CDS Scientific Council

R&D activities at CDS

Large surveys, CDS in the VO, Web 2.0/3.0

Thomas Boch
Sébastien Derrière
André Schaaff

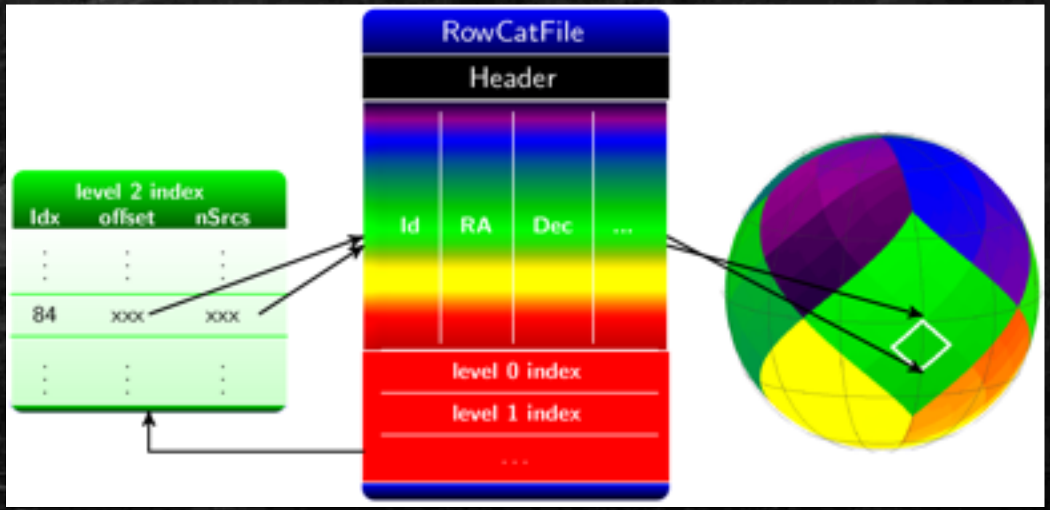
Outline

1. Handling large surveys
   - fast positional queries
   - fast multi-criteria queries
   - VizieR global index
2. CDS in the Virtual Observatory (VO)
   - TAP services for VizieR and SIMBAD
   - MOC, spatial signature of datasets
3. Web 2.0/3.0 and other collaborations
   - HTML5 visualizer
   - Mobile platforms
   - Annotations
   - Smart portal

1. Handling large surveys

Fast positional queries (1/2)

- Generic framework for querying large tables by position (initially developed for the cross-match service)
- Data structure
  - 3 parts : header + data + indexes
  - Data sorted by HEALPix cell number
    - Sources close on the sky are close on the file
  - (Compressed) binary data stored row by row of fixed length
    - implicit record number
    - direct access to a source data from its record number



- Advantages
  - generic method (one software to access all tables)
  - fast (can deliver 300 queries/second)
    - reads (almost) only the needed data
  - low complexity (fixed-length rows)

Fast positional queries (2/2)

- Limitations
  - disk space not always optimized (but disk is cheap)
    - NULL values
    - no support for variable-length strings
- Challenges
  - efficient sorting of 2+ billion rows
  - streaming input/output
- in production in cross-match service
- in production in VizieR for large surveys
  - easier and faster ingestion procedure
  - 10 large surveys ingested in VizieR using this method (including WISE, UKIDSS LAS/GPS, GAIA GUMS) since January 2012
- Improvements for version 2:
  - additional metadata in header
  - better compression support
  - binary output

Multi-criteria query (1/2)

- «find in 2MASS all sources having J-H<0.3 and H-K<0.3»
- solutions for fast non-positional queries:
  - full scan (already implemented)
    - no need for extra space
    - suited if a large fraction of the sources matches the query
  - **indexes**
    - suited if a small fraction of the sources matches the query
- Which type of index ?
  - uni-dimensional ?
    - binary-search tree
    - b-tree (used in DBMS)
  - multi-dimensional
    - kd-tree
    - R-tree (used in DBMS)

- bs-tree and kd-tree
  - very good performances
  - slow to update

Suited for static datasets

- b-tree and R-tree
  - less performant than bs and kd-trees
  - easier to update

Suited for changing datasets

Multi-criteria query (1/2)

- «find in 2MASS all sources having J-H<0.3 and H-K<0.3»
- solutions for fast non-positional queries:
  - full scan (already implemented)
    - no need for extra space
    - suited if a large fraction of the sources matches the query
  - **indexes**
    - suited if a small fraction of the sources matches the query
- Which type of index ?
  - uni-dimensional ?
    - binary-search tree
    - b-tree (used in DBMS)
  - multi-dimensional
    - kd-tree
    - R-tree (used in DBMS)

- bs-tree and kd-tree
  - very good performances
  - slow to update

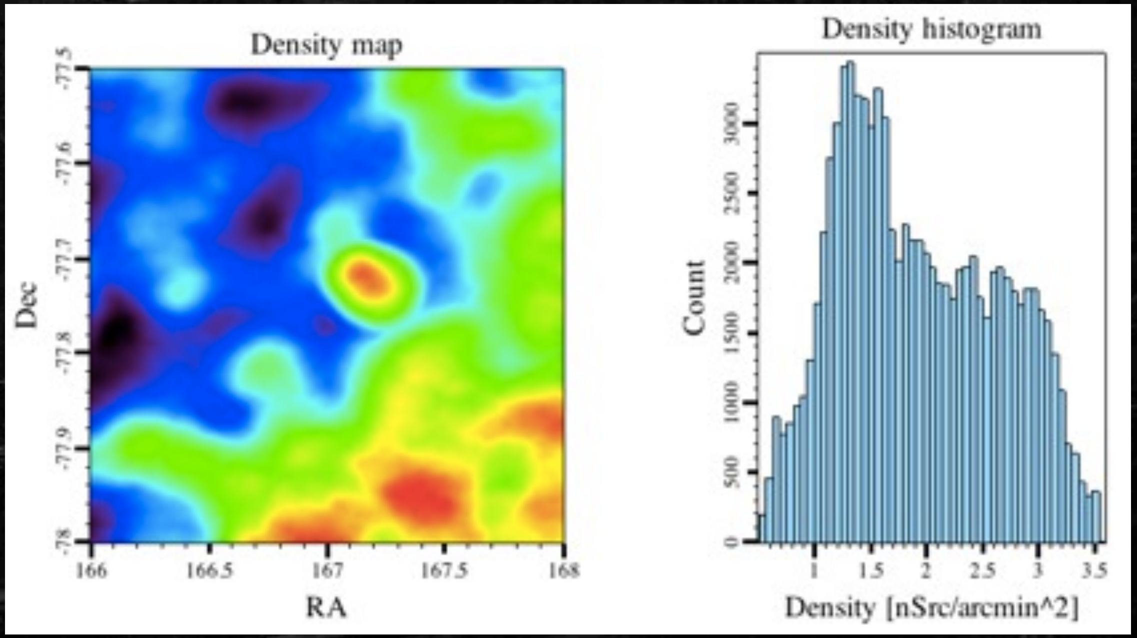
Suited for static datasets

- b-tree and R-tree
  - less performant than bs and kd-trees
  - easier to update

Suited for changing datasets

Multi-criteria query (2/2)

- Implementation using a kd-tree index
  - stored by blocks in a file
  - no size limit (tested on the 2+ billion GAIA GUMS records)
  - generic (can store any data type: float, integer, complex, ...)
  - support of multithreaded queries
- Applications:
  - fast multi-criteria queries
    (e.g. indexes on the SDSS 4 colors u-g, g-r, r-i, i-z)
  - density maps (e.g. GAIA density map generator)
  - classification algorithms
    (Kernel Density Classification, Mean Shift, ...)



- Status
  - working prototype

VizieR global index (1/2)

- **Goal**: *retrieve as fast as possible VizieR sources included in a given region of the sky*
- How ? build an index storing positions of all VizieR sources
  - 9 billion entries
  - one entry contains
    - VizieR table identifier
    - Source identifier in the table
    - Object position
- Implementation
  - hierarchy of directories and files
  - sources are sorted by HEALPix index
  - one file = one HEALPix cell
  - Fast : 8 seconds to retrieve basic info for 14 million sources in a 2 degrees cone around SMC

VizieR global index (2/2)

- **Challenge**: updating the index with sources of new tables
- possible extensions:
  - very fast SED building
  - find out all knowledge about what is observed and published at any position in the sky
    (full characterisation of a small sky area)
- Status:
  - working prototype
  - ongoing work for updating the index

2. CDS in the VO

TAP interfaces for CDS services

- TAP = *Table Access Protocol*
  a Virtual Observatory protocol
  for accessing tabular data
  - allows one for complex queries («find me all sources around Orion having J-K 2MASS color smaller than 1 and Hipparcos proper motion greater than 10mas/year»)
  - relies on ADQL query language

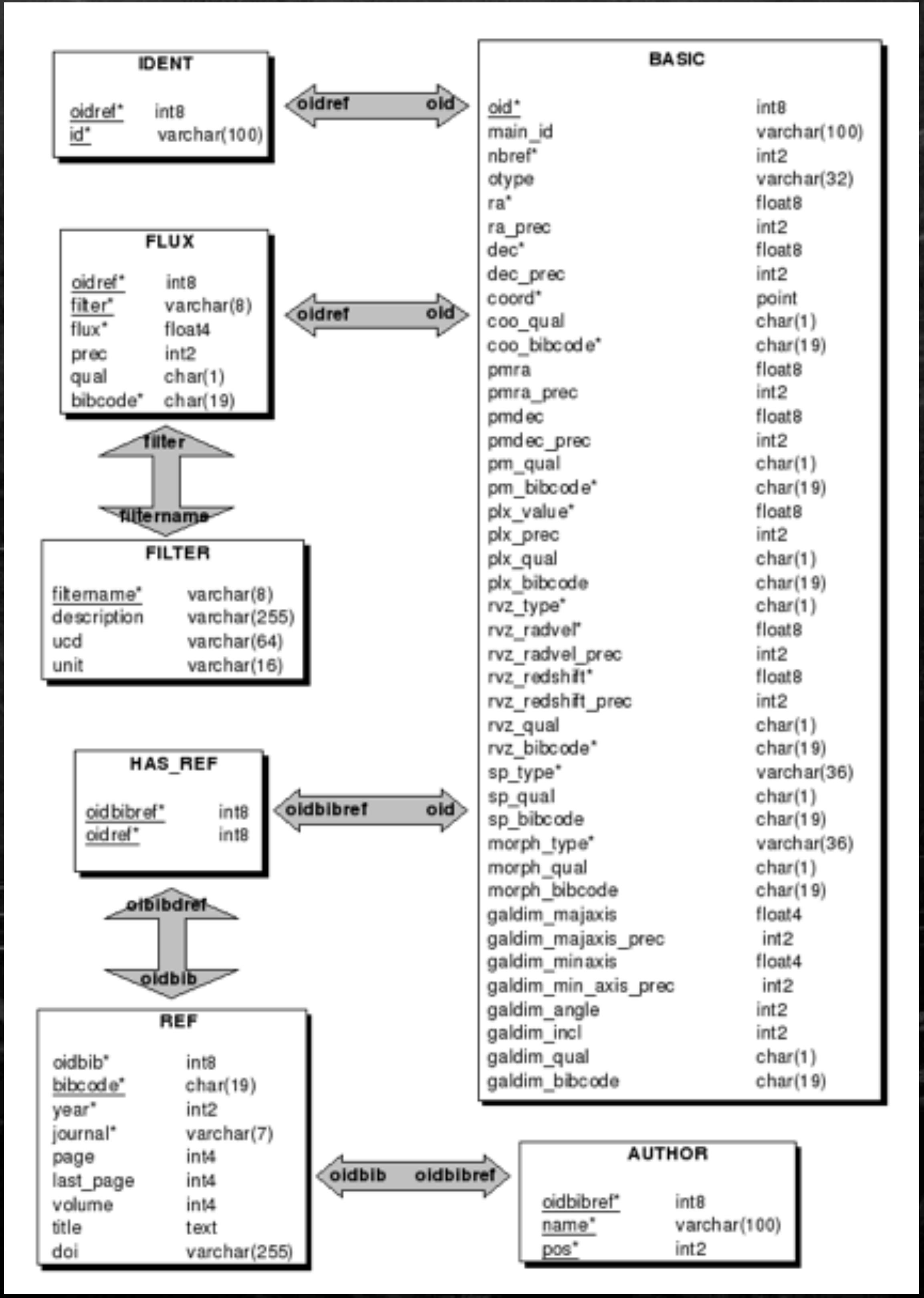
```
SELECT *
FROM t2MASS, tHipparcos
WHERE t2MASS.Jmag−t2MASS.Kmag<1 AND ...
```

- TAP implementations at CDS are based on TAP library and ADQL parser developed in-house and publicly available

Simbad TAP interface

- released last February
  http://simbad.u-strasbg.fr/simbad/sim-tap
- exposes a simplified view of SIMBAD database
- more powerful than existing query interfaces (by coordinates, by script, by criteria) but syntax more difficult to learn
  - documentation
  - examples, syntax *cheat sheet*
  - query checker

Simbad TAP interface



IDENT	
oidref*	int8
id*	varchar(100)

FLUX	
oidref*	int8
filter*	varchar(8)
flux*	float4
prec	int2
qual	char(1)
bibcode*	char(19)

FILTER	
filtername*	varchar(8)
description	varchar(255)
ucd	varchar(64)
unit	varchar(16)

HAS_REF	
oidbibref*	int8
oidref*	int8

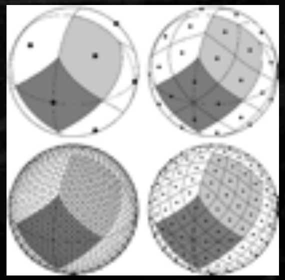
REF	
oidbib*	int8
bibcode*	char(19)
year*	int2
journal*	varchar(7)
page	int4
last_page	int4
volume	int4
title	text
doi	varchar(255)

BASIC	
oid*	int8
main_id	varchar(100)
nbref*	int2
otype	varchar(32)
ra*	float8
ra_prec	int2
dec*	float8
dec_prec	int2
coord*	point
coo_qual	char(1)
coo_bibcode*	char(19)
pmra	float8
pmra_prec	int2
pmdec	float8
pmdec_prec	int2
pm_qual	char(1)
pm_bibcode*	char(19)
plx_value*	float8
plx_prec	int2
plx_qual	char(1)
plx_bibcode	char(19)
rvz_type*	char(1)
rvz_radvel*	float8
rvz_radvel_prec	int2
rvz_redshift*	float8
rvz_redshift_prec	int2
rvz_qual	char(1)
rvz_bibcode*	char(19)
sp_type*	varchar(36)
sp_qual	char(1)
sp_bibcode	char(19)
morph_type*	varchar(36)
morph_qual	char(1)
morph_bibcode	char(19)
galdim_majaxis	float4
galdim_majaxis_prec	int2
galdim_minaxis	float4
galdim_min_axis_prec	int2
galdim_angle	int2
galdim_incl	int2
galdim_qual	char(1)
galdim_bibcode	char(19)

AUTHOR	
oidbibref*	int8
name*	varchar(100)
pos*	int2

VizieR TAP interface (1/2)

- Challenges
  - Volumetry : large surveys make large DB tables
    - SDSS8: 673GB
    - Total size in PostgreSQL: 3.5 TB
  - Ingestion in DB from original VizieR catalogues
    - different storage mechanisms: DBMS, binary files
    - conversion from fields stored value to actual value
  - managing heterogeous coordinates systems
    - computation of J2000 positions at epoch 2000.0
- Implementation
  - one dedicated server
  - PostgreSQL database with H3C indexing (based on HEALPix tessellation)

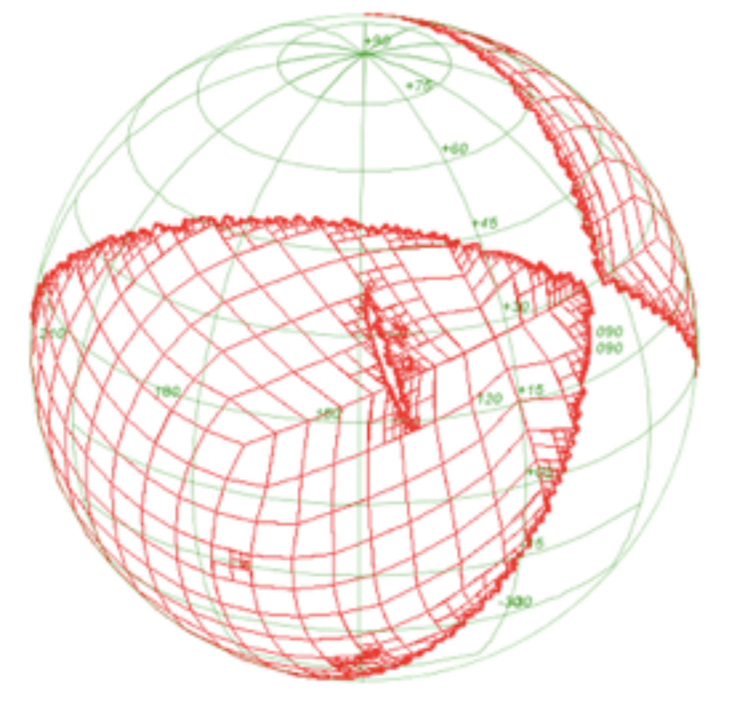


VizieR TAP interface (2/2)

- Feedback on limitations of the TAP standard
  - TAP schema requires to describe all columns exposed by a TAP service
    - not suited to VizieR 300,000 columns
      ⟶ TAP should provide with an alternative way to describe columns of a given table
  - no function in ADQL to perform conversion between coordinate systems
- Status
  - currently in beta test for Strasbourg people

MOC:
Multi Order Coverage map

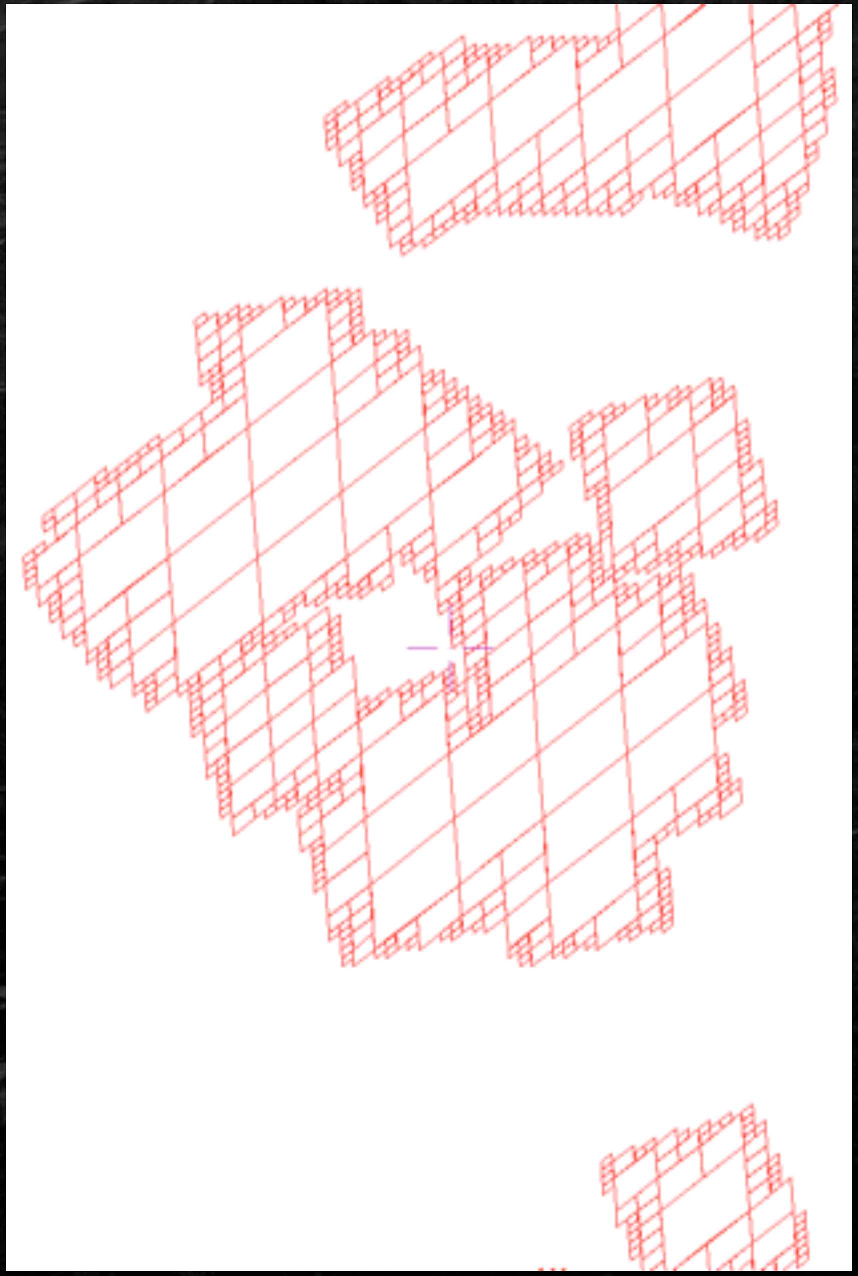
- The idea:
  - A MOC: a simple and powerful method to specify sky regions.
  - Goals : provide very fast comparisons and data access methods.
- Principle: based on HEALPix sky tesselation
- History:
  - Jan 2011 : the idea + prototype in Aladin
  - May 2011 : Napoli Interop presentation + VizieR catalog MOCs generation
  - Nov 2011 : ADASS poster
  - Dec 2011 : TOPcat implementation (multicone-search)
  - Apr 2012 : IVOA note (Boch,Donalson,Fernique, O'Mullane, Reinecke, Taylor)



MOC in action

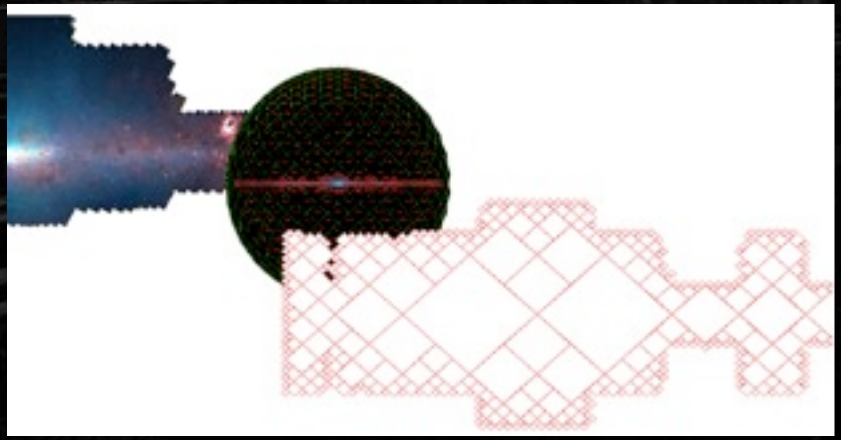
« Provide me HST images in which there are interesting quasars »

- Load the MOC of the HST images (computed from the HEALPix HST allsky survey)
- Compute the intersection with MOCs of VizieR quasar tables
- Provide the quasar measurements in this MOC intersection (query by MOC)
- Load the « progenitors » (original HST images) in this MOC intersection



MOC status

- MOCs for 7,000 VizieR catalogues with positions available from: http://alasky.u-strasbg.fr/footprints/tables/vizier/
- Java libraries for generation and manipulation of MOCs
- Oral presentation at ADASS 2012 (collaboration CDS/CADC)



3. Web 2.0/3.0 and other collaborations

HEALPix visualization

- Development of a Web prototype to visualize HEALPix surveys
  - based on HTML5/Canvas, a solution adapted to both desktop and mobile devices
  - web application running in browser, no plugin needed
- Involved in a CNES R&T action which has just started
  - developement by CNES contractors of a Web HEALPix visualizer based on HTML5/WebGL (efficient on desktops but not well implemented on mobile devices)
  - production of a Javascript HEALPix library which should be usable in our own developments

Scientific workflows

- Work started in July 2011 through an IVOA Note
  - existing projects, tools, needs in this domain
- IVOA discussion list
- Presentations during IVOA Interop sessions
- BoF at ADASS Paris in november 2011 co-animated with Juande Santander Vela (ESO)

iCORDI
INTERNATIONAL COLLABORATION
ON RESEARCH DATA INFRASTRUCTURE

- FP7- Infrastructures - 2012 - 1, started since 1/09/2012 (2 years) and coordinated by CSC-IT Center For Science Finland
  - We have started a work with the University of Edinburgh (EPCC) who is developing OGSA-DAI (a framework for distributed data access and management)
    - implementation of OGSA-DAI at CDS (to learn it)
    - supporting them in TAP access

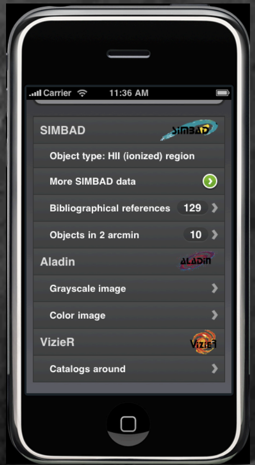
Mobile apps

- Developement of a new application (SkyObjects) for both iPhone/iPad, android and HTML5 to evaluate the human cost and the conversion tools (e.g. HTML5 to native applications)
- Improvement of SkySurveys (the android visualizer of HEALPix surveys)
- Poster at ADASS Paris 2011 and an oral presentation at the next ADASS in Urbana Champaign in November 2012

Illustrations

- Some images
- Video clips

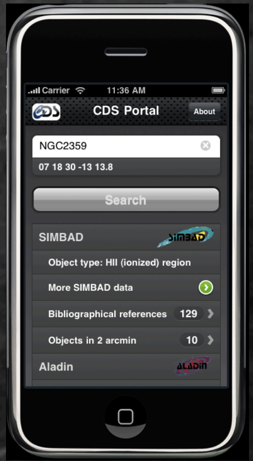
SAMSUNG
SkyObjects
Go to...
Helps you to find the position of an object in the Sky around you...
Return
Bookmarks


.ıll Carrier 11:36 AM

SIMBAD
- Object type: HII (ionized) region
- More SIMBAD data
- Bibliographical references 129
- Objects in 2 arcmin 10

Aladin
- Grayscale image
- Color image

VizieR
- Catalogs around


.ıll Carrier 11:36 AM

CDS Portal — About

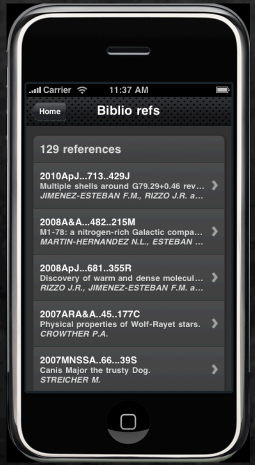
NGC2359

07 18 30 -13 13.8

Search

SIMBAD
- Object type: HII (ionized) region
- More SIMBAD data
- Bibliographical references 129
- Objects in 2 arcmin 10

Aladin


.ıll Carrier 11:37 AM

Home — Biblio refs

129 references

2010ApJ...713..429J
Multiple shells around G79.29+0.46 rev...
JIMENEZ-ESTEBAN F.M., RIZZO J.R. a...

2008A&A...482..215M
M1-78: a nitrogen-rich Galactic compa...
MARTIN-HERNANDEZ N.L., ESTEBAN ...

2008ApJ...681..355R
Discovery of warm and dense molecul...
RIZZO J.R., JIMENEZ-ESTEBAN F.M. a...

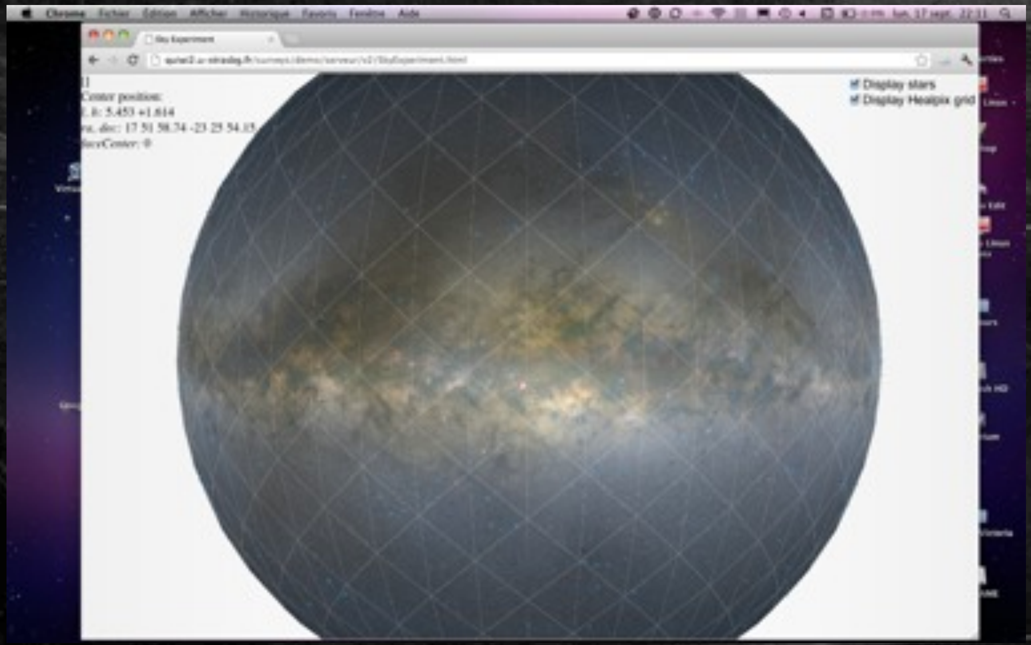
2007ARA&A..45..177C
Physical properties of Wolf-Rayet stars.
CROWTHER P.A.

2007MNSSA..66...39S
Canis Major the trusty Dog.
STREICHER M.

.ıll Carrier 11:38 AM

Home — Color image

View gray image



CDS Centre de Données astronomiques de Strasbourg
Strasbourg astronomical Data Center
Other services
Latest news
Hosted services
Keep in touch
Featured news

http://simbad.u-strasbg.fr/simbad/sim-id?Ident=m51&NbI...
M 51 -- Seyfert 2 Galaxy
ARCHOS 101 internet tablet

http://cdsweb.u-strasbg.fr/SimPlay/
ARCHOS 101 internet tablet

http://vizier.u-strasbg.fr/viz-bin/VizieR
VizieR Service
Search Criteria
Find catalogs among 10500 available
Search by Position across 10059 tables
Find Catalogs
ARCHOS 101 internet tablet

Video clips

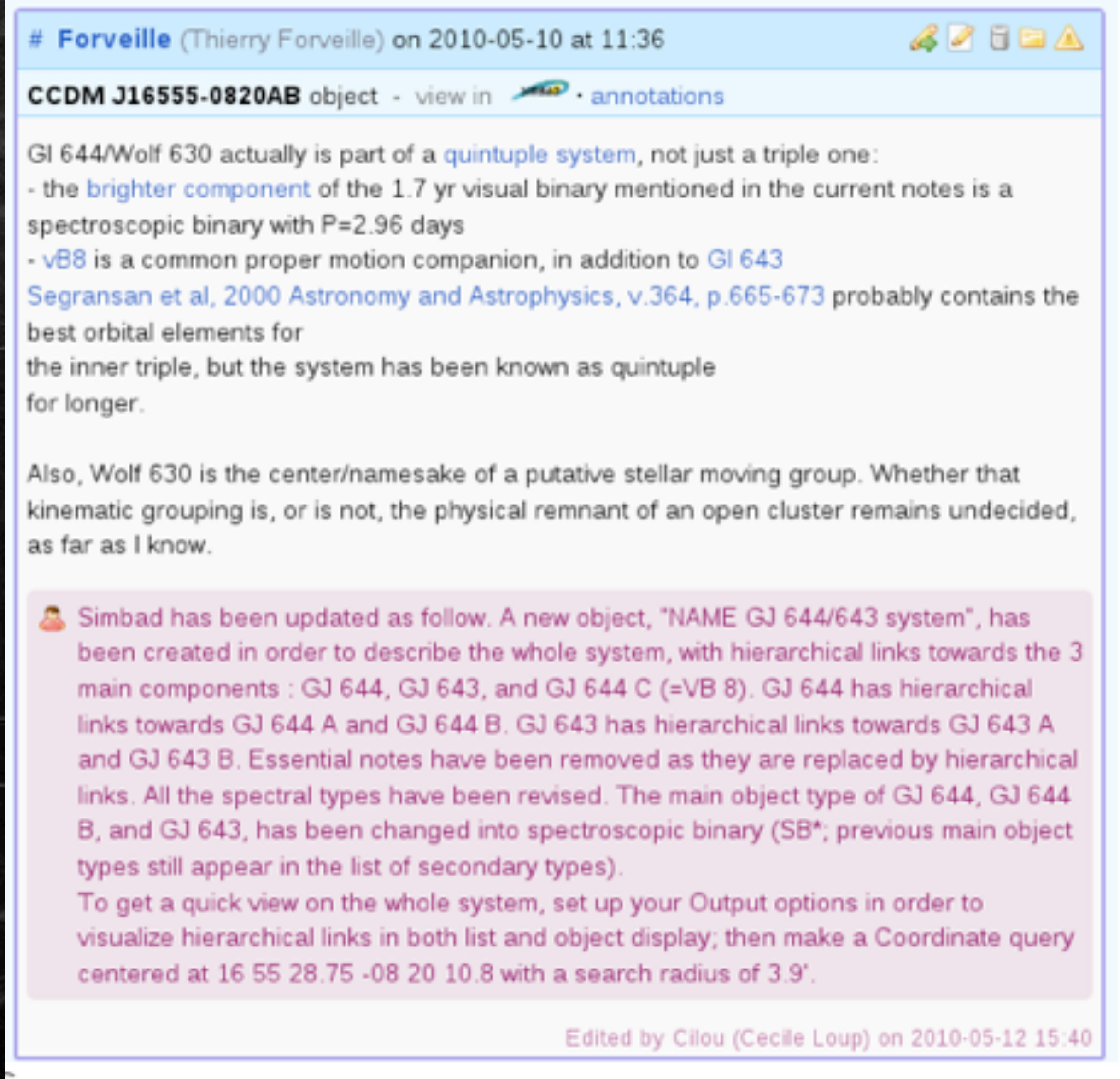
- …

Planned developments

- Improve the HEALPix HTML5/Canvas visualizer
- Finalize SkyObjects and publish it
- Explore the possible use of Multitouch and 3D screens in astronomy (meeting with people from VO France)
- Evaluate the use of Clouds in CDS services through an implementation

CDS annotations

- Statistics since March 2010
  - 465 users have subscribed
    - 167 users have posted at least one annotation
  - 1000 annotations have been posted (~1/day)
    - 815 different SIMBAD objects
    - 123 private
    - 407 error notifications
  - 14,000 pages viewed
  - no spam



Forveille (Thierry Forveille) on 2010-05-10 at 11:36

CCDM J16555-0820AB object - view in · annotations

Gl 644/Wolf 630 actually is part of a quintuple system, not just a triple one:
- the brighter component of the 1.7 yr visual binary mentioned in the current notes is a spectroscopic binary with P=2.96 days
- vB8 is a common proper motion companion, in addition to Gl 643
Segransan et al, 2000 Astronomy and Astrophysics, v.364, p.665-673 probably contains the best orbital elements for
the inner triple, but the system has been known as quintuple
for longer.

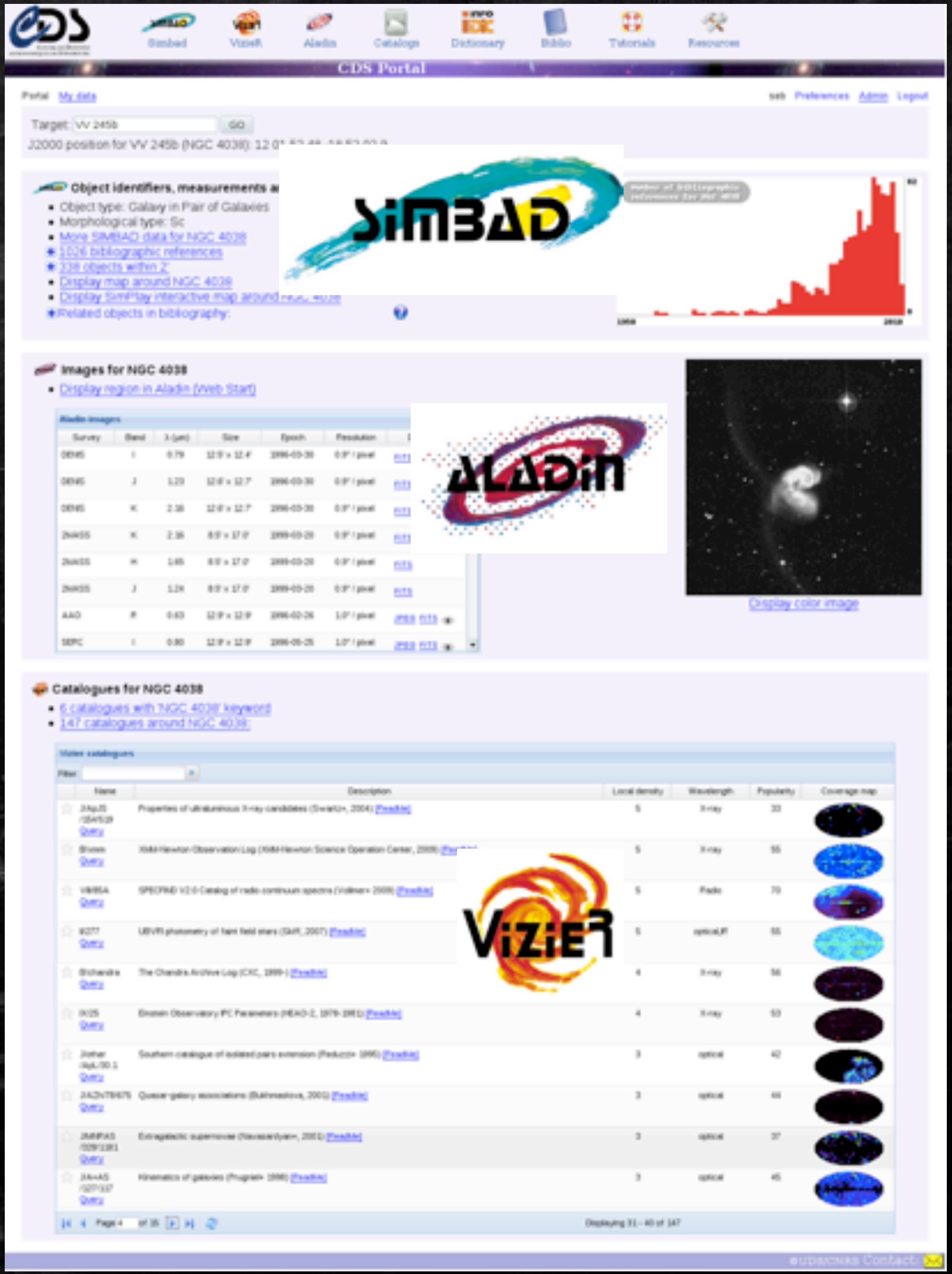
Also, Wolf 630 is the center/namesake of a putative stellar moving group. Whether that kinematic grouping is, or is not, the physical remnant of an open cluster remains undecided, as far as I know.

Simbad has been updated as follow. A new object, "NAME GJ 644/643 system", has been created in order to describe the whole system, with hierarchical links towards the 3 main components : GJ 644, GJ 643, and GJ 644 C (=VB 8). GJ 644 has hierarchical links towards GJ 644 A and GJ 644 B. GJ 643 has hierarchical links towards GJ 643 A and GJ 643 B. Essential notes have been removed as they are replaced by hierarchical links. All the spectral types have been revised. The main object type of GJ 644, GJ 644 B, and GJ 643, has been changed into spectroscopic binary (SB*; previous main object types still appear in the list of secondary types).
To get a quick view on the whole system, set up your Output options in order to visualize hierarchical links in both list and object display; then make a Coordinate query centered at 16 55 28.75 -08 20 10.8 with a search radius of 3.9'.

Edited by Cilou (Cecile Loup) on 2010-05-12 15:40

CDS Portal

- Planned improvements
  - Data & service discovery using additional inputs
  - Smart interpretation of input
  - Keep it simple



Portal My data

seb Preferences Admin Logout

Target: VV 245b GO

J2000 position for VV 245b (NGC 4038): 12 01 52.48 -18 52 02.9

Object identifiers, measurements a

- Object type: Galaxy in Pair of Galaxies
- Morphological type: Sc
- More SIMBAD data for NGC 4038
- 1026 bibliographic references
- 338 objects within 2'
- Display map around NGC 4038
- Display SimPlay interactive map around NGC 4038
- Related objects in bibliography:

Images for NGC 4038

- Display region in Aladin (Web Start)

Aladin images

Survey	Band	λ (µm)	Size	Epoch	Resolution	D
DENIS	I	0.79	12.5' x 12.4'	1996-03-30	0.9" / pixel	FITS
DENIS	J	1.23	12.6' x 12.7'	1996-03-30	0.9" / pixel	FITS
DENIS	K	2.16	12.6' x 12.7'	1996-03-30	0.9" / pixel	FITS
2MASS	K	2.16	8.5' x 17.0'	1999-03-20	0.9" / pixel	FITS
2MASS	H	1.65	8.5' x 17.0'	1999-03-20	0.9" / pixel	FITS
2MASS	J	1.24	8.5' x 17.0'	1999-03-20	0.9" / pixel	FITS
AAO	R	0.63	12.9' x 12.9'	1996-02-26	1.0" / pixel	JPEG FITS
SERC	I	0.80	12.9' x 12.9'	1996-05-25	1.0" / pixel	JPEG FITS

Display color image

Catalogues for NGC 4038

- 6 catalogues with 'NGC 4038' keyword
- 147 catalogues around NGC 4038:

VizieR catalogues

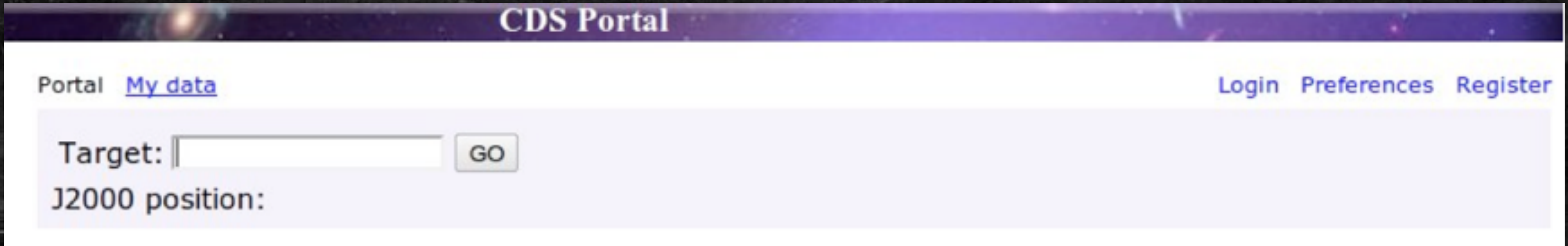
Filter:

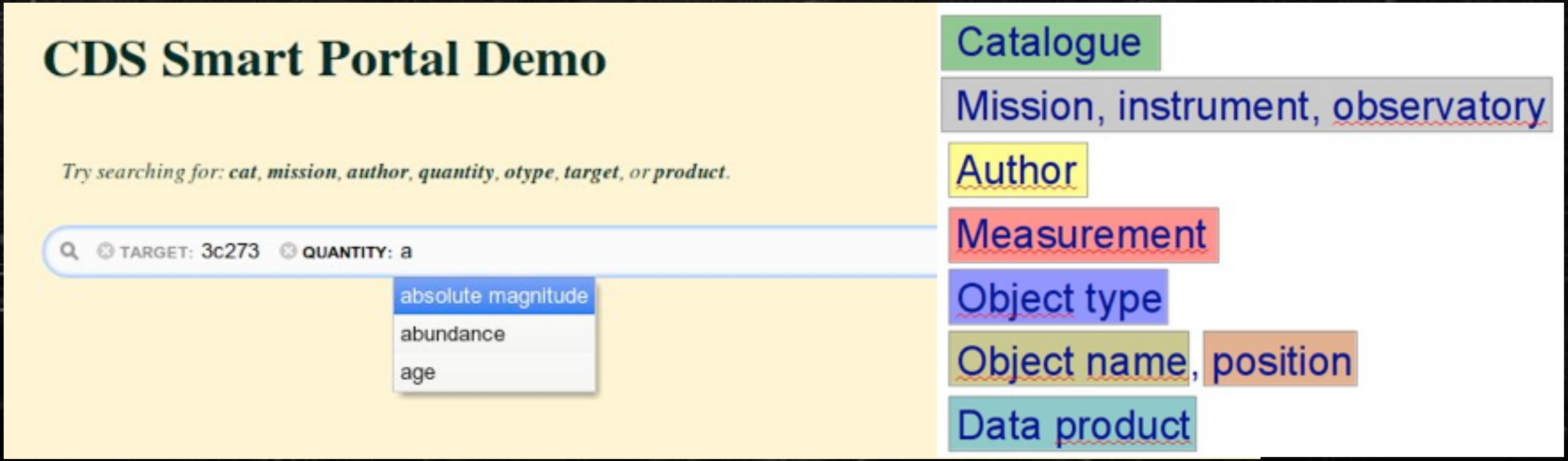
Name	Description	Local density	Wavelength	Popularity	Coverage map
J/ApJS/154/519 Query	Properties of ultraluminous X-ray candidates (Swartz+, 2004) [ReadMe]	5	X-ray	33	
B/xmm Query	XMM-Newton Observation Log (XMM-Newton Science Operation Center, 2009) [Re	5	X-ray	55	
VIII/85A Query	SPECFIND V2.0 Catalog of radio continuum spectra (Vollmer+ 2009) [ReadMe]	5	Radio	70	
II/277 Query	UBVRI photometry of faint field stars (Skiff, 2007) [ReadMe]	5	optical,IR	55	
B/chandra Query	The Chandra Archive Log (CXC, 1999-) [ReadMe]	4	X-ray	58	
IX/25 Query	Einstein Observatory IPC Parameters (HEAO-2, 1978-1981) [ReadMe]	4	X-ray	53	
J/other/ApL/30.1 Query	Southern catalogue of isolated pairs extension (Reduzzi+ 1995) [ReadMe]	3	optical	42	
J/AZh/78/675 Query	Quasar-galaxy associations (Bukhmastova, 2001) [ReadMe]	3	optical	44	
J/MNRAS/328/1181 Query	Extragalactic supernovae (Navasardyan+, 2001) [ReadMe]	3	optical	37	
J/A+AS/127/117 Query	Kinematics of galaxies (Prugniel+ 1998) [ReadMe]	3	optical	45	

Page 4 of 15 — Displaying 31 - 40 of 147

«Smart» CDS portal

http://cdsportal.u-strasbg.fr/





CDS Smart Portal Demo

Try searching for: ***cat, mission, author, quantity, otype, target,*** *or* ***product.***

Oral presentation at AstroInformatics 2012, 09/2012