VO-Tech DS4 User Tools Study Report

The European Virtual Observatory - VO Technology Centre

A DESIGN STUDY

implemented as

SPECIFIC SUPPORT ACTION

Editor :	Markus Dolensky European Southern Observatory on behalf of DS4 co-workers

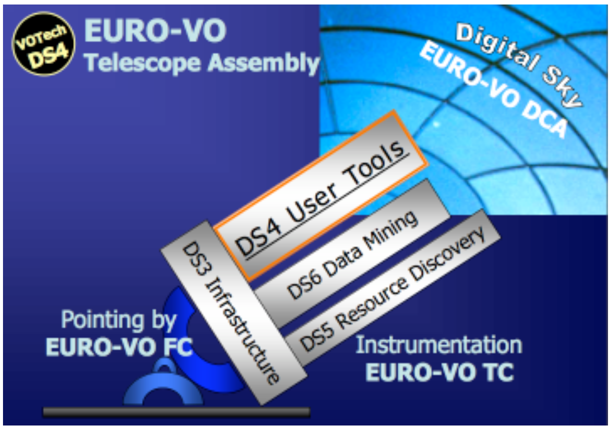
Contract Number :	011892
Project website :	http://eurovotech.org/

Table of Contents

1. Introduction & Context

This is the report of the Design Study 4 *User Tools* (DS4) of the VOTech project. It covers the efforts in the period from Jan 1st, 2005 until June 30th, 2007. The total investment of manpower until the end of the project in 2008 will be 6 FTEs. Due to the continued efforts an update is planned at the end of this project.

The task of the Design Study 4 is best described as designing and prototyping the instrumentation of a virtual telescope.
The telescope assembly rests on the foundation of the EURO-VO Facility Centre http://www.euro-vo.org/ which provides scientific guidance and user support. In metaphorical terms FC is pointing the instruments at observational targets in the digital sky.



The vaults of the EURO-VO Data Centre Alliance (DCA) are the roams that await their exploration by archival researchers. A further similarity to a a real observatory is the set of various types of instruments complementing each other. Research and development on the advancement of VO instrumentation is coordinated by the EURO-VO Technology Centre (TC). The common instrument platform is the topic of the DS3 Infrastructure. It provides the framework for interoperating data mining techniques (DS6), resource discovery methods (DS5) and finally the specialized user tools (DS4) which are the subject of this report

Acknowledgement

Many thanks to all co-workers of this design study. Special thanks to Mark Taylor and Keith Noddle (Astrogrid), Thomas Boch, Pierre Fernique, François Bonnarel, Sébastien Derriere, François Ochsenbein, André Schaaff, Mark Allen, Laurent Cambrésy, Bernd Vollmer (all CDS), Patrizia Manzato and Riccardo Smareglia (INAF), Fabien Chéreau, Nausicaa Delmotte, Francesco Pierfederici, Jean-Christophe Malapert, Bruno Rino, Remco Slijkhuis (all ESO).

2. Scope of Design Study 4

The intention of this study was to create designs and prototypes for new VO-compliant end-user tools. It started by producing a list of priority tools. These application areas were derived from feedback of precursor projects, the input from the scientific user community and the VOTech science team.

Work was conducted in close collaboration with data providers, e.g., members of the EURO-VO Data Centre Alliance, current and future missions, simulations as well as tool builders. A goal is to help existing facilities with technology take-up and to have a lasting impact on Astronomy data centres and their user community by exploring new designs and deriving generalized IVOA standards.

3. Scientific Drivers

3.1 VOTech Science Team

The VOTech Science Team is a group of astronomers recruited from all partner organizations. It assesses the relevance of implementations being derived from the DS groups to science problems, and arrange the scientific development of science-services, based on emerging DS designs/ implementations.

3.2 AVO Science Reference Mission

The Science Working Group of the Astrophysical Virtual Observatory project (EC contract HPRI-CT-2001-50, 2001-2004) generated a list of key science cases, the AVO Science Reference Mission [8]. Its analysis helped to identify the seven application areas outlined in the next chapter.

3.3 Astrogrid Science Cases

The AstroGrid project contributed a set of science drivers [9] as a component of its Phase-A stage - laying the foundations for the construction of its VO system.

3.4 INAF Science Cases

Bridging the Grid computing world and the more liberal design of interacting EURO-VO data centres on one hand and building interfaces between the theoretical Astronomy and observational Astronomy on the other hand led to the formulation of some specialized use cases:

- cosmological N-body+SPH simulations performed with Gadget2 code;
- cosmological N-body+AMR simulations performed with Enzo code;
- cosmological N-body simulations performed with FLY code;
- stellar evolution tracks and isochrones simulations perform with BaSTI code.

4. Compilation of Priority Tool List

The AVO Science Reference Mission (SRM) in conjunction with the other above described science drivers helped to compile a list of tools or application areas. All partners were involved in the discussion in the kick-off meeting, technical advisory panel meetings etc..

To give concrete examples below a mapping of 8 AVO SRM scenarios (S1-8) against required tools (T1-7):

#	SRM Scenario
S1	Circumstellar disks: from pre-Main Sequence stars to stars harboring planets
S2	Intermediate Velocity Clouds
S3	Which star will go Supernova next?
S4	Initial Mass Function within 1 kpc: from planetary to stellar masses
S5	Initial Mass Function for massive stars
S6	The contribution from low and intermediate mass stars to the interstellar medium
S7	Galaxy Formation and Evolution from z = 10 to z = 0.1
S8	Build-up of supermassive black holes

4.1 Scenarios vs. Tools

#	Required Tools by Scenario	S1	S2	S3	S4	S5	S6	S7	S8	Count
T1	multi-waveband image tool	+	+	+	+	+	+	+	+	8
T2	best fitting	+	-	+	-	+	+	+	+	6
T3	SED builder from archival data	-	-	-	+	-	+	+	+	4
T4	positional cross matcher	-	+	+	+	-	+	+	+	6
T5	access to theoretical data	-	-	+	-	-	-	-	+	2
T6	robotic telescope interface	-	-	-	-	-	+	-	-	1
T7	generic source extraction	-	+	-	+	-	+	+	+	5

After identifying application areas the next step was to map available manpower and expertise of each partner project and to plan specific tasks accordingly.

Partner	T1	T2	T3	T4	T5	T6	T7
ASTROGRID	+	+	-	+	-	-	+
CDS	+	-	+	+	-	-	-
ESO	+	-	+	+	-	-	+
INAF	-	+	-	-	+	-	-

Above table summarizes the situation with respect to DS4 only regardless of additional resources which were invested in the other VOTech design studies.

5. Application Areas & Matrix of Pre-Existing Tools

After identifying application areas a simple matrix about existing tools was compiled. Below summary reflects the status as of 2005. Obviously, such a *small* study cannot be fully complete and comments are solely intended to highlight aspects of possible DS4 work. It must not be taken as a criticism of existing software.

5.1 Multi-Waveband Image Tool

This is about software which can process and combine image-like data taken in different electromagnetic bands.

Tools for generating multi-colour images from individual FITS files

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
Aladin	3.026	Mature	Java	YES	YES	Free	
DS9	3.03	Mature	None	YES	NO	Free	Fast, works with IRAF
Skyview	3.5b	Beta	POSIX	YES	YES/NO	Free	Fast, works with IRAF, can run in server mode.
Photoshop	CS2	Mature	Mac/Win	YES	NO	Commercial	8 bit and 16 bit data only. Needs FITS plug in and WCSTools
GIMP	2.2.6	Mature	None	YES	NO	Free	8 bit and 16 bit data only. Needs FITS plug in and WCSTools
SkyView	1.1	Beta-level	None/Java	YES?	YES	Free	Can it be run standalone?
SDSS IDL code		Mature	IDL	YES	NO	Free	
SDSS C code		Mature	POSIX	YES	NO	Free	WCS info in JPEG images.

Tools able to create RGB (false-) color images (no display)

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
IRAF	2.12.2	Mature	POSIX	YES	NO	Free	Can generate RGB images, does not display them
IDL	6.1	Mature	Most platforms supported.	YES	YES	Commercial	Requires external libraries.

Other relevant tools

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
XImtool	1.3.1	Mature	POSIX	YES	NO	Free	8 bpp only. Cannot create RGB images yet.
SkyCat	2.7.3	Mature	POSIX?	YES	NO	Free	Cannot create RGB images.
JSky	2.5	Mature	POSIX	YES	NO	Free	Cannot create RGB images.
ATV	1.5	Mature	IDL	YES	NO	Free	Can it create RGB images?
Montage	2.2	Mature	POSIX	YES	NO	Free	Grid ready. Can project images. Cannot create RGB images. No display.
WCSTools	3.6.1	Mature	POSIX	YES	NO	Free	Can be used to project images. No display.

5.2 Best Fitting

Automated classification of millions of celestial objects by comparison of observational data and theoretical models is an important and frequent use case.

Astronomical Model Calculation & Fitting Packages/Libraries

Name	Vers.	Status	Reqs.	Stand-alone	Web-based	Licen.	Notes
Pegase	2.0	Mature	POSIX	YES	NO	Free	PÉGASE is a code to compute the spectral evolution of galaxies. Requires VOTable plugin.
starburst99	5.0	Mature	None	NO	YES	Free	Starburst99 is a web based software and data package designed to model spectrophotometric and related properties of star-forming galaxies.
GALAXEV		?	YES	NO	POSIX	Free	GALAXEV is a library of evolutionary stellar population synthesis models. Requires VOTable plugin.
GAVO SED Classifier		?	Java	NO	YES	Free	Compares observed and model SEDs resulting in a rectangular classification matrix of dissimilarities.

General Data Analysis Tools offering Fitting Support

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
ROOT	4.04/02	Mature	POSIX	YES	NO	Free	Generic data analysis sw.
R	2.1.0	Mature	None	YES	NO?	Free	Extensive statistics package.
IRAF	2.12.2	Mature	POSIX	YES	NO	Free	ICFIT package.
IDL	6.1	Mature	Most platforms supported.	YES	YES	Commercial	Generic data analysis tool.
SpecView	2.11	Mature	Java	YES	YES	Free	

5.3 SED Builder

In this document the term spectral energy distribution (SED) denotes a plot of brightness, or flux density versus frequency or wavelength. It is widely used in Astronomy to characterize and to study astronomical sources. The output of an SED builder can be input to a fitting tool. Conversely, the results of a source extractor, multi-waveband tool and cross matcher are all potentially input to an SED builder.

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License
SpecView	2.11	Mature	Java	YES	YES	Free
VOSpec	1.4	Mature	Java	YES	YES	Free

Generic plotting tools/libraries

IDL - Gnuplot - PGPLOT - Matplotlib - Mirage

5.4 Positional Cross Matcher

This tool assesses the probability that two distinct measurements observed the same object. Although this sounds trivial it is obviously not because it may require data fusion across detectors with very different properties (e.g. resolution) at different epochs and so forth.

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License
VizieR	?	Mature	None	NO	YES	Free
Open SkyQuery[2]	1.16	Mature	None	NO	YES	Free
Aladin	3.026	Mature	Java	YES	YES	YES
TOPCAT	1.5	Mature	Java	YES	NO	Free
AstroGrid X-Corr / Federation prototype			None	NO	YES	Free

Databases offering spatial query support

MySQL - PostgreSQL - DB2 - Oracle - Sybase

Other relevant software

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
R	2.1.0	Mature	None	YES	NO?	Free	Extensive statistics package.
Mirage	0.3+	Beta	Java	YES	NO	Free	Needs VOTable/Registry and FITS plugins.

Relevant Resources

- Requirements for a Cross Matching Tool
- Report on Cross Matching Catalogues

5.5 Access to Theoretical Data

Name	Web-based	License	Notes
Planck Satellite Pipeline Simulator	YES	Free	
Simulation query tool	YES	Free	
ITVO@Trieste	YES	Free	Simulation query tool
ITVO@Catania	YES	Free	Simulation query tool

5.6 Robotic Telescope Interface

The most notable recent development in this context is a standard on Sky Event Reporting *VOEvent* [7]. Contributions to this standard, however, were not coordinated by this design study. None of the project partners had a dedicated specialist available for related work in the context of DS4. Consequently, there are no activities related to this application area to report. In all other application areas there had been dedicated efforts though.

5.7 Generic Source Extraction

Typical detectors observe many objects simultaneously. A typical scenario is to break down an image into a list of individual stars and galaxies.

Name	Vers.	Status	Reqs.	Stand alone	Web-based	License	Notes
SExtractor	2.3.2	Mature	POSIX	YES	NO	Free	Has been wrapped into a couple of WebServices[7].
ACE	0.4	Beta	IRAF	NO	NO	Free	Fast. Not released yet.
DAOPHOT II	1.3.2	Mature	POSIX	YES	NO	Free	Available for may platforms (e.g. IRAF, IDL).
DOPHOT	3.1	Mature	POSIX	YES	NO	Free	Web site not accessible.
MOPEX/APEX	1.6	Mature?	Solaris/Linux	YES	NO	Free	Developed for Spitzer data.

6. Designs and Prototype Implementations

6.1 Multi-waveband Image Tool

Requirements for a multi-waveband image tool were fed back to DS3 and led to improvements of infrastructure components such as the server side version of the Astrogrid Runtime library (ASR) and the design of the VOSpace storage system. Independently, a number of new features were added to the Aladin Interactive Sky Atlas, as developed in the next section.

Aladin improvements

The following developments have been pushed both by scientific drivers and user requirements.

Improved color image manipulation

- Color images can now be converted to grayscale images, allowing to use features not available on RGB planes, *e.g. computing isocontours, extracting sources, etc*.

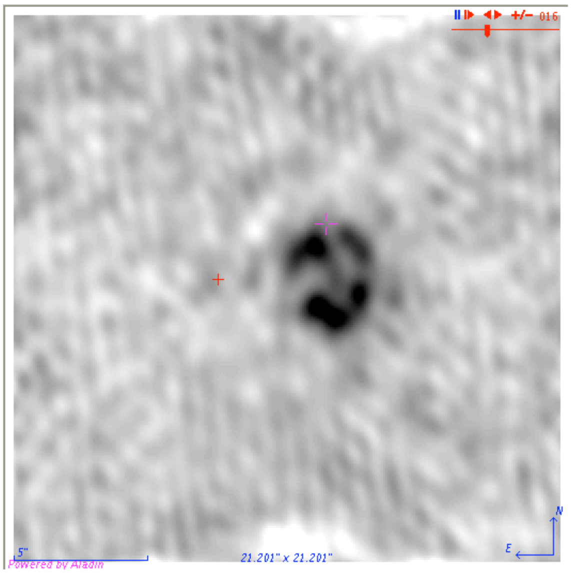
- Color images can also be saved as FITS files, thus preserving astrometry (WCS) information. The convention adopted in the generated FITS is to use the following key/value pairs : *NAXIS=3, CTYPE3='RGB'*. This is compatible with DS9, meaning that color images created and saved in Aladin as RGB FITS can be reopened in DS9, and vice versa.

3D data access and manipulation

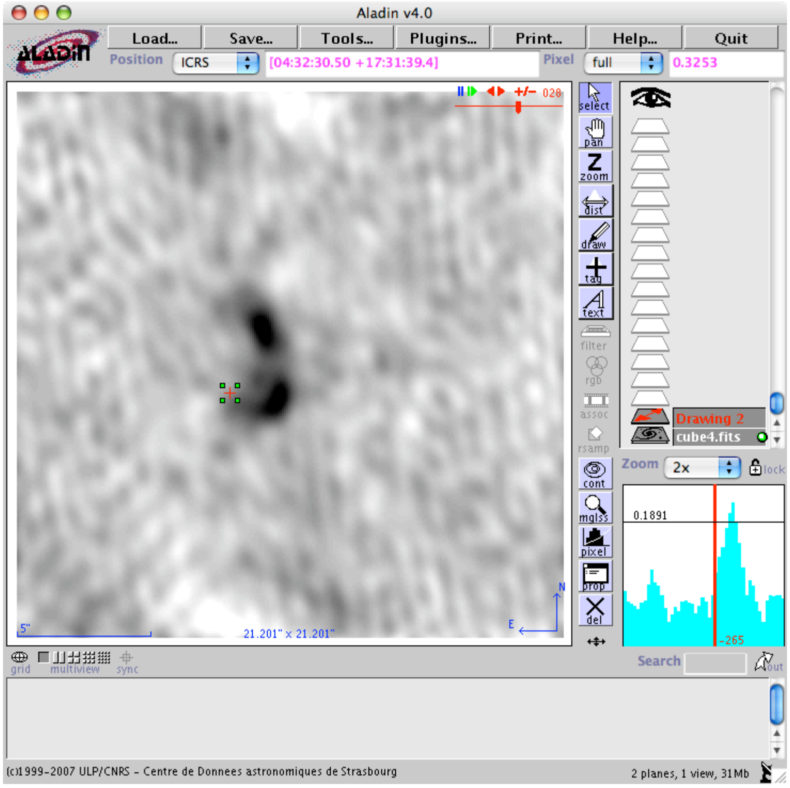
Support of data cubes and MEF (Multi FITS Extension) has been added to Aladin. This feature has been tested in collaboration with ESO and CADC (Daniel Durand) with CGPS and SINFONI cubes.

Some basic cube manipulation have been developed :

- once loaded, each slice of the cube is displayed as an animation. The user can choose to pause the animation, and go to a specific frame.
- the cube plane acts as a regular Aladin plane : contrast can be adjusted for the whole cube, objects from tables can be overlaid.
- when tagging a given position of the cube, Aladin builds and displays the spectrum corresponding to the pixel value of each slice at the chosen position.

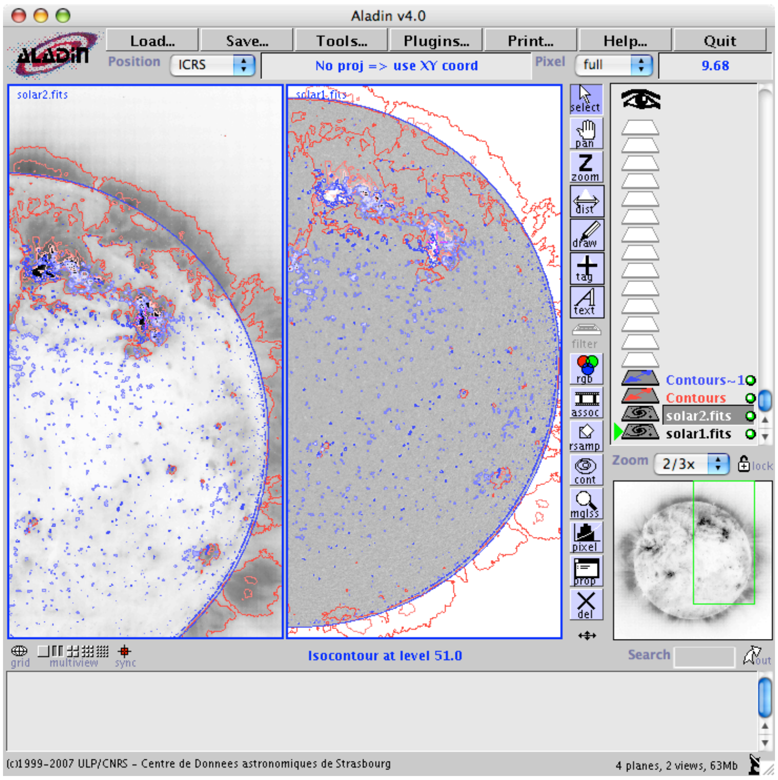


On this screenshot, Aladin displays a slice of a loaded data cube. Click on the thumbnail to visualize how Aladin displays the cube as an animation.



Clicking on a given position generates the corresponding spectrum, displayed in the small view at the bottom right of the window. The current position in the cube is shown by the red bar

Support for solar images



Support for solar images has been improved. A new projection, "XY linear transformation", is now taken into account. It allows computation of isocontours, creation of RGB or blink planes for solar images loaded in Aladin.

Instrument footprint visualization

Aladin is a testbed for the instrument footprint standard (sect. 7.1): It allows to load user-defined description of instrumental footprints, to display and overlay them on top of images. The early implementation of this emerging standard allowed to provide useful feedback which led to some updating and refining of the format.

This development has been performed with 2 use cases in mind :

- preparation of telescope observation proposals. In this case, footprints can be moved, rotated, and the position and roll angle can be easily retrieved.
- visualization of the coverage of a set of observations *(coming from instance from a SIAP service)*

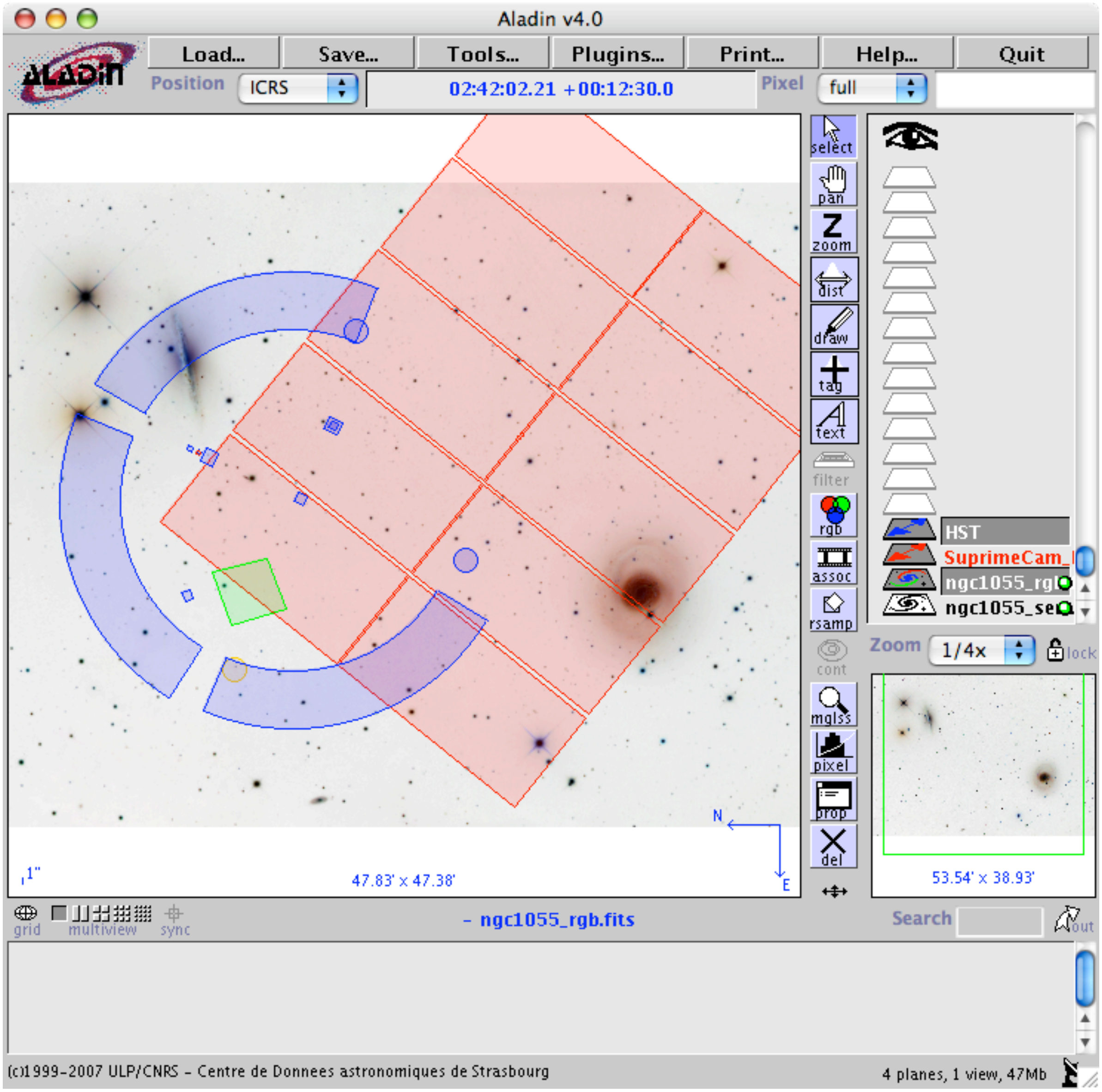
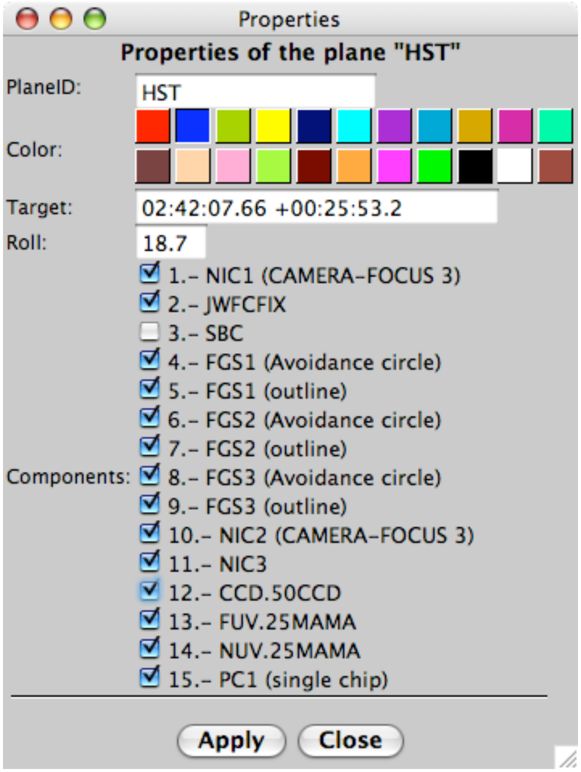


Figure 1: In this example, Aladin displays the footprint of the Subaru Suprime instrument (in red) as well as the footprint of the Hubble Space Telescope on top of a SDSS background image.



The properties window of a footprint plane allows to retrieve the current position angle and the roll angle. The user can also choose which parts of the footprint he wants to hide/show.

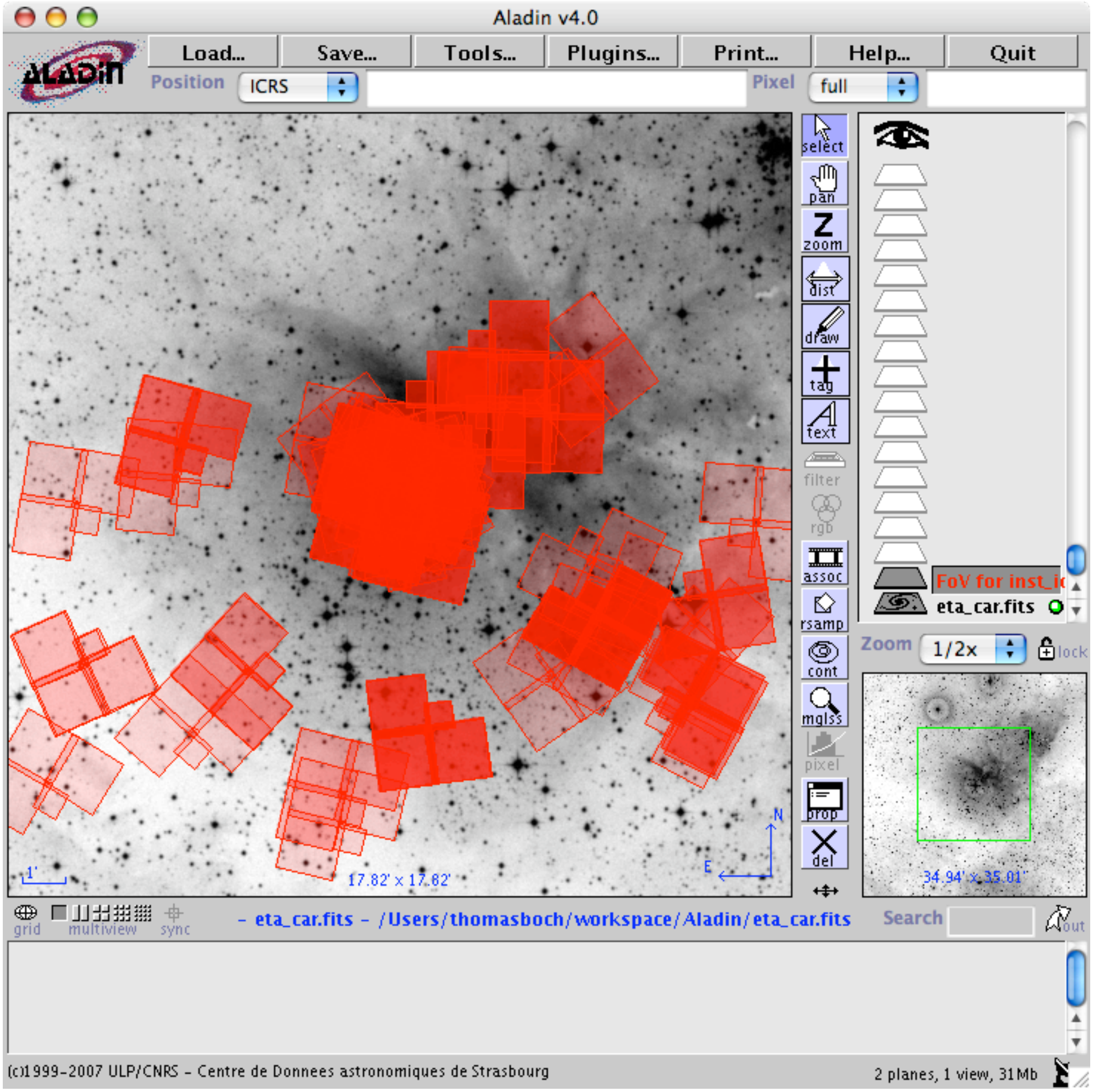
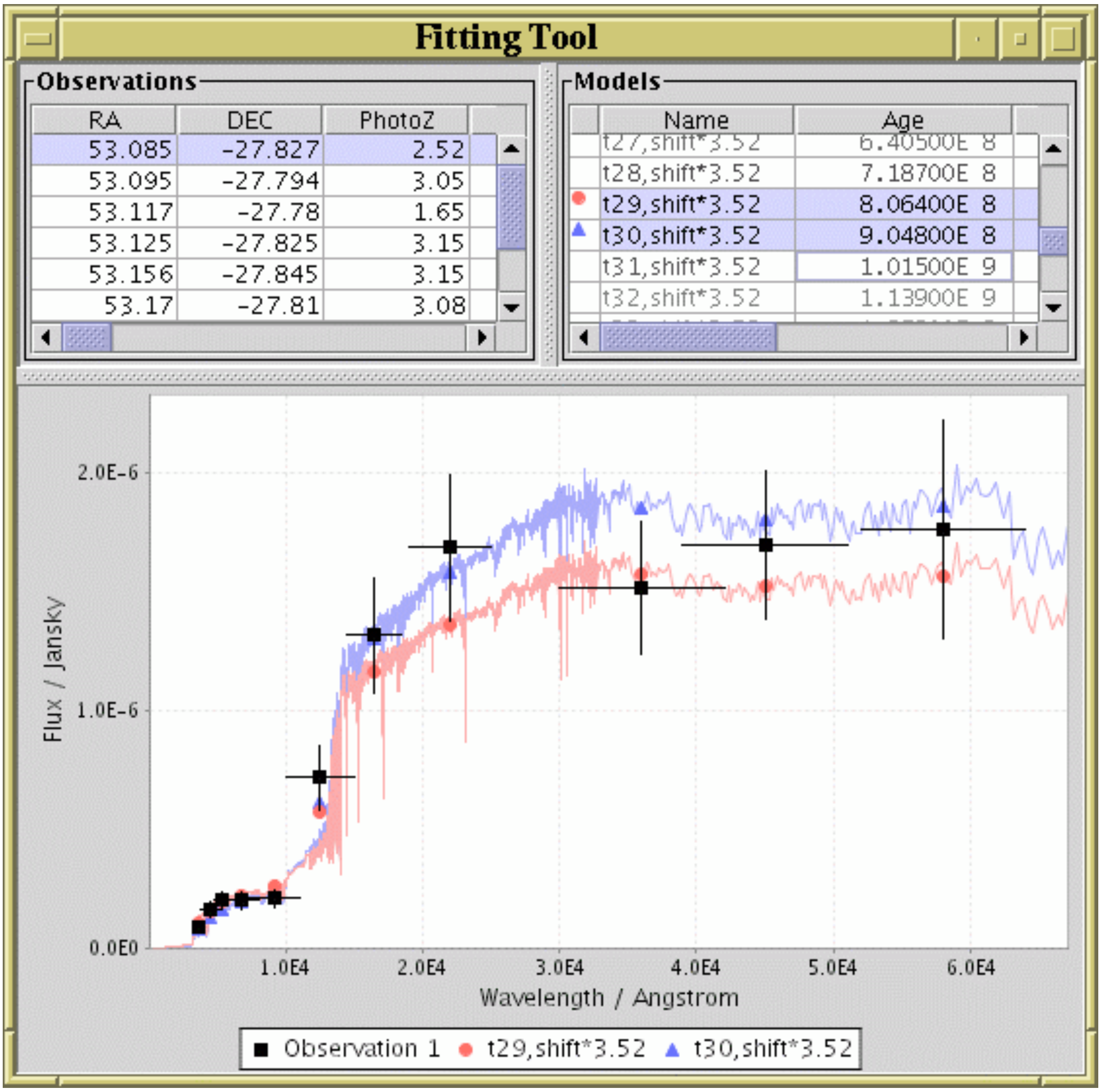


Figure 2: When visualizing the coverage of a set of observations, displaying translucent footprints proves to be quite useful as it gives a visual indication of which regions have been observed most.

All those improvements are already available in current version of Aladin or will be available soon in a next release.

6.2 Best Fitting

A prototype best fitting tool called Yafit has been developed. It is capable of comparing model and observational data to identify the best fits, suitable for interactive or batch use. The screenshot below shows Yafit comparing a list of observed SEDs (data in the top left table) with a list of library model spectra (names and attributes in top right table). The zoomable lower panel shows two of the model spectra (red and blue curves) which form the best fits to the selected observed SED (black points), allowing a visual inspection of the goodness of fit.



An extensible I/O framework provides compatibility with a number of observational and model data formats. Currently supported model formats are Galaxev, Starburst99, SVO-format VOTable archives and user-prepared VOTables. Observational data is supplied as a table (VOTable or other) plus an additional key file. Support for additional formats, including ones related to the IVOA Spectral Data Model, can be added in future.

The tool is documented, tested and usable, but more functionality needs to be added before it can satisfy a wide range of scientific use cases.

6.3 SED Builder

Work towards spectral energy distribution (SED) building entailed contributions to an IVOA standard spectral data model and its utilization for a public service converting photometric magnitudes to physical flux densities. Furthermore, a colour cutter for the selection of objects based on photometric colours has been implemented as a configurable workflow to the Astrogrid workbench.

SED Workflow. The SED builder is a workflow encompassing several stages. The first one is a data access stage. The second one generates a catalogue of objects using a source extractor. The cross matcher will then cross identify objects across several catalogues. The individual values need to be put on a common scale with a given physical unit. In the optical domain this can often be achieved by converting Vega or AB magnitudes into physical fluxes. Such a magnitude to flux converter has been set up for several ESO instruments, but it also provides a (limited) service for general conversions.

Magnitude to Flux Converter. The conversion of Vega or AB magnitudes to physical fluxes is based on instrument specific transmission curves. They are retrieved on demand from the ESO Exposure Time Calculators. The CDS's unit conversion library [3] computes the value in one of the many physical units that are customary in the various subdisciplines of Astronomy. A web interface is available here http://archive.eso.org/apps/mag2flux/ [5].

Instrument Filters. One of the limitations of current SED builders is the limited programmatic access to instrument filter characteristics. For the aforementioned prototype service 200 such data sets have been made available:

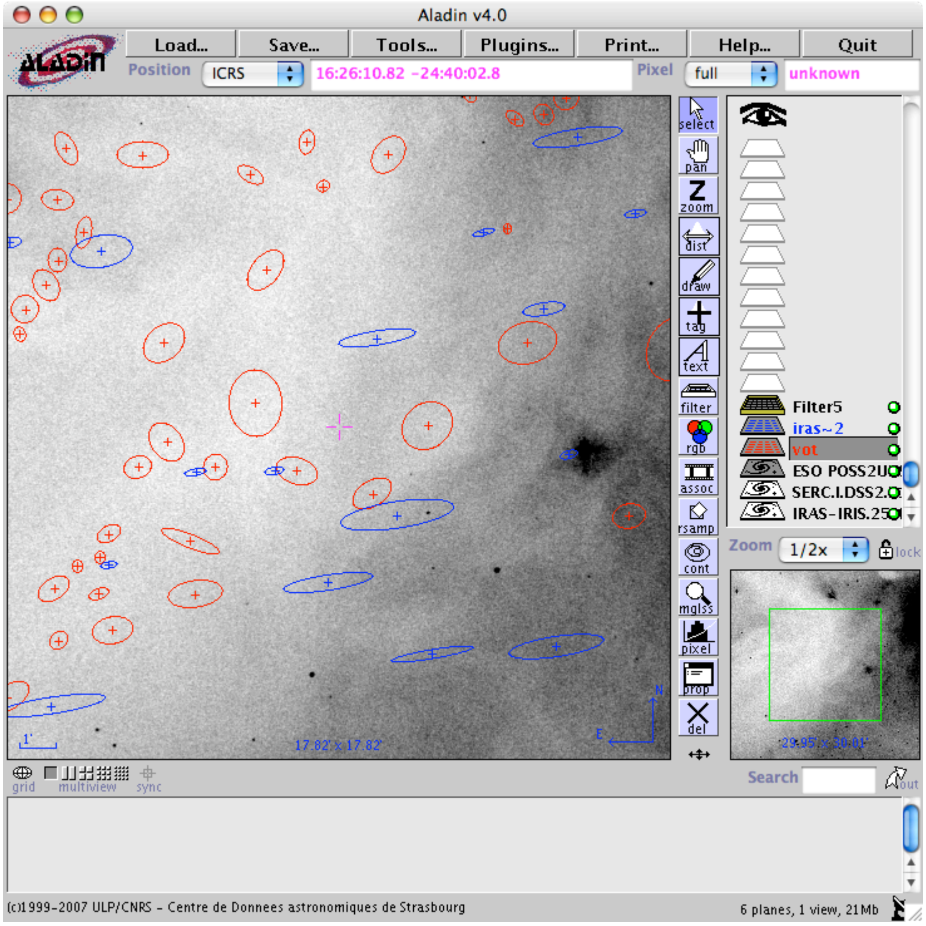
Instrument	No. of filters	Instrument	No. of filters	Instrument	No. of filters
SOFI	3	ISAAC_SW	20	WFCAM	5
FORS1	44	ISAAC_LW	10	WFI	47
FORS2	44	VIMOS	6		

6.4 Positional Cross Matcher

In VO era, being able to cross-identify objects coming from different data sources has become a crucial topic, as it allows to merge multi-wavelength information. The community is demanding cross-matching facilities and several components were prototyped.

6.4.1 Improved cross-matcher in Aladin

The cross-match tool available in Aladin, initially developed for the AVO demo, has been improved to take into account *error ellipses*, i.e. uncertainty on positions.



In this screenshot, we display the position uncertainty ellipses for the two catalogues to be cross-identified.

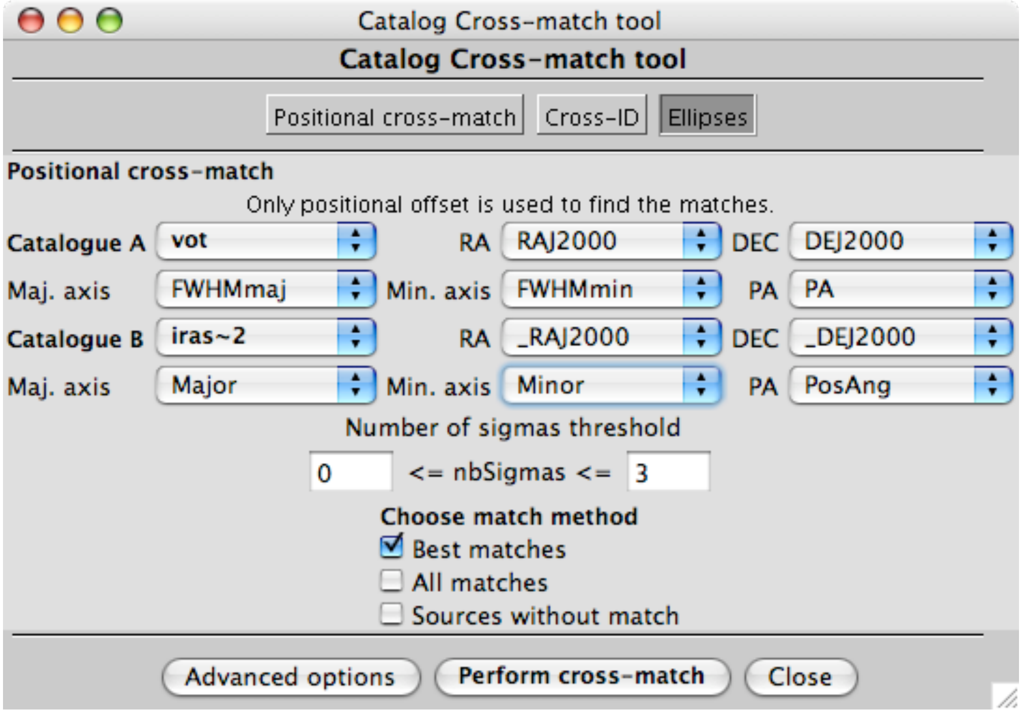
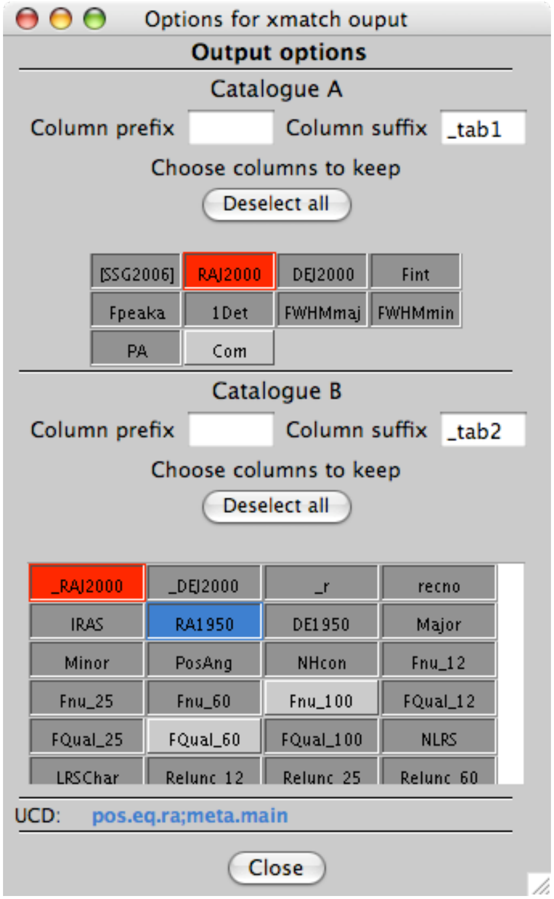


Figure 3: cross match dialog

In the dialog (Figure 3), one can specify for each catalogue :

- columns holding the position (right ascension and declination)
- columns holding the uncertainty ellipse major axis
- columns holding the uncertainty ellipse minor axis
- columns holding the uncertainty ellipse position angle

In most cases, if the tables came with sufficient metadata, this information is already correctly picked, on the basis of UCDs. Eventualy, the user enters, as a *number of sigmas*, the minimum and maximum threshold to be used for the computation.



The user also has the ability to specify which columns he wants to keep in the cross-match output.

As a result, a new catalogue plane is created, with all kept columns from the first table, all kept columns from the second table, and with the computed *number of sigma*.

6.4.2 Asynchronous REST cross-match service

In order to let external tools or services perform positional cross-matching, we have developed a prototype cross-match service using the REST protocol pattern, based on the cross-matching algorithm used in Aladin. It allows positional cross-match of any 2 VOTables accesible on-line through a URL, and follows recommendations of the first UWS for REST document. For large tables, the asynchronous nature of the service is essential. Here is an overview of how a user initiates a cross-match computation and retrieve its results :

- POSTing VOTable URLs and xmatch parameters to *xmatch/jobs* will initiate a new cross-match task. The user gets back a job identifier.
- This identifier can then be used to query the service (GET request to *xmatch/jobs/<job-id>*) in order to get the status of the job (running/failed/completed)
- When the status of the job is set to **completed**, the result can be retrieved by a GET request to *xmatch/jobs/<job-id>/result*

Future development plans include the distribution of the cross-match task on a cluster of PCs, thus improving its scalability.

6.4.3 Positional index for VizieR

The design of a positional index on several thousand source (VizieR) catalogues started. Such an index will enable more powerful and efficient positional operations, and will notably allow :

- Fast search by position
- Fast cross-match between any 2 VizieR catalogues
- Easy creation of density maps for any catalogue with position

Current VizieR architecture. The VizieR database is made of 12,500 tables, half of them have got coordinates on the sky for a total of more than 5 billion positions. Tables with sky position are organized in 2 parts :

- tables with less than 10 million rows are stored in a relational database (Sybase or PostgreSQL)
- larger tables are stored as Unix binary files associated to a specific search program

With the current index architecture, the query *"give ALL sources in any of the 5300 tables around a position"* is a 2 steps process:

- Query the index to get the list of tables with potential matches
- Query each table to retrieve the sources around the position

Querying many tables at a time is compute intense.

New positional index design. In the new positional index scheme, full-precision positions of all catalogues will be merged and stored in a single dataset with an approximate size iof 100 GB. Along with the positions, this index will only store the corresponding position accuracy of the table.

Indexing will be performed using *Qbox*, a hierarchical partitioning scheme allowing to split the celestial sphere into Qbox cells.

The main difficulty will be the maintenance of the index. As the re-creation of the whole index is a slow process, the index will most likely stand in a centralized place instead of being recreated on every VizieR mirror.

The first implementation of this index is planned for VOTech stage 5 (March-September 2007).

6.5 Access to Theoretical Data

Linking theoretical and observational Astronomy is the aim of three combined and ongoing activities. Firstly, the definition of a data model specifying terms for interaction between the two domains. Secondly, the definition of the Simple Numerical Access Protocol (SNAP) to access the theoretical data and also making a different cutout of the snapshot boxes. These two activities happen in close collaboration with the IVOA Semantics and Theory working groups. Thirdly, the implementation of a prototype Theory Virtual Observatory TVO (9) for this selected scenario: a tool to create on-the fly maps and profiles of some quantities for the simulated galaxy cluster, the possibility of opening via the Aladin tool also simulated data.Work in progress are: the developing of a Web Services tool that via SNAP protocol is able to access and cutout the output file: Fly , HDF5, Gadget; and ADQL requests processing. To be considered as a goal by the end of the project, the implementation of a view on hydrodynamic simulations as Chandra satellite observations.

These activities allow to validate standards, protocols and data models defined within the IVOA collaboration. The tools built for the different implementation shall became of general use for the access and manipulation of theoretical data.

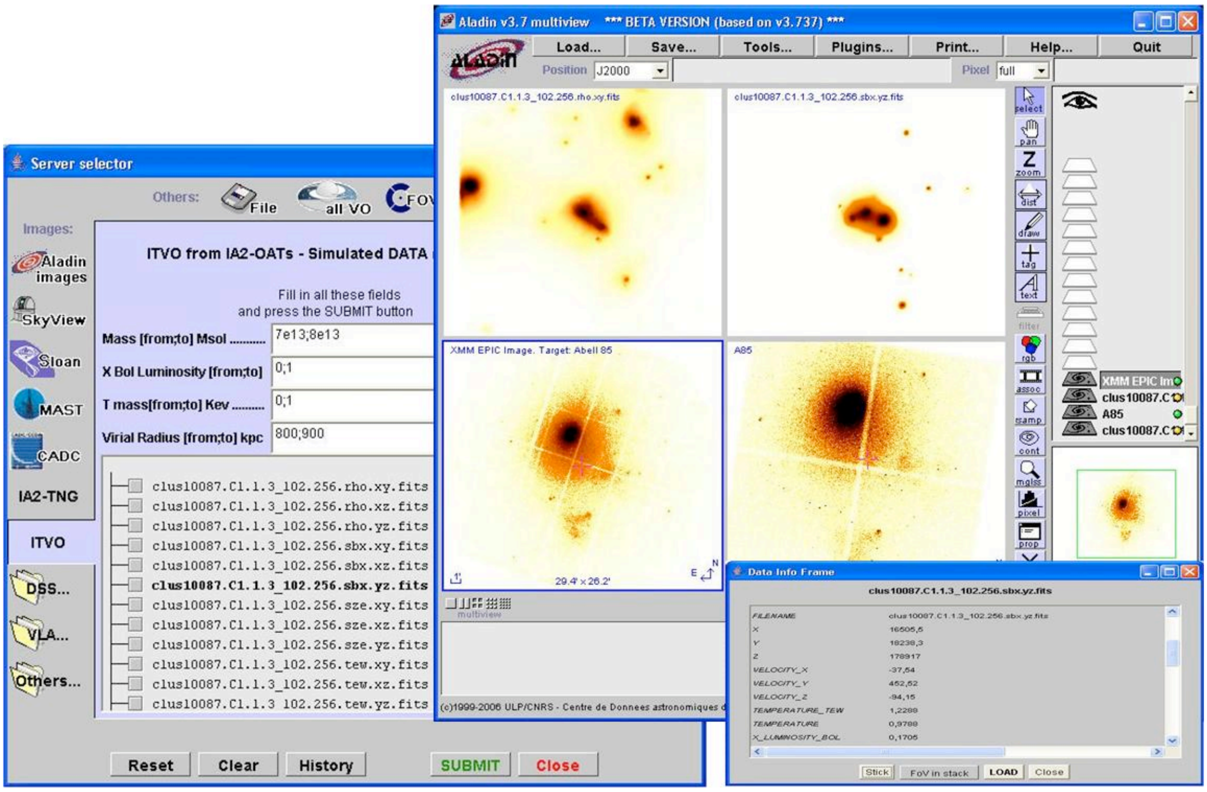
These activities are performed in coordination with the VO-DCA project, WP4 and WP5.

The access to theory data is given by multi level post-processing Web Portals:

- ITVO@Trieste (http://wwwas.oats.inaf.it/IA2/) The features are:
    - multilevel DB based on Oracle 10g ;
    - web interface written in PHP code;
    - personalize the SQL query (for all 3 levels);
    - single or multiple data download(for all 3 levels);
    - DVD data request (for all 3 levels);
    - Header view (only for the 2 level);
    - Maps and Profiles Preview (in future graphics of profiles and maps will be create on-the-fly);
    - VOTable missing UCD(at the moment only for the 2 level but in future also for the other levels);

- ITVO@Catania (http://www.astrocomp.it/itvo) The features are:
    - multilevel DB based on PostgreSQL[?] 8.1;
    - web portal in PHP with pre-configured or custom SQL queries;
    - direct "get" link to download simulation files via anonymous FTP;
    - database schema currently modeled on cosmological simulations;

- New feature add to the Aladin tool for opening simulated 2-D data:
  - Search via web services the 2-D maps of the simulated galaxy clusters;
  - Visualized the header of the galaxy groups or clusters maps;
  - Open the fits file maps (projection on one plane) of simulated galaxy clusters for the quantities:
    - rho = density
    - sbx = X luminosity
    - sze = Sunyaev-Zel'dovich emission
    - tem = emission-weighted Temperature
    - tmw = mass-weighted Temperature
    - tsl = spectroscopic-like Temperature



6. 6 Generic Source Extraction

The Redshift Maker workflow available through the Astrogrid workbench is a service performing automated extraction of object properties from observational data. It derives photometric redshifts from optical imaging based on data from the INT Wide Field Survey.

6.7 Further Achievements

Worklfows. A number of workflows became available through the Astrogrid workbench. Apart from above mentioned Colour Cutter and Redshift Maker there is the Movie Maker for assembling solar image time series, GouldBelt[?] for the identification of components of the Gould Belt in multi wavelength catalogues and a workflow investigating solar Coronal Mass Ejections (CME) and their influence on geomagnetic storms at the Earth.

Robustness. In order to improve overall robustness, versatility and interchange capabilities across tools several further software updates and documentation items were released. They are not specific to a particular of above applications but complement at least one of them. Some examples are mentioned below:

Aladin Prototype. New features of the Aladin Interactive Sky Atlas prototype include support of the PLatform for AStronomical Tool InterConnection[?] protocol (PLASTIC), a programmatic interface to the widely used commercial IDL computing language, an improved cross-matcher, improved support of data cubes as well as colour images and solar images as well as integration of an experimental workflow engine (JLOW). Furthermore, it acts as a testbed for a newly defined format for encoding instrumental footprints of observations and for visually representing ranked query results.

SAVOT. Binary data support was added to the SAVOT library for reading and writing data files in VOTable format.

WebStart. Similarly a study of the Java Web Start technology resulted in a new deployment method for software updates.

TOPCAT. The Tool for OPerations on Catalogues And Tables - now supports also solar time series data, additional plot and data formats, enhancements to the expression language, as well as the PLASTIC interface. Visualisation enhancements including error bar plotting will be available shortly. STILTS - the Starlink Tables Infrastructure Library Tool Set - the sister package of Topcat for command-line tools has added support for most of its facilities through the Common Execution Architecture (CEA) web service framework. A number of other enhancements were also made to both packages.

Astroscope. A browser for dynamic discovery of data available for further processing via a VO interface - has numerous new features available through its graphical user interface and can now dynamically perform registry look-up queries.

PLASTIC. PLASTIC capabilities were added to the third party (ex-Starlink) tools GAIA and SPLAT.

6.8 MEx: metadata extraction tool

Rationale. In order to search and find data of interest it is necessary to describe and catalog them in a homogeneous way. The MEx utility is supporting this task for astronomy data products like images and spectra that are stored in FITS format.

MEx extracts and transforms keywords and thereby removes the instrument and observatory signature. This is achieved by converting values to physical units using the CDS unit conversion library [3] and mapping them to standard vocabularies (UCD) and concepts (utype). Users may supply their own mapping and data models. Special purpose s/w modules can be hooked in where simple mapping expressions are not sufficient to compute the desired values. This makes MEx very flexible and versatile.

Architecture. To achieve the proposed goal in a generic way, usable by any data centre, MEx is split in two components, executed in sequence: keyword mapping and persistence.

The keyword mapping component processes the FITS files, applies the mapping definition to them, and produces an in-memory list of all files and normalized keyword values.

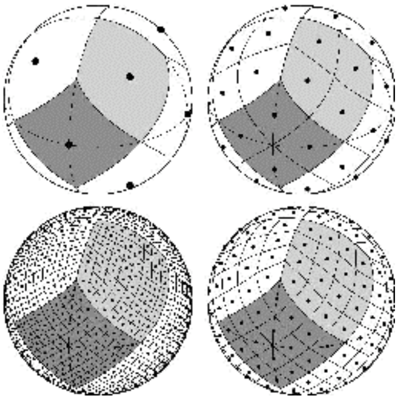
The persistence component persists this list in whatever format/database/etc a specific application might require; a particular data centre will customize this component to meet its data model.

This tool is to be installed and customized by data centres, and used by project scientists. For example, it can be used to populate a database that will serve data for a SIAP service.

6.9 Specialized SIA server

Intersection problem. In a typical SIA usage scenario all images within a given search region are returned. For irregular detector shapes and depending on the projection it can be demanding to compute an accurate result set.

Resolving a positional query becomes much more efficient when pre-computing the four 3D vectors of the corners of a rectangular shaped detector with respect to the camera's center on the tangent plane.



It is still too demanding to select from millions of observations in an interactive session. Therefore, the *Healpix* algorithm [6] (Hierarchical Equal Area isoLatitude Pixelization) is used for partitioning the celestial sphere and to pre-select small subsets which dramatically speeds up the calculation.

The optimal size of a sky index for wide field imaging cameras is about 3 square degrees.

WCS2FOV Library. A typical FITS file header contains Astrometric information to locate an image on the sky thanks to the WCS keywords. However, these keywords do not represent directly the detector shape. A WCS2FOV library has been developed to convert the WCS keywords into a proper instrumental footprint description. For visual browsing of multi-instrument or even multi-mission archives it is desirable to plot such instrumental footprints. It is then natural to combine this geometric information within the result set of an SIA query. A technical specification has been defined for that very purposes (see section about IVOA).

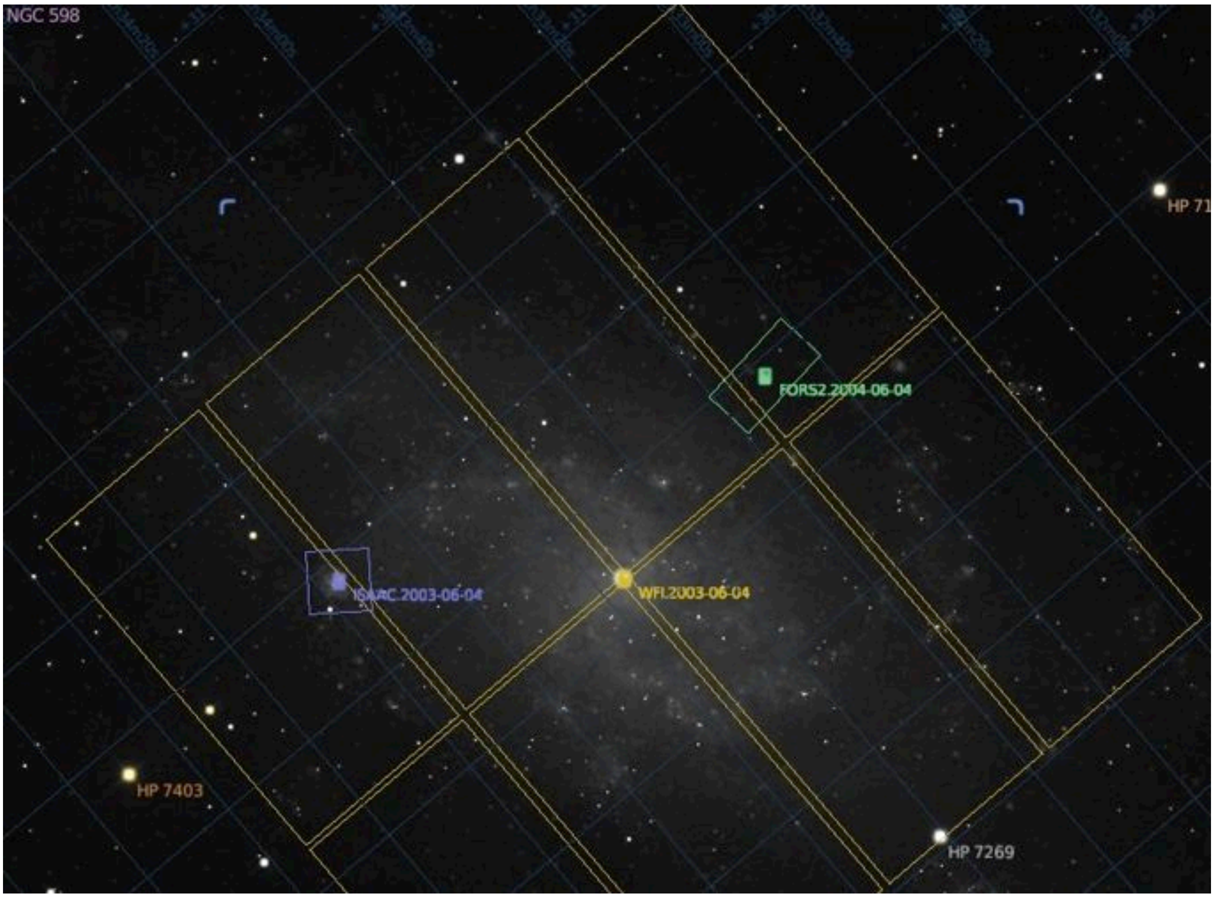


Figure 4: Visual browsing scenario; 8 chip CCD camera (yellow), ISAAC instrumental footprint overlay (blue), FORS2 observation (green).

6.10 Interoperability between tools

Interoperability between tools is another key topic. This includes not only interaction between VO tools themselves, but also interaction between VO tools and legacy software, as IDL, MIRAF, AIPS, etc. The latter is a real challenge as it will allow astronomers to access and use VO services from their everyday tools. The following sections relate CDS effort on the topic of interoperability between Aladin and some other tools.

6.10.1 PLASTIC-compatible Aladin

Aladin developers have been part of the development of the PLASTIC protocol, and Aladin has been one of the first application being *PLASTIC compatible*.

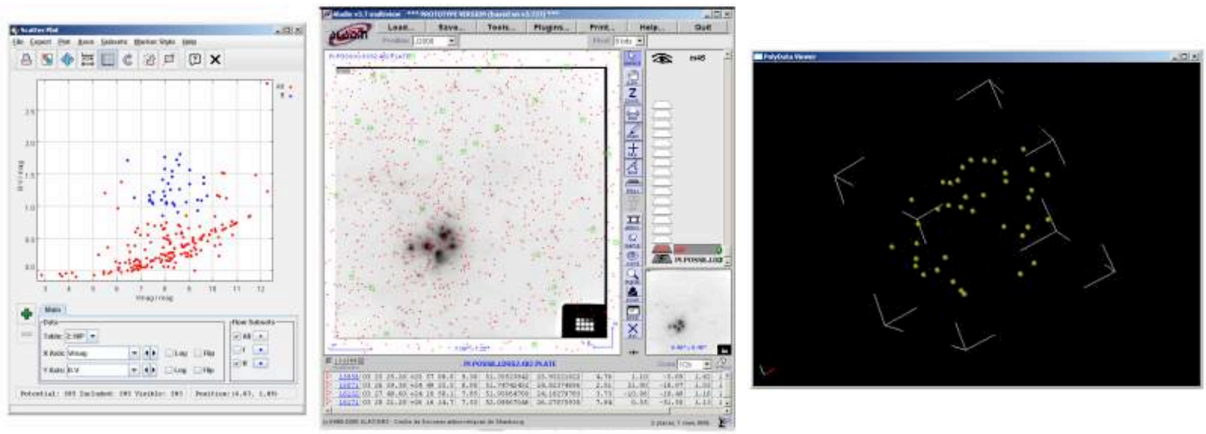


Figure 5: Supporting PLASTIC allows Aladin to interact with other PLASTIC-compatible tools, and extends its capabilities. In this example, the same set of points has been cross-selected in Aladin, TOPCAT and VisIVO, enabling exploration of data from different perspectives.

For more information about motivations and lessons learnt from PLASTIC integration in Aladin, please refer to section 5.4.4.2 of the DS6 Study Report

6.10.2 IDL library to control Aladin

IDL (Interactive Data Language) is a data analysis package widely used in the astronomical community. Some astronomers have expressed the need for being able to use Aladin features from their IDL scripts. To fulfill this request, we have developed a set of IDL functions allowing to control Aladin. This library relies on IDL Java Bridge, a mechanism allowing to call Java methods from IDL. It allows to perform the following operations :

- launch Aladin
- load a table in Aladin from a set of IDL variables
- load an image in Aladin
- retrieve an Aladin image plane as IDL variables
- retrieve table values from an Aladin catalogue plane
- select some objects in Aladin
- retrieve pixel value at a given position
- modify Aladin color table
- set the reticule at a given position on the sky

Full documentation and installation instructions are available on the project's wiki. A Flash exercise is there as well.

6.10.3 Plugin mechanism in Aladin

Beginning with version 4 Aladin can be extended by plugins written by 3rd party developers. A set of methods allowing to access and manipulate Aladin internal objects is provided.

Gains from this modular approach are :

- Simplicity : Developers can easily develop their own Aladin extensions for their own purposes without having to interact with Aladin developers
- Robustness : Aladin core code stays untouched, and is not bloated by external developments
- Flexibility : Users choose to install only plugins of their interest

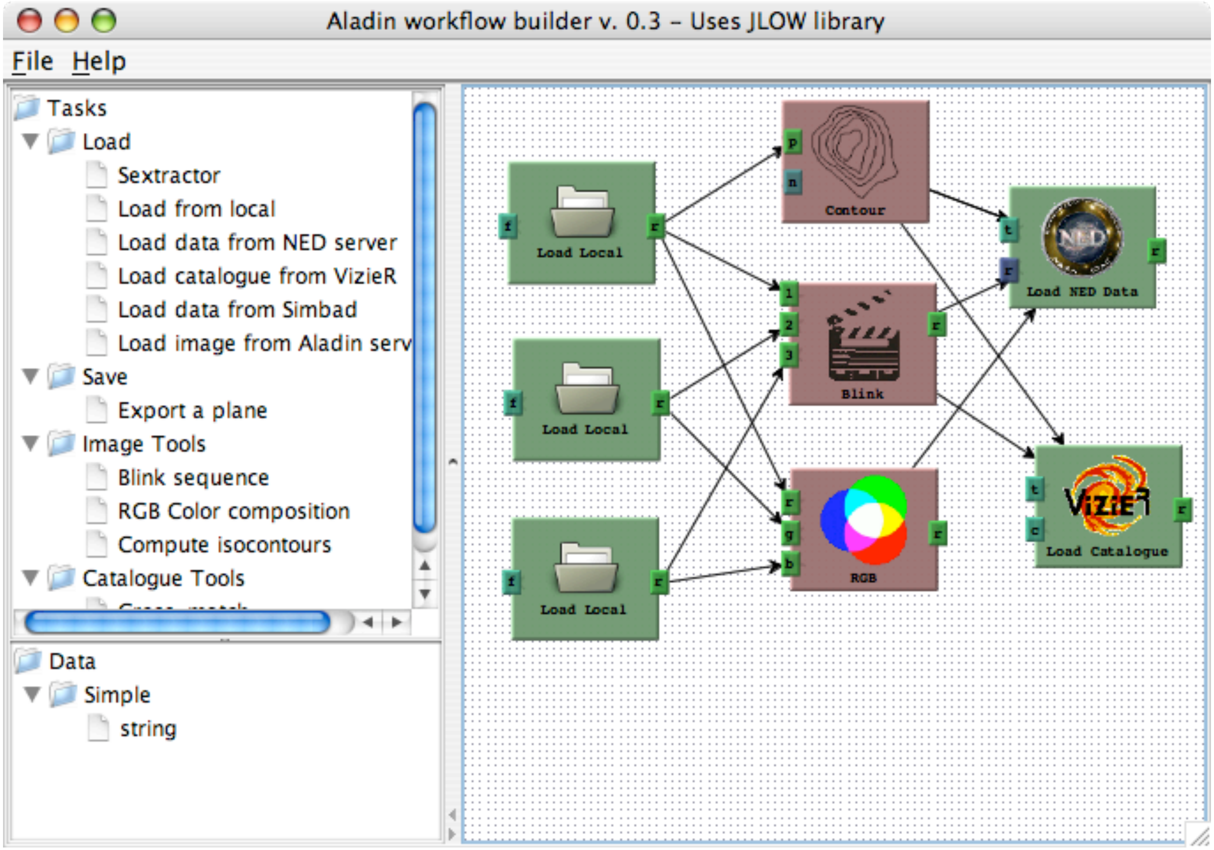


Figure 6: Here is plugin example. Called "Aladin workflow builder", it is based on the JLOW library and allows to build and execute a sequence of script commands from a graphical interface, just by dragging and dropping some boxes (representing tasks) and linking them

Details about development and installation of a new plugin are available [on Aladin plugins page].

7. Interaction with global IVOA Community

A number of activities were carried out in collaboration with the International VO Alliance (www.ivoa.net) and resulted in technical standard specifications, proof of concept type of software prototypes and experimental compute services:

For instance, a specification for exchanging instrumental footprints of observed regions on the sky has been drafted and is already supported by various tools such as the Aladin suite and a new visual browser to the ESO archive (release in Q3/2007). Another example is the first use and adaptation of the Simple Image Access (SIA)[16] protocol to attach and transfer world coordinate system information together with well supported industry standard formats like .gif and .png.

Two standards for theorectical data were drafted: One defining a data model [10] for simulated data and one about the Simple Numerical Access Protocol (SNAP)[11] for accessing and subsetting such data.

7.1 IVOA Note: Footprint Overlay Specification

Most observing tools, archive browsers and proposal preparation tools provide means to graphically overlay instrumental footprint information onto some celestial background image with a given world coordinate system [12]. The exact description of detector characteristics and the observation context can be truly complex. Therefore, it is extremely helpful to provide a quick look capability which mimics a human observer glancing at the sky and drawing a pattern of the aerial coverage as one would see it when peeking through various oculars.

Notwithstanding the merits of other efforts in this topical area [13] the Footprint Overlay Specification is special in that it can be directly plugged into query results of image and spectrum data services [14,15] using extensions to a DAL query response [16,17]. Furthermore, it provides a container to embed application specific attributes to exploit specific features of a given legacy client software (e.g. Hubble Space Telescope - Astronomer's Proposal Tool, HST - APT). Care has been taken to make clear separation between the geometry and the application specific attributes. Therefore, it is possible to reuse the respective IVOA data models [18,19] and still add, for instance, a representation layer for rendering.

This convention provides means to define the geometry of an instrumental footprint such that it can be projected onto a sky coordinate grid and rendered by a visual archive browser (Aladin, new ESO archive browser, ...). It is also a step towards the creation of spatial search engines. The instrumental footprint may be that of an actual or planned or simulated observation. A footprint may consist of one or several STC shapes. A given footprint can be associated with a set of observations. A rule set is available to assign rendering attributes in a flexible manner. The document is exhaustive regarding the elements describing the footprint geometry but it is deliberately open to customization for application specific features related to the presentation layer. In particular, it is not the purpose of this document to replace graphic formats such as Scalable Vector Graphics [20]. It does, however, give means to link existing visualization capabilities of legacy tools to selected footprint elements.

7.2 WCS Augmented Graphics

Usual graphics formats as JPEG, PNG or GIF appear as quite interesting for data publishers as they allow to deliver lightweight color images, sufficient for quick-look or preview visualization. But they lack a metadata container mechanism able to store WCS information, crucial in astronomy since allowing to link a (x,y) position on the image to a (alpha,delta) position on the sky. We have studied 2 approaches allowing to associate the WCS info with those images.

SIA response with WCS for Graphic Files. The idea is to carry the WCS information in the SIAP query response. Here is a list of the needed FIELDs and the corresponding FITS WCS keyword :

```
ucd="VOX:WCS_CoordRefPixel"              CRPIX
ucd="VOX:WCS_CoordRefValue"              CRVAL
ucd="VOX:WCS_WCS_CDMatrix"               [CD matrix values]
ucd="VOX:WCS_Image_Naxis"                NAXIS
ucd="VOX:WCS_CoordProjection"            [projection type]
ucd="VOX:STC_CoordEquinox"
ucd="VOX:STC_CoordRefFrame"
```

Then, when an image is loaded, the client can easily build the astrometry from the value of those fields. A trial implementation has been developped by CDS and ESO : ESO has built a prototype SIA service delivering images along with their WCS information, and CDS updated Aladin so that it can consume this service.

This solution will work with any image format, as long as it is supported by the client. Drawback : WCS information is separated from the data file.

Graphic files with embedded WCS. In order to keep the astrometry information in the same file as the data, we tried to integrate FITS keywords and values directly in the comment field of JPEG files. It has been successfully implemented and validated in the Aladin image server (able to produce JPEG files with embedded WCS) and in the Aladin client (able to consume such files). This development is compatible with SDSS initiative to deliver JPEG images with WCS information. In other words, Aladin reads and supports WCS information coming with SDSS JPEG files.

7.3 Messaging Protocols

The PLASTIC protocol for inter-application messaging has become popular, and various applications both within and without VOTech have now been enhanced with PLASTIC facilities, enabling a substantial degree of transparent interoperability between tools on the astronomer's desktop. Efforts have been made to progress to an IVOA standard for applications messaging. These have taken the form of vigorous discussions within the IVOA Applications Working Group to specify a successor protocol to PLASTIC with support from all the IVOA member organizations. The outline of such a protocol, tentatively named Simple Applications Messaging Protocol, has now been agreed, and a draft standard document, as well as infrastructure and applications which implement it, are expected before October 2007.

7.4 Ranking Query Result Sets

The Problem. The IVOA Simple Spectrum Access Protocol [SSA IF] introduces the concept of a scoring mechanism to rank matching records by relevance. In collaboration with international partners an IVOA note [2] has been released which provides guidelines for design and implementation. Sample code is available from the VOTech download page.

Design Considerations. In the context of an SSA service one can view the scoring method as an optional sub-service that can be switched on and off without affecting the core functionality.

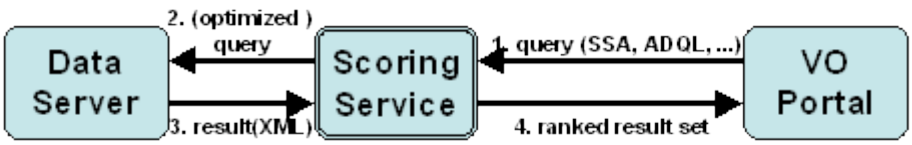


Figure 7: Service achitecture.

A ranking mechanism may gather input from a number of sources (Figure 7).

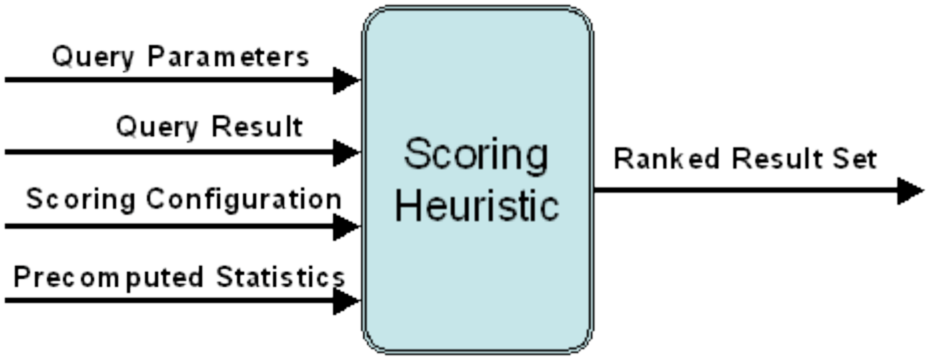


Figure 8: I/O of a scoring heuristic.

Most importantly there is the actual set of query parameters. Then there is the query result document including the score column that gets computed by the ranking system (Figure 8).

```
<FIELD name="score" datatype="float" ucd="stat.likelihood"
utype="ssa:Query.Score"/>
```

Figure 9: Score column definition in a VOTable V1.1 with UCD 1+.

There is also a priori knowledge which may have various origins such as:

- domain expert (archive scientist)
- pre-computed statistics on the data collection
- worklog captured from previous user sessions (feedback loop)

One possibility to capture knowledge of a domain expert is to apply global weights to each supported query attribute and to define a measure to normalize the proximity of a query and matching records.

The algorithm should be modular such that the total score is a sum over contributing attributes or related groups of attributes (Figure 10).

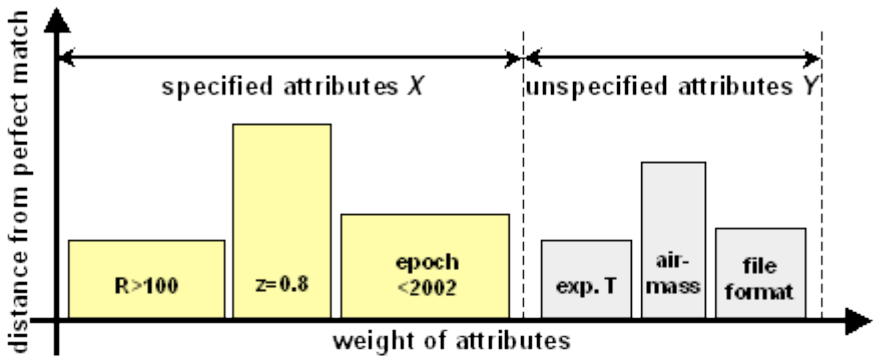


Figure 10: Matching a given tuple with some given query constraints. The score is indirectly proportional to the integrated area of the boxes: Each attribute has got a weight (width) whereby wider means more important. User specified parameters X are generally more relevant than 'guessed' or unspecified query parameters Y which may be ignored completely in a simple implementation. The smaller the distance from a perfect match (vertical axis), the better.

Note: Normalizing the total score is not recommended as this would create the impression to the end user that the score is absolute and could be used to compare the score of different services; in reality the score is strictly relative to a single query from a single service. A client application may always choose to normalize later on.

The general problem can be broken down into patterns for certain classes of DB attributes.

The design should allow for shortcuts when a query yields a trivial result like for instance no matching records, or everything matches or there is one perfect match. The bigger the result set, the more relevant is the score. Hence, performance does matter.

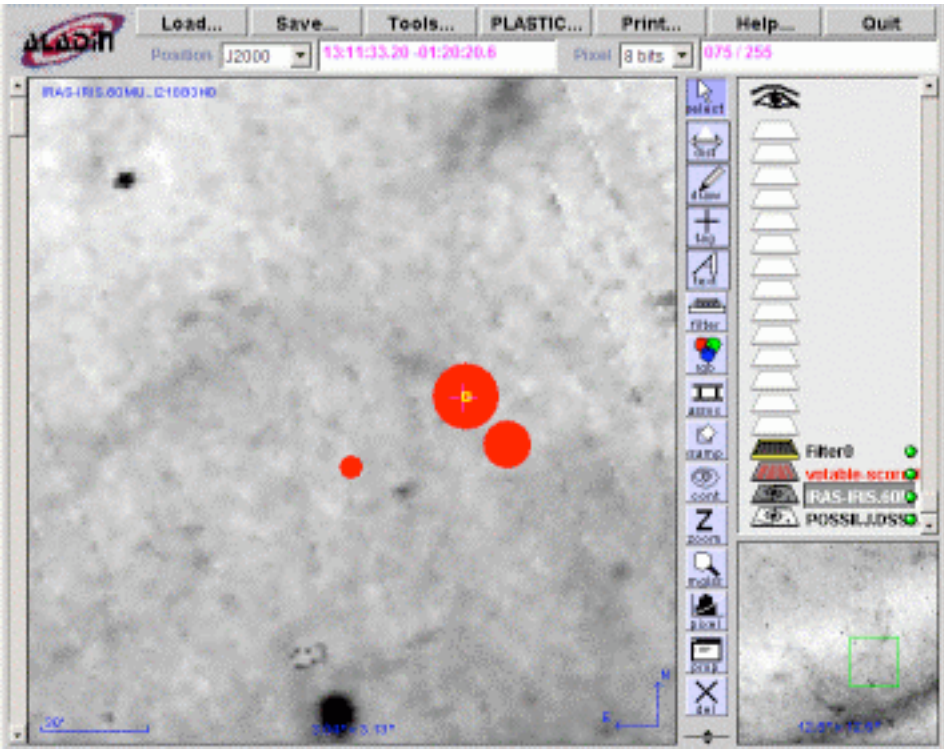


Figure 11: Red markers visualize the score. The larger the radius the higher the probability that a given observation is relevant to a user defined query.

8. Dissemination

8.1 Software

- download (prototype) software [1]: http://wiki.eurovotech.org/twiki/bin/view/VOTech/SoftwareDownload

8.2 Scientific Events & Technical Workshops

The following scientific and technical events were supported.

- http://wiki.eurovotech.org/twiki/bin/view/VOTech/IVOAHands on exercises and software demos were set up with other EURO-VO partners for the occasion of the Workshop on how to publish data in the VO, 25-29 June 2007 at ESAC, Spain.
- DAA International Workshop (http://www.math.unipa.it/~daa_erice07/) 15-22 April 2007 Erice -Italy
- "VObs.it – Osservatorio Virtuale Italiano" Workshop (http://vobs.astro.it/index.php?option=com_content&task=view&id=17) 23-24 november 2006 Rome, Italy.
- Exhibition booth at the General Assembly of the International Astronomical Union (IAU); set up on behalf of and in coordination with the 16 IVOA partner projects, 14-25 Aug., 2006, Prague, Czech Republic.
- The conference pages http://www.ivoa.net/pub/VOScienceIAUPrague/ of IAU special session no. 3 "The Virtual Observatory in action: New science, new technology, and next generation facilities" were designed and hosted.
- Hands on exercises and software demos were set up with other EURO-VO partners for the occasion of the Workshop on VO Standards and Systems for Data Centres and Large Projects, 27 June – 1 July 2005 at ESO, Germany.
- 'Journée briques logicielles OV France' 19 December 2006 http://vo.obspm.fr/seminaire/journee-logiciel-programme.html

8.3 Publications

The following are publications wholly or partly due to VOTECH work.

- M. Leoni, M. Dolensky, et al., 2006, Multi-Purpose Metadata Repository for a Real & Virtual Ovservatory, ASP Conf. Ser., 351, 414
- R. Slijkhuis, M. Dolensky, et al., 2006, Feeding VO Data Products into the ESO Archive, ASP Conf., Ser., 351, 425
- N. Delmotte, M. Dolensky, et al., 2006, ESO Science Archive Interfaces, ASP Conf. Ser., 315, 690
- P. Manzato et al. 2007 "An archive of cosmological simulations and ITVO multilevel database" Proceedings book by the World Scientific Press,

Singapore
- M. Comparato et al. 2007 "Visualization and data mining in the Virtual Observatory framework" Proceedings book by the World Scientific Press, Singapore
- T. Boch, P. Fernique et al., 2006, Aladin, a portal for the Virtual Observatory, SF2A-2006, Proceedings of the annual meeting of the French Society of Astronomy and Astrophysics
- M. Dolensky et al., 2006, Status of the VOTech Design Study about User Tools, ASP Conf. Ser., 351, 410
- F. Bonnarel, T. Boch, P. Fernique, Aladin, VO standards, Spectra and data cubes, Astronomical Spectroscopy and the Virtual Observatory Workshop 2007
- L. Cambrésy et al., 2005, ALADIN multiview, Proceedings of the 2005 meeting of the French Society of Astronomy and Astrophysics

9. Recommendations

Below recommendations are very much in synch with the goals of the IVOA roadmap 2007 [21]. This should not surprise given the overlap of contributors to the two documents and the fact that technical aspects are actively aligned with comparable global efforts of the VO- and Grid-communities.

Data Access Layer. Data access protocols should be re-considered in terms of their ability to handle multiple requests and bulk data. For instance, SIA services are limited to single requests or short lists. The protocols need to be expanded to support also the more complex data set types as produced by modern instrumentation. This includes 3d cubes, integral field unit data and in general it requires a container concept for associating science products, calibration items and ancillary data such as quality control plots and documentation. Access protocols need also support user authentication, logging and asynchronous operations. The respective design of the SIA version 2 protocol has been initiated.

Theory Data. It will be important to agree on a standard format similar to FITS in the observational Astronomy. HDF5 (Hierarchical Data Format) could play this role for output from simulation code. Another interesting approach is to structure cosmological simulation code according to the SNAP Data Model schema. The SNAP protocol should be the standard way for selecting subsets of data from big simulations which are otherwise too big to download and where cutting out sub-zones would limit the possibilities of subsequent analysis.

Interaction between VO and analysis software.Great care should be taken about studying how popular analysis softwares used daily by astronomers could be interfaced with VO tools and VO services. If those interfaces do exist, it will greatly facilitate VO take-up, as scientists will be able to access VO facilities from software they are used to.

Bundling of astronomy software such as the Scisoft and ex-Starlink collections provides a convenient way of distributing many packages at once to ease the burden of installation. Bundled distributions of VO software would assist in up-take of VO tools.

Applications Messaging. Most astronomers do most of their work with software packages like IDL, IRAF, DS9, MIDAS, Sextractor, etc. It is highly desirable that these be interoperable with the VO framework through use of VO services and desktop messaging.

Therefore, application authors are encouraged to provide interoperability with other tools using the existing PLASTIC protocol or, in due course, its successor(s) as specified by the IVOA Applications Working Group. The experience of tool authors has been that adding such functionality is not a great burden and can lead to considerable improvements in usefulness by providing access to facilities from other tools - the whole application suite can thereby become greater than the sum of its parts. The loose coupling used in PLASTIC messaging means that these gains can be realised without the, perhaps impractical, requirement for close coordination or collaboration between

Physical Units.Most scientific quantities carry units, and data returned by VO services should also carry explicit unit information when not clear implicitly. Units should follow the IAU recommendation, and/or the VOTable recommendation. As trivial as this may sound it has to be understood that different services may store and return data in different units and that automated conversion when collating results can have side effects. One can distinguish two cases: When a user makes a query based on a quantity, units can either be user-defined or fixed. In the former case, the user has the freedom to express the quantity in arbitrary units or an enumerated choice. In the case of fixed units, the data model of the query is bound to specific units, e.g., all angles must be in decimal degrees.

Towards SED Building.Progress towards SED building is a particular challenge due to the different nature of observational data being combined. One requirement is to record the history of instrumental setups, for instance, by systematically logging optical filter changes. Whereas this is usually done for space missions it is harder to keep track on the ground where support engineers and observers can lay their hands on the hardware. The project started an effort to define photometric systems and filter curves in a generally agreed format. In order to make also historic object catalogs 'SED worthy' most used and popular filters & photometric systems need to be identified. A programmatic and heterogeneous access method needs to be established. The project recommends the to adoption of the IVOA spectral data model for the electronic transfer. This will allows code reuse of spectral access services and to naturally evolve their functionality.

10. References

#	Paper or On-line Resource
[1]	VOTech Software: http://wiki.eurovotech.org/twiki/bin/view/VOTech/SoftwareDownload
[2]	Dolensky M., Rino B., IVOA Note: Ranking Query Result Sets V1.00, 2006, http://www.ivoa.net/Documents/latest/Ranking.html
[3]	CDS's Unit conversion library: http://cdsweb.u-strasbg.fr/cdsdevcorner/units.gml
[4]	Exposure Time Calculators: http://www.eso.org/observing/etc/
[5]	Mag2flux interface: http://vo.eso.org:8080/mag2flux/
[6]	Healpix software: http://healpix.jpl.nasa.gov/
[7]	Sky Event Reporting Metadata (VOEvent), Seaman R., Williams R. et al., 2006, http://www.ivoa.net/Documents/latest/VOEvent.html
[8]	AVO Science Reference Mission: http://www.euro-vo.org/internal/Avo/AvoSRM/srm.pdf
[9]	Astrogrid's Ten Science Drivers http://wiki.astrogrid.org/bin/view/Astrogrid/ScienceProblems
[10]	Domain model for simulations, their results and post-processing, G. Lemson et al. 2007 http://www.ivoa.net/internal/IVOA/IVOATheorySimulationDatamodel/SimulationDomainModel.doc
[11]	Simple Numerical Access Protocol (SNAP) for theoretical data, C. Gheller et al. 2007 http://www.ivoa.net/internal/IVOA/IVOATheorySNAP/Snap-0.5-1.doc
[12]	Greisen & Calabretta, A&A, 395, 1077-1122, Representations of spectral coordinates in FITS
[13]	Budavari et al., Johns Hopkins University Footprint Service, http://www.voservices.net/footprint/
[14]	Bonnarel et al., DAL query Response with Extensions: Use cases and implementation rules. Example of SIAP, http://www.ivoa.net/Documents/latest/DalResp.html
[15]	Tody et al., Simple Spectral Access Protocol V1.01, interface definition, http://www.ivoa.net/Documents/latest/SSA.html
[16]	Tody & Plante, Simple Image Access Specification V1.0 or higher, http://www.ivoa.net/Documents/latest/SIA.html
[17]	Ochsenbein et al., VOTable Format Specification V1.1 or higher, http://www.ivoa.net/Documents/latest/VOT.html
[18]	Rots, Space-Time Coordinate Metadata for the Virtual Observatory, V1.30, http://www.ivoa.net/Documents/latest/STC.html
[19]	Louys et al., Data Model for Astronomical DataSet? Characterisation V1.11, http://www.ivoa.net/Documents/latest/CharacterisationDM.html
[20]	Scalable Vector Graphics 1.1 Specification, http://www.w3.org/TR/SVG/
[21]	The IVOA in 2007: Assessment and Future Roadmap, Williams et al., http://www.ivoa.net/internal/IVOA/TechnicalMilestones/IVOARoadMap-2007-final.pdf

11. Glossary

Acronym	**Description**
ADQL	Astronomy Data Query Language
ASR	Astrogrid Server Runtime library
AVO	Astrophysical Virtual Observatory
CADC	Canadian Astronomy Data Centre
CDS	Centre de Données Astronomiques de Strasbourg
CEA	Common Execution Architecture
CGPS	Canadian Galactic Plane Survey
CME	Coronal Mass Ejection
DCA	Data Centre Alliance
DS4	Design Study 4 - (New) User Tools
ESO	European Southern Observatory
FITS	Flexible Image Transport System
FTP	File Transfer Protocol
FORS	FOcal Reducer and low dispersion Spectrograph
HDF5	Hierarchical Data Format
HEALPIX	Hierarchical Equal Area isoLatitude Pixelization
IDL	Interactive Data Language
INAF	Istituto Nazionale di Astrofisica
ISAAC	Infrared Spectrometer And Array Camera
IVOA	International Virtual Observatory Alliance
MEF	Multi Extension FITS
PLASTIC	PLatform for AStronomical Tool InterConnection
REST	Representational State Transfer
SED	Spectral Energy Distribution
SIAP	Simple Image Access Protocol
SINFONI	Spectrograph for INtegral Field Observations in the Near Infrared
SNAP	Simple Numerical Access Protocol
SOFI	Son of ISAAC
SQL	Structured Query Language
SRM	AVO Science Reference Mission
STILTS	Starlink Tables Infrastructure Library Tool Set
TVO	Theory VO
UCD	Unified Content Descriptor
UWS	Univeral Worker Service
VIMOS	VIsible MultiObject Spectrograph
VO	Virtual Observatory
WCS	World Coordinate System
WFI	Wide Field Imager
WFCAM	Wide Field Camera