VO-Tech Science Framework Document

The European Virtual Observatory - VO Technology Centre

A DESIGN STUDY

implemented as

SPECIFIC SUPPORT ACTION

Editor: Nicholas Walton
Institute of Astronomy
The University of Cambridge, UK

on behalf of the VO-Tech Science Team

Contract Number: 011892

Project Website: http://www.eurovotech.org

Table of Contents

Executive Summary....4
Introduction & Context....4
The VOTC Context within the Euro-VO....4
Euro-VO Data Centre Alliance....4
The Euro-VO Facility Centre....4
The Euro-VO Technology Centre....5
Science Environment and Key Input Drivers....5
Key Current Science Priorities....5
Key Current and Future Astronomical Missions with European Leadership....6
European Wide Science Horizons....6
The AVO Science Reference Mission....6
The AstroGrid Phase A Study and 'Top Ten'....7
Euro-VO Community Drivers....7
The VOTC Science Framework....7
Science Benefits....7
Science area....7
Scale of computational use....8
Expected number/type of users....8
Individual Astronomers....8
Small to medium size collaborations....8
Project Teams....9
Types of Science....9
Key Science Themes....10
Exemplar Science Themes....10
ExtraGalactic Case: Clustering of clusters (of galaxies)....10
Background....10
Required Data and Applications....10
VO Workflow....11
End Results....12
The Initial Mass Function of massive stars....12
Background....12
Data Requirements....13
Theoretical Models....13
Tools for data manipulation and comparison....13
Initial investigations....14
Future Developments....16
Individual Objects : 3C295 and its cluster....18
Multi-wavelength data....18
Spectral Energy Distribution of 3C295....18
Spatially resolved Radio Spectral Index and X-ray properties....19
Identification of neighbouring galaxies by colour....20
Detailed spectral analysis....22
Time Domain Cases....22
Requirements extracted from the AVO SRM....23
Data....24
Tools....25
Technical Demands....26
Science focus on technical capabilities....26

Operational focus on technical capabilities 27
Acknowledgements...27
Appendices and Links 27
Glossary 28

Executive Summary

This document sets out a number of science cases which have been considered in scoping the VOTECH activities. The document also notes initial prototype solutions utilising emerging VOTECH Virtual Observatory technologies which have been used for a number of the VOTECH science cases.

Introduction & Context

The VO-Tech project is a European Commission funded, Framework 6 programme, Design Study which aims at completing all technical preparatory work necessary for the construction of the European Virtual Observatory (Euro-VO). The concept of the Virtual Observatory (VObs) is that all the world's data should feel like it sits on the astronomer's desk top, analysable with a user selected workbench of tools, made available through a standard interface. Internationally this is set to transform and re-structure the way astronomy is done. Euro-VO is a specifically European implementation of this idea, and will produce a world leading infrastructure providing a unified virtual data resource and the ability to perform complex data discovery and manipulation tasks across the whole range of astronomy. Access to data and tools will be equally good across Europe, regardless of location.

This document lays out a number of key current and emerging science drivers which are defining the scope and capabilities required to be considered and defined by the VO-Tech project for eventual deployment Euro-VO wide.

This document is scoped to be relevant to a number of key stakeholders of the Euro-VO and VOTC

- The Euro-VO Science Advisory Committee
- The VOTC/ VOTECH project teams
- A summary for the Euro-VO community of the key science drivers behind the VOTC.

The VOTC Context within the Euro-VO

This section describes the current structure of the Euro-VO and the relationship of the VOTC to the other constituent components of the Euro-VO.

The Euro-VO is formed of three main components, all being distributed activities, which together create the Euro-wide Virtual Observatory in Europe.

Euro-VO Data Centre Alliance

The Data Centre Alliance (DCA) provides coordination for the main European resource centres who will provide much of the content for the Euro-VO. The DCA project is currently funded through a EU Framework 6 [Communication Network Development](#) programme - see http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/WebHome for details.

The Euro-VO Facility Centre

The Euro-VO Facility Centre (VOFC) provides a focus for scientific operations of the Euro-VO, including acting as a resource for end user astronomers. The Euro-VO portal can be found at http://www.euro-vo.org

The Euro-VO Technology Centre

The Euro-VO Technology Centre (VOTC) provides the technical infrastructure which underpins the Euro-VO operational system. The VOTECH project currently (2005-2008) funds much of the VOTC activity.

Science Environment and Key Input Drivers

This document sets out a number of key drivers (astronomical research programmes) as determined as priorities by the member communities of the Euro-VO, taking note of the key science underpinning the development of major new astronomical missions such as the Extremely Large Telescope (ELT), ESA's GAIA and the Square Kilometer Array (SKA).

To a significant extent these define the upcoming 'science horizon', the key science programmes together with the key science data streams and applications which the Euro-VO technical infrastructure, when deployed Europe wide, must be capable of supporting in a relevant fashion.

This section lists a number of pre-existing science drivers for the VO which have been generated by pre-cursor Euro-VO programmes, or by the national VO projects which together will participate in the Euro-VO. Additionally a number of other key European wide initiatives which scope the Euro-VO landscape are given.

Key Current Science Priorities

The VO must be able to deliver real capability to support those science programmes that the European community has identified as of high importance.

For instance, in Europe, the Astronomy community has identified the following in the ESA 'Cosmic Vision' plan:

1. What are the conditions for life and planetary formation?
2. How does the Solar System work?
3. What are the fundamental laws of the Universe?
4. How did the Universe originate and what is it made of?

Key Euro-VO science themes are linked to these, with the science requirements from those, in turn needing to be met by the Euro-VO technical system and service. Major ground-based projects such as ALMA, ELT and SKA have adopted similar science drivers.

An example of where VOs already offer support is in the identification of protoplanetary disc candidates and their environments:

1. Some Spitzer, ISO, JCMT and other FIR/sub-mm images and spectra are now accessible to VO tools
2. Interoperable packages for handling and modelling or comparing spectra include TopCat, SPLAT and VOSpec.
3. The IVOA Spectral Line Access Protocol, implimented in VOSpec, allows identification of atomic and molecular lines by reference to several international databases (including NIST).
4. Multi-wavelength data from catalogues or via a source extraction tool can be used to

construct spectral energy distributions in order to identify candidates (see example Stellar case below for more details of the processes).

5. Basic black-body fitting is already available for any temperature. Mocassin (Ercolano 2005MNRAS.362.1038E) is an example of a VO accessible application which provides models suitable for dusty discs.

Key Current and Future Astronomical Missions with European Leadership

It is clear that the Euro-VO must be able to support the science processes involved in the exploitation of those facilities and missions which have significant or exclusive European leadership and involvement. These include (the list being non-exhaustive):

- Current:
  - ESO-VLT, Gemini, WHT, MERLIN, VLBI, XMM, UKIRT
- Near-Future: 2007-2010
  - Vista, VST, e-MERLIN, Grantecan, Planck, Herschel, ALMA, Lofar
- Future: >2010
  - ELT, GAIA, SKA

Liason with these major projects is occurring across the Euro-VO. The VOTECH project has been considering technical issues associated with the support of these major facilities. The Euro-VO DCA activity is working with the major data providers in informing them as to the technical issues related to deploying the VO infrastructure.

It is noted that the Euro-VO plays a significant partnership role working with these major new facilities such as UKIDSS, VISTA, LOFAR, e-MERLIN, ALMA and various ESO projects which have already committed to making their science data available to the community via the VO. There is already (2007) some progress in also making tools and elements of the data flow systems from these missions available through the VO.

European Wide Science Horizons

A number of key initiatives are currently under-way with the aim of setting the agenda for a unified roadmap for astronomy in Europe. These include the EU suppported AstroNet initiative and the ESF roadmap of key facilities in europe - ESRI. The Euro-VO/ VOTECH is working to include these roadmaps in its technical planning, in particular is providing input in the generation of the AstroNet infrastructure roadmap.

The 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 for the European Astrophysical Virtual Observatory VO, the AVO Science Reference Mission (SRM)[AVO-SRM]. These science cases have been broken down to set requirements on elements of the Euro-VO. An analysis of the needs expressed in these cases has been carried out within the DS4 Tools area of VOTECH:

Science Reference Mission case and resulting tools requirements.

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

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

The AstroGrid Phase A Study and 'Top Ten'

The UK based AstroGrid project generated a set of key science use cases as a component of its Phase-A stage - laying the foundations for the construction of its VO system. These can be found at The AstroGrid Ten key science drivers. The analysis of these use cases has been incorporated into this work.

Euro-VO Community Drivers

The Science Advisory Committee (SAC) of the Euro-VO was consitituted in 2006. It has a remit to provide science advice to the Euro-VO project as a whole. A number of the science use cases outlined in this document have been discussed at Euro-VO SAC meetings.

The VOTC Science Framework

Science Benefits

This section comments on the various types of science users and communities that will make use of the Euro-VO systems, and the issues these raise including benefits.

Science area

- individual analysis of objects on a per object basis: the VO should offer the possibility to repeat for a significantly larger number of objects.
- time domain astronomy, e.g. SWIFT, LSST, Solar Physics. Here the VO must enable access to temporal data, in some cases 'real-time' access to data must be facilitated.
- multi-scale astronomy, e.g. Cosmic Microwave Background research. The VO must provide

access to data over all spatial scales upto and including all sky maps.

Scale of computational use

- desktop scale
- medium scale scripted services
- large scale heavy computing: e.g. cosmological simulations

Expected number/type of users

The VOTC must develop a system suitable to support the research activities of the European astronomical reseach communities. In terms of number of active astronomers this figure is <10,000 potential users,

It is noted that some capabilities developed through the VOTC may be of more general interest to the wider public - thus the potential for usage of the system could be much greater.

Individual Astronomers

Individual astronomers will be able to use VO tools and services in much the same way as they use current astronomy tools. Existing astronomy tools and services are becoming VO aware, allowing VO use via known and familiar interfaces but with significant increases in power and capability enabled by VO systems. New and innovative tools are being developed to provide new and previously impossible capabilities. As such, via use of VO enabled astronomy tools, or new VO built tools we can expect ALL astronomers to be scientific users of the VO.

The main benefits for individual astronomers will be interoperability, ability to use large data volumes, and dramatically increased information and data content. Interoperability allows the diverse tools and different types of data to be easily combined for scientific analysis, for example combining multi-wavelength data to extract SEDs and colour maps. High volume VO data storage that works as simply as local files will allow processing of survey size data sets. Information published via VO registries opens up new possibilities for data searching and will be used by astronomers in a multitude of ways.

Individual astronomers will expect maximal flexibility in using data and services. Scientific analysis requires cross checking and validation of results via various methods, and often using techniques outside the scope of standard tools. As such the ability to apply custom written algorithms in VO tools, or to connect user code via simple VO interfaces will be of great interest for rigorous use of VO tools.

Small to medium size collaborations

Much Astronomy is done in small to medium size collaborations of a few to perhaps 10 members spread over a number of institutes and countries. These collaborations are driven by a common scientific project and usually involve data sets or models shared by the members, with the focus on publishing scientific results in journals. VO systems promise to provide simple ways of sharing data within a collaboration, and of publishing detailed scientific products associated with journal papers.

Important issues will be ease of use, and how the data are managed and hosted on VO systems including access control. Small to medium sized collaborations often have little technical support so making data available on the VO with minimal curation effort will need to be easy. There is also likely to be a need for VO hosting of collaborative projects at the level of national VOs.

Project Teams

The VO will be of importance to scientific project teams of surveys and large observing programs in terms of generating, delivery and use of the scientific products. Pipeline processing of large project data streams requires many scientific (as well as technical) checks during processing. VO systems have a role to play here in providing external information using during processing, for example to flag known variable objects, or to compare calibrations against existing data. VO compatible archiving and publishing of scientific data will be a priority so that the user community may use the data.

Types of Science

VO systems are expected to be applicable to many if not all areas of astronomy because the underlying principles of data description and interoperability will make it possible to use many types of astronomy data with a wide range of tools and services.

The types of science where we expect immediate benefits are those that deal with large data volumes, require detailed interoperability, compare models to observations, and rapid time response astronomy.

The use of imaging surveys for cosmological investigations of large scale structure, stellar population studies, and rare object detection all require dealing with large volumes of imaging and catalogue data. VO systems that allow re-processing of imaging survey data, and cross matching of billion object catalogues will open up new capabilities currently not available. Algorithms and tools for rare object detection are already well developed, their application to large volumes of data via the VO will bring immediate benefits, and will be an essential part of analysing the future large sky coverage surveys.

Physical understanding of astrophysical objects requires observations across the entire electromagnetic spectrum as different radiative processes dominate in different wavebands. Making use of current and future multi-wavelength data will be made easier with VO tools because the data can be described with a common data model. Such interoperability between different types of data will greatly benefit studies of AGN which emit at all wavelengths. Detailed multi-wavelength studies of individual objects that are possible now will become feasible for large samples of objects, and combination of direct imaging and interferometry data should become routine. Studies of Galaxies, Planetary Nebulae, Supernova remnants and stars with disks, are also poised to benefit from multi-wavelength combinations of data for large samples of objects.

Astronomy where simulations and models and compared to observational data stand to benefit from VO systems. This is because the simulations and models are now often generated in a manner that allows the results to be 'observed' with virtual instruments, so that real and simulated observations can be handled in a similar way in the VO, facilitating their detailed comparison. Large N-body cosmology simulations such as the Millennium Run are prime examples of this scenario. Stellar models are a major resource for stellar population studies of star clusters and galaxies. Studies of this kind will be significantly helped with common interfaces to the models.

The VO will help make new discoveries about transient objects. VO systems provide the framework for rapid notification of transient events in the sky. Known information can quickly be obtained and combined to assess the importance of an event, and launch coordinated and immediate follow-up observations by dynamically scheduled, and/or robotic observatories.

Key Science Themes

A number of example use cases were decided upon as being representative of key VO areas of use. These are discussed in more detail in the following section, highlighting the issues that the VO will be required to address.

Exemplar Science Themes

1. large scale statistical problem: Clustering of clusters (of galaxies), Gal form from z=10 to z=0.1
2. mid scale : Initial Mass Function of massive stars
3. individual objects: 3C295 and its cluster e.g. focus on the Crab Nebula, Cass-A
4. temporal: SN pre-cursors

ExtraGalactic Case: Clustering of clusters (of galaxies)

Background

The formation and evolution of clusters of galaxies is an important key to understand the formation of large scale structure, the formation and evolution of galaxies, the reionization of the Universe and the origions of the metallicities in the intergalactic medium.

The abundance of clusters beyond z=1 will be a field of intense activity over the next years. Much of the emphasis will be on X-ray surveys with XMM (e.g. Romer et al., 2001), and the emerging field of Sunyaev-Zeldovich (SZ) survey astronomy (Holder et al., 2000). These surveys will cover large areas, tens of square degrees, and target redshifts beyond z=1.

Required Data and Applications

The following key data sets will be required to support this programme.

- WFCAM (IR)
- SDSS (Opt)
- XMM-Newton (Xray)
- Spitzer (mid- & far-IR)

The status (end 2006) of XMM-Newton programs aiming at large scale structure studies is as follows:

COSMOS

- 26 observations performed so far among them:
    - 6 are public
- 13 will be public before March 2006
- 7 will be public after March 2006
- Work done on clustering of clusters (Finoguenov et al, 2007) , and AGN (Miyaji et al 2007).
- Only the 1.4 Ghz catalogue for a pilot project is available (J/AJ/128/1974) and the K-band AO imaging is available (J/A+A/468/937). Images in several bands - X-rays included - are available

XMM-Newton Large Scale Survey (LSS)

- 29 public observations
- catalogue of X-ray sources in the central 1 squared degree published in Vizier (J/A+A/439/413/xmds1) - Paper published in Astronomy & Astrophysics (Chiappetti et al. 2005)
- 12 spectroscopically identified clusters at z<0.6 published by Willis et al. 2006. The catalogue of a larger and complete sample of clusters is still to come.
- catalogue of spectroscopically identified AGN expected to to published on longer timescales
- Spitzer data already available.

Others:

Southern Abell Redshift Survey (local universe comparison)

- Paper with catalogue of spectroscopy of 3440 galaxies in 39 galaxy clusters.

Applications

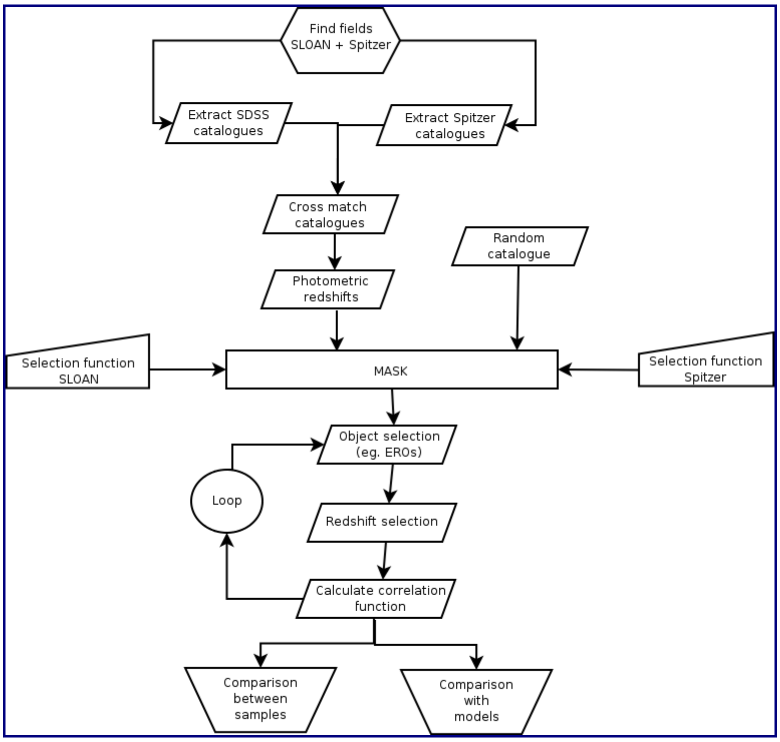
- SWARP, Sextractor, Cross Match, R, VisIVO, Aladin
- Cluster finding algorithm (Astroneural)
- Tools to calculate angular correlation function of the selected sources/clusters, inversion (VisIVO)
- Theoretical simulations tools (or use the datasets already available from e.g. White et al).

VO Workflow

The outline workflow that would be required to support this case is as follows:

- Find areas of sky covered by SDSS and Spitzer.
- For each of those retrieve the catalogues of objects from their catalogue servers or download the catalogues.
- Cross match the catalogues.
- Determine photometric redshifts.
- Generate a mask file from the coverage of the surveys, taking into account bad regions, areas not covered by both surveys, etc. (e.g. using Mangle (http://casa.colorado.edu/~ajsh/mangle)).
- Generate a random catalogue (positions and redshifts).
- Apply the mask to object and random catalogue.
- Now do different selection of objects, e.g. EROs, blue objects, QSOs, ...
- Do a second selection on redshift or/and luminosity.
- Calculate 2D correlation function. Also other correlation functions like 3D, sigma-pi would be useful.
- Do the calculation for the different samples.
- Compare results for different samples, compare with theoretical models of galaxy evolution.

A schematic view of the workflow is attached. Click on the image for a full resolution view.



End Results

This section notes the typical outputs from the above work processes.

Colour images of clusters selected using photometric redshifts. See description in this paper (Dahlen, Fransson and Naeslund, 1991, arXiv:astro-ph/0110419v1)

The Initial Mass Function of massive stars

Background

The Initial Mass Function (IMF) of the bright end of the mass scale is a crucial parameter for determining the upper limit of stellar mass and its connection to the environment such as metallicity. In order to determine the IMF, it is necessary to know the mass of the stars (as inferred from their position in the colour-magnitude diagram) and their absolute magnitude (i.e. distance). While the distance is not difficult to determine in extragalactic systems, for our own Galaxy the best way is to consider stellar associations, where distance and extinction is found for all stars in each system. High-mass star forming regions are identified by the presence of UCHII (ultra-compact HII) regions, molecular and spectral diagnostics such as the signature of hot dust. Very massive stars evolve so fast that supernovae and their rapidly-expanding remnants (SNR) are also indicators of violent starburst activity, especially in distant and obscured galaxies. For more detail see the original case from the AVO SRM.

This describes an initial approach using low-resolution Spectral Energy Distributions, followed by potential extensions using a wider range of resources and applications to other populations.

Data Requirements

- Locations and properties of Galactic open stellar clusters
- **Identification based on IR-optical-UV stellar characteristics**
  The simplest way to identify massive stars is by colour. F A B and O stars have temperatures from 8 000 to 30 000 K or hotter (see Colour Indices). This gives (approximately):
    - B-V < 0.4
- U-B < 0.0
- Optical/IR data up to U-band: SDSS, 2MASS and other catalogues now available using an AstroGrid science workflow which send a list of positions (e.g. clusters) as multiple cone searches to one or more catalogues. Catalogues also available via ColourCutter which selects data based on colour around a single position.
- UV photometric data with high spatial resolution are not yet easily available to the VO but access to VLT-UVES is expected early in 2007.

Theoretical Models

- **Theoretical stellar spectral energy distributions** are available from the Spanish Virtual Observatory (SVO) Theoretical Models Web Server. These can also be selected directly from inside VOSpec .

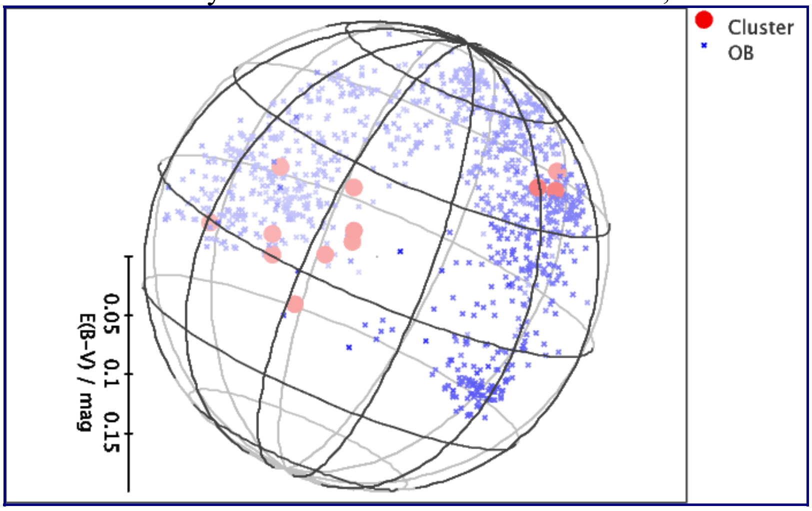
Tools for data manipulation and comparison

- The data processing requires:
    - Cross-matching of data in each region based on position;
- Substitution of effective wavelengths for filter band names;
- Conversion of magnitudes to physical units;
- Homogenisation of units (e.g. Jy to erg cm-2 s-1 A-1)
- Conversion of data formats such as IPAC format to VOTable;
- Transposition of tabular columns and rows in order to convert data ordered by object/position to data ordered by spectral coordinate;
- Comparison between observed data and models.
- Data selected using VO tools can be sent directly to catalogue and spectral manipulation tools such as TopCat and VOSpec using PLASTIC, or saved to MySpace and used in multi-stage workflows (see future developments below). For example, all operations performed inside TopCat can also be scripted using the STILTS library.

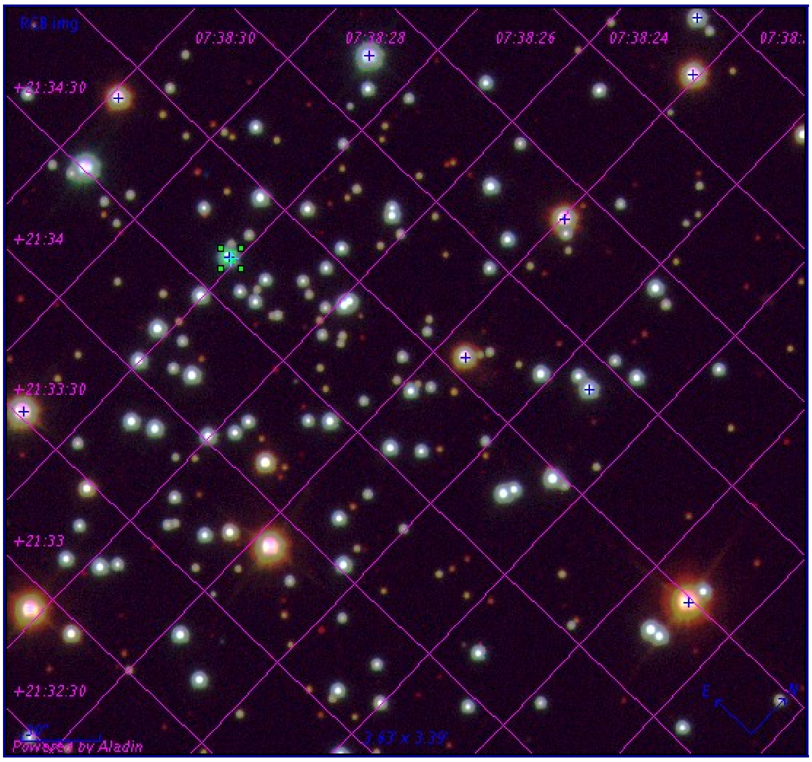
Initial investigations

This description details Euro-VO capabilities relevant to the demands of this use case:

1. *Optically visible open clusters and Candidates* (Dias+ 2002-2005) VII/229A contains 1697 entries. Select 12 clusters with
    1. Low reddening e.g. E(B-V) < 0.2
2. Distance > 500 pc
3. Angular size < 0.5 degree
4. Over 30 known members
    1. The clusters under investigation are shown in red along with OB stars with V < 9 measured by Yasuda+ 1986 I/249, shown in blue.



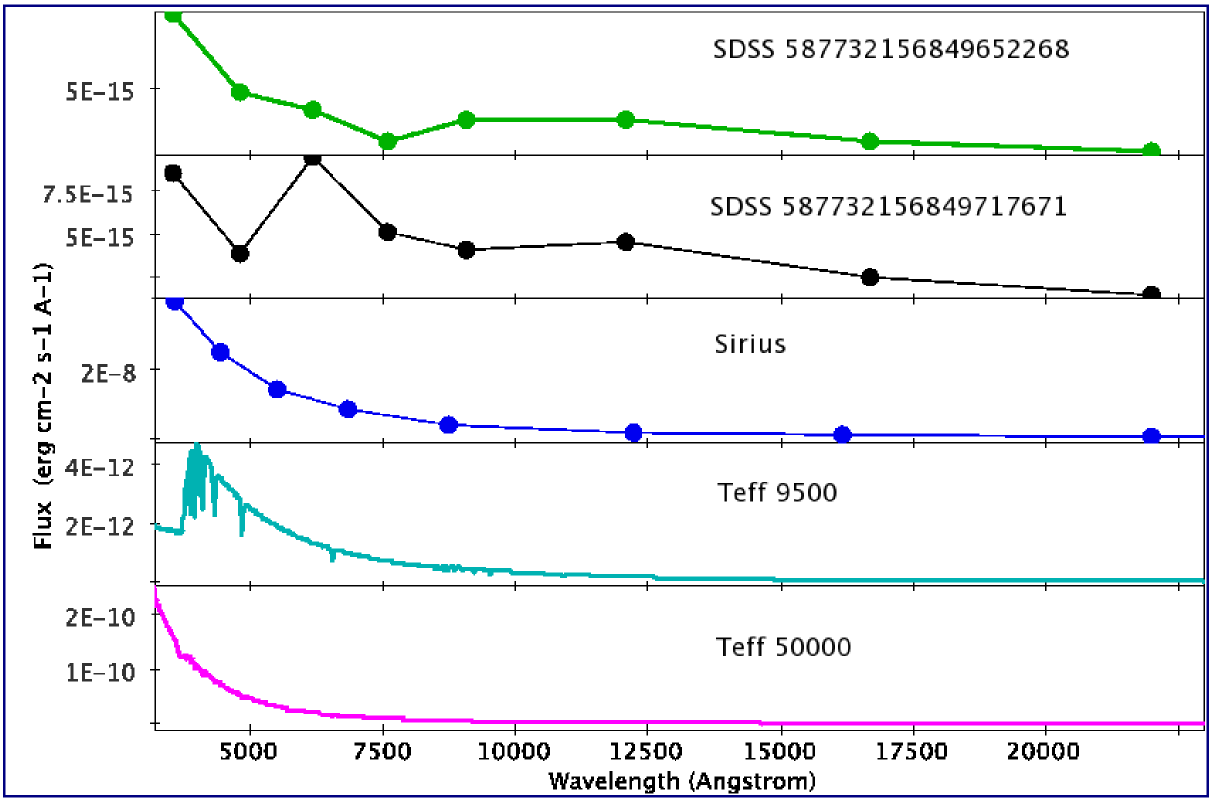
5. Utilise the AstroGrid Cone Search Science Workflow to search for SDSS and 2MASS data around each cluster. Further investigate one cluster covered by both catalogues, NGC 2420.
    1. Extinction correction: for NGC 2420, E(B-V)= 0.029 and thus E(U-B) = 0.72*E(B-V) = 0.021 (Turner 1989AJ.....98.2300T).
        - Extinction corrections can be applied automatically in VOSpec.
    2. Colour conversion: the relationship between the hot star seletion criteria given above and SDSS colours is given by
        - u-g = 0.75(U-B) + 0.77(B-V) + 0.72 (Jordi et al. (2006) astro-ph/0609121), leading to the extinction-corrected selection criterion
6. u-g < 1.02
    1. Use TopCat to make u-g and to select 37 stars in NGC 2420 with u-g < 1.02 (also selected as star-like using codes provided in the SDSS and 2MASS catalogues, with valid detections in all bands). The accumulated uncertainties (see references) are up to about 10%
        - The selection process could be scripted using the STIL library for VOTable manipulation.
7. Chose two bright hot objects, SDSS OBJID 587732156849717671 and 587732156849652268 (the star with the most extreme u-g is also faint (g<18) and is possibly a white dwarf candidate (von Hippel & Gilmore 2000AJ....120.1384V).
8. The hot star candidates are plotted as blue crosses on an SDSS image displayed in Aladin, constructed from u g and z images retrieved using the AstroGrid SDSS SIAP query workflow. SDSS 587732156849652268 is marked.



1. Converting the data to a suitable form for spectral display:
    - Use built-in functions in TopCat to convert the optical magnitudes to erg cm-2 s-1 A-1 (assuming input AB magnitudes). Corrections are not yet available for the 2MASS bands.
        - This could be scripted using the STIL library. Interactive conversion is also possible using VOSpec, based on the SSA protocol.
    - Insert the effective wavelength for each band.

9. Transpose rows and columns.
    1. Comparison with models:
        - Obtain similar data for Sirius for comparison.

10. Download models with hot star parameters from the Spanish-VO web site in VOTable format.
11. Use Topcat to display the stacked spectra for comparison.



Future Developments

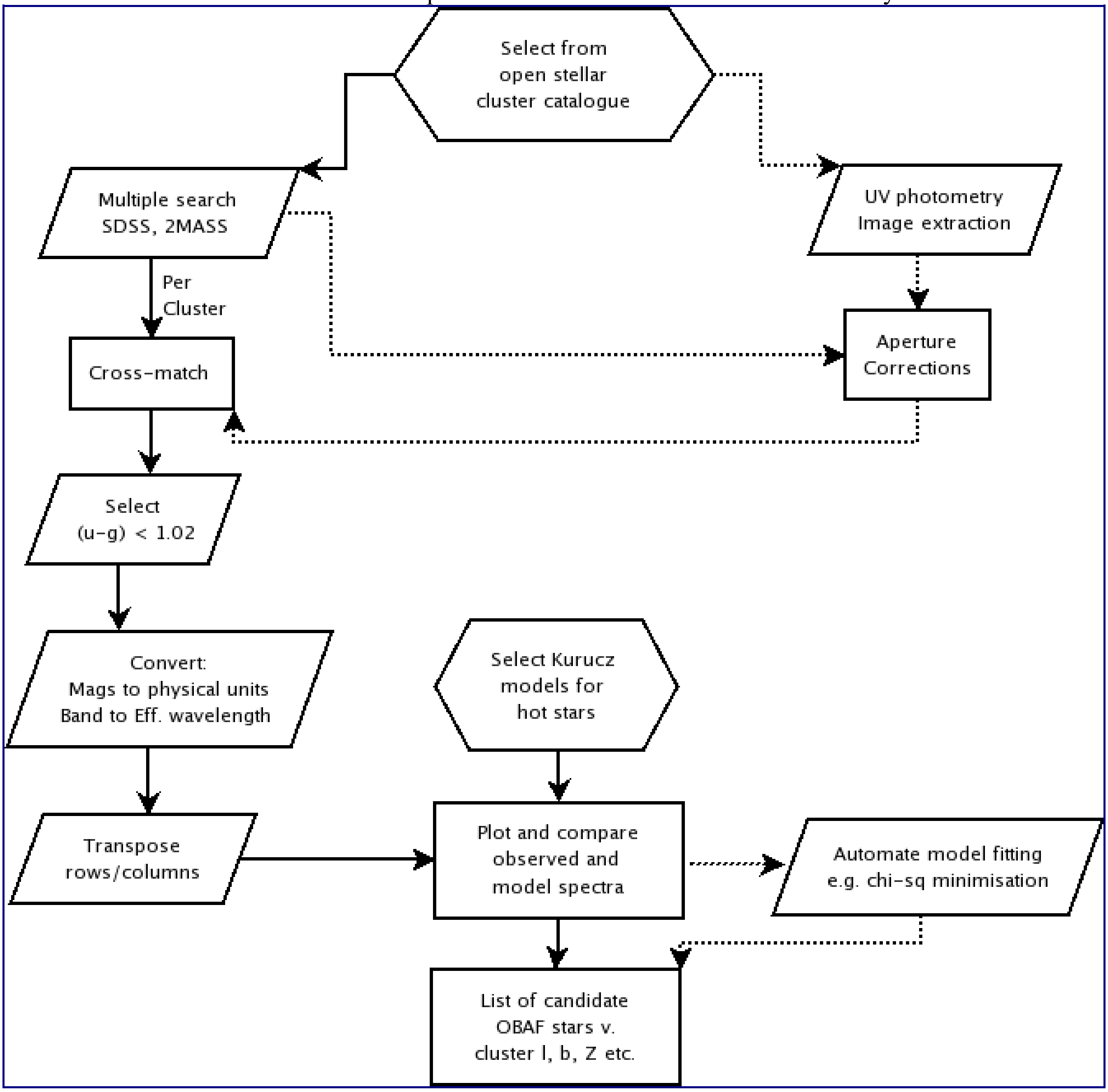
Requirements for further work include:

- Wider data selection for SED-building
    - Make more catalogues available to the Multiple Cone Search workflow. Needs improvements in the quality of archive registry entries and VO interfaces; possibly more efficient to search a box.
- Extract measurements from images as well as catalogues. An SED Builder tool is under development by Malapert et al. at ESO. This runs SExtractor on images to add to both low- and high- resolution photometry and spectra.
- Data from the ESO archive e.g. UVES and from new surveys like UKIDSS
- Unit conversions
    - At present, selected magnitude-to-flux conversions are available:
        - TopCat/STILTS provides exrepssions for AB and a few other optical bands;
- VOSpec provides generic conversions;
- An [http://vops1.hq.eso.org:8080/SedUI/][ESO]] tool provides accurate conversions for ESO instruments; this could eventually be generalised and PLASTICised
    - More standard functions would be useful such as converting
        - band labels to the effective wavelength (using look up tables);
- more band magnitudes to physical units, expecially in the IR;
- between physical units on the spectral and flux axes;
- X-ray units.
    - Using such expressions in workflows would be streamlined by accurate use of ucd1+
- Model fitting
    - A routine to fit spectral models to numerical data is being developed in STILTS. For each input spectrum, each model is sampled at each point from the observational

spectrum and scaled appropriately, and a goodness-of-fit figure is calculated. This can involve smoothing, use of supplied errors etc. and a choice of quantifiers including chi-squared minimisation.

- Extension to other cases
  - IMF and other properties of star-forming regions, see Gould Belt
- Stellar clusters in the Magellenic Clouds, M31 and other galaxies

This flow chart shows current steps with future additions connected by dotted lines.



For original plan see VotcSciImfMassiveStars

Individual Objects : 3C295 and its cluster

The properties of 3C295, a large FR2 radio galaxy located near the centre of a cluster forms a suitable use case. This makes use of packages developed by VOTECH and its member partners for:

- Data discovery (from Europe, USA and Japan)
- Interactive manipulation and exchange of data between different packages (PLASTIC)
- Extracting customised images and specialised selections from multiple catalogues using interfaces which do not require any specialised knowledge (Workbench, MERLINImager, ColourCutter)
- Automatic alignment and manipulation of images and manual astrometric correction (Aladin)
- Conversion of units including magnitude-to-flux (STILTS/TopCat)
- Spectral line identification, including a redshift correction (VOSpec).

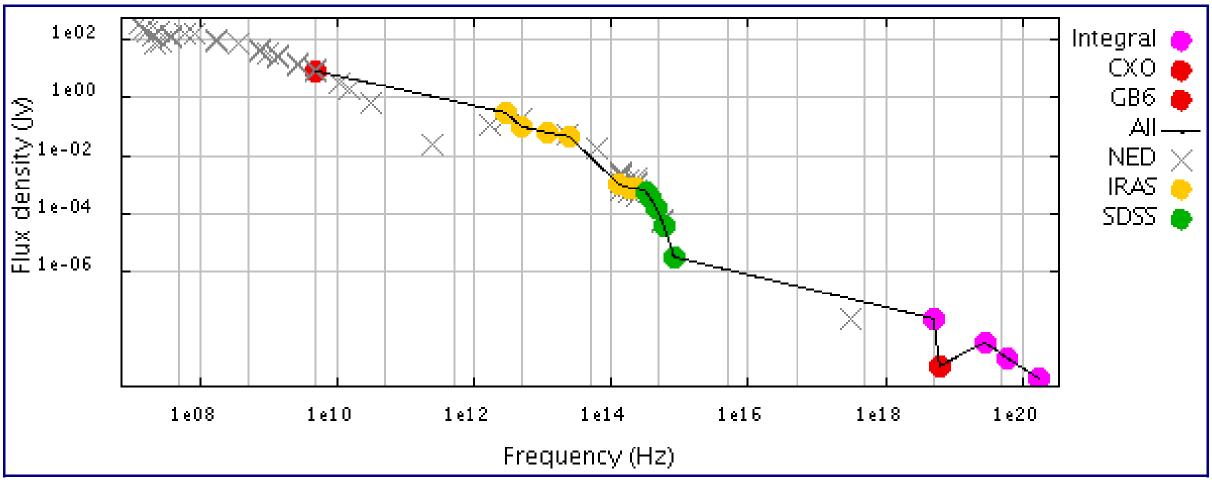
Multi-wavelength data

- AstroScope is used to find images, spectra and catalogue entries for 3C295.
    - Selected images are displayed in Aladin, spectra in VOSpec and tables in TopCat.
- This rapidly delivers a standard or static selection from each archive. These are in compatible formats for easy comparison.
- More specialised tools are available via the VO (such as the AstroGrid ColourCutter and MERLINImager) to obtain customised images and catalogue selections.

Spectral Energy Distribution of 3C295

- A selection of photometric catalogue entries from AstroScope was sent to TopCat and and then the built-in magnitude-to-flux converter is used to turn SDSS magnitudes into Jy. Also construct functions for the simpler transformations between other flux density and spectral axis units.

- In this test no sophisticated corrections were applied (e.g. for the photon index). Comparison of the resulting SED with the NED values, shows that, in many cases, the very simple transformations are adequate for identifying the main features of an SED over a large frequency range. The highest-frequency NED point is for the *Chandra* (CXO) 0.2-2.0 keV band; the point is for the full band. The CXO measurements are significantly fainter than the *Integral* points, probably due to the larger beam of the latter detecting more extended cluster gas, as seen in the images in the next Section.

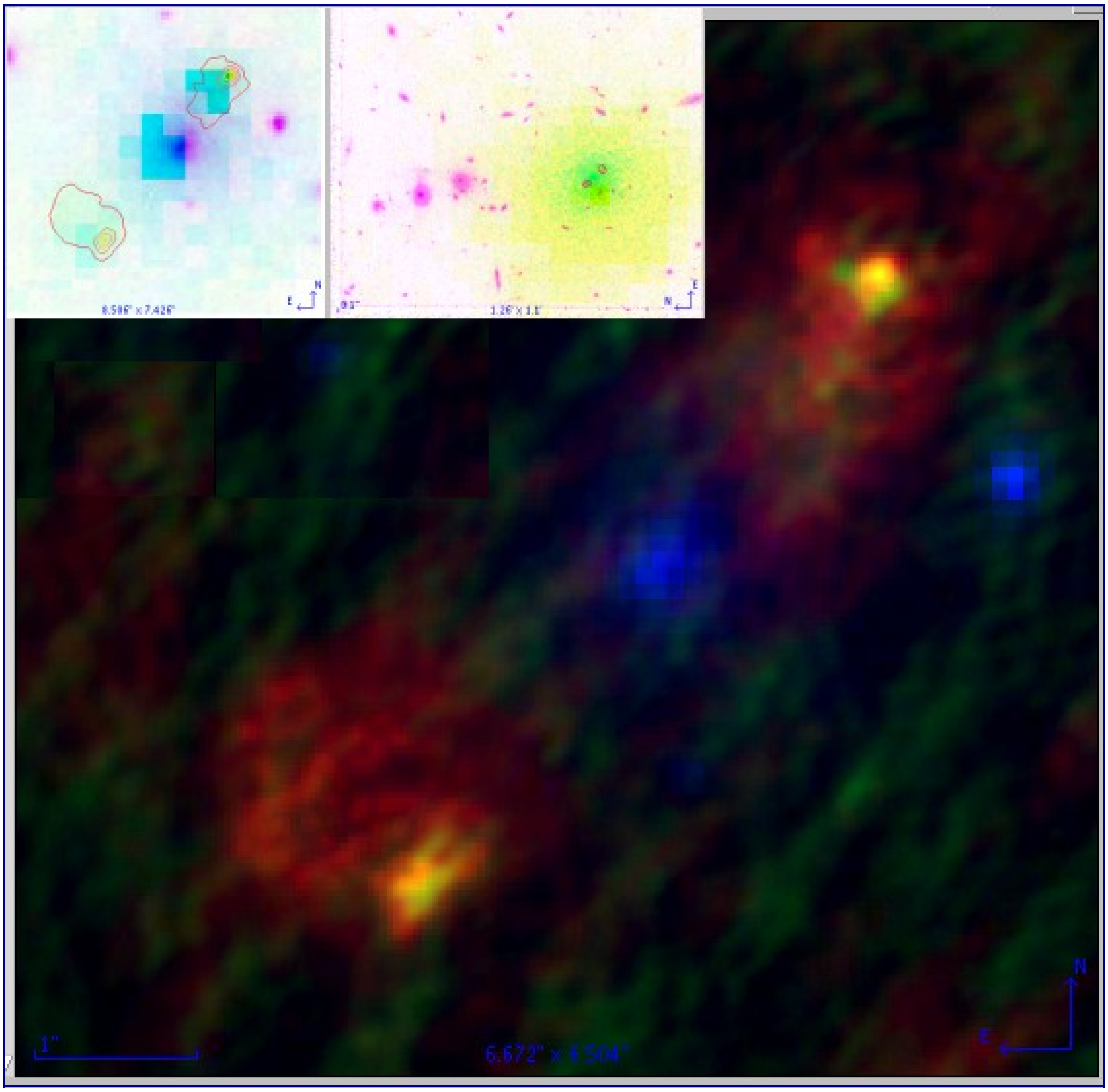


The Spectral Energy Distribution of 3C295

Improved methods for SED construction are described in the 'Hot Star' case above.

Spatially resolved Radio Spectral Index and X-ray properties

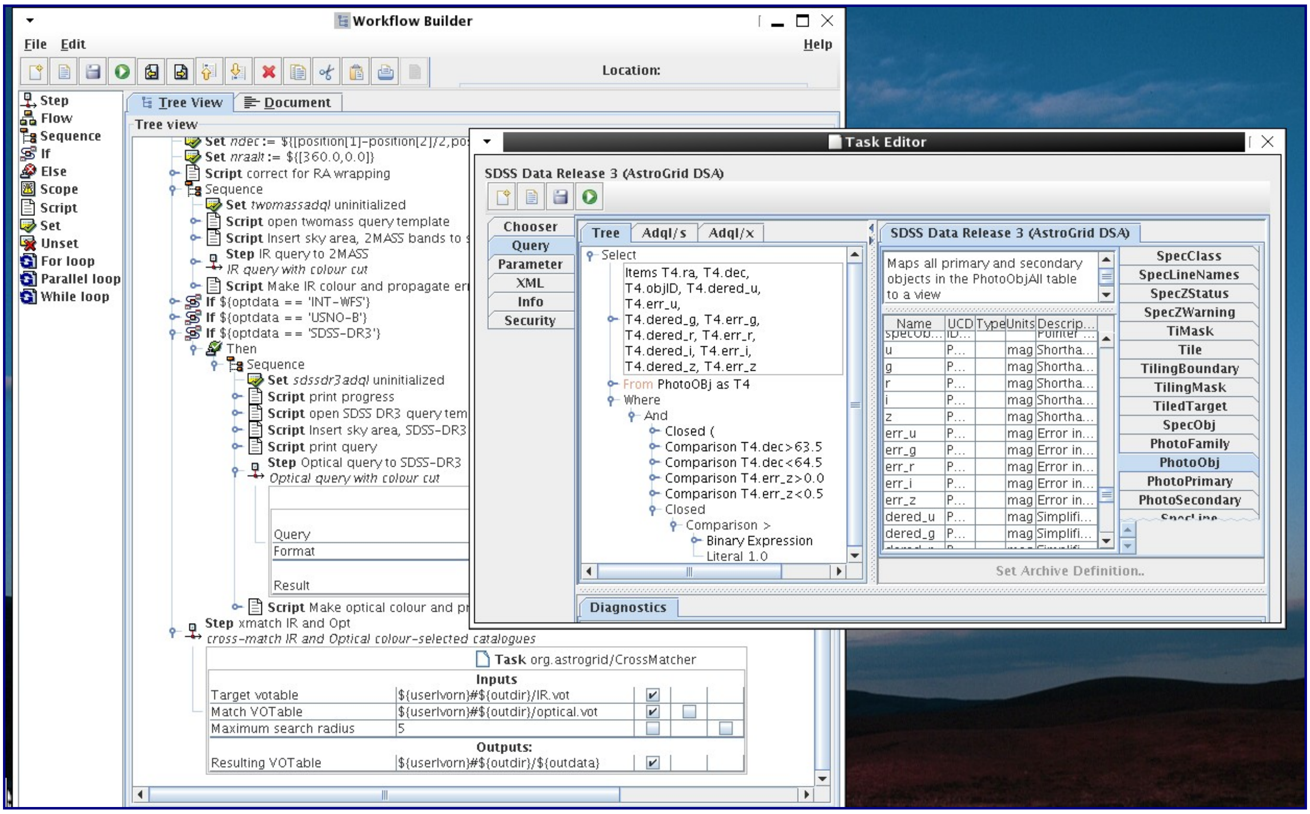
- The MERLINImager was used to extract 21- and 6-cm wavelength images at a matching resolution of 0".1, from the MERLIN visibility data archive, to construct a spectral index image. The 21- and 6-cm data are shown in red and green respectively in the large image and an HST image is shown in blue, highlighting the core of 3C295. This core is also just visible in a high-resolution (0".05) 6-cm radio image (not shown), which was used to align the radio and optical images, whilst even more extended 21-cm lobe emission appears at 0". 25 resolution. The insets show the use of *Chandra* high-resolution X-ray data (left, blue) and *ASCA* data at lower resolution (superimposed in yellow, right), allowing the X-ray emission associated with the AGN and its jets and with the cluster-centre IGM to be disentangled. In both insets the optical image is in magenta and the 21-cm contours are in red.



Radio, X-ray and optical images of 3C295

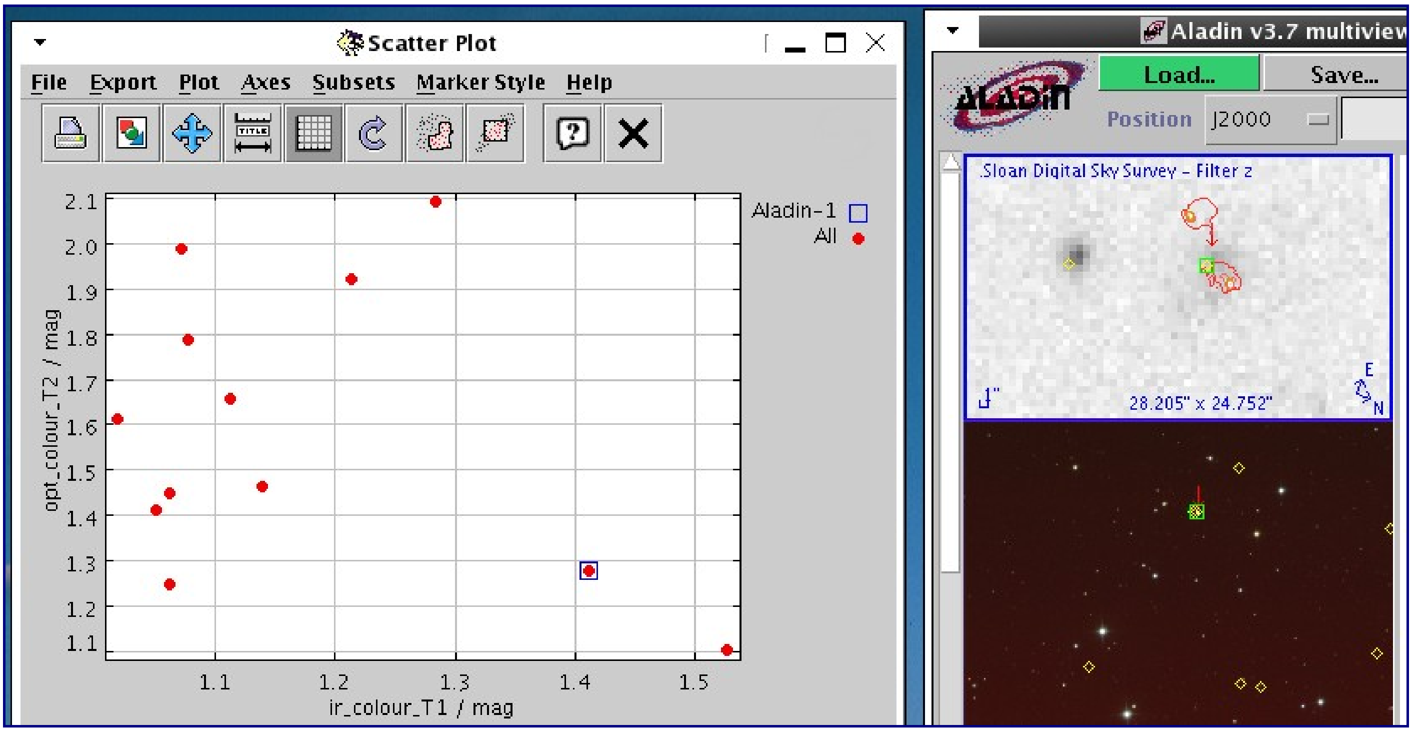
Identification of neighbouring galaxies by colour

- The AstroGrid Colour Cutter was used to select tabular data in particular colour ranges, cross-matching optical and IR catalogues.



Example workflow for selecting data by colour from more than one table (Right-click to open any image full-size in a new window)

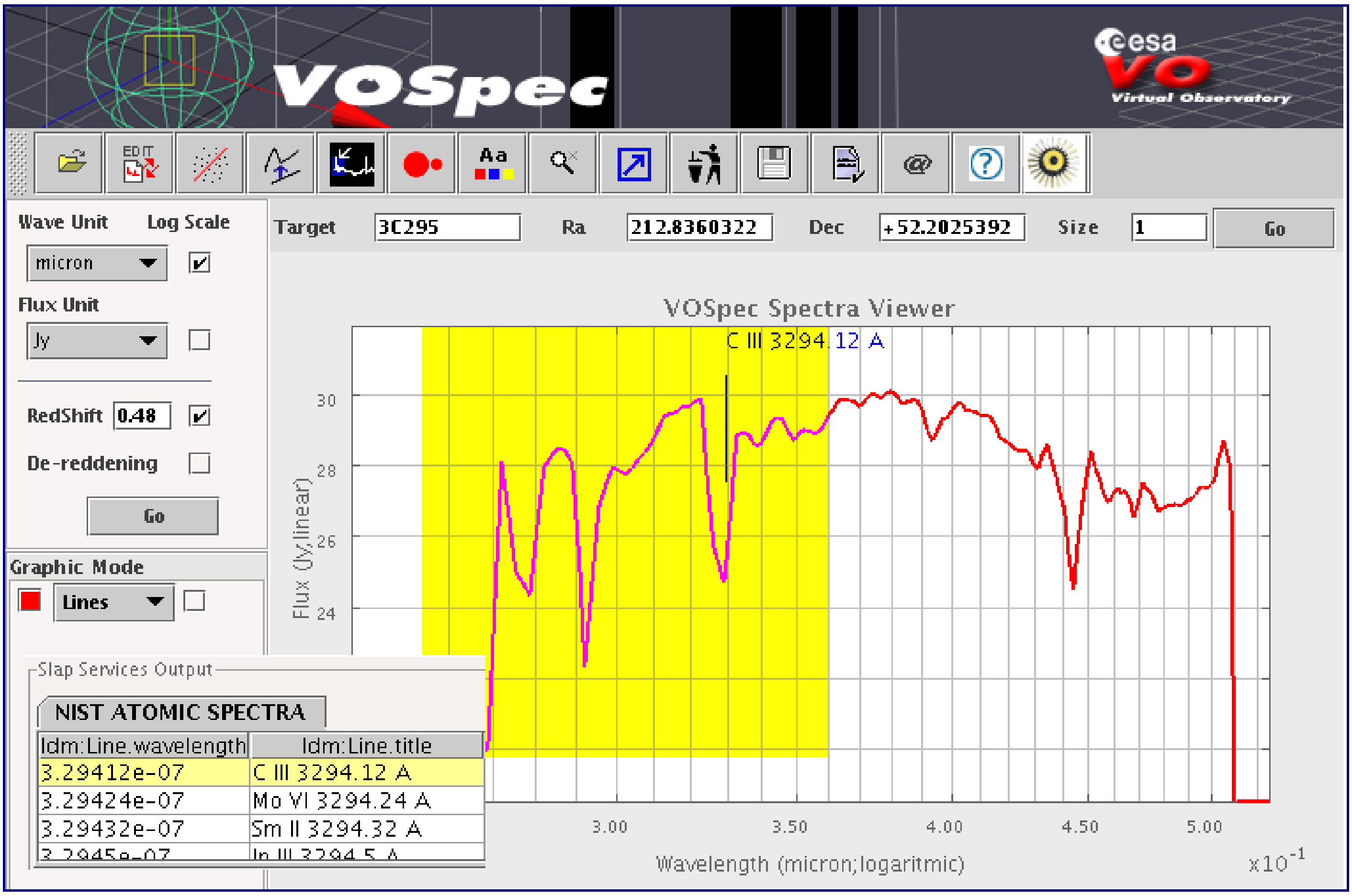
- PLASTIC allows objects to be identified simultaneously in a colour-colour plot (left, i-z against J-K) and on an image, revealing the proximity of an interesting cluster galaxy to 3C295 itself (the red contours on the upper image). The lower SDSS image shows a larger region with the positions of other cluster members overlaid.



Selecting a point in the TopCat plot highlights the same data in the image display

Detailed spectral analysis

- We used VOSpec to display Migale spectra (from the Hyperleda archive) for objects in the direction of 3C295, corrected to the its redshift of 0.48. We used the SLAP line identification tool to see, for example, whether absorption lines belonged to the cluster or to foreground stars.



Spectral line identification at z=0.48

Time Domain Cases

The inclusion of the time domain into Virtual Observatory (VO) applications is lagging with respect to the more mature treatment of the imaging and spectral domains. Still, the VO concept can give a significant contribution to new, better and more efficient science. Two of the areas where major advancements can be achieved are the discovery and follow-up of (suitable candidates for) transient events, and the study of the time evolution (and casual relation thereof) of spectral/imaging properties in individual sources.

The IVOA context is mature. VOEvent has now reached the status of standard IVOA recommendation. According to it charter: "The objective of the VOEvent effort is to define the content and meaning of a standard information packet for representing, transmitting, archiving, and publishing a discovery of an immediate event in the sky. We will call this packet VOEvent. The objective is to drive robotic telescopes, to drive archive searches, to alert the community, and to build interoperable archives. The effort will include not just "photon" events, but also gravitational waves, neutrinos, air showers, etc.". VOEvent will be used as a broadcast for alert by the Heterogeneous Telescope Network (HTN). VOEvent feeds are being also used by the eStar project, which is also actively participating in the definition of the IVOA standards. At the level of protocols, the VO language is being already spoken by the key actors in this field.

The AVO Science Reference Mission case (Kerber 2005) had already highlighted one of the most intriguing applications of the time domain in the VO as a discovery tool. Statistical and stellar evolutionary arguments suggest that the next Galactic supernova is overdue. The discovery, prompt follow-up, and historical (pre-explosion) study of the next supernova would allow at the same time a unique testbed for theories of the last phases of stellar evolution in massive stars, and of the chemical enrichment of the interstellar medium.

In order to establish a list of good potential candidates for the next supernova, one needs to collect multiwavelength-, multi-epoch data cubes for a sufficiently large number of bright stars, as well as for the circumstellar environment, the latter providing insights on the star mass loss history. These data need to coves the largest possible wavelength range, as different energy bands carry the imprinting of different aspects of the object: star, stellar wind, circumstellar medium. Moreover, the observational data need to be linked and interpreted through stellar synthesis models.

For the brightest stars in our Galaxy huge databases of multi-wavelength, multi-epoch data exist. The compilation of these data is a typical example of added value provided by a VO approach to the exploitation of existing archives. Seamlessly dealing with different apertures, resolutions, color corrections etc. requires a treatment of meta-data, which is possible only in the context of the VO, if one wishes to fulfill this task within a typical astronomers' life span.

In order to be able to reconstruct evolutionary sequences, applications designed to deal with data in different domains shall be able to "read", "interpret" and filter data according to their location on the time axis. There is no need to tell that this capability can be applied to any class of sources displaying a variable behavior: Galactic compact objects, active stars, active galactic nuclei, Gamma Ray Burst afterglows.

In order to achieve these goals, VO applications need to be able to:

- cross-match data from different archives for each individual object (e.g.: for each star above a certain magnitude limit)
- compare Spectral Energy Distributions with theoretical models (e.g. stellar synthesis models)
- dealing with the time domain in a data cube. This implies both extracting data in the time domain, and filtering data in other domains according to their position on the time axis
- dealing with different resolutions and apertures of observations at different wavelengths

Links:

- eStar: http://www.estar.org.uk/wiki/index.php/Main_Page
- HTN: http://www.telescope-networks.org/wiki/index.php/Main_Page
- VOEvent: http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOEvent

Requirements extracted from the AVO SRM

The AVO Science Working Group composed a number of use cases for the VO - this forming the

AVO Science Reference Mission - see http://www.euro-vo.org/pub/fc/cases/srm.pdf. These cases have been analysed to determine technical requirements. The SRM cases are:

- Circumstellar disks: from pre-Main Sequence stars to stars harbouring planets
- Intermediate Velocity Clouds
- Which star will go Supernova next?
- Initial Mass Function within 1 kpc: from planetary to stellar masses
- Initial Mass Function for massive stars
- The contribution from low and intermediate mass stars to the interstellar medium
- Galaxy Formation and Evolution from z = 10 to z = 0:1
- Build-up of supermassive black holes
- The Formation and Evolution of Galaxy Clusters
- Correlation of Cosmic Microwave Background, radio/mm and optical/NIR Galaxy Surveys

Data

The data needs to be science-ready or provided alongside tools to make it (the concept of virtual data - deriving science data products on-demand from the raw data products) thus e.g. convert to physical units. Characterisation e.g. PSF/resolution in all dimensions, coverage etc. will be needed to meet requirements. Assume that image and spectra standards access is available and that it is possible to perform cone searches. Data not likely to be available before 2008 are not listed. In most cases both images and spectra have been requested where available and often also timing information.

1. Selection of data based on any/all of position and spectral and temporal criteria
2. Already-accessible (or partly so) data mentioned explicitly:
    - Vizier catalogues
3. SIMBAD
4. IUE
5. ISO
6. XMM-Newton and Chandra
7. HST
8. AAVSO
9. EXPORT (Eiroa et al. ASP conf series 219, 3)
10. SPITZER
11. DIRBE
12. Various H surveys (e.g. see CGPS):
    - HI: IAR, others
13. Halpha: SHASSA, others
14. Radio
    - High-resolution images (VLA, MERLIN, ATCA)
15. Spectra with high position accuracy (Nancay, Efflesburg, Puschino, Mopra, correlated with interferometric images).
16. VLT and other ESO data
17. SDSS
18. SCUBA and other (sub)-mm - APEX?
19. GALEX (UV)
20. General need for IR and optical spectroscopy and data from which SEDs can be extracted
21. Multi-wavelength deep-field data
    - Extra-galactic e.g. GOODS

22. Galactic SFR

Tools

In addition to existing tools e.g. TopCat, VOPlot, VOSpec, ROI generator. Many cases imply a need to set up a series of steps by hand and then repeat the same workflow with adjustment of inputs in response to diagnostic output.

- **Image analysis**
    1. Astrometric alignment
        1. x-y-theta fit to astrometric catalogue
- Corrections for stretch etc.
- Bootstrapping in sparse fields or for resolved/different structures.
- Automation
    1. Photometry
        1. Checking absolute accuracy wrt reference catalogues/images
            - Variations across an image
        2. Conversion to physical/interoperable units
            - Web service to give access to databases of filter profiles, central/effective wavelengths, zero points, colour corrections, magnitude dependence etc.
- Should work albeit at lower accuracy (reported) if not all information is available
- Extend from radio to X-ray (gamma ray?)
    1. Source extraction (in addition to current SExtractor implimentation)
        1. PSF/aperture corrections to measured flux densities for data of different origins
- Measurement of extended/irregular sources
            - Galactic clouds and nebulae
- Morphology of resolved galaxies - CAS parameters (Abrahams et al., Conselice et al.), surface brightness profiles (GALFIT, GIM2D, GASPHOT).
        1. Source extraction: at exact catalogue positions or best detection close to catalogue positions.
    2. Image manipulation
        1. Convolution to matching resolution
- Maths - + - / * log etc. (incl. weighting, using many images etc.)
- Measurement of angular clustering, pixel or larger correlation scales, 'power' spectral analysis etc.
- **Catalogue manipulation**
    1. See issues similar to those for images e.g. astrometry, photometry, aperture corrections
- Cross-matching large (>10^6) sources, optimisation of operations, accuracy.
- Multi-catalogue, multi-wavelength, multi-resolution cross-matching using SED criteria (e.g. SpecFind[?] for all wavelength domains).
- Scriptable/workflow-based manipulation e.g.
        1. Colour creation/comparison
- Library of common functions e.g. calculate luminosity, distance etc. for given cosmology or for Galactic sources.
- **SED construction**
    1. Photometric/astrometric alignment for input images/catalogues
- Corrections for extinction and reddening
- Automatic SED construction and parameterisation

- Redshift compensation including K-correction (various templates)
- **Time domain**
    1. Selection of data by epoch
- Sorting, display and manipulation using time e.g. d(flux)/d(t)
- **Spectral analysis**
    1. Long-slit spectra
        1. Slit angle - use information in extracting measurments
- P-V image display/measurement
    1. Data cube display, analysis, extracting 1-D etc. spectra, calculating moments and fitting rotation curves etc.
- Automatic line classification, measurement, velocity assignment
- Clickable links between images overlaid with spectral catalogues and the spectra
- **Model-fitting**
    1. Libraries and/or generation of synthetic stellar population SEDs
        1. Galaxies of various ages, metallicities, SFRs, IMFs - Brunzal & Charlot, Jimenez, PEGASE etc.
- Galactic star forming regions
- Individual YSO e.g. D'Alesio (2001 ApJ[?] 553 321) and stellar evolution
    1. Statistical analysis tools for comparison between data and models (and other uses) including Bayesian methods
- Neural network or similar learning-based automatic classification
- Photometric redshift packages
        1. Documented, able to adjust templates/input parameters e.g. extinction, extend outside optical, extract or fix other parameters e.g. luminosity, SFR...
    2. Comparison with galaxy formation simulations e.g. VIRGO, GALICS
- Multi-wavelength comparisons to estimate absorption and extinction
        1. Optical/IR e.g.
- X-ray/HI

Technical Demands

The system is required to have functionality to meet the science needs of the community. There must also be sufficient functionality to enable the system to be run in a robust, reliable and cost effective fashion.

Science focus on technical capabilities

Key implications can be charaterised into a number of headings.

- data
    - location - USA vs Europe etc
- applications
    - client side vs server side
- volume requirements
    - size of data
- speed requirements
    - asynchronous vs synchronous activites
- access controls
    - proprietory vs public data

The analysis of these issues is being undertaken in VOTECH DS3 - Infrastructure.

Operational focus on technical capabilities

The science cases set certain constraints in the operational area:

- system availability
    - 24/7
- network connectivity
    - best geometry for location of primary data / application resources
- complexity
    - impacts on technical knowledge required to provide VO service
- use of common standards
    - interoperability with non Euro-VO services

Consideration of these issues will be furthered in VOTECH DS3 - Infrastructure

Acknowledgements

This document was prepared with input from the VO-Tech science team: Mark Allen (CDS), Eduardo Gonzalez-Solares (AstroGrid), Matteo Guianzani (ESA), Anita Richards (AstroGrid) and Andreas Wicenic (ESO). Thanks also to Paolo Padovani (ESO) and Evanthia Hatziminaoglou. (ESO).

Bibliography

AVO-SRM: AVO SWG, The AVO Science Reference Mission,

Appendices and Links

- ALMA:
- CDS:
- e-Merlin:
- ESO-VLT: ESO Very Large Telescope
- Euro-VO: European Virtual Observatory: http://www.euro-vo.org
    - Euro-VO DCA: Data Centre Alliance: see http://
- Grantecan:
- Herschel:
- Lofar:
- Merlin:
- Planck:
- WHT: 4.2m William Herschel Telescope
- UKIRT: United Kingdom Infra Red Telescope
- Vista: ESO IR survey telescope
- VLBI:
- VST: VLT Survey Telescope
- XMM: XMM-Newton X-ray telescope.

Glossary

- AVO: Astrophysical Virtual Observatory - a EU FP5 funded design study as a pre-cursor to the Euro-VO activities.
- CDS:
- ESA: European Space Agency
- ESO: European Souther Observatory
- VO: Virtual Observatory