

Metadonnées, UCD, sémantique dans le VO

Sebastien Derriere

CDS

derriere@astro.u-strasbg.fr

Plan

- Objectifs du VO
- Metadonnées en astronomie

Exemple des UCDs

- Problématique
 - Utilisation
 - Maintenance
- Vers une sémantique pour l'astronomie

(quelques) Objectifs du VO

- Accès à des données distribuées, hétérogènes
Observations, modélisations
- Offrir de nouveaux outils
Traitement, analyse, codes de simulation
- Mots-clés:
Interopérabilité
Définitions de standards
Protocoles d'accès

Metadonnées en astronomie

- Exemples de métadonnées:
 - FITS keywords (NAXIS, CRPIX, CDELT...)
 - Thesaurus IAU
 - Mots-clés des journaux
- Problèmes :
 - Consensus pour l'adoption
 - Respect des standards
 - Maintenance sur le moyen/long terme

Pourquoi les UCDs?

- Comment trouver des données pertinentes ?
- Comment savoir si des données peuvent être comparées?
- Comment automatiser certains traitements?
- ... sans une description standardisée des contenus?

Pourquoi les UCDs?

- Les UCD ont été créés à l'origine pour le service VizieR de catalogues au CDS

+ de 5000 catalogues différents

Origines diverses

- Numérisation de plaques photos
- Observations dans tout le spectre EM
- Publication électronique de tables des journaux

Les descriptions et noms de colonne varient énormément

- Plus de 140 noms différents pour une magnitude V !!



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
214.267771	-57.766321	17.388	15.712	99.999	0.13	0.18	9.99	16.713	15.023	99.999
214.532174	-57.766314	16.119	14.323	13.031	0.06	0.11	0.16	16.099	14.255	13.102
214.56529	-57.766304	17.219	15.955	99.999	0.11	0.20	9.99	17.213	17.242	99.999
214.25634	-57.766279	16.465	14.660	13.241	0.07	0.12	0.18	16.426	14.683	12.860
214.49974	-57.766278	17.009	14.855	13.531	0.10	0.13	0.22	16.628	14.731	13.337
214.59738	-57.766200	17.174	99.999	99.999	0.11	9.99	9.99	16.903	99.999	99.999
214.31693	-57.766161	15.516	14.837	99.999	0.05	0.12	9.99	15.496	14.697	99.999
214.34276	-57.766060	99.999	15.804	99.999	9.99	0.19	9.99	99.999	15.705	99.999
214.27851	-57.766037	16.967	15.725	99.999	0.10	0.18	9.99	17.014	16.130	99.999
214.58847	-57.765921	15.941	13.637	12.231	0.06	0.09	0.12	15.858	13.620	12.295
214.42628	-57.765900	99.999	16.214	13.921	9.99	0.23	0.27	99.999	16.593	13.955
214.41001	-57.765870	15.641	14.793	99.999	0.05	0.12	9.99	15.604	14.415	99.999
214.51790	-57.765842	99.999	16.320	99.999	9.99	0.24	9.99	99.999	18.163	99.999
214.46307	-57.765814	17.931	15.865	99.999	0.17	0.19	9.99	17.819	15.194	99.999
214.55482	-57.765781	15.537	12.979	11.401	0.05	0.08	0.09	15.540	12.948	11.385
214.47425	-57.765746	18.131	99.999	99.999	0.20	9.99	9.99	18.831	99.999	99.999

RA

DEC

magI magJ magK

Que sont les UCDs ?

- UCD = Unified Content Descriptors
- UCD : métadonnées associées aux données
- Donner une **description sémantique** des quantités astronomiques = ce que c'est, avec un vocabulaire standardisé
- UCDs particulièrement utiles pour
 - Homogénéiser des collections de données hétérogènes
 - Interoperabilité lors des échanges de données

Première version: UCD1

- ESO-CDS data mining project

Exploration manuelle de 100,000 colonnes

creation de nouveaux mots pour chaque quantité

~1500 mots créés

- PHOT_EXTINCTION_ISM = Interstellar extinction
- POS_GAL_LAT = Galactic Latitude

- UCD donne une description sémantique du contenu des colonnes

Permet de rechercher des catalogues d'après leur contenu

Permet des comparaisons automatisées

IVOA: UCD1+

- UCD1 définis avant le VO
- L'IVOA définit des standards permettant l'interopérabilité entre les projets VO: intérêt pour les UCDs
- Migration vers UCD1+
 - re-utiliser une base existante (UCD1)
 - syntaxe plus flexible
 - vocabulaire contrôlé par IVOA

IVOA: UCD1+

- Même but que UCD1

Décrire des quantités: "ce que c'est"

Avec une précision raisonnable

Permettre l'interopérabilité

- Nouvelle liste de mots (<500) avec définitions (IVOA Rec. 31/12/2005)

- Nouvelle syntaxe (IVOA Rec. 19/08/2005)

Pas sensible à la casse

Possibilité de combiner plusieurs mots dans 1 UCD

Le 1er mot porte l'essentiel du sens

Exemples d'UCD1+

- `pos.eq.dec;meta.main` = déclinaison
- `phot.mag` = magnitude
- `phot.mag;em.opt.V` = magnitude mesurée dans l'optique entre 500 et 600nm
- `stat.error;phys.temperature` = erreur de mesure sur la temperature
- `phot.color;em.opt.B;em.opt.V` = couleur B-V

UCDs dans le VO (1)

- Standards sur le site IVOA
<http://www.ivoa.net/Documents/latest/>
- Discussions: ucd@ivoa.net
- Vocabulaire susceptible d'évoluer:
 - Amélioration de la liste, maintenance
 - Ouverture à d'autres domaines
- Scientific Board: vocabulaire
- Technical board: support et outils

UCDs dans le VO (2)

- Utilisation dans d'autres standards:

VOTable (attribut "ucd" de <FIELD> et <PARAM>)

Registry: description du contenu des colonnes
<column> pour TabularSkyService

VOEvent: utilisation dans <param> et <group>

- Utilisation dans Cone Search

- ucd= "ID_MAIN"

- ucd= "POS_EQ_RA_MAIN"

- ucd= "POS_EQ_DEC_MAIN"

- Utilisation dans applications (Aladin, TopCat)

Utilisation dans VOTable

```
- <FIELD datatype="float" name="plx" precision="2" ucd="pos.parallax.trig" unit="mas" width="5">
  <DESCRIPTION>? Hipparcos <I/239> trigonometric parallax</DESCRIPTION>
</FIELD>
- <FIELD arraysize="11*" datatype="char" name="SpType" ucd="src.spType">
  <DESCRIPTION>Spectral type and luminosity class</DESCRIPTION>
</FIELD>
- <FIELD datatype="float" name="LDD" precision="2" ucd="phys.angSize" unit="mas" width="5">
  <DESCRIPTION>Limb-darkened disc diameter</DESCRIPTION>
</FIELD>
- <FIELD datatype="float" name="e_LDD" precision="3" ucd="stat.error" unit="mas" width="6">
  <DESCRIPTION>Error on LDD</DESCRIPTION>
</FIELD>
- <FIELD datatype="short" name="Teff" ucd="phys.temperature.effective" unit="K" width="4">
  <DESCRIPTION>Effective temperature</DESCRIPTION>
</FIELD>
- <FIELD datatype="float" name="logg" precision="2" ucd="phys.gravity" unit="[cm/s2]" width="5">
  <DESCRIPTION>Surface gravity: log(g) with g in cm/s^2</DESCRIPTION>
</FIELD>
- <FIELD datatype="float" name="Kmag" precision="2" ucd="phot.mag;em.IR.K" unit="mag" width="5">
  <DESCRIPTION>Johnson's K magnitude</DESCRIPTION>
</FIELD>
- <DATA>
  - <TABLEDATA>
    - <TR>
      <TD>1</TD>
      <TD>00 11 34.42</TD>
      <TD>-27 47 59.1</TD>
      <TD>720</TD>
      <TD>Simbad</TD>
      <TD>HIP</TD>
      <TD>00 11 34.4201</TD>
      <TD>-27 47 59.052</TD>
```

Utilisation dans le registry

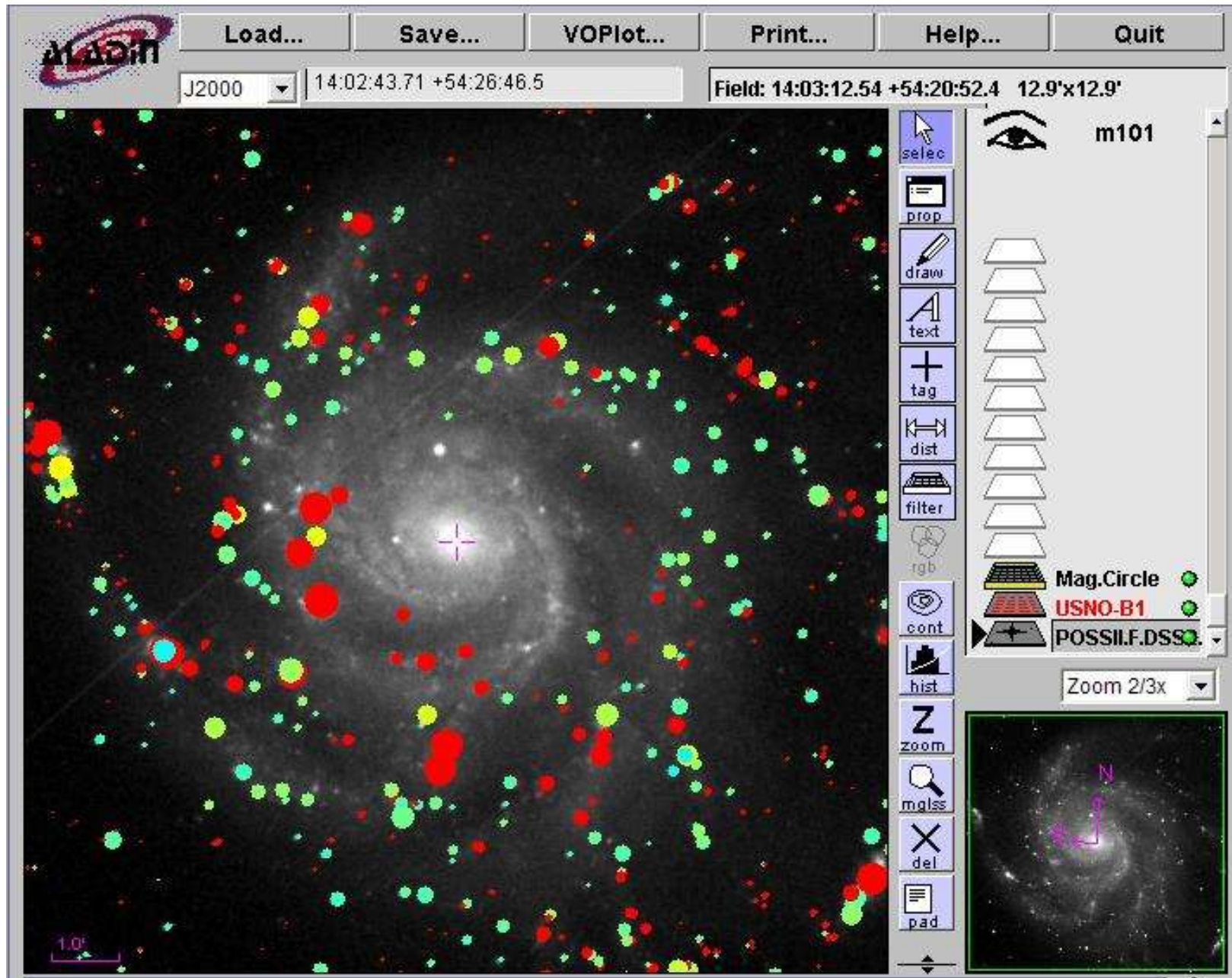
- Rechercher des VO resources d'après leur contenu

Plus efficace que par mot-clé

- Indispensable pour trouver des données pertinentes dans de grands ensembles hétérogènes!

Dans les applications

- Avec des UCDs présents dans un VOTable, on peut faire des traitements automatiques:
 - Identifier les colonnes contenant les positions
 - Appliquer des actions dans connaître la structure de la table
 - ex: filtres Aladin



UCDs et Services VO

- 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)

Maintenance des UCD1+

- <http://vizier.u-strasbg.fr/UCD/>
documents, outils...
- "Scientific board" chargé de maintenir la liste des mots, et d'étudier les nouvelles propositions ou modifications
ucd-sci@ivoa.net
- Groupe technique, en charge du développement d'outils
ucd-tech@ivoa.net

Outils UCD1+

- Ensemble d'outils disponibles pour:
 - Faciliter le passage UCD1 -> UCD1+
 - Manipuler les UCD1+
- Outils sous forme de scripts CGI et Web Services
 - <http://vizier.u-strasbg.fr/UCD/tools.htx>
 - <http://cdsweb.u-strasbg.fr/cdsws/ucdClient.gml>

UCD builder



UCD builder



Use this interface to find the UCD corresponding to a description:

Enter a description in natural language:

Last updated Thu May 12 15:54:16 2005

Suggested complete UCD: **phot.mag;em.opt.R;stat.max**

Refine your search:

The following words matched your query. You might force the selection of words matching precisely your query, and rebuild a UCD using these words:

word	definition	flag	score
<input type="checkbox"/> phot.mag	Photometric magnitude	E	60
<input type="checkbox"/> phot.mag.bol	Bolometric magnitude	Q	50
<input type="checkbox"/> phys.magAbs.bol	Bolometric absolute magnitude	Q	30
<input type="checkbox"/> em.opt.R	Optical R band between 600 and 750 nm	S	15
<input type="checkbox"/> stat.max	Maximum or upper limit	S	30

using selected words.

A part les UCD

- Définition dans le VO de modèles de données: STC, characterization, catalogue/source, quantity, ...
- Concepts définis de manière très précise
Identification par des **utypes**

UCD -> sémantique

- Extension du groupe UCD à un cadre plus large: sémantique liée à l'OV
- Répondre aux nouveaux besoins, ex: VO Event
 - Décrire les types d'objets astronomiques
 - Décrire les phénomènes physiques
- Travaux en cours...