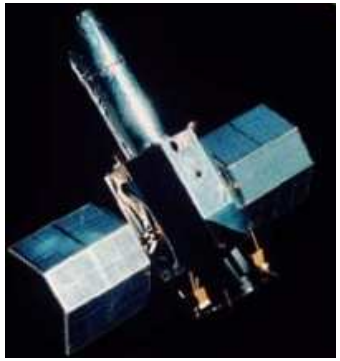
Heterogeneous data fusion and data mining in astronomy

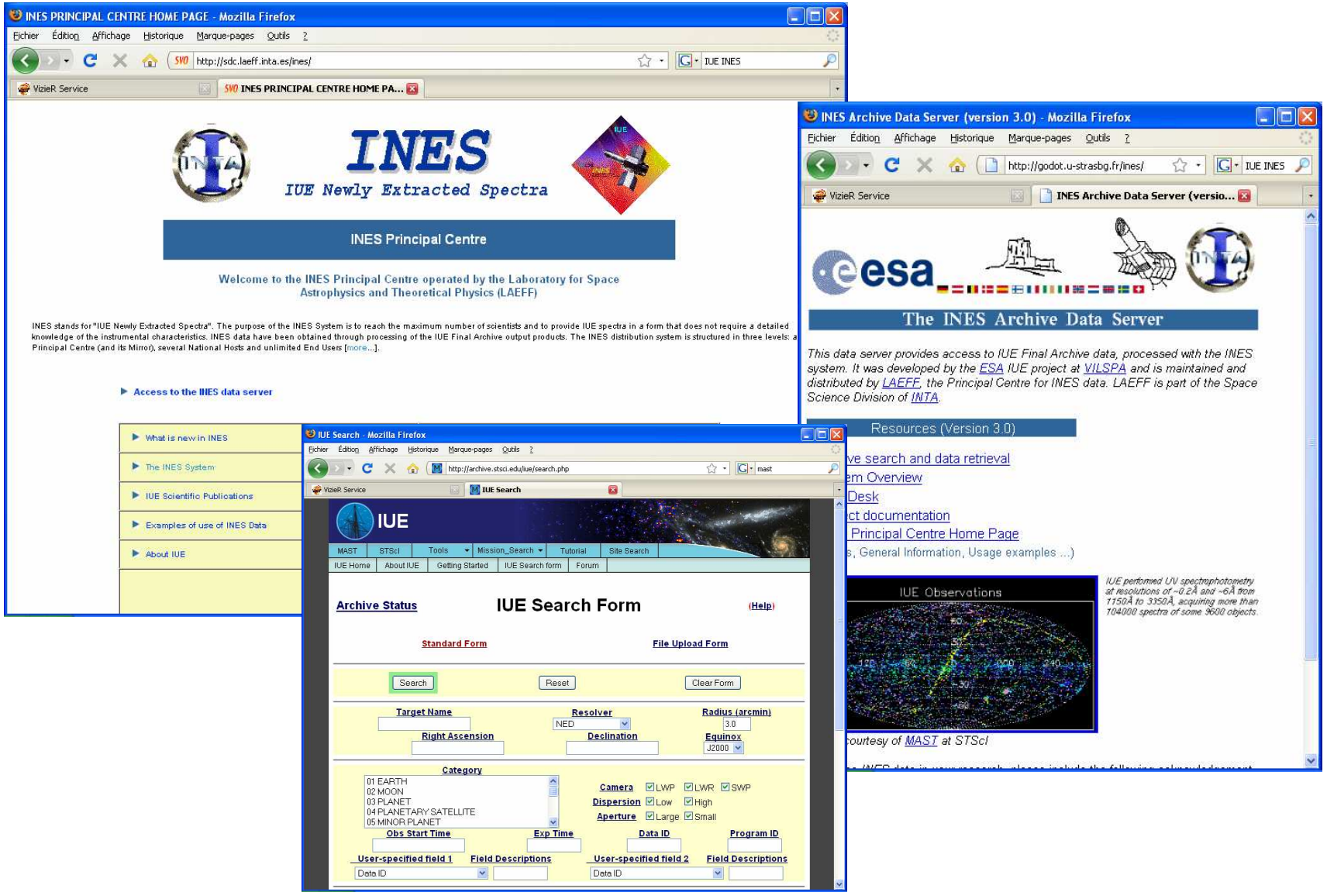
Sharing astronomical data

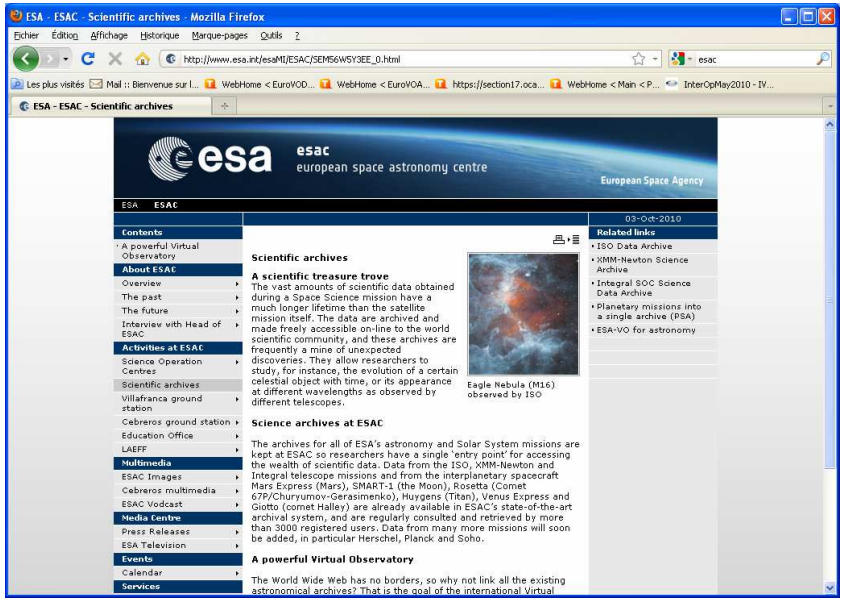
- The astronomy model: competitive AOs to get observation time on facilities
- Re-using data for scientific objectives different from the original ones, i.e. optimize the science return of large ground- and space-based instruments and of large surveys



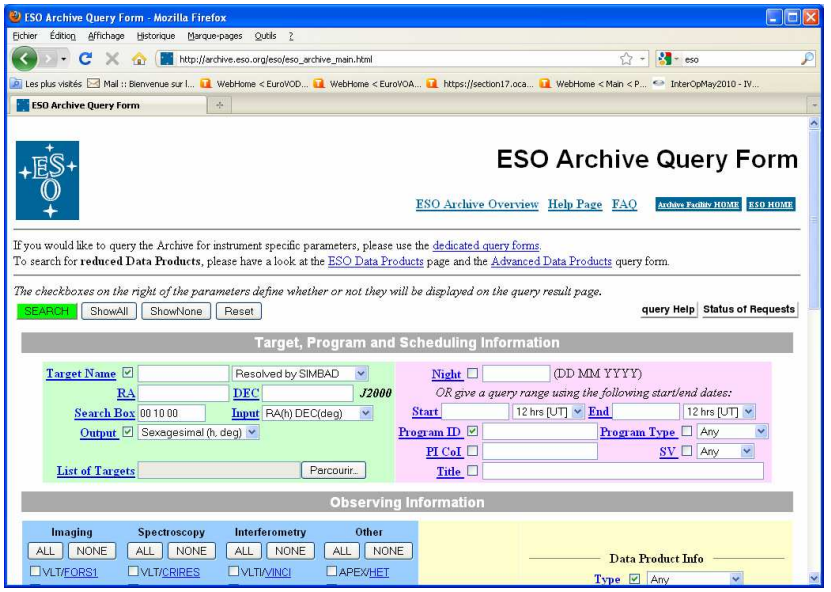
IUE (1978-1996): five times more publications from data retrieved in the archive than from the selected observing teams (Wamsteker, Griffin, 1995) – a major precursor

Access to IUE data





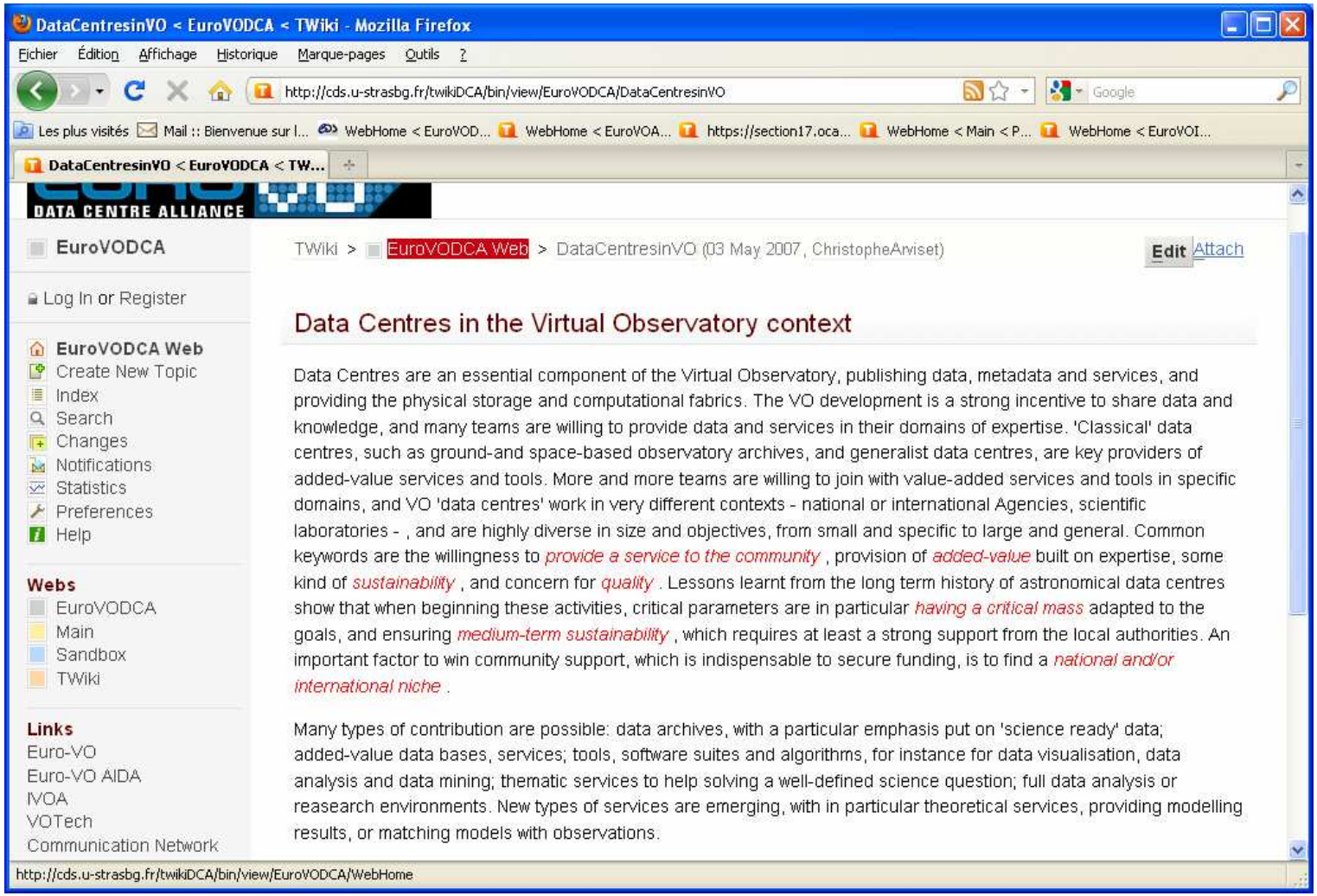
European Agency
data archives



And MANY others

DCA/AIDA Census of European data Centres

- Inclusive definition
- Keywords: provide a service to the community, added-value, sustainability, quality, national and/or international role
- Characterizes the population of European data centres (70 answers from 16 different countries, > 230 forms filled describing a resource)
- Lively snapshot of a very diverse landscape
  - Covers all disciplines of astronomy
  - Large variety of approaches and sizes, from large data centres maintained by European or National Agencies to small teams maintaining a specific service


DataCentresinVO < EuroVODCA < TWiki - Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils ?
http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/DataCentresinVO
DATA CENTRE ALLIANCE
EuroVODCA
Log In or Register
EuroVODCA Web
Create New Topic
Index
Search
Changes
Notifications
Statistics
Preferences
Help
Webs
EuroVODCA
Main
Sandbox
TWiki
Links
Euro-VO
Euro-VO AIDA
IVOA
VOTech
Communication Network
TWiki > EuroVODCA Web > DataCentresinVO (03 May 2007, ChristopheArviset)
Edit Attach
Data Centres in the Virtual Observatory context
Data Centres are an essential component of the Virtual Observatory, publishing data, metadata and services, and providing the physical storage and computational fabrics. The VO development is a strong incentive to share data and knowledge, and many teams are willing to provide data and services in their domains of expertise. 'Classical' data centres, such as ground-and space-based observatory archives, and generalist data centres, are key providers of added-value services and tools. More and more teams are willing to join with value-added services and tools in specific domains, and VO 'data centres' work in very different contexts - national or international Agencies, scientific laboratories - , and are highly diverse in size and objectives, from small and specific to large and general. Common keywords are the willingness to provide a service to the community , provision of added-value built on expertise, some kind of sustainability , and concern for quality . Lessons learnt from the long term history of astronomical data centres show that when beginning these activities, critical parameters are in particular having a critical mass adapted to the goals, and ensuring medium-term sustainability , which requires at least a strong support from the local authorities. An important factor to win community support, which is indispensable to secure funding, is to find a national and/or international niche .
Many types of contribution are possible: data archives, with a particular emphasis put on 'science ready' data; added-value data bases, services; tools, software suites and algorithms, for instance for data visualisation, data analysis and data mining; thematic services to help solving a well-defined science question; full data analysis or reasearch environments. New types of services are emerging, with in particular theoretical services, providing modelling results, or matching models with observations.
http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/WebHome

A multipolar world

- Many facilities and authorities
- Large diversity
- Common VObs framework

Registry of Resources

- Key element: the ‘yellow pages’
- Compliant with OAI-PMH (allows interoperability with digital libraries)
- Dublin core + disciplinary extensions
- Not a unique registry, but
  - Several harvestable registries (+ publishing registries)
  - + A Registry of Registry

Euro-VO Registry : Main Page

Demo

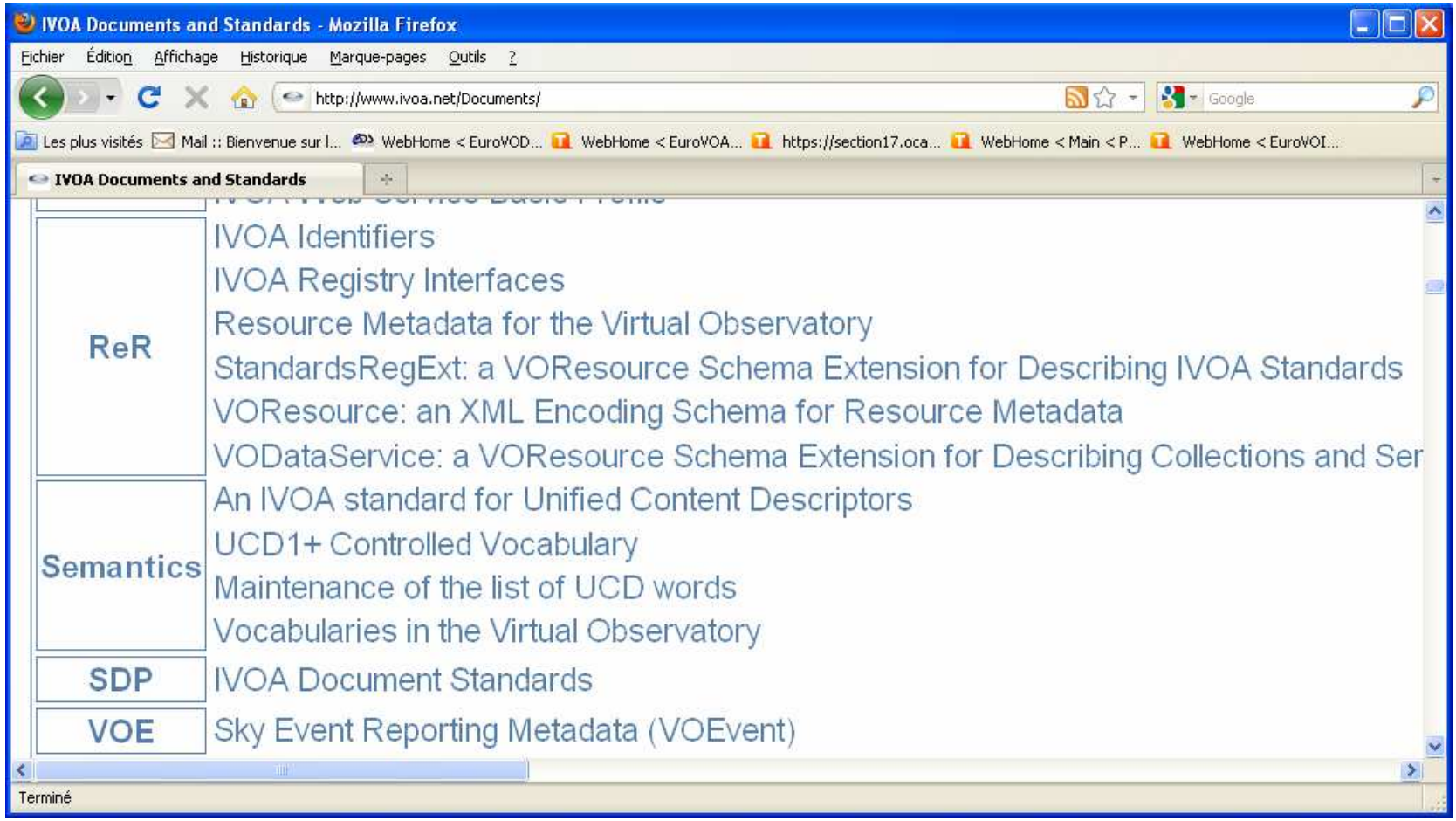
http://registry.euro-vo.org/

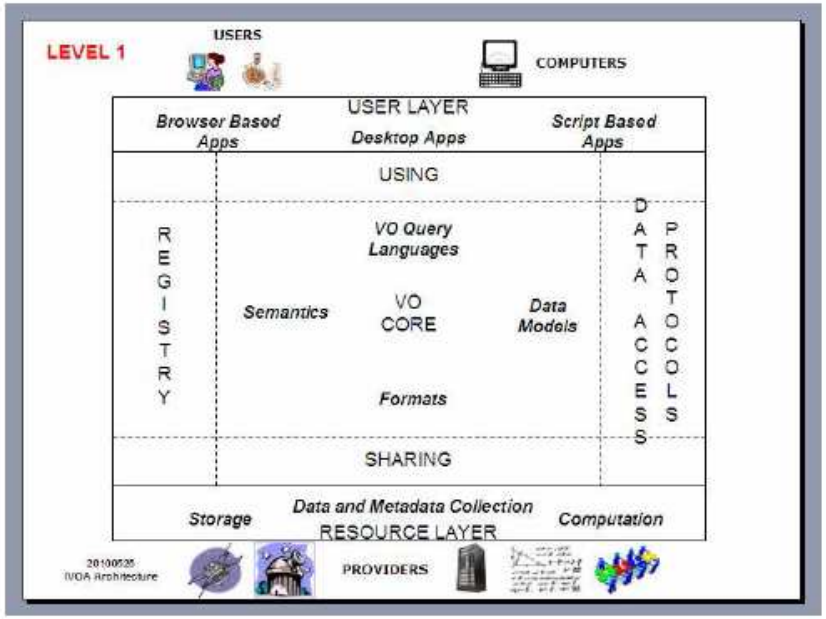
Current issues on registry

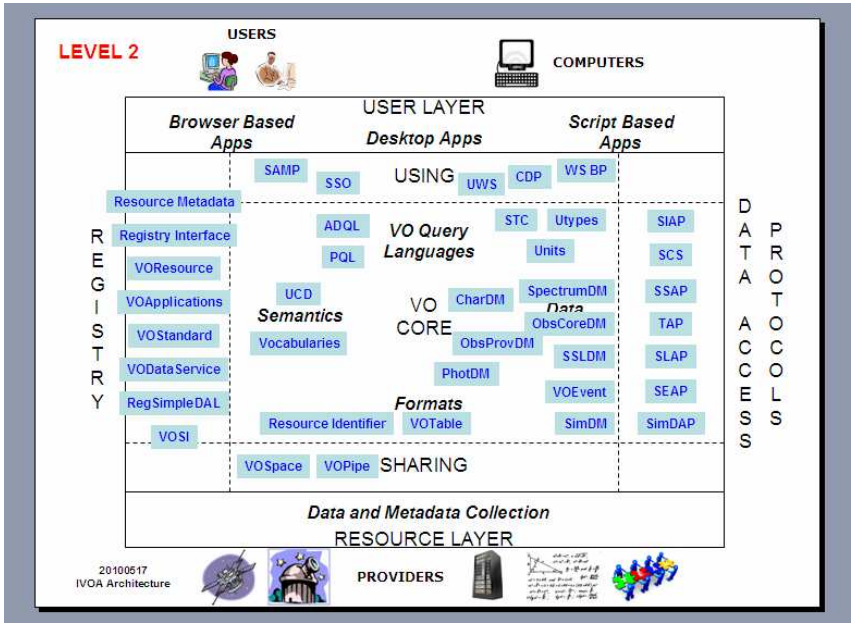
- Validation of entries (rules, tools)

  which implies to validate the validators!!
- International effort

- Add extensions to describes the different types of services


IVOA Documents and Standards - Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils ?
http://www.ivoa.net/Documents/
Google
Les plus visités
Mail :: Bienvenue sur l...
WebHome < EuroVOD...
WebHome < EuroVOA...
https://section17.oca...
WebHome < Main < P...
WebHome < EuroVOI...
IVOA Documents and Standards
ReR
IVOA Identifiers
IVOA Registry Interfaces
Resource Metadata for the Virtual Observatory
StandardsRegExt: a VOResource Schema Extension for Describing IVOA Standards
VOResource: an XML Encoding Schema for Resource Metadata
VODataService: a VOResource Schema Extension for Describing Collections and Ser
Semantics
An IVOA standard for Unified Content Descriptors
UCD1+ Controlled Vocabulary
Maintenance of the list of UCD words
Vocabularies in the Virtual Observatory
SDP
IVOA Document Standards
VOE
Sky Event Reporting Metadata (VOEvent)
Terminé


LEVEL 1
USERS
COMPUTERS
USER LAYER
Browser Based Apps
Desktop Apps
Script Based Apps
USING
REGISTRY
VO Query Languages
Semantics
VO CORE
Data Models
Formats
DATA ACCESS
PROTOCOLS
SHARING
Storage
Data and Metadata Collection
Computation
RESOURCE LAYER
20100526
IVOA Architecture
PROVIDERS


LEVEL 2
USERS
COMPUTERS
USER LAYER
Browser Based Apps
Desktop Apps
Script Based Apps
SAMP
SSO
USING
UWS
CDP
WS BP
Resource Metadata
Registry Interface
VOResource
VOApplications
VOStandard
VODataService
RegSimpleDAL
VOSI
REGISTRY
ADQL
PQL
VO Query Languages
STC
Utypes
Units
UCD
Semantics
VO CORE
CharDM
SpectrumDM
ObsCoreDM
ObsProvDM
SSLDM
PhotDM
VOEvent
Vocabularies
Formats
Resource Identifier
VOTable
SimDM
SIAP
SCS
SSAP
TAP
SLAP
SEAP
SimDAP
DATA ACCESS
PROTOCOLS
VOSpace
VOPipe
SHARING
Data and Metadata Collection
RESOURCE LAYER
20100517
IVOA Architecture
PROVIDERS

Identity metadata	
Title	Sloan Digital Sky Survey
ShortName	SDSS
Identifier	ivo://stsci.edu/mast/sdss
Curation metadata	
Publisher	Space Telescope Science Institute/MAST
PublisherID	ivo://stsci.edu/mast
Creator	Sloan Digital Sky Survey Consortium
Creator.Logo	http://archive.stsci.edu/images/sdss_logo.gif
Contributor	Sloan Digital Sky Survey Consortium
Date	2003-02-01
Version	SDSS EDR
Contact.Name	Archive Branch, Space Telescope Science Institute
Contact.Address	3700 San Martin Drive, Baltimore, MD 21218 USA
Contact.Email	archive@stsci.edu
Contact.Telephone	+1-410-338-4547
General content metadata	
Subject	galaxies, quasars, stars, CCD photometry, spectroscopy, redshift, sky surveys
Description	The Sloan Digital Sky Survey is using a dedicated 2.5 m telescope and a large format CCD camera to obtain images of over 10,000 square degrees of high Galactic latitude sky in five broad bands (u', g', r', i' and z', centered at 3540, 4770, 6230, 7630, and 9130 Å, respectively). Medium resolution spec-

Source	2002AJ….123..485S
ReferenceURL	http://archive.stsci.edu/sdss/index.html
Type	Survey, Catalog, EPOResource
ContentLevel	Research
Relationship	mirror-of
RelationshipID	ivo://sdss.org/sdss/edr

Collection and service content metadata

Facility	Apache Point Observatory, Sloan 2.5-m Telescope
Instrument	Five-band clocked CCD camera
Coverage.Spatial	PositionInterval FK5 145.17 –1.25 235.9 1.25 PositionInterval FK5 250.71 52.15 267.0 66.29 PositionInterval FK5 350.43 –1.25 359.99 1.17 PositionInterval 0.0 –1.25 56.37 1.17
Coverage.RegionOfRegard	0.0001
Coverage.Spectral	Optical
Coverage.Spectral.Bandpass	u', g', r', i', z'
Coverage.Spectral.MinimumWavelength	400.e-9
Coverage.Spectral.MaximumWavelength	850.e-9
Coverage.Temporal.StartTime	1999-12-25
Coverage.Temporal.StopTime	2001-07-15
Coverage.Depth	3.e.6

Data quality metadata

DataQuality	A	
ResourceValidationLevel	4	[provided by registry curator]
ResourceValidatedBy	ivo:/us-vo.org/registry	
Uncertainty.Photometric	3.e-7	
Uncertainty.Spatial	0.00003	
Uncertainty.Spectral	1.e-11	
Uncertainty.Temporal	0.1	

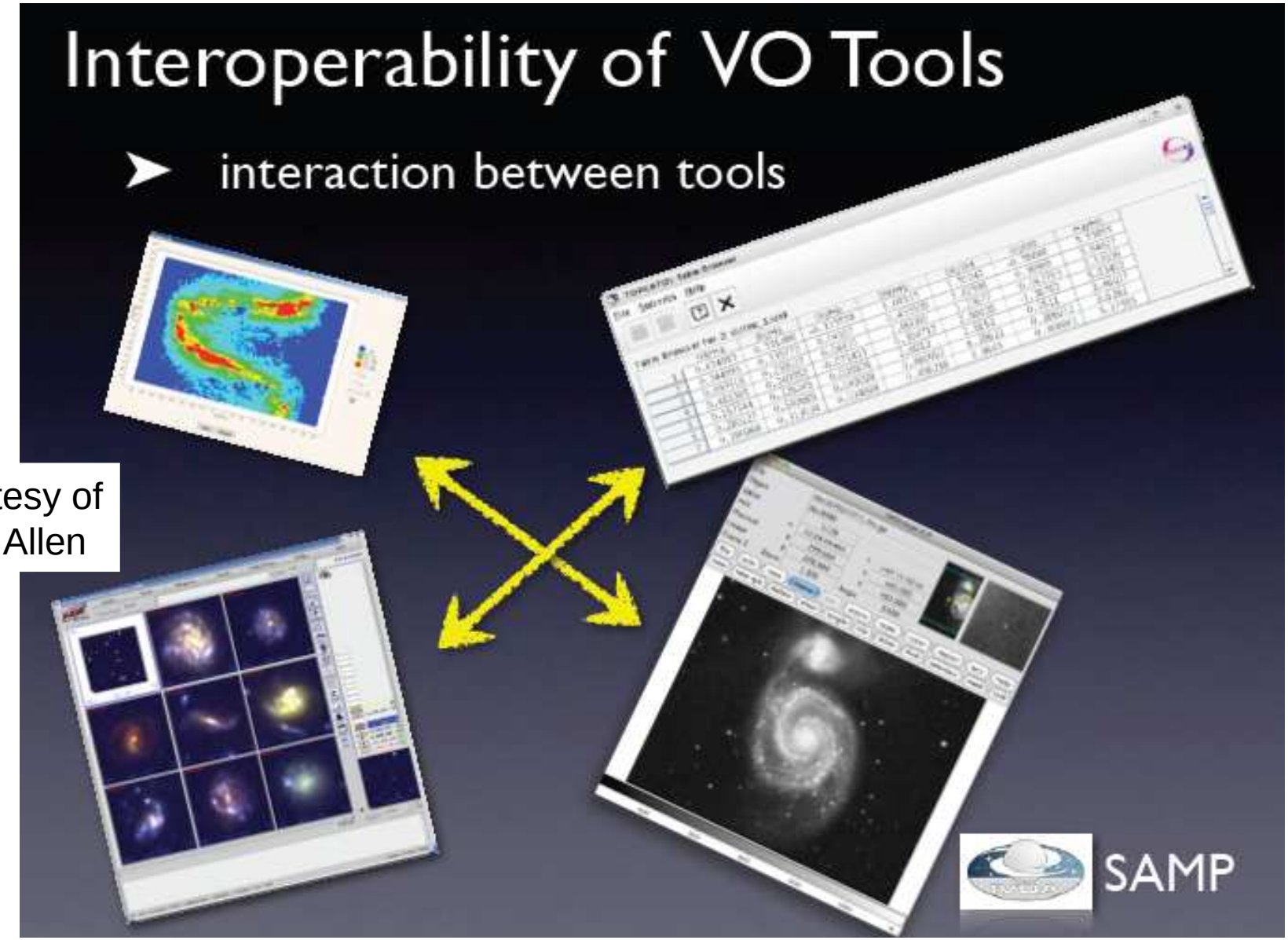
Service metadata

Service.AccessURL	http://archive.stsci.edu/cgi-bin/sdss/catalog
Service.InterfaceURL	http://archive.stsci.edu/sdss/catalog.html
Service.BaseURL	http://archive.stsci.edu/cgi-bin/sdss/catalog
Service.HTTPResultsMIMEType	text/xml
Service.StandardID	ivo://ivoa.net/Services/ConeSearch
Service.MaxSearchRadius	0.2
Service.MaxReturnRecords	5000
Service.MaxReturnSize	5.e8

SAMP

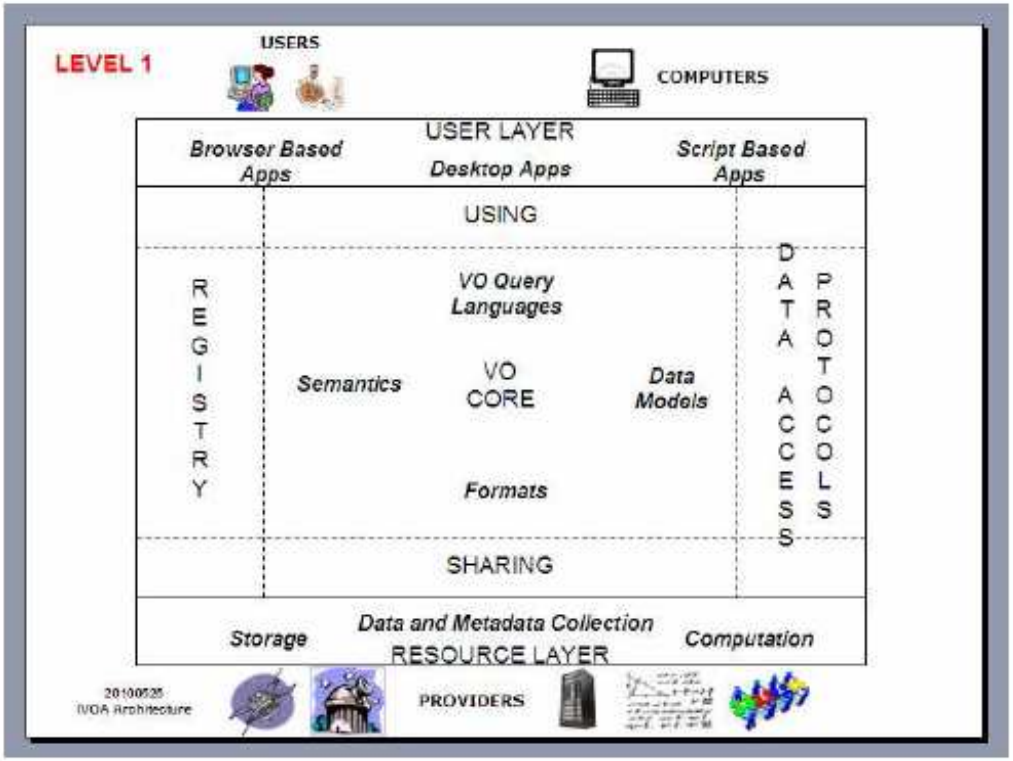
- The Simple Application Messaging Protocol
- Enables software tools to interoperate and communicate

  IVOA members have recognised that building a monolithic tool that attempts to fulfil all the requirements of all users is impractical, and it is a better use of our limited resources to enable individual tools to work together better. One element of this is defining common file formats for the exchange of data between different applications. Another important component is a messaging system that enables the applications to share data and take advantage of each other's functionality.



Courtesy of
M.G. Allen

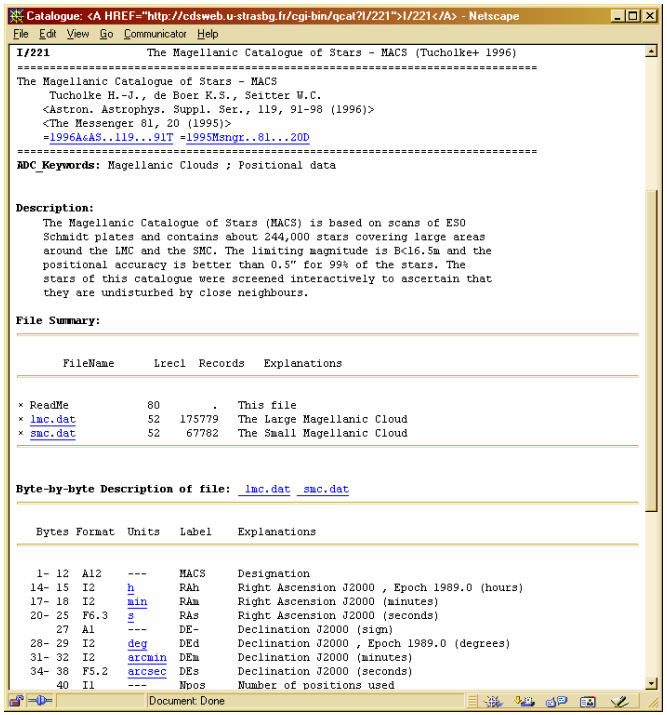
Semantics


LEVEL 1
USERS
COMPUTERS
USER LAYER
Browser Based Apps
Desktop Apps
Script Based Apps
USING
REGISTRY
VO Query Languages
Semantics
VO CORE
Data Models
Formats
DATA ACCESS
PROTOCOLS
SHARING
Storage
Data and Metadata Collection
RESOURCE LAYER
Computation
20100525
IVOA Architecture
PROVIDERS

Catalogues and published tables

CDS VizieR service

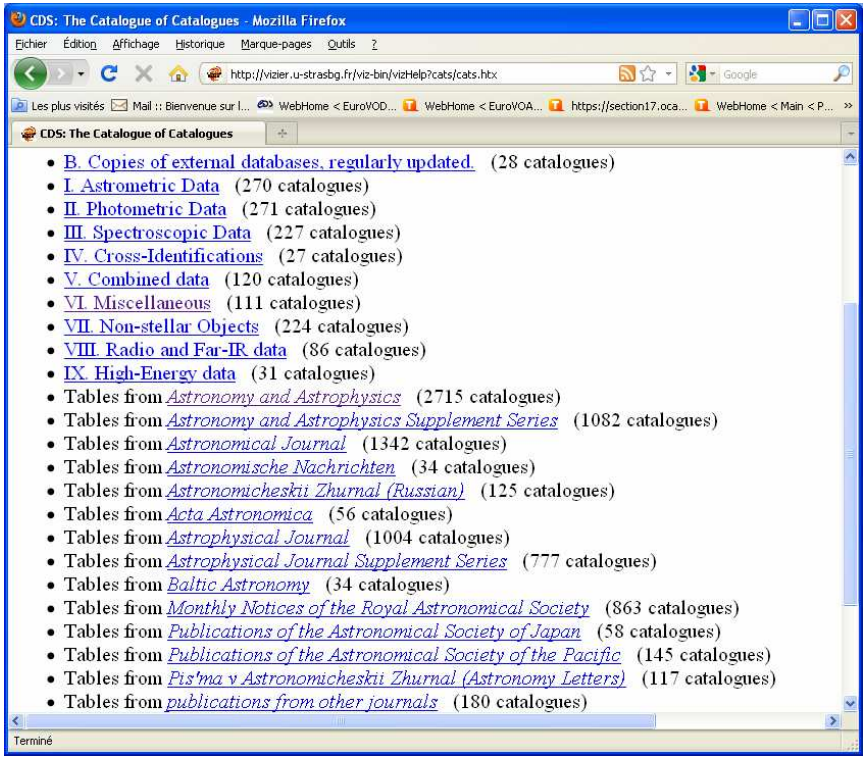
A single standard



Catalogues

A single view of the tables
published in journals;
collab. journals + data centres

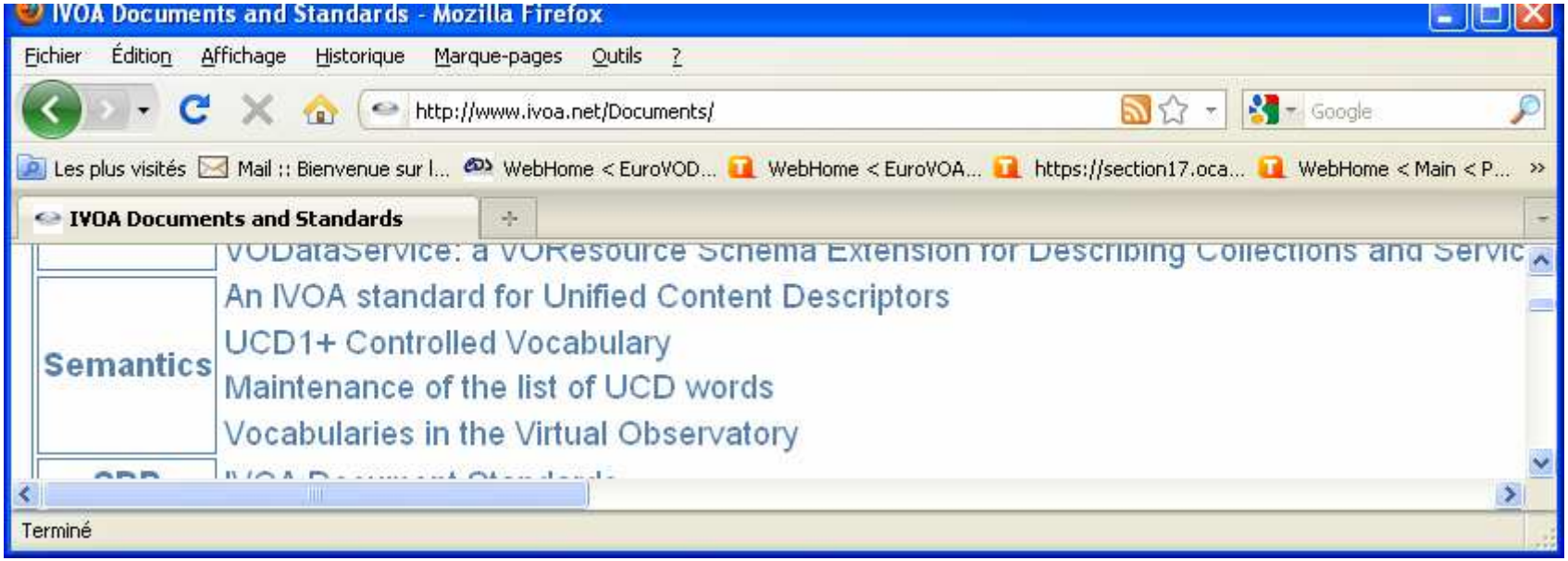
Lists of observations



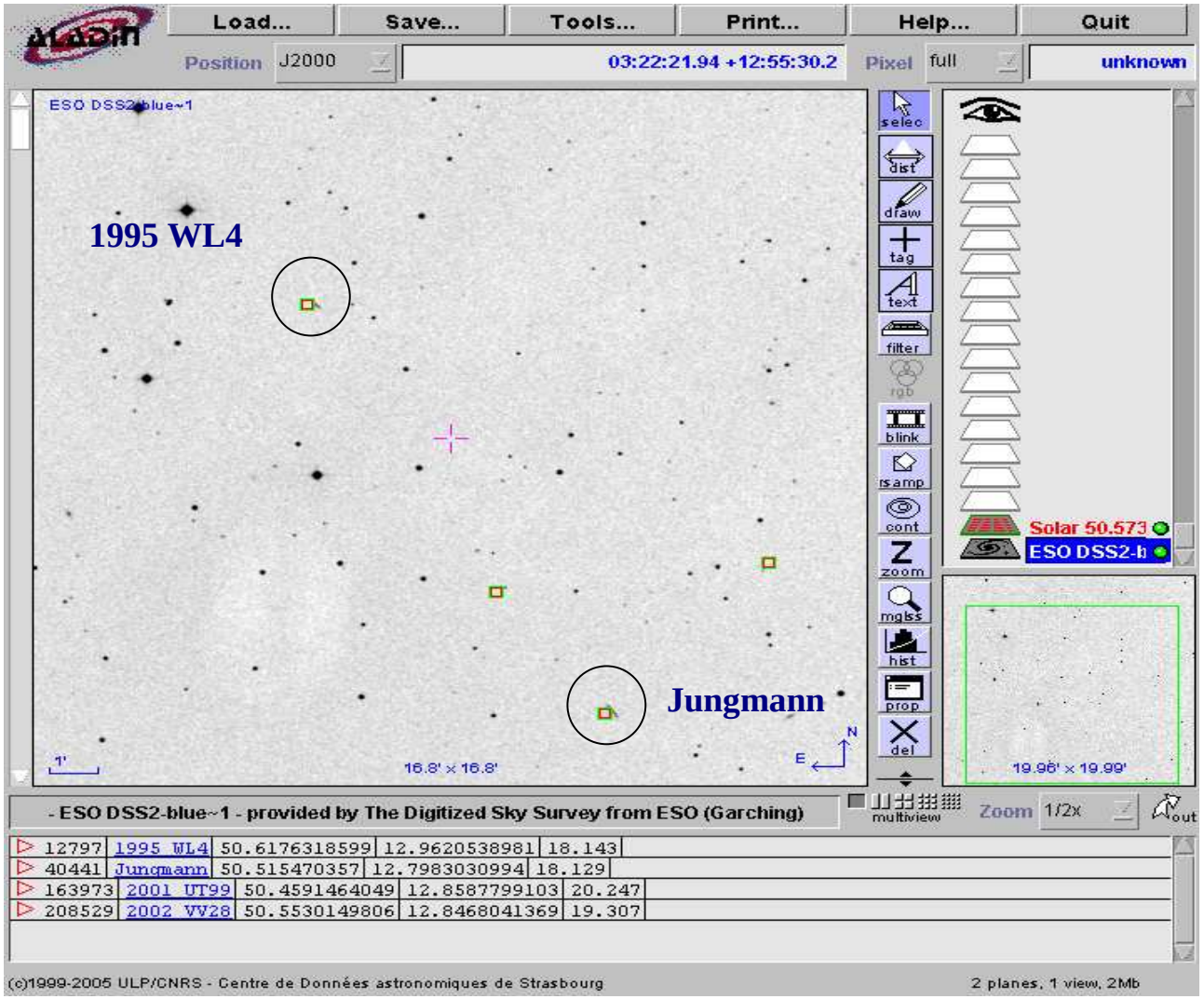
The UCDs

- Describes quantities in astronomy
- Starting point: the CDS catalogue service VizieR
- 100.000 columns in VizieR

> IVOA standard Unified Content Descriptor



- Maintenance of UCDs: a procedure to add new ones
- Standard for vocabularies: W3C (SKOS, RDF)
- Notes on ontologies (ontology of object types, use case)



SkyBot (IMCCE, Paris Observatory)
+
Aladin (CDS, Strasbourg Observatory)
+
VO standard (VOTable)

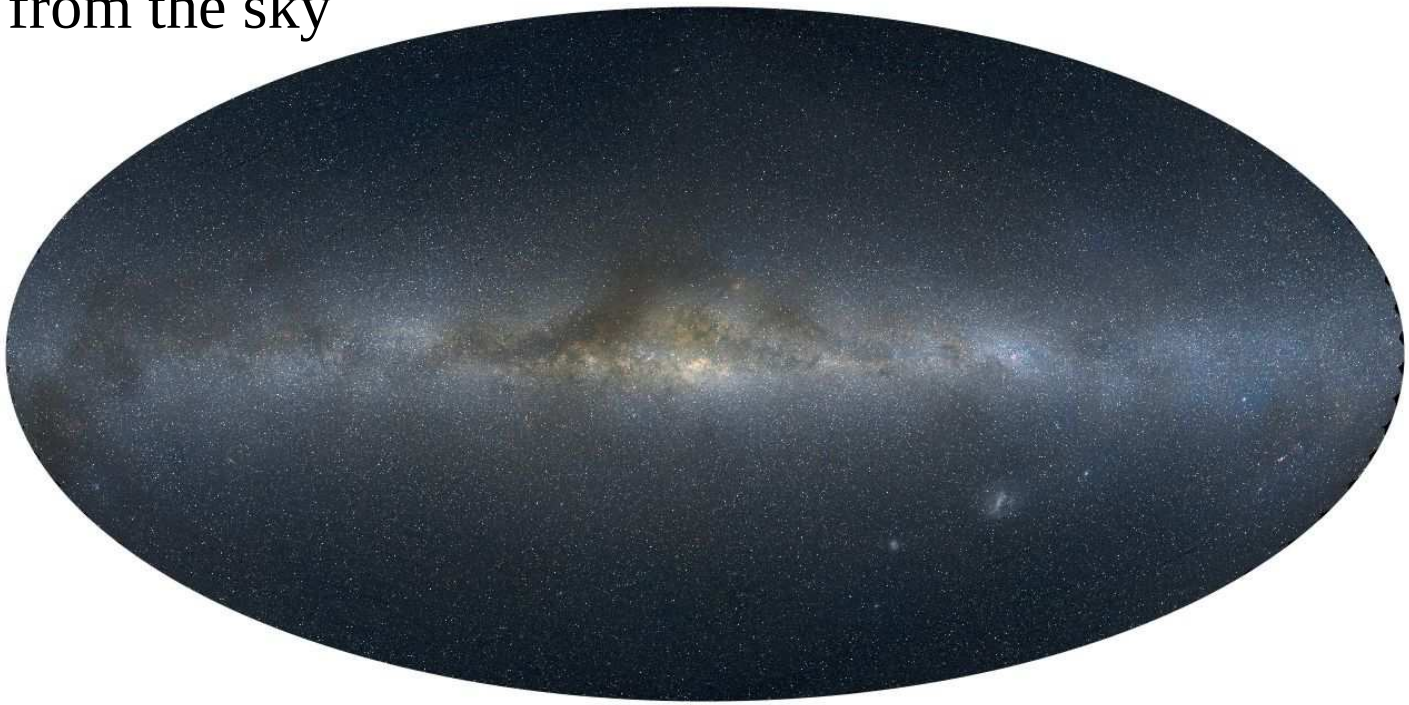
Fast solar system moving targets

- Also, e.g., VOEvent – LSST! (Ray's talk)

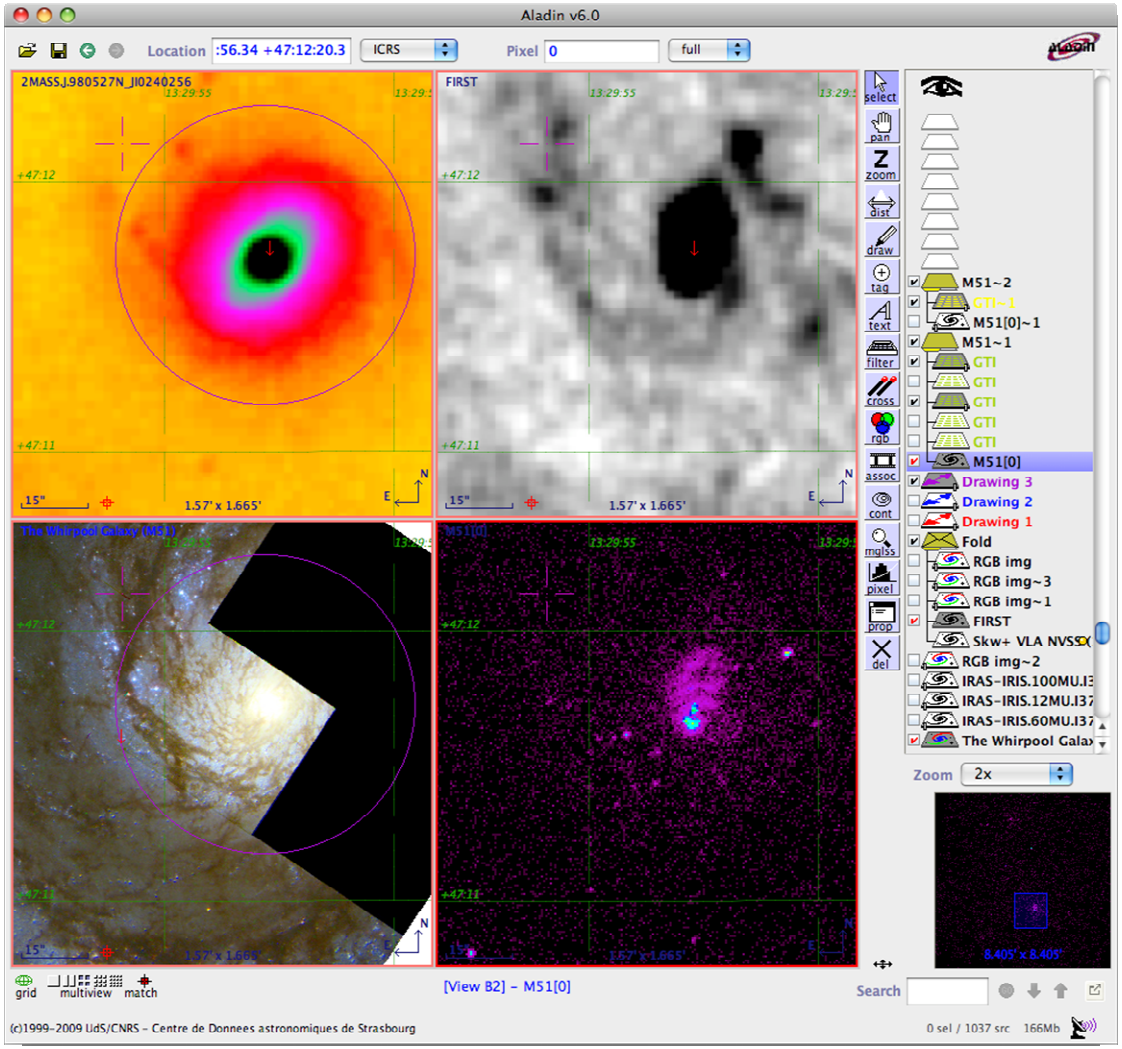
Long term data accumulation and re-use

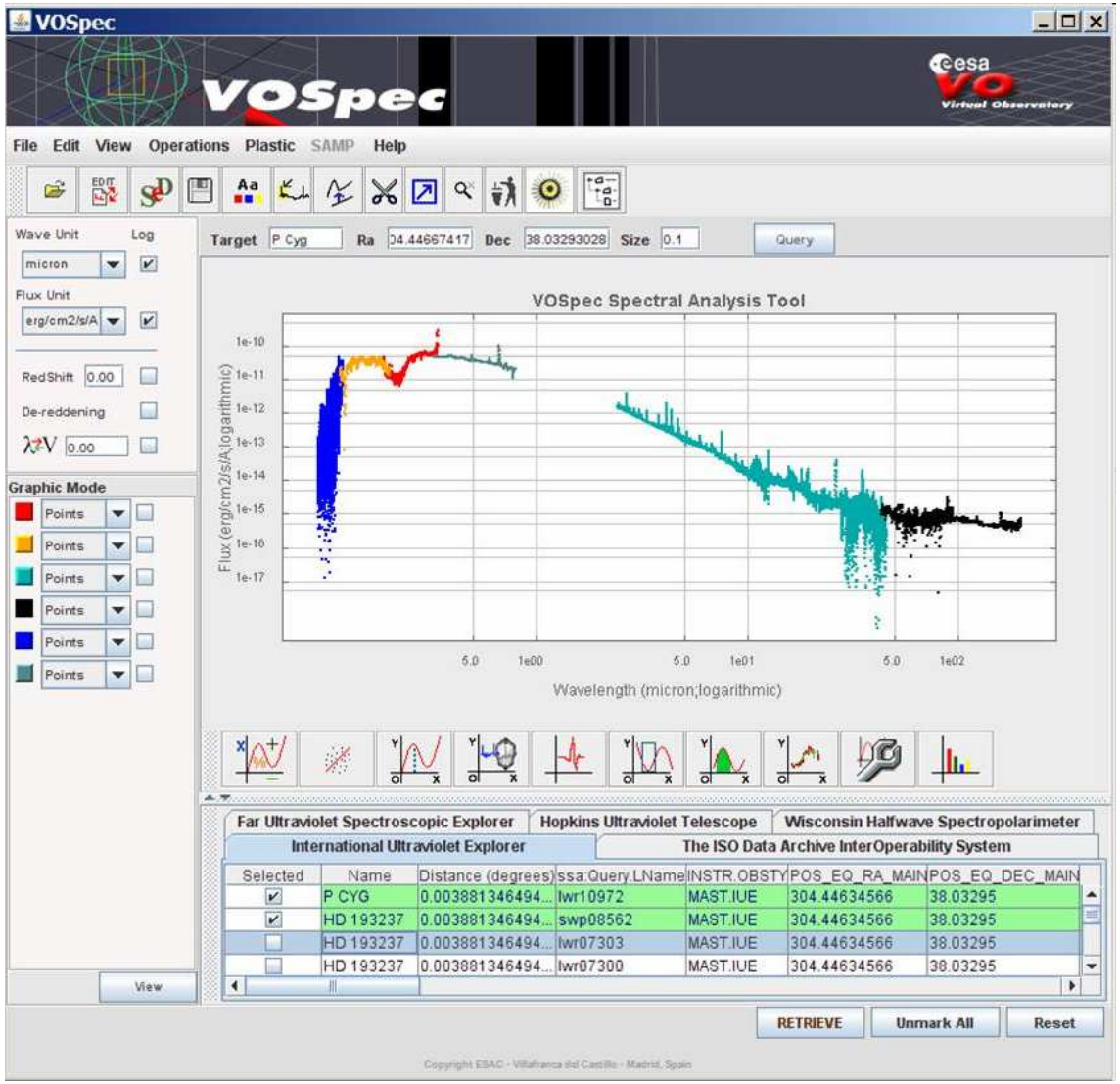
Mellinger colored

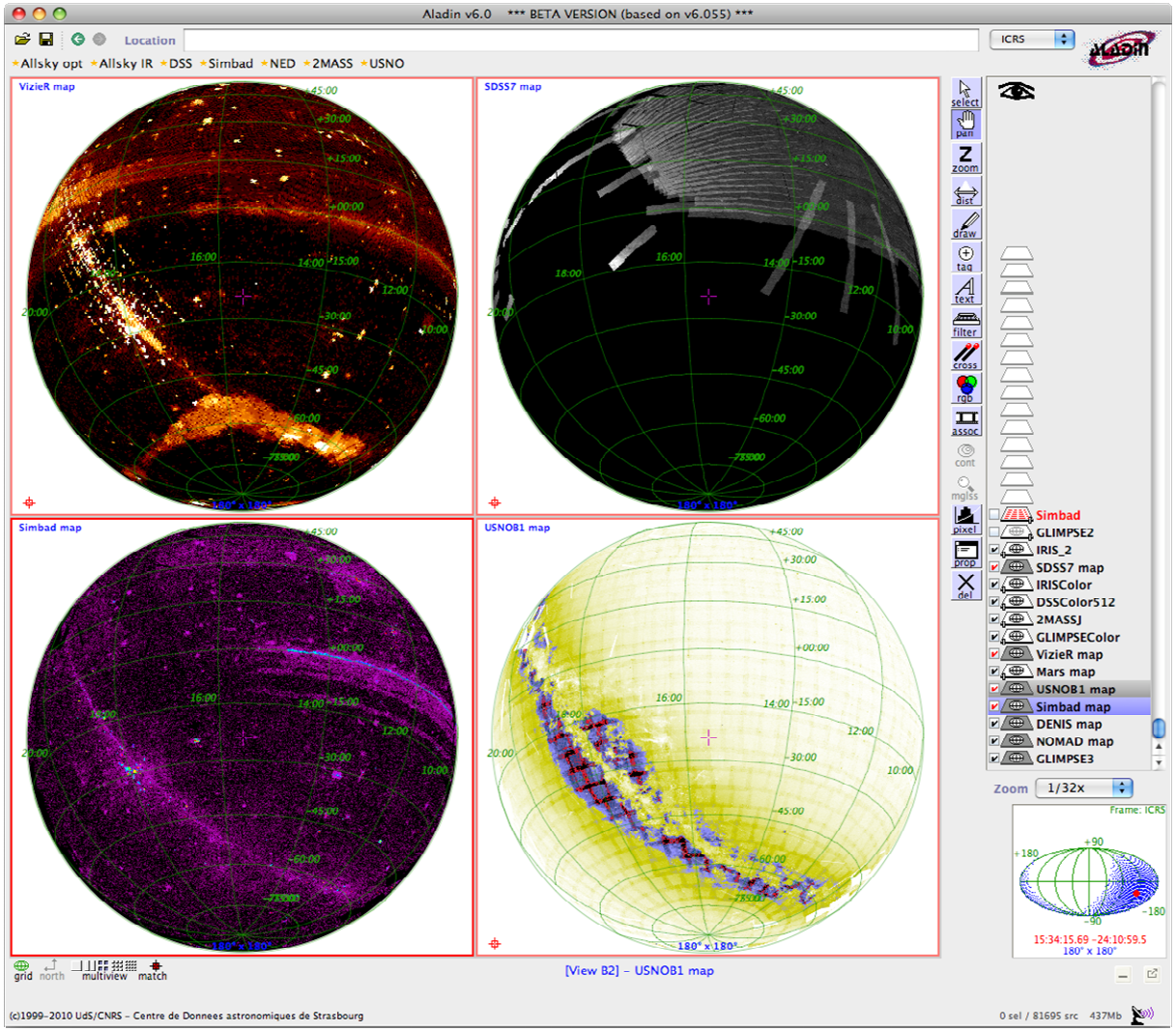
A view from the sky



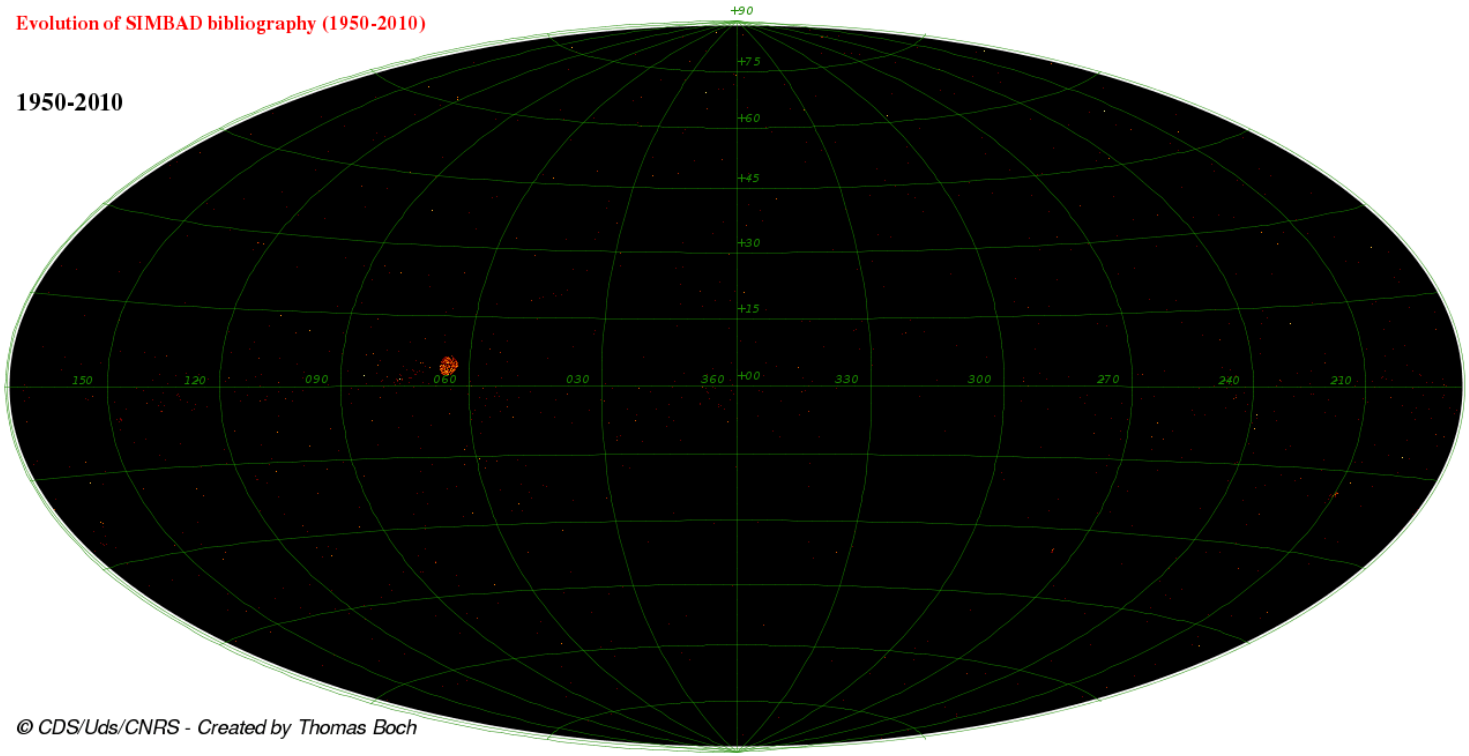
360° x 180°

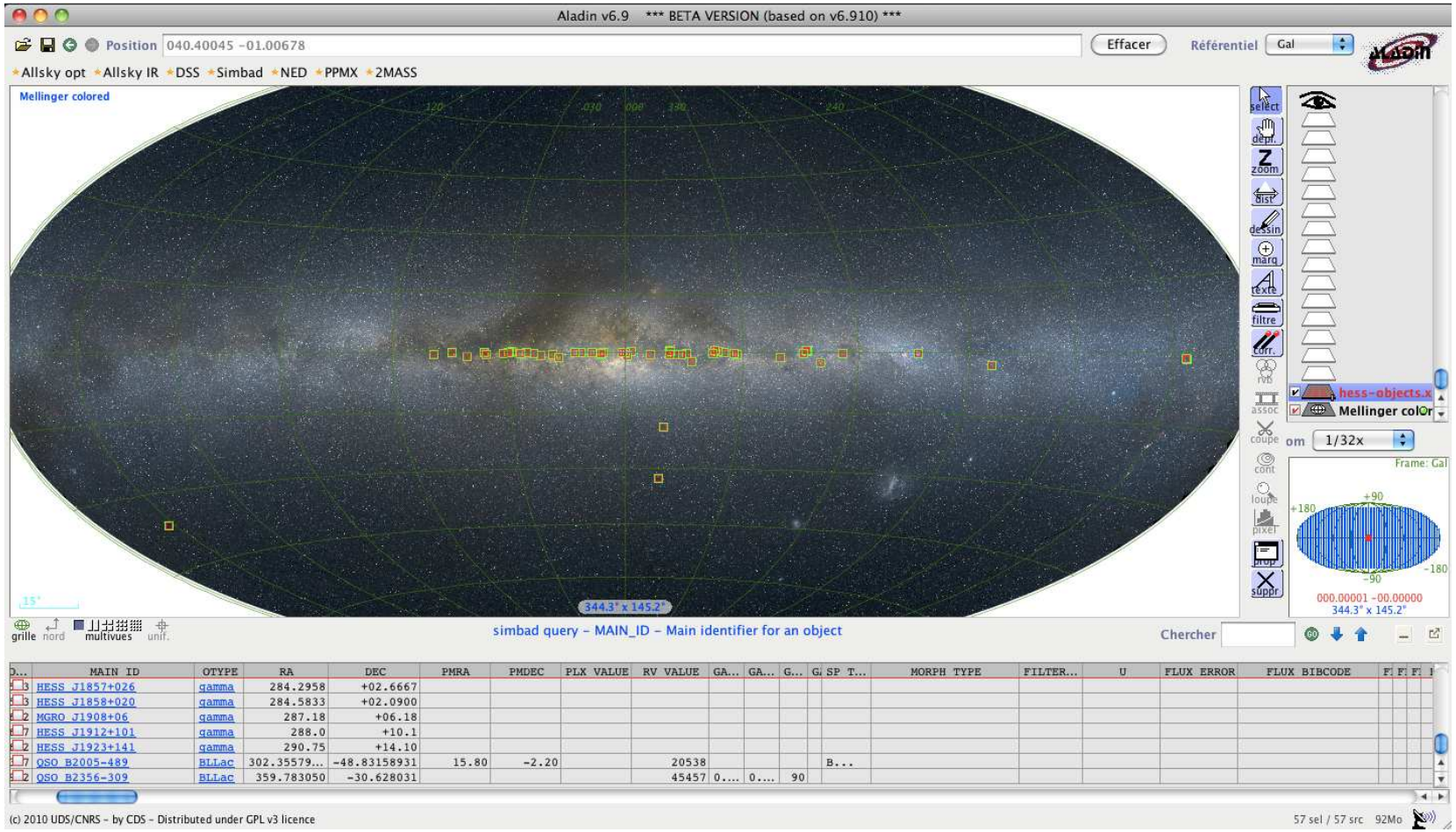

Aladin v6.0
Location :56.34 +47:12:20.3
ICRS
Pixel 0
full
2MASSJ.980527N_JI0240256
FIRST
M51~2
GTI~1
M51[0]~1
M51~1
GTI
M51[0]
Drawing 3
Drawing 2
Drawing 1
Fold
RGB img
RGB img~3
RGB img~1
FIRST
RGB img~2
The Whirpool Gala
Zoom 2x
[View B2] - M51[0]
Search
(c)1999-2009 UdS/CNRS - Centre de Donnees astronomiques de Strasbourg


VOSpec
VOSpec
esa
VO
Virtual Observatory
File Edit View Operations Plastic SAMP Help
Wave Unit
micron
Log
Flux Unit
erg/cm2/s/A
RedShift 0.00
De-reddening
0.00
Graphic Mode
Points
Points
Points
Points
Points
Points
View
Target P Cyg
Ra 04.44667417
Dec 38.03293028
Size 0.1
Query
VOSpec Spectral Analysis Tool
Flux (erg/cm2/s/A;logarithmic)
1e-10
1e-11
1e-12
1e-13
1e-14
1e-15
1e-16
1e-17
5.0
1e00
5.0
1e01
5.0
1e02
Wavelength (micron;logarithmic)
Far Ultraviolet Spectroscopic Explorer
Hopkins Ultraviolet Telescope
Wisconsin Halfwave Spectropolarimeter
International Ultraviolet Explorer
The ISO Data Archive InterOperability System
Selected
Name
Distance (degrees)
ssa:Query.LName
INSTR.OBSTY
POS_EQ_RA_MAIN
POS_EQ_DEC_MAIN
P CYG
0.003881346494...
lwr10972
MAST.IUE
304.44634566
38.03295
HD 193237
0.003881346494...
swp08562
MAST.IUE
304.44634566
38.03295
HD 193237
0.003881346494...
lwr07303
MAST.IUE
304.44634566
38.03295
HD 193237
0.003881346494...
lwr07300
MAST.IUE
304.44634566
38.03295
RETRIEVE
Unmark All
Reset


Aladin v6.0 *** BETA VERSION (based on v6.055) ***
Location
ICRS
Allsky opt
Allsky IR
DSS
Simbad
NED
2MASS
USNO
VizieR map
SDSS7 map
Simbad map
USNOB1 map
select
pan
zoom
dist
draw
tag
text
filter
cross
rgb
assoc
cont
mass
pixel
prop
del
Simbad
GLIMPSE2
IRIS_2
SDSS7 map
IRISColor
DSSColor512
2MASSJ
GLIMPSEColor
VizieR map
Mars map
USNOB1 map
Simbad map
DENIS map
NOMAD map
GLIMPSE3
Zoom
1/32x
Frame: ICRS
15:34:15.69 -24:10:59.5
180° x 180°
grid
north
multiview
match
[View B2] - USNOB1 map
(c)1999-2010 UdS/CNRS - Centre de Donnees astronomiques de Strasbourg
0 sel / 81695 src
437Mb

An example of construction on the long term
Constructing knowlegdge from bibliography (i.e., research results)
SIMBAD view




Aladin v6.9 *** BETA VERSION (based on v6.910) ***
Position 040.40045 -01.00678
Effacer
Référentiel Gal
Allsky opt Allsky IR DSS Simbad NED PPMX 2MASS
Mellinger colored
hess-objects.x
Mellinger color
1/32x
Frame: Gal
344.3° x 145.2°
simbad query - MAIN_ID - Main identifier for an object
Chercher
MAIN ID OTYPE RA DEC PMRA PMDEC PLX VALUE RV VALUE MORPH TYPE FILTER... U FLUX ERROR FLUX BIBCODE
HESS J1857+026 gamma 284.2958 +02.6667
HESS J1858+020 gamma 284.5833 +02.0900
MGRO J1908+06 gamma 287.18 +06.18
HESS J1912+101 gamma 288.0 +10.1
HESS J1923+141 gamma 290.75 +14.10
QSO B2005-489 BLLac 302.35579... -48.83158931 15.80 -2.20 20538
QSO B2356-309 BLLac 359.783050 -30.628031 45457
(c) 2010 UDS/CNRS - by CDS - Distributed under GPL v3 licence
57 sel / 57 src 92Mo