Virtual Observations lectures – Kapteyn Institute - Groningen

15/12/2006

CDS and the VO

Sébastien Derriere
derriere@astro.u-strasbg.fr

Outline of the talk

- Introduction
- The main CDS services:

- The Virtual Observatory and the CDS
- Practical demos using Aladin

CDS - over 30 years of history

- **1972**: creation of the **C**entre de **D**onnées **S**tellaires.
  - Electronic data
  - Expertise on data
  - International center
  - Goal: research
- **1983**: extension to galaxies an other non-stellar bodies: **C**entre de **D**onnées astronomiques de **Strasbourg**

Collect, homogenize, distribute, preserve astronomical information for the whole community.

CDS - over 30 years of history

- **1993/94: WWW starts**
- **2001**: CDS participates in Virtual Observatory projects
- CDS today:
  - Collaboration INSU/ULP (Institut National des Sciences de l'Univers; Université Louis Pasteur)
  - ~25 staff members in Strasbourg team + a few other ones in France:
    - 'documentalists'
    - astronomers
    - software developers
  - Many collaborations (national and international)
  - Several 10^4 hits/day on our services

A data center (1)

- Why care about data preservation in astronomy?
  - Instruments are more and more powerful, why keep « old » less accurate data?
- Mandatory mission to preserve data:
  - variable phenomenons require observations over a time range: changes in luminosity, positions, with possibly very long periods (centuries)
  - statistics
  - re-use of data: saves time and money (VO: archive=telescope)

A data center (2)

- Multi-wavelength astronomy
  - better analysis of physical processes, usage of data by non-specialists
- Large surveys
  - complexity and volumes are growing
- WWW
  - direct access to distributed data

A data center (3)

- What data in astronomy?
  - compilation databases (SIMBAD/NED)
  - observations (reference images, surveys, ground or space-based observation archives)
  - spectras
  - catalogues (observation, compilation)
  - bibliography (journals, ADS, preprints)
  - yellow pages, softwares
  - personal data
- Exponential growth of data volumes

Data volumes

- 1801, Uranographia: 17,200 stars;
- 1924-1936, HD: 272,150 stars;
- 1989, IRAS: 500,000 sources;
- 1997, Tycho: 1,000,000 sources;
- 1997, USNO-A1 488,006,860 sources;
- 2003, USNO-B1 1,045,913,669 sources.
- > 5000 new bibliographic references/year
- 4-500 new catalogues/ year (including electronic publications of journal tables)
- modern sky surveys: several Tb of images

Data storage

1975 – Punch card

1974 – hard drive 400kb

1980 - magnetic tapes

CDrom 700Mb

2006 - HD 500Gb

A data center (4)

- A data center is not only a data warehouse
- A validation work is needed
- Data must come with quality metadata describing
  - their nature
  - their origin
  - the processing history
  - quality
  - ...

Activities of CDS (1)

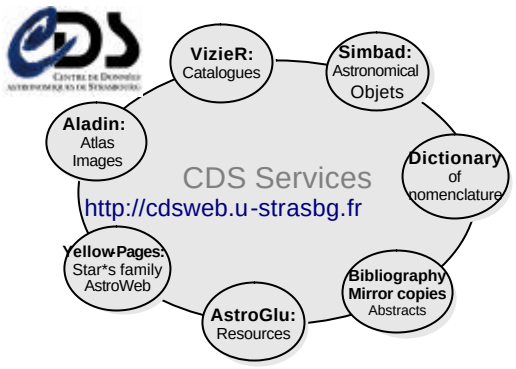
- Reference services with high value-added for the astronomical community
- Technology watch, R&D / pluridisciplinary actions
- Evolution of the services
  - daily work on contents (+5000 bib. refs, +500 catalogues/year) and features
  - contents validation
  - operational constraints on services

Activities of CDS (2)

- Selective collect and distribution of astronomical data (tables, logs, surveys, publications, reference images):
  - value added to raw data through evaluation and comparisons ;
  - distribute results to the community;
  - do research on these data.

Activities of CDS (3)

- Develop databases and access interfaces
- Miror copies (journals, bibliography)
- Support for users (*question@simbad*)
- Knowledge diffusion
- Participation in projects
- Virtual Observatory
  - evolution of reference services
  - standards and tools
  - national action (OV-France)



The main CDS services

Astronomical Objects
identifications, bibliography, measurements

Catalogue service
catalogues, published tables, observation logs, surveys

Information integration
images, databases, catalogues, archives, personal data

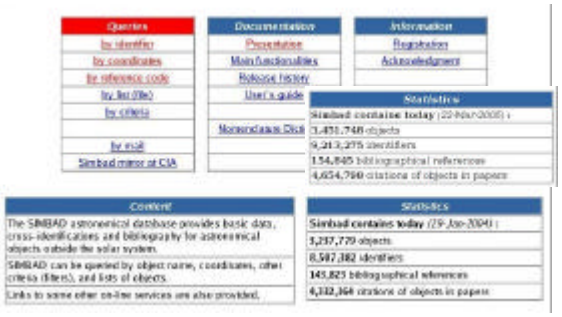
~25,000 queries/day

Identifiers, basic data, bibliography and measures from:
- published literature
- selected catalogues

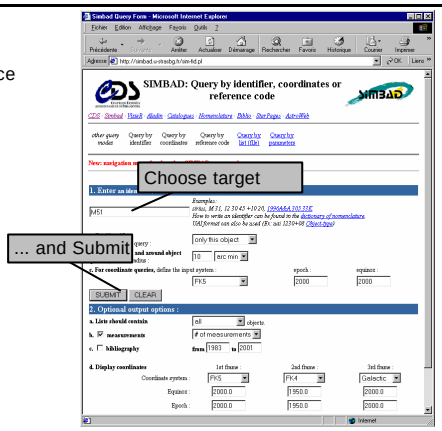
Dictionary of nomenclature (collab. GEPI)
Systematic cross-identification

Contents (updated daily) from:
- Bibliography (90 journals, collaborations Obs. Paris, IAP, ...)
- Catalogues (multi-wavelength coverage, major projects) – in collaboration with experts
- Collab. from OMP, GRAAL, ...

SIMBAD = **S**et of **I**dentifications, **M**easurements and **B**ibliography for **A**stronomical **D**ata.



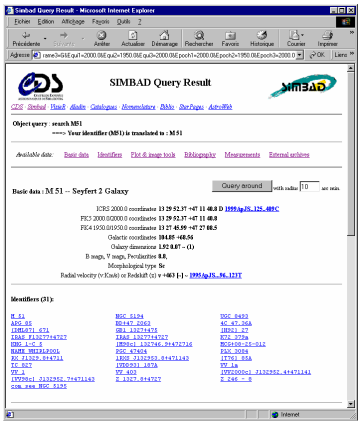
SIMBAD3
Web Interface



Result...

Data

Identifiers



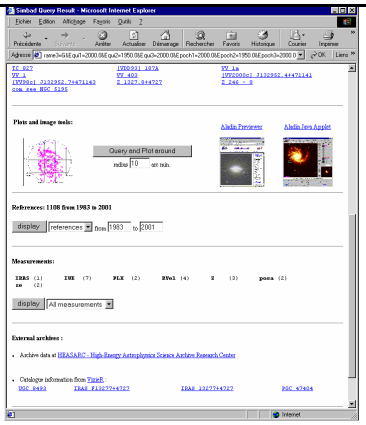
Result...

Links to images, graphs

Bibliographic references

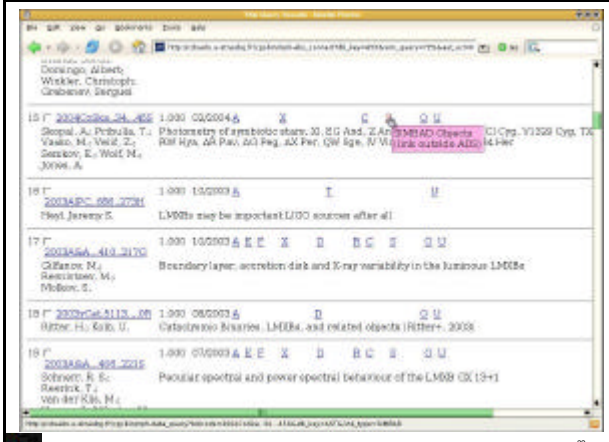
Measurements

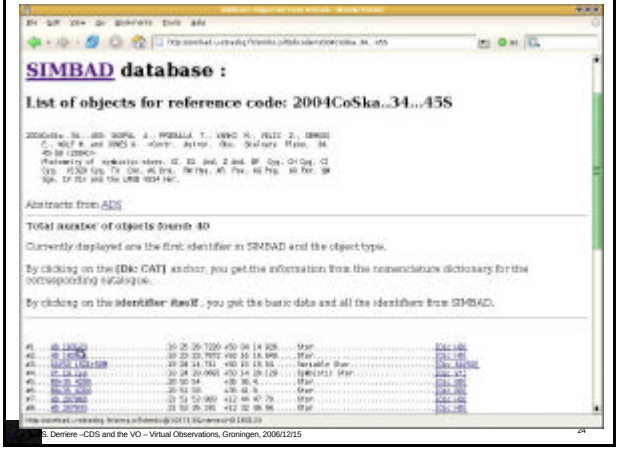
Links to external archives and VizieR catalogues

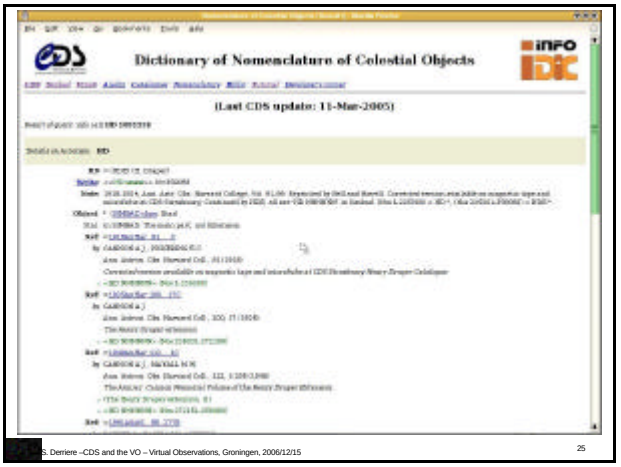


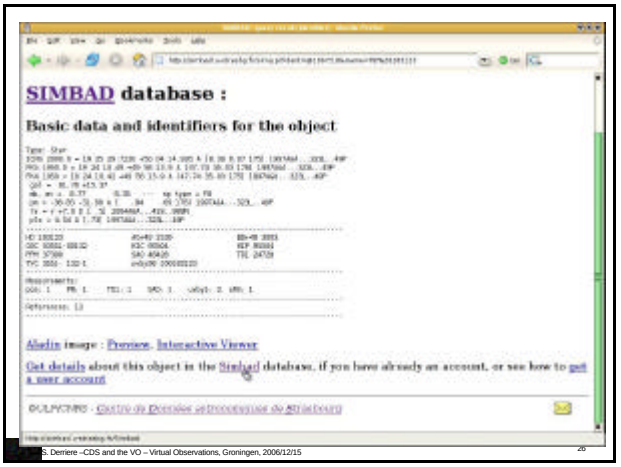
SIMBAD

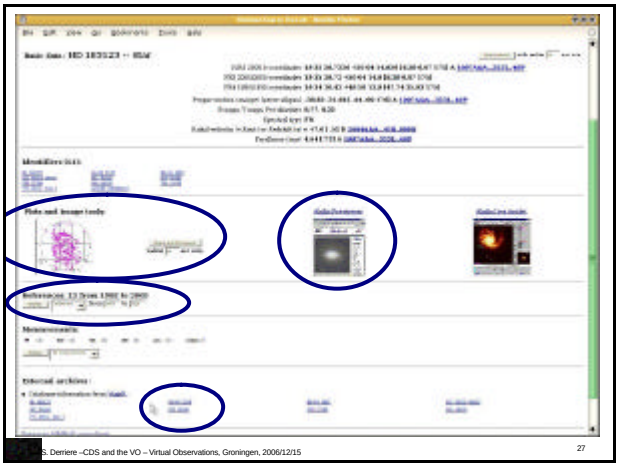
- The identifiers syntax is controlled by the dictionary of nomenclature
- Associated service : Name Resolver
  - returns coordinates for an identifier
  - used in many applications
- There are links between SIMBAD, the bibliographic services (ADS, with bibcodes), VizieR, Aladin...



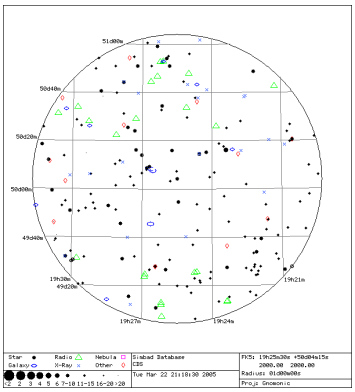


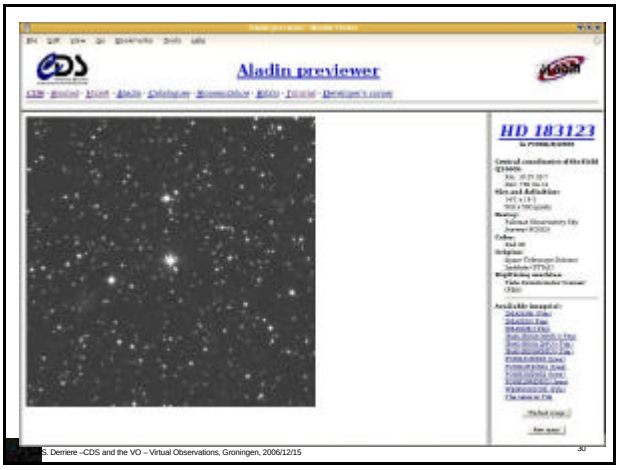


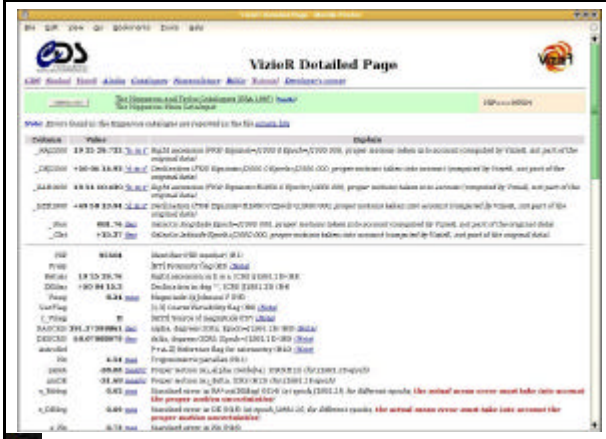




Field map for
SIMBAD objects

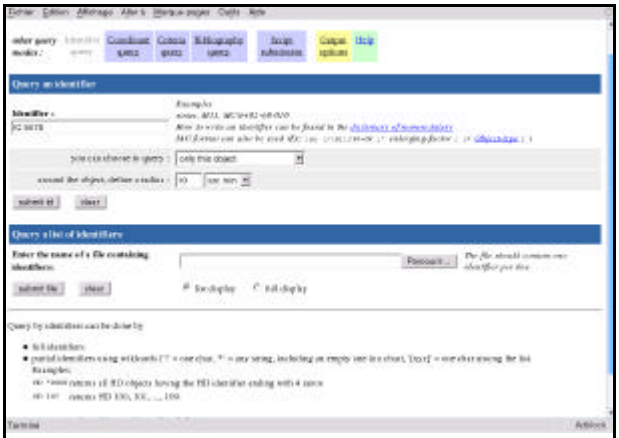


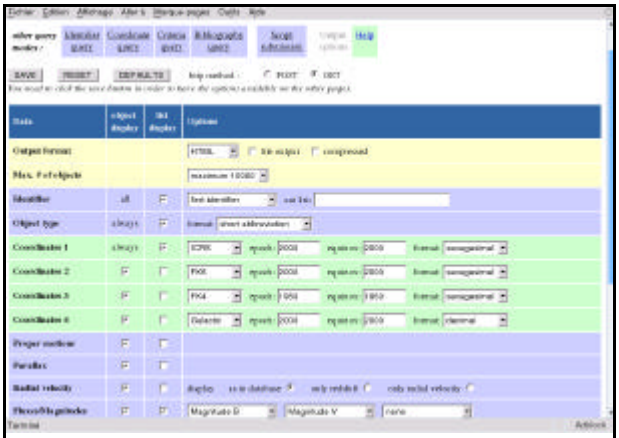




SIMBAD4

- Released december 2006
- Complete rewrite: Java + PostgreSQL
- Advanced output options and formats (VOTable)
- Query on any parameter (e.g. redshift>5)
- Easy to query by list of targets
- Improved script queries





```
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
214.503814|  -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
214.382821|  -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
214.532274|  -57.766314| 16.179| 14.323| 13.036| 0.06| 0.11| 0.16| 16.099| 14.255| 13.102
214.565229|  -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| 14.660| 13.244| 0.07| 0.12| 0.18| 16.426| 14.683| 12.860
214.490574|  -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
214.316793|  -57.766161| 15.516| 14.837| 99.999| 0.05| 0.12| 9.99| 15.496| 14.697| 99.999
214.342976|  -57.766060| 99.999| 15.804| 99.999| 9.99| 0.19| 9.99| 99.999| 15.705| 99.999
214.278851|  -57.766037| 16.967| 15.725| 99.999| 0.10| 0.18| 9.99| 17.014| 16.130| 99.999
214.588047|  -57.765921| 15.941| 13.637| 12.239| 0.06| 0.09| 0.12| 15.858| 13.620| 12.295
214.426628|  -57.765900| 99.999| 16.214| 13.928| 9.99| 0.23| 0.27| 99.999| 16.593| 13.955
214.410201|  -57.765870| 15.641| 14.798| 99.999| 0.05| 0.12| 9.99| 15.604| 14.415| 99.999
214.517290|  -57.765842| 99.999| 16.320| 99.999| 9.99| 0.24| 9.99| 99.999| 18.163| 99.999
214.463607|  -57.765814| 17.931| 15.866| 99.999| 0.17| 0.19| 9.99| 17.819| 15.194| 99.999
214.554082|  -57.765781| 15.537| 12.979| 11.402| 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
```

~30,000 queries/day

- Over 5000 catalogues (1cat = n tables)
- Metadata: calibrated and homogenized
- Standardized description (ReadMe).
- Partnership with journals (A&A since 1993).
- Same interface to access large surveys (USNO, GSC, DENIS, 2MASS, UCAC, ...).
- Many mirrors (USA, India, China, Japan, ...).

Each catalogue has a ReadMe file: simple ascii description of contents

- 1994 : 680 catalogues - 3Gb
- 2004 : 4000 catalogues - (210 Gb, up to 1 billion sources for USNOB1.0)

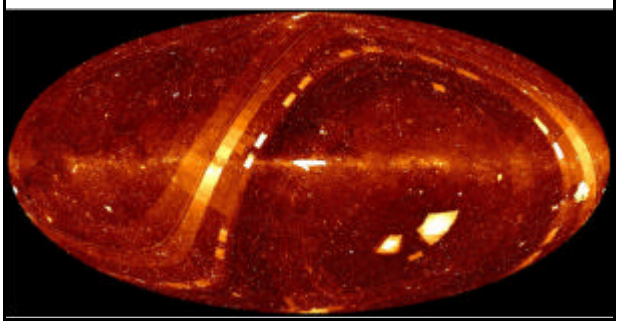
VizieR

- Catalogues from:
  - electronic publication of journal's tables
  - digitized photographic plates
  - digital surveys
  - observations, archive logs
- Catalogues can come with additional material (spectra, images, time series, ...), or external links

Many ways to use VizieR

- Catalogue discovery (find new relevant catalogues)
  - search by keywords, author, topic, mission, UCD (i.e. columns content)
  - constraint on sky coverage
- Catalogue(s) query
  - FTP download
  - Position query (cone, box)
  - Contrain parameters
  - List
- Batch mode: cdsclient package (Dev. corner)

The VizieR mine

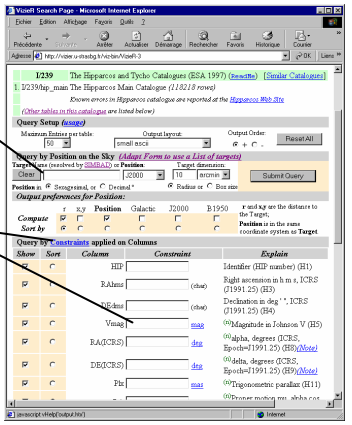


Ex: HIPPARCOS Catalogue

Search by position or name

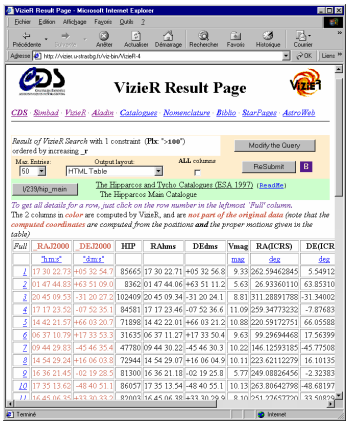
Contraints on columns

Possibility to query by list of targets.



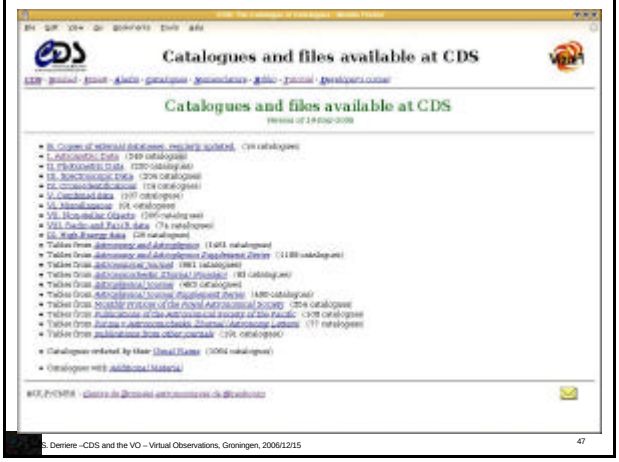
Result :

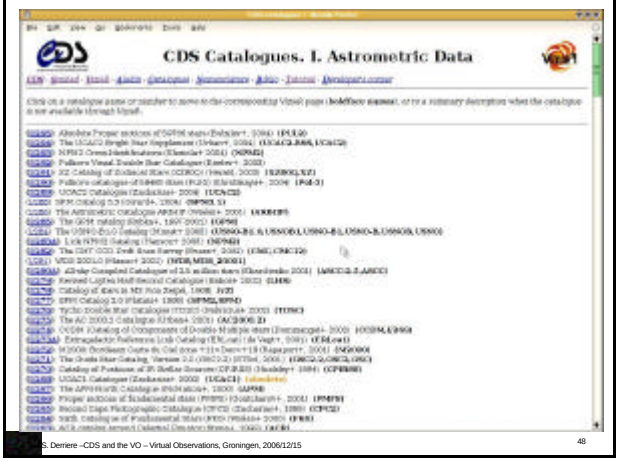
HTML
TSV
VOTable
ASCII

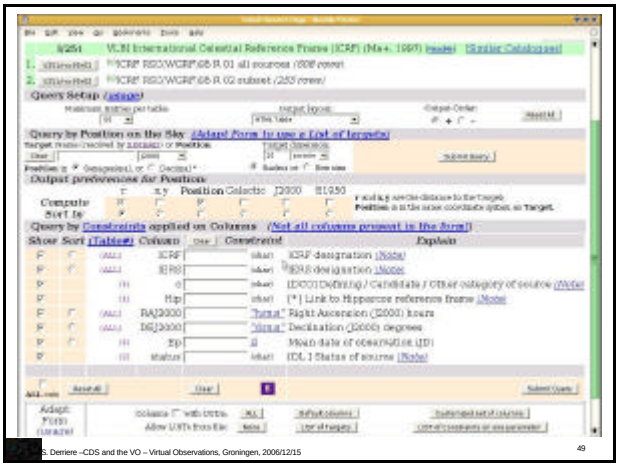


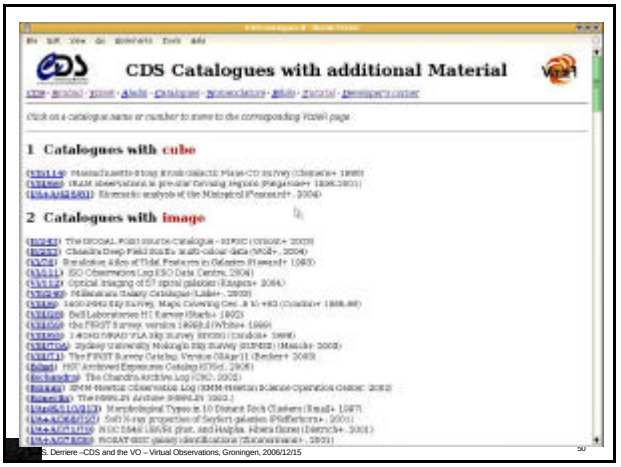
VizieR

- Astrometric reference catalogues available
  - VLBI / ICRF
  - Hipparcos, Tycho
  - 2MASS, USNO
- Many links to additional data
- Many output formats (+vizualisation)
- Some regularly updated catalogues (minor bodies in Solar System, observation logs)

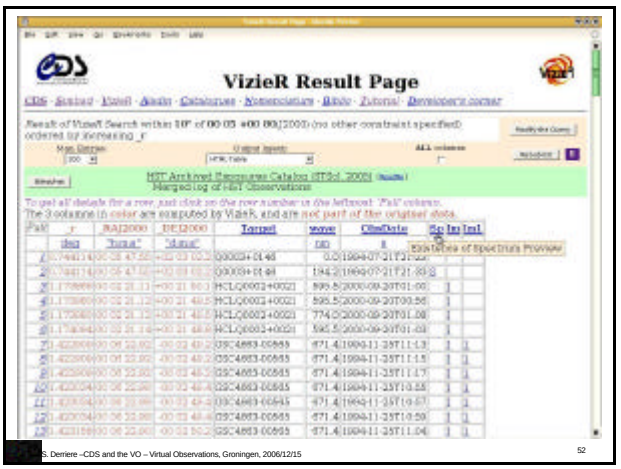




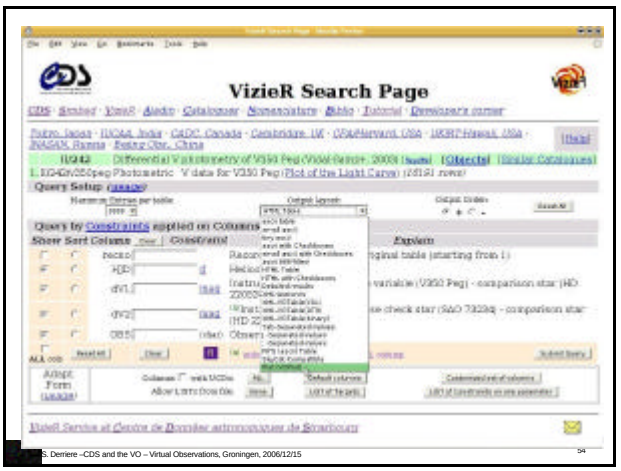

VLBI International Celestial Reference Frame (ICRF) (Ma+, 1997)
Query Setup
Query by Position on the Sky
Output preferences for Position
Query by Constraints applied on Columns
Show Sort Column Constraint Explain

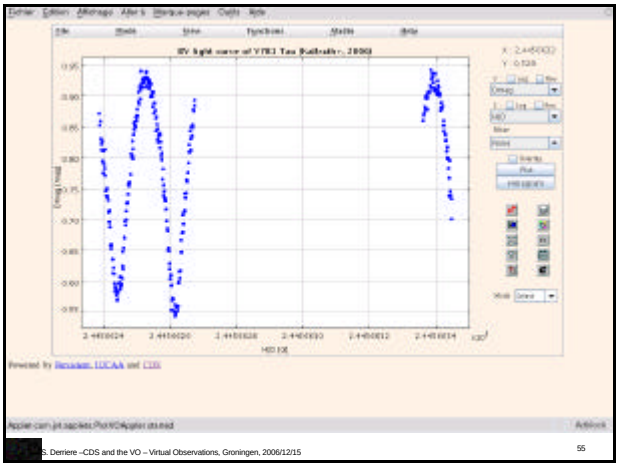

CDS Catalogues with additional Material
1 Catalogues with cube
2 Catalogues with image

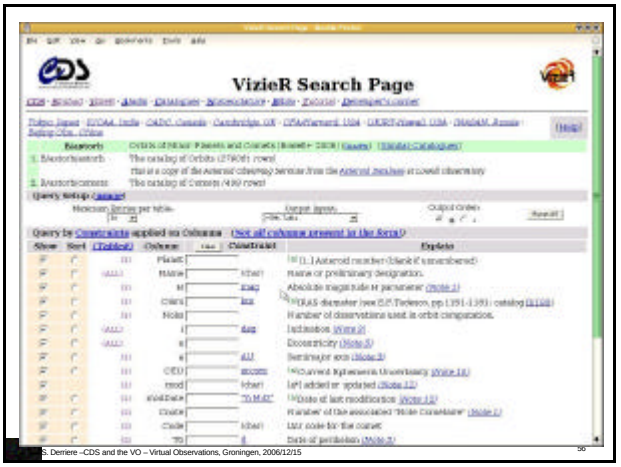
3 Catalogues with model
4 Catalogues with spectrum (or SED, line profile)

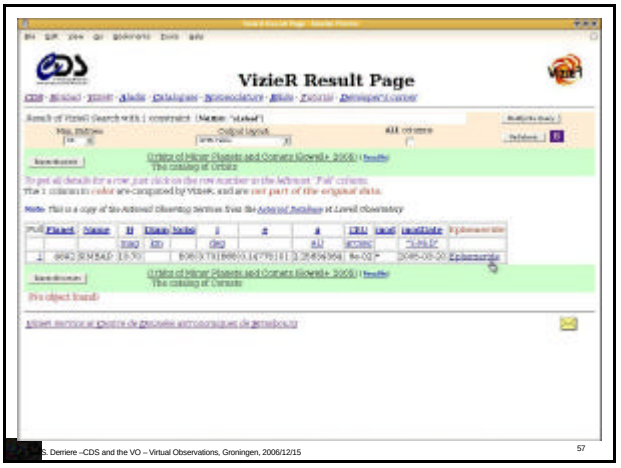

VizieR Result Page

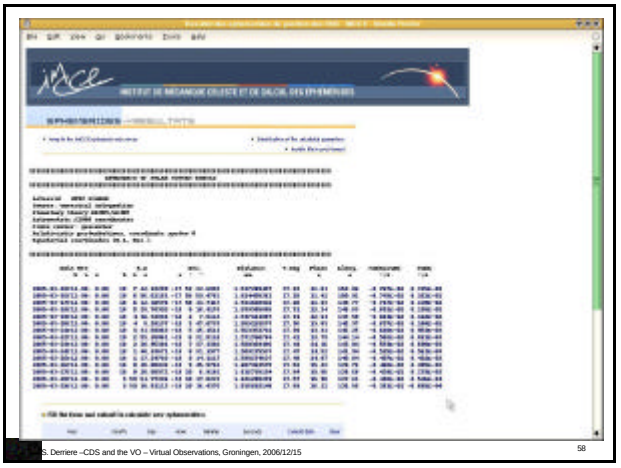
5 Catalogues with timeSerie

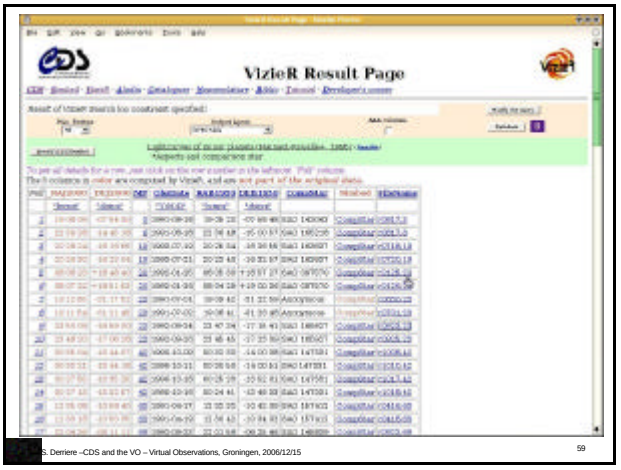

VizieR Search Page

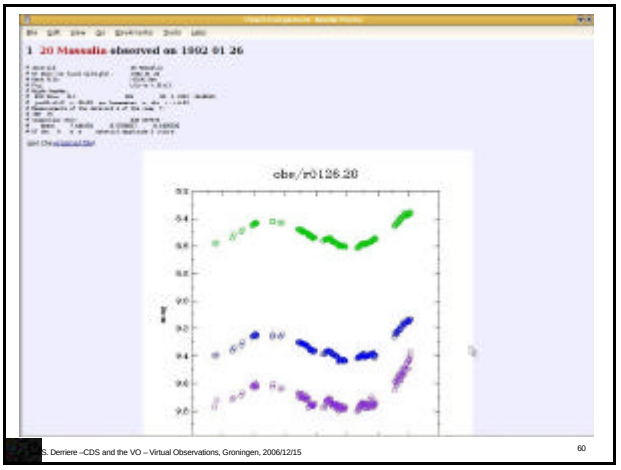



VizieR Search Page


VizieR Result Page




VizieR Result Page


1 20 Massalia observed on 1902 01 26
obs/r0126.20

4-5000 queries/day

- Younger than SIMBAD and VizieR: 1997
- Two aspects:
  - image server
  - Java interface (was X-windows at first)
- Three levels of usage:
  - Previewer (very simple)
  - Java Applet in Web page
  - Aladin Standalone : full application
- 2005: Aladin v3.0 – 2007 v4.0

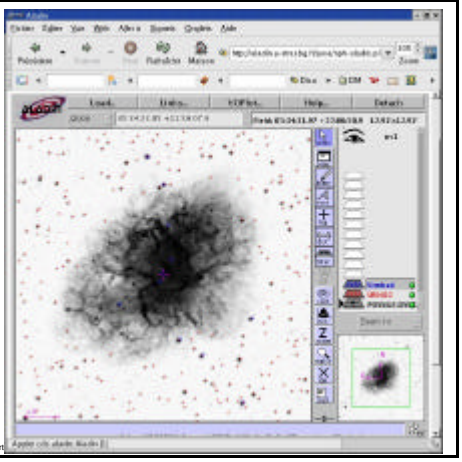
Data integration tool:

- Images in the Strasbourg image server:
  - MAMA, DSS1, DSS2, 2MASS, EROS1, IRAS
- Access to remote services:
  - HST, VLA/FIRST, SuperCOSMOS, Skyview, SDSS, ...
- Access to local data (images, catalogues)
- MANY features:
  - overlay on images: contours, catalogues, NED, SIMBAD, field of view, filters...
  - astrometric calibration, resampling, color composition, blinking, multi-view, scripting, cross-match, diagrams (VOPlot)...

Interactive sky atlas

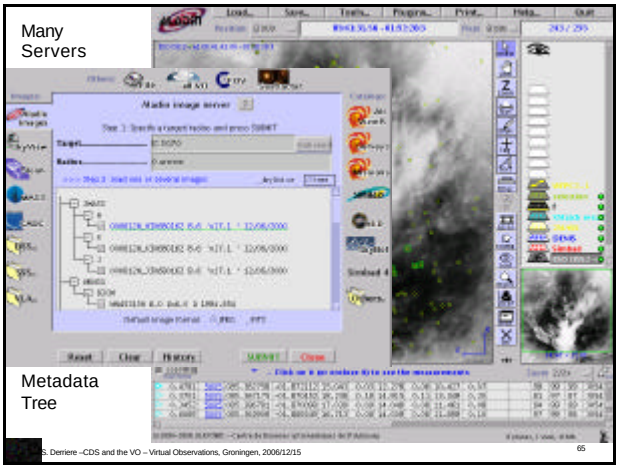
Integration of local/remote image and catalogue data.

Free !

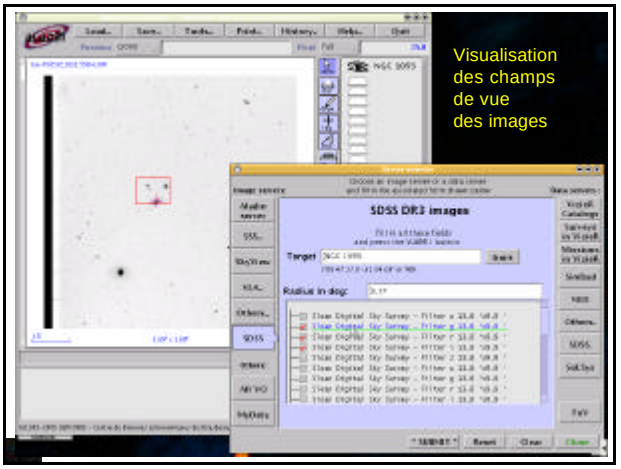


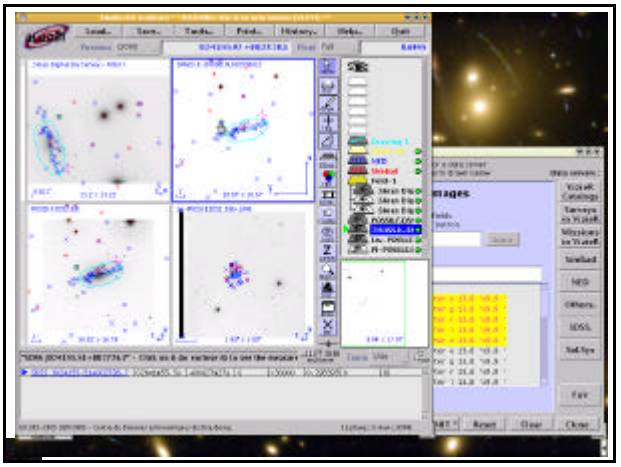
Many Servers

Metadata Tree

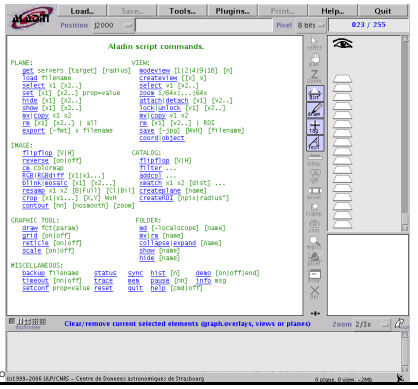


Visualisation des champs de vue des images



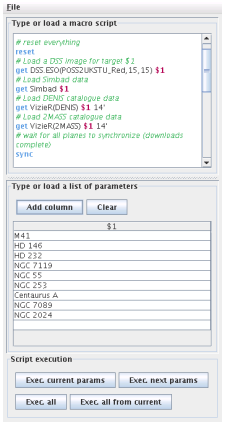


Aladin script capabilities



Aladin script capabilities

- Macro controller to easily perform a list of tasks on a list of targets



International Virtual Observatory Alliance

The Virtual Observatory

- A Virtual Observatory for astronomy:
  - Provide astronomers with tools to locate, combine and exploit all the available data and services.
  - International Virtual Observatory Alliance (IVOA)
- Need for INTEROPERABILITY between the services
  - Definition of standards to exchange data.
  - Process metadata with data.

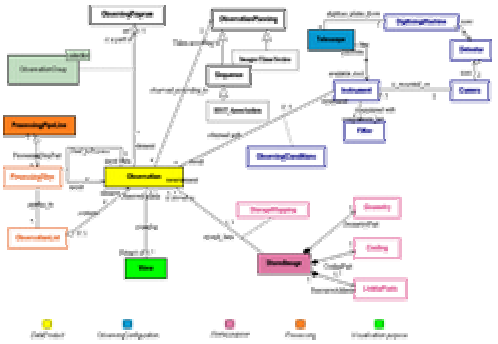
Working Groups (WGs)

1. DM (Data Models)
2. WS (Web Services) & GRID
3. Registry
4. DAL (Data Access Layer)
5. VOQL (Query Language)
6. (UCD Unified Content Descriptors) ->Semantics
7. VOTable
8. Standards and processes

(1) DM - objectives

- Abstract description of concepts and their relationships:
  - define names and meaning of concepts
  - structure concepts and relations
- Production of the WG:
  - White papers
  - UML Diagrams
  - XML Schemas (serialization)
- DM = conceptual layer of the VO

(1) DM – UML Diagram



(2) WS & GRID

- Communication with/between services:
  - HTTP GET/POST
    - human-service (ex: web form)
  - SOAP (Simple Object Access Protocol)
    - WSDL: self -describing
    - service-service (ex: SESAME Name Resolver)
  - GRID
    - SOAP + authentication, state management, ...

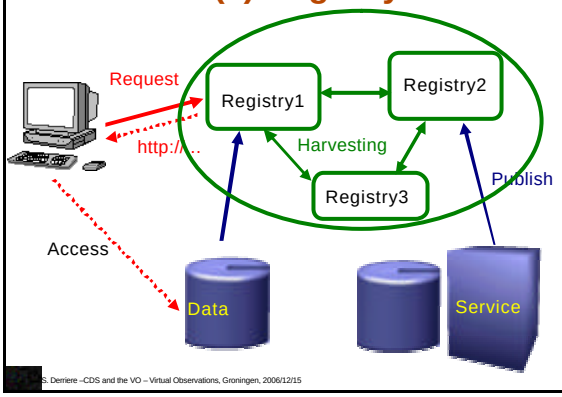
(2) WS & GRID

- Each service choses its communication mode
- GRID: computing grid – data grid
- Collaboration VO - GRID:
  - transfer programs rather than data
  - data replication
  - authentication, autorization
  - VOSpace (personnel storage, computing space). ex: mySpace
  - workflow construction (graph of services)

(3) Registry

- Description of VO Resources (data & services) by metadata
  - XML schema (*curation*, coverage)
  - similar to OAI (*library of Congress*)
  - unique identifier ivo:// for 1 resource
- 3 actions:
  - publish (data or service producer)
  - harvest (exchange between registries)
  - query (astronomers find the resources)

(3) Registry



(4/5) DAL / VOQL

- Data access protocols
- Simple Services:
  - Cone Search – catalogues with positions
  - SIAP – Simple Image Access Protocol
  - SSAP – Simple Spectrum Access Protocol: 1D spectra
- OpenSkyQuery et ADQL
  - SQL / XML requests
  - distributed execution plan (SkyNode)

(6) UCD

- UCD = Unified Content Descriptors
- Sémantic description of contents
- Origin (1997) in VizieR
  - UCD1
- VO reused it for:
  - describing tables (VOTable)
  - registry
  - DAL (Cone Search)
- New scheme: UCD1+

(7) VOTable

- XML format for exchanging tabular data
  - data and metadata in the same file
- Adoption of VOTable 1.0 april 2002; now: VOTable 1.1
- Most used standard in VO! (many tools: VOPlot, Aladin,Mirage, ...)
- Many librairies to read/write VOTable

Example of VOTable 1.1:

METADATA

DATA

```
<?xml version="1.0"?>
<VOTABLE version="1.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:noNamespaceSchemaLocation="http://www.ivoa.net/xml/VOTable/VOTable/v1.1">
  <COOSYS ID="J2000" equinox="J2000." epoch="J2000." system="eq_FK5"/>
  <RESOURCE name="myFavouriteGalaxies">
    <TABLE name="results">
      <DESCRIPTION>Velocities and Distance estimations</DESCRIPTION>
      <PARAM name="Telescope" datatype="float" ucd="phys.size;instr.tel"
             unit="m" value="3.6"/>
      <FIELD name="RA"   ID="col1" ucd="pos.eq.ra;meta.main" ref="J2000"
             datatype="float" width="6" precision="2" unit="deg" />
      <FIELD name="Dec"  ID="col2" ucd="pos.eq.dec;meta.main" ref="J2000"
             datatype="float" width="6" precision="2" unit="deg" />
      <FIELD name="Name" ID="col3" ucd="meta.id;meta.main"
             datatype="char" arraysize="8*"/>
      <FIELD name="RVel"  ID="col4" ucd="src.veloc.hc" datatype="int"
             width="5" unit="km/s"/>
      <FIELD name="e_RVel" ID="col5" ucd="stat.error;src.veloc.hc"
             datatype="int" width="3" unit="km/s"/>
      <FIELD name="R" ID="col6" ucd="phys.distance" datatype="float"
             width="4" precision="1" unit="Mpc">
        <DESCRIPTION>Distance of Galaxy, assuming H=75km/s/Mpc</ DESCRIPTION>
      </FIELD>
      <DATA>
        <TABLEDATA>
        <TR>
          <TD>010.68</TD><TD>+41.27</TD><TD>N  224</TD><TD>297</TD><TD>5</TD><TD>0.7</TD>
        </TR>
        <TR>
          <TD>287.43</TD><TD>-63.85</TD><TD>N  6744</TD><TD>839</TD><TD>6</TD><TD>10.4</TD>
        </TR>
        <TR>
          <TD>023.48</TD><TD>+30.66</TD><TD>N  598</TD><TD>182</TD><TD>3</TD><TD>0.7</TD>
        </TR>
        </TABLEDATA>
      </DATA>
    </TABLE>
  </RESOURCE>
</VOTABLE>
```


CDS and the VO

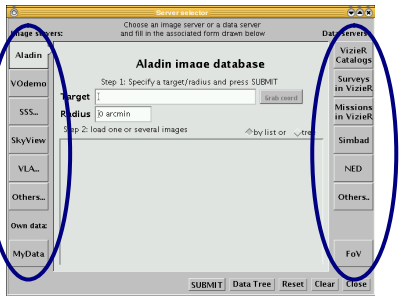
- CDS plays an active role in the VO projects (VO-France, AVO, VOTech, DCA, IVOA, ...)
- CDS is participating to the definition of the VO standards (VOTable, UCD, ...)
- The VO progresses impact on the CDS services

Aladin: a VO portal

- VO access for astronomers:
  - discover available data and services
  - access and query
  - manipulation and analysis
  - publish, disséminate results
- The AVO prototype was an extension of Aladin
- Aladin has integrated several new features developed in VO projects

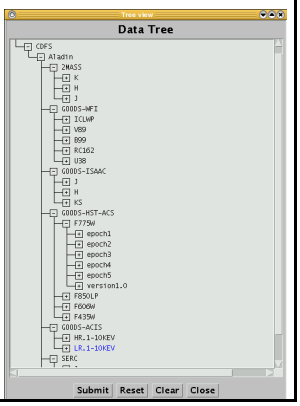
ex: AVO prototype

- Data access:



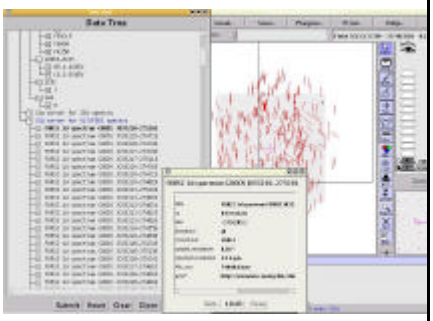
ex: AVO prototype

- Image access
  - DAL: SIA
  - the Aladin image server returns the metadata of the images
  - construction of the 'metadata tree'



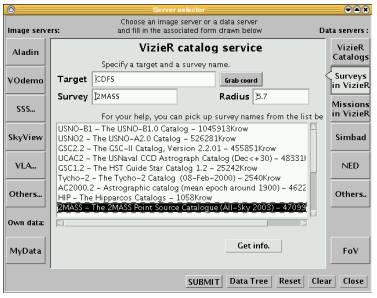
ex: AVO prototype

- Spectrum access
  - DAL: SSA
  - ex: metadata of the ESO server



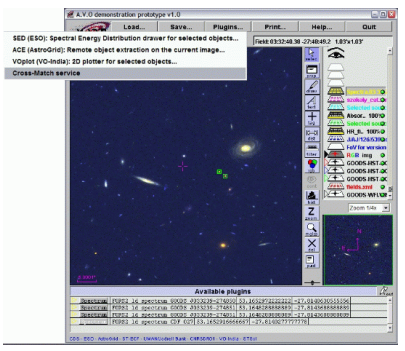
ex: AVO prototype

- Catalogue access
  - Cone Search
  - ex: VizieR



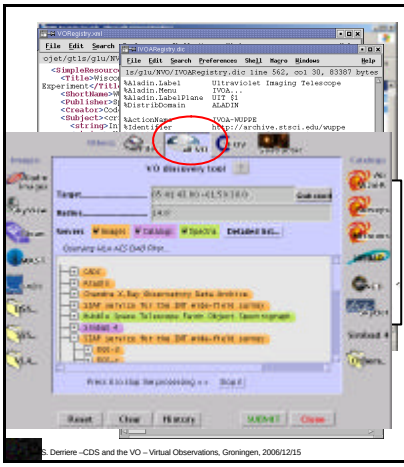
ex: AVO prototype

- Access to services
  - SExtractor
  - Cross-match
- Protocols HTTP, WS...
- FITS, VOTable



Recent VO developments

- PLASTIC (VOTech)
  - PLAtform for Astronomical Tools Inter Connection
  - Various applications exchange messages through a hub
  - Aladin, TOPCAT, VisIVO, SPLAT, ...
- Instrumental FoVs (VOTech)
- All-VO action in Aladin
  - built from the VO Registry



1.Harvesting from STScI

2. Conversion and filtering into a GLU dic

3. GLU site synchronization

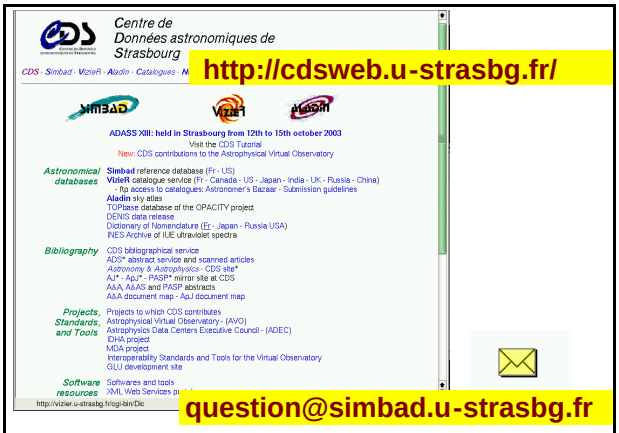
4.New Aladin form managing these VO GLU records...

5…to produce general metadata trees

6… to choose and manipulate images and catalogs..

Other VO-related evolutions

- VizieR access:
  - ConeSearch
  - SkyNode
- VOTable and UCDs used everywhere
  - VizieR, SIMBAD, Aladin (filters...)
- Catalogue cross-match in Aladin
- Many things to get your hands on during this afternoon's session !



Links (1)

- CDS web site:
  - http://cdsweb.u-strasbg.fr
- SIMBAD:
  - http://simbad.u-strasbg.fr/
- VizieR
  - http://vizier.u-strasbg.fr/
  - http://vizier.u-strasbg.fr/viz-bin/VizieR
- Aladin
  - http://aladin.u-strasbg.fr/

Links (2)

- CDS Tutorials
  - http://cdsweb.u-strasbg.fr/Tutorial/
  - http://aladin.u-strasbg.fr/java/aladin.pdf
- IVOA
  - http://www.ivoa.net/
- Action Spécifique OV-France
  - http://www.france-ov.org/