

# UCD - lessons learned

What was learned from trying to assign UCDs to:

- large catalogues/databases
- specific domain in astronomy
- data models
- FITS headers



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

- « Structure » of UCD
- Assignment
- Application



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

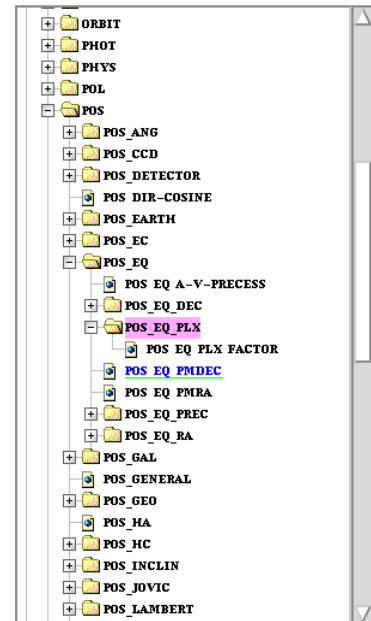
# The « structure » of UCDs

The presentation of UCDs is misleading:

**The tree structure is not mandatory!**

It is only:

- a convenient way of grouping similar elements from a given point of view
- a specification of the context (make implicit information explicit)



# The « structure » of UCDs

UCDs are (standard, unique) names for concepts

e.g. we find a new(!) concept : « temperature »  
we name it **temperature**

We forgot to mention that it was the « effective temperature of a star », because it sounded obvious in our context.

What if we find a new concept : « temperature of an instrument » ???



## The « structure » of UCDs

1. We call the 2 concepts **temperature**
2. We add a little something to distinguish the 2 kind of temperatures, defining more elaborated words:
  - **effective-temperature-of-a-star**
  - **temperature-of-an-instrument**

In the process of elaborating UCDs, it was just convenient to group concepts relative to physical quantities, or instrument, together.. that's just how UCDs were defined...  
**PHYS\_TEMP\_EFFEC** and **INST\_TEMP\_SYST**

The structure could be different.



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## The « structure » of UCDs

- The UCDs are NOT a universal data model, they are not an ontology, they do not impose a structured view of the universe
- But UCDs can be used to **name** attributes of data models
- The data model is structured, hierarchical -- not UCDs
- The data model carries the structure and describes links between its components -- UCDs are used to name the components



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

# Assignment of UCDs

Given a dataset: how to describe it, how to assign UCDs ?  
(translate my own description into something more standard that can be understood by others...)

This requires :

- A list of existing UCDs, with their definitions
- A set of decision rules for assignment
- Re-use the knowledge of already assigned data
- Build new terms following standard syntax

**AT LEAST**

**IMPROVES  
EFFICIENCY**



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

# Assignment of UCDs

The original descriptions of elements can consist of:

- a name
- a description
- a unit

```
RAdeg  deg  alpha, degrees (ICRS, Epoch=J1991.25)
DEdeg  deg  delta, degrees (ICRS, Epoch=J1991.25)
Plx     mas  Trigonometric parallax
pmRA   mas/yr Proper motion mu_alpha.cos(delta)
pmDE   mas/yr Proper motion mu_delta, ICRS
e_RAdeg mas  Standard error in RA*cos(DEdeg)
e_DEdeg mas  Standard error in DE
e_Plx  mas  Standard error in Plx
e_pmRA mas/yr Standard error in pmRA
e_pmDE mas/yr Standard error in pmDE
```



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

# Assignment of UCDs

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

Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Application 1: SDSS (A. Szalay)

1300 columns for the complete SDSS database.  
Input file for assignation built from SQL DB schema.

### Results:

- Need for manual verification in all cases (not automatic 1-to-1 assignation).
- Relatively few new concepts (not described by existing UCDs):

```

STAT_STDEV, _VARIANCE, _COVARIANCE
FIT_PARAM_COVARIANCE
ID_VERSION
CODE_HTM
INST_SKY_SIGMA
PHOT_TRANS_PARAM
POS_EQ_CART_X, _Y, _Z
POS_SDSS_MU, _NU, _LAMBDA, _ETA
METADATA_ID, _DESCRIPTION, _VERSION, _TABLE, _COLUMN,
_UNIT, _NAME, _COMMENT
    
```



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Application 2: Radio Data (A. Richards)

MERLIN database.

Application to a specific domain (Radio).

### Results:

Concepts specific to the radio/interferometry domains :

CHANNELWIDTH, VISIBILITY, BASELINE, deconvolution, beam, ...

Same words with other definitions: « extension », FoV, position (source / field)



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Application 3: Data model (M. Louys)

The IDHA data model:

- ~120 model attributes with their definitions

### Results:

Except for INST quantities, direct assignation is rare.

But very often, a proper UCD exists.

Descriptions of data model attributes **and** UCDs have to be checked/improved

Missing UCDs are related to:

- image format
- pixel coding
- software description
- data reduction process



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

Note: SIAP and VOX elements

The assignation program found some relevant already existing UCDs for some VOX elements.

(without exploring the whole UCD list !)

To be continued...



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Application 4: FITS keywords (A. Preite Martinez)

FITS headers from different surveys:

- list all keywords.

### Results:

In most cases, some relevant UCDs are suggested.

If not:

- the FITS keyword definition is not accurate
- the FITS keyword definition is cryptic (abbreviations, even human assignation of UCD is very difficult)
- it is a very specific parameter (a given instrument configuration)
- it is related to software domain



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Conclusions (1)

Automatic assignation of UCDs is not easy, because...

### **The UCD list is not complete:**

- there are missing UCDs in specific domains: missing terms must be defined by small representative groups (not only one project to keep it general -- distinguish specific parameters from « core » ones)
- UCDs are missing to describe software-related parameters, and pipeline processing



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16

## Conclusions (2)

### **The UCD list is not complete:**

- No UCD for very specific parameters

Which is the level of granularity for « core » UCDs ?  
How specific should the UCD description itself be (use of GROUP/parameters and atoms)?

### **Transforming language into UCD for assignation is not easy !**

- We must be flexible on the input (allow to describe things in natural language)
- But we must have enough information to guess what we're talking about (column name + unit + description)



Interoperability meeting

S. Derriere, Cambridge, 2003 May 12-16



## Conclusions (3)

- Understand how to provide « efficient » descriptions on the assignation side, and on the data provider side.
- Provide examples

### How do we do this in an evolving world ?

- define a « core » UCD list
- share the UCD list and definitions / update the list (curator) ?
- keep track of version of the UCD list, of the assignation tool version, of deprecated UCDs, etc...
- distribute the list of existing implementations (parameter descriptions and already assigned UCDs) ?

