of the observation	active
phys.temperature;site	Temperature of the site	active
phys.temperature;atm	Temperature of the atmosphere	active
phys.temperature;sky	Temperature of the sky	active
phys.temperature;star	Temperature of the star	active
phys.temperature;galaxy	Temperature of the galaxy	active
phys.temperature;cluster	Temperature of the cluster	active
phys.temperature;group	Temperature of the group	active
phys.temperature;field	Temperature of the field	active
phys.temperature;region	Temperature of the region	active
phys.temperature;sector	Temperature of the sector	active
phys.temperature;zone	Temperature of the zone	active
phys.temperature;band	Temperature of the band	active
phys.temperature;filter	Temperature of the filter	active
phys.temperature;channel	Temperature of the channel	active
phys.temperature;beam	Temperature of the beam	active
phys.temperature;pixel	Temperature of the pixel	active
phys.temperature;arcsec	Temperature of the arcsec	active
phys.temperature;mas	Temperature of the mas	active
phys.temperature;microarcsec	Temperature of the microarcsec	active
phys.temperature;nanarcsec	Temperature of the nanarcsec	active
phys.temperature;microarcmin	Temperature of the microarcmin	active
phys.temperature;nanarcmin	Temperature of the nanarcmin	active
phys.temperature;microarcsecond	Temperature of the microarcsecond	active
phys.temperature;nanarcsecond	Temperature of the nanarcsecond	active

Utilisation des UCD1+

- VOTable (<FIELD ucd="pos.eq.ra">)
- Registry (<ucd> element in table description for TabularSkyService)
- Nombreux outils (avec VOTable...)

Utilisation des UCDs par les fournisseurs de données

Pas obligatoire d'utiliser les UCDs en interne: chacun gère ses données comme il l'entend.

Il faut juste une couche de transition, qui est capable:

- d'associer au vol les UCDs aux données (pour un producteur de données)
- d'interpréter les UCDs en termes de paramètres attendus (pour un service)

Les UCD sont uniquement nécessaires dans les **échanges de données** ! (interopérabilité)

Questions ?

TP ?