Usage statistics and archiving process of VizieR data in the VO context

VO Implementation status
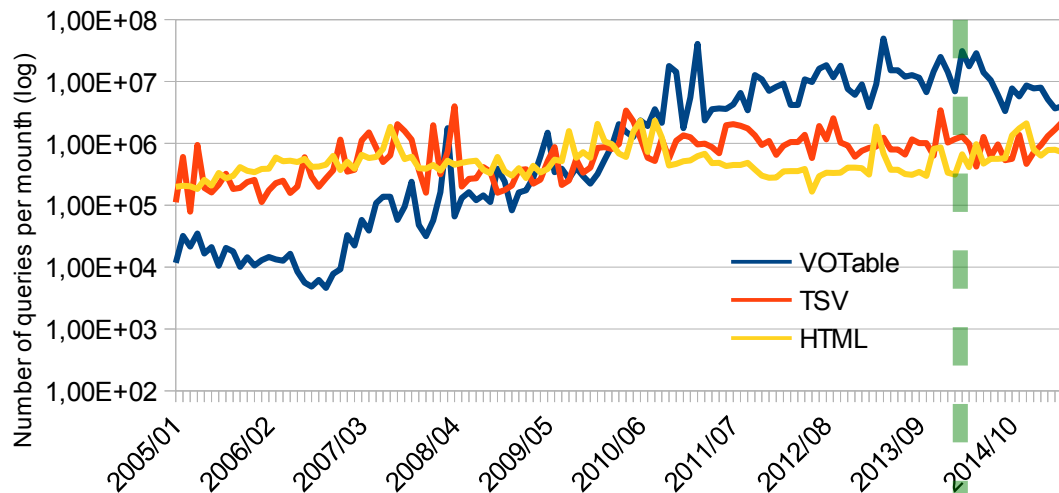
Application	✓	VOTable (1.1 1.2 1.3)	
	✓	MOC	*Available through Aladin and to query tables using a MOC*
	✓	SAMP	
Semantic	✓	UCD1 UCD1+	
Data Access Layer	✓	Simple Cone Search	
	⚠	TAP 1.1	*Need some arrangements to facilitate client access!*
	TODO	SIA (V1)	
	TODO	SSA (V1)	*beta release planned before the end of the year*
	TODO	ObsTAP	
Data Model	✓	Photometry Model (IVOA note)	*"Providing Photometric Data Measurements Description in VOTables" (S.Derriere)*
	TODO	ObsCore	*mandatory items only*

VizieR catalogues and the TAP VizieR service are in registries

VOTable output statistics

Output type evolution

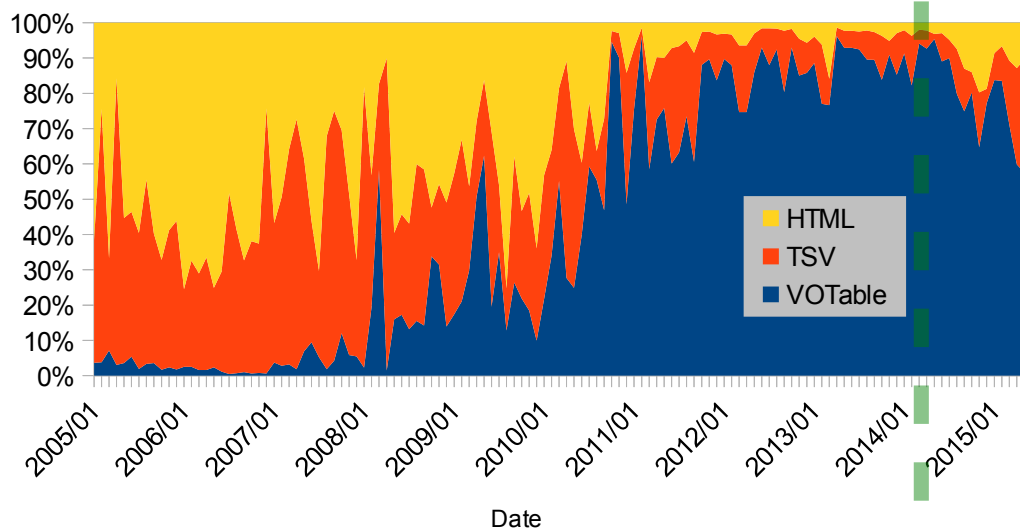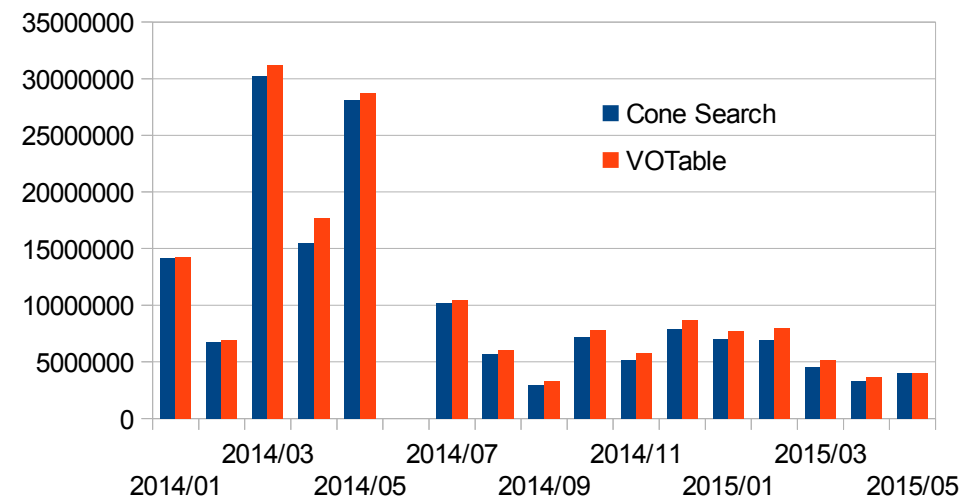### CDS Xmatch API

- ➜ 700 IP/day
- ➜ Number queries average (VOTable)
  2014 ~410,000 queries/day
  2015 ~190,000 queries/day
- ➜ 92% of VOTable output comes from "Simple Cone Search" queries
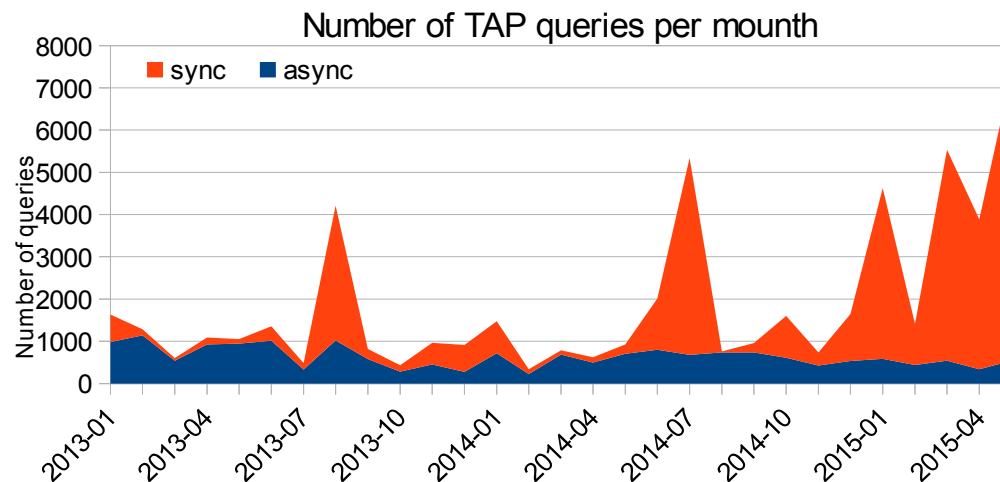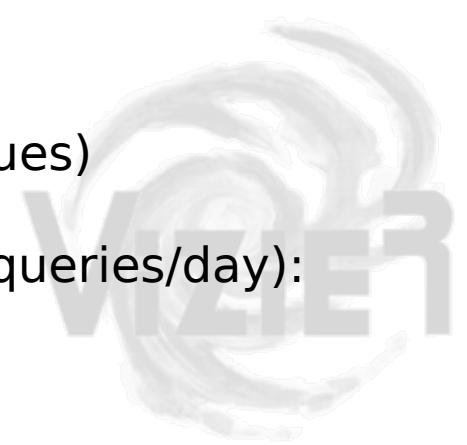
Output type repartition

Cone search in VOTable output

TAPVizieR statistics

TAPVizier contains all VizieR tables (except obsolete catalogues)

Statistics 2015, without bots and registry queries (~26,000 queries/day):

➔ 10 IP/day
➔ ~150 queries/day
➔ 7→ 10% queries/day contain ADQL geometrical functions

Need some arrangements to facilitate client access!

Difficulties to work with TAPVizieR, because:
- Custom use of schema due to the the important number of columns
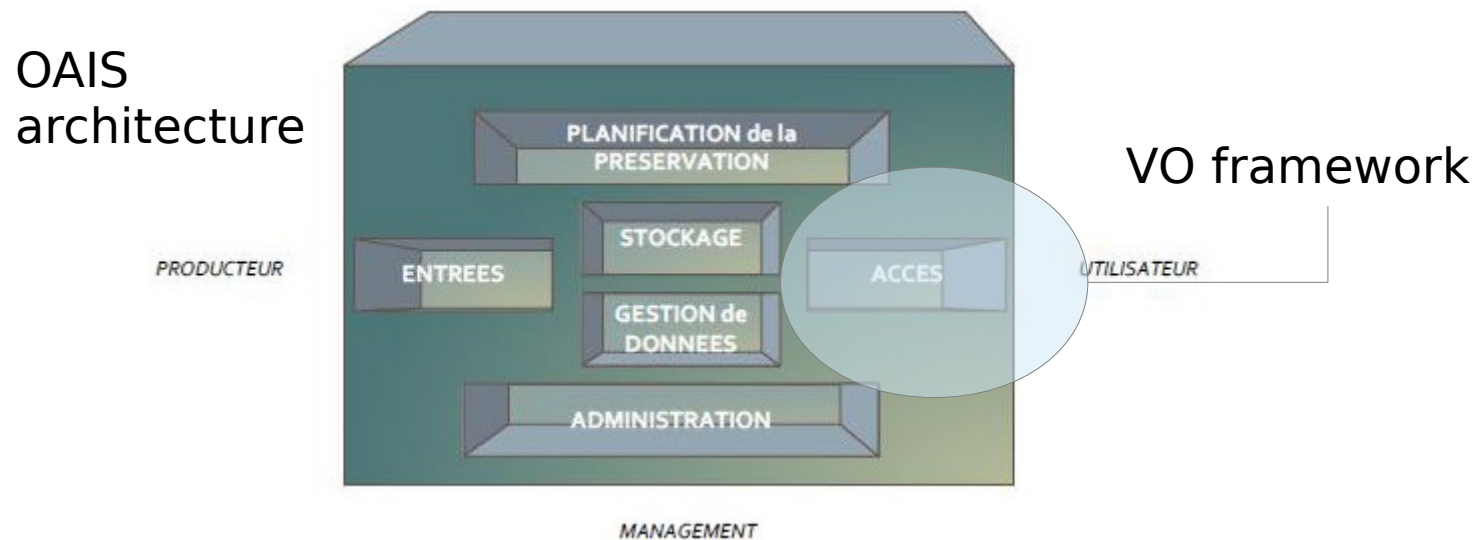- Columns/tables names which need quotes.

⟹ To provide quoted names in TAP_SCHEMA and VOSI output ?

The VO visibility

The VO is an adapted framework to provide data in the preservation context
➔ VO standards (protocols, formats, registries (OAI-PMH)) guaranty reusable data
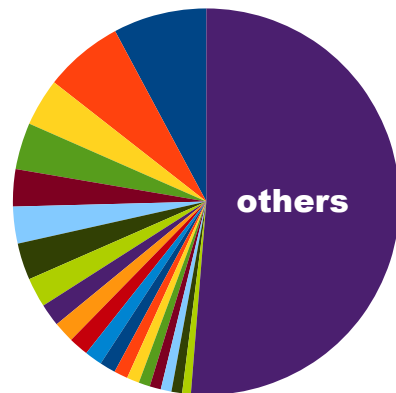➔ Matches with definition of the Access layer of OAIS (Open Archive Information system)

OAIS architecture

VO framework

PLANIFICATION de la PRESERVATION

STOCKAGE

PRODUCTEUR

ENTREES

GESTION de DONNEES

ACCES
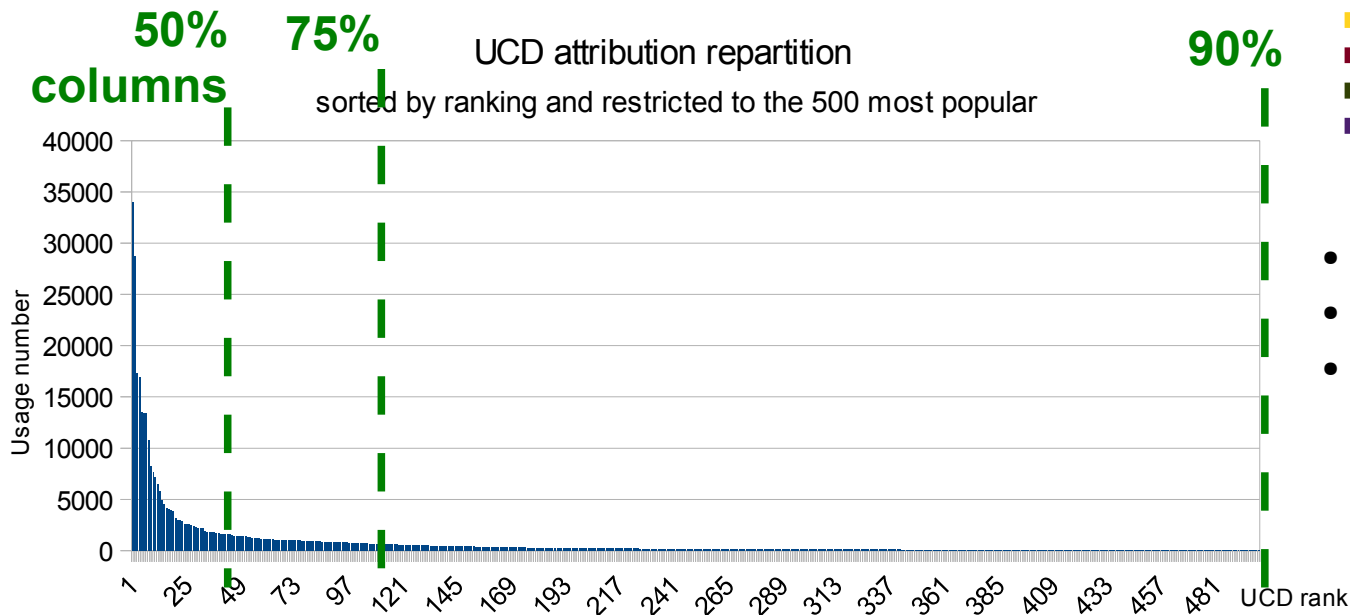
UTILISATEUR

ADMINISTRATION

MANAGEMENT

Assigning UCD

➜ CDS documentalists pay particular attention to UCD attribution
➜ A distribution with a long tail
➜ Main UCD (position, magnitude) are well assigned
➜ But, important usage of generic UCD and sometimes not optimum

⟹ The difficulty to have a perfect matching!

The most popular UCD in VizieR

- stat.error
- meta.id;meta.main
- meta.id
- pos.eq.dec;meta.main
- meta.code
- meta.ref.url
- phot.mag;em.opt.V
- stat.fit.param
- meta.ref
- phot.mag;em.opt.I
- others
- meta.record
- meta.note
- pos.eq.ra;meta.main
- meta.code.error
- meta.number
- time.epoch
- phys.abund
- spect.line.eqWidth
- phot.mag;em.opt.B
- meta.ref;pos.frame

50% columns   **75%**   UCD attribution repartition   **90%**

sorted by ranking and restricted to the 500 most popular

- 434,867 columns
- 3350 different UCD
- 6,5% columns with no UCD

The UCD assignment

➔ CDS documentalists set UCD1 for each columns using a UCD1 builder
➔ UCD1+ is constructed from UCD1 and other meta-data

The reason to assign UCD1:

➔ The "Simple Cone Search" needs UCD1 for main positions
➔ Easier to work with a "simple and restricted list" than to construct UCD1+
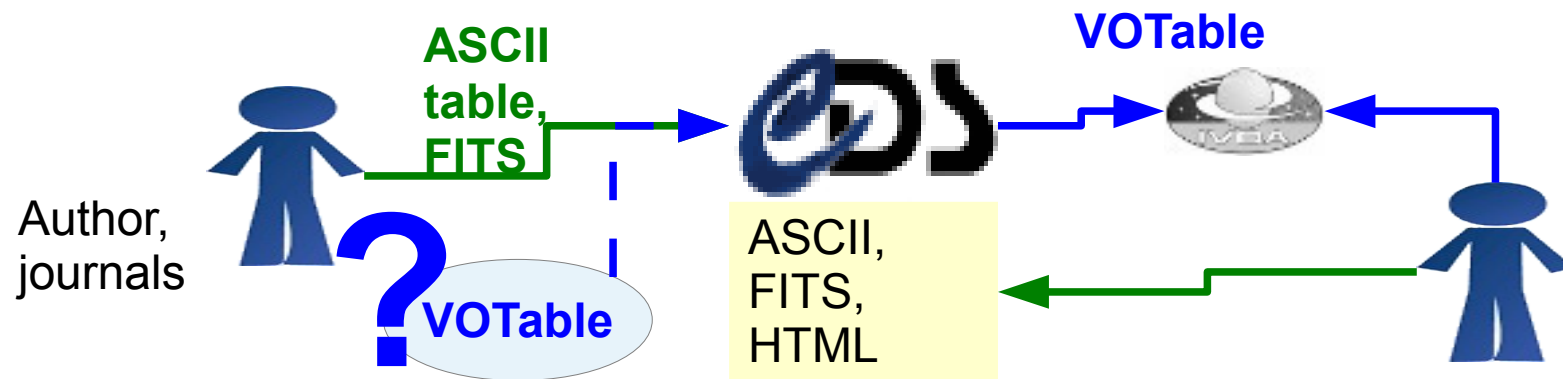➔ UCD1 photometry is used to describe filters of the magnitude columns
example:

UCD1	Filter
PHOT_JHN_B	Johnson, B
PHOT_COUS_I	Cousins I
PHOT_HST_F170W	HST/WFPC2, F814W
PHOT_WLRV_W	Walraven, W

To provide and preserve through the VO

Data preservation :
- ➔ the original data in input containing meta-data (ex: FITS)
- ➔ and the data provided (ex: VOTable)

To accept Votable in input?

➔ Currently no VOTable are stored in the VizieR repository
➔ FITS headers are not well standardized

> · VOTable containing rich metadata could improve the pipelines
> · To provide the VOTable in their original format using SIA, SSA

To promote VO standards in input

➔ VizieR encourage today the space agencies to provide VO standards for the VizieR logs pipeline (B/* catalogues updated weekly..)
➔ Nextly, VOTable will be soon accepted in VizieR for Associated data (spectra/time series/images using Saada and indexed with ObsCore)

Note: Saada extract the metadata from simple VOTable

Maintenance/actions needed to preserve and provide rich meta-data

➔ Adding new mandatory items in **output** DM requires significant efforts for CDS (documentalists, engineer) to search the informations. In particular for old catalogue
So Please, we would like to have the possibility of "notSet" or "noInfo"

➔ In **Input**, rich meta-data in VOTable (DML, utype or other):
  ➔ Needs libraries for authors/instruments for the VOTable generation
  ➔ Need to continue to provide classicals header for current clients
  ➔ Increases the maintenance for preservation because:
    ➔ The evolution of the VO standards
    ➔ Obsolete VOTable needs migration which includes obsolete search processing, format migration and search adding information