The Astrophysical Virtual Observatory Science Reference Mission

Astrophysical Virtual Observatory
Science Working Group

April 2005

edited by
Paolo Padovani
European Southern Observatory
AVO Science Manager

Contents

1 Introduction

1.1 The Astrophysical Virtual Observatory

The Astrophysical Virtual Observatory initiative (AVO) was a project to create the foundations of a European scale infrastructure by conducting a research and demonstration programme on the Virtual Observatory (VO) scientific requirements and necessary technologies. The AVO was jointly funded by the European Commission (under the Fifth Framework Programme [FP5]) with six European organisations participating in a three year Phase-A work programme from November 2001 to November 2004. The partner organisations were the European Southern Observatory in Munich, the European Space Agency, AstroGrid (funded by PPARC as part of the United Kingdom's E-Science programme), the CNRS-supported Centre de Donnees Astronomiques de Strasbourg (CDS) and TERAPIX astronomical data centre at the Institut d'Astrophysique in Paris, the University Louis Pasteur in Strasbourg, and the Jodrell Bank Observatory of the Victoria University of Manchester. The AVO was the definition and study phase leading towards the Euro-VO (`http://www.euro-vo.org`) – the development and deployment of a fully fledged operational VO for the European astronomical research community. A Science Working Group (SWG; `http://www.euro-vo.org/twiki/bin/view/Avo/ScienceWorkingGroup`) was established to provide scientific advice to the project.

The AVO project was driven by its strategy of regular scientific demonstrations of VO technology, held on an annual basis in coordination with the IVOA. For this purpose progressively more complex AVO demonstrators were constructed. The last one was an evolution of Aladin, developed at CDS, and has become a set of various software components, provided by AVO and international partners, which allows relatively easy access to remote data sets, manipulation of image and catalogue data, and remote calculations in a fashion similar to remote computing. The latest version of the AVO prototype is available for download at `http://www.euro-vo.org/twiki/bin/view/Avo/SwgDownload`.

1.2 The Science Reference Mission

The purpose of this document is to provide the AVO Science Reference Mission (SRM). The SRM defines the key scientific results that the full-fledged EURO-VO should achieve when fully implemented and consists of a number of science cases covering a broad range of astronomical topics, with related requirements, against which the success of the EURO-VO will be measured. The SRM was put together based on the input received by the SWG. The original contributors are acknowledged in the document.

The various cases have a similar structure: first the science case is briefly described, then the required data are mentioned, and finally the VO implications are discussed.

2 Science Reference Mission Cases

2.1 Circumstellar disks: from pre-Main Sequence stars to stars harbouring planets

E. Solano et al. (LAEFF)

`Science Case`

The study of protoplanetary disks is currently undergoing an exciting phase partly propelled by the discovery of extrasolar planetary systems following the detection of 51 PegB by Mayor & Queloz (1995, Nature 378, 355). The possible discovery of terrestrial planets in the near future, in addition to the Jovian-like planets already detected, will pose interesting questions on the formation of extrasolar planetary systems. Knowledge of the properties of protoplanetary disks and how they evolve to debris disks around main-sequence (MS) stars is, therefore, a crucial step in such a process.

Theories of dust evolution have given the general outline on how the disks evolve from the optically thick phase around the youngest stars to the debris disks possibly harbouring planets around the main sequence Vega-type stars. There are, however, still many open questions concerning the driving mechanisms and the factors that influence disk evolution.

The goal of this case is to study a statistically significant number of stars covering a wide range of masses and evolutionary stages (HAeBes, ETTs, CTTs, PTTs, A-shell and Vega-type stars) to track and characterise from an observational point of view the evolution of protoplanetary disks. A methodological approach based on the comparison of the observed spectral energy distribution (SED) with a grid of synthetic models including both the stellar and disk contributions will allow us to derive the physical parameters of the star and the disk. Theoretical modelling of SEDs has proven to constitute an invaluable tool for understanding the structure and properties of protoplanetary disks (e.g., Merín et al., 2004, A&A, 419, 301; Chiang & Goldreich, 1997, ApJ, 490, 368; 1999, ApJ, 519, 279), D'Alessio et al. 1998, ApJ, 500, 411; 1999, ApJ, 527, 893; 2001, ApJ, 553, 321; Dullemond et al. 2001, ApJ, 560, 957).

`Requirements`

- Archival Data. The building of SEDs needs to collect multiwavelength information across the electromagnetic spectrum. This will require access to on-line astronomical archives with absolute calibrated data (IUE, ISO) as well as astronomical catalogues (IRAS, 2MASS, Strömgren photometry, etc.) through the Vizier service and, desirably, other information taken from publications.

- EXPORT Data. In addition to the archive data we will make use of the EXPORT observations (Eiroa et al. 2000, ASP Conf. Series, vol. 219, p. 3) consisting of photometry from optical to the near infrared, and optical medium and high resolution spectroscopy of a sample of 70 stars (both pre-main sequence [PMS] and MS stars). The use of EXPORT data will permit to take advantage of the fact that the optical and near-IR photometry were obtained simultaneously. This observational approach is appropriate since T Tauri and HAeBe stars vary markedly in these spectral regimes and a significant part of the total luminosity of the object is radiated by the PMS stellar photosphere at these wavelengths.

- Models. The models from D'Alessio et al. (e.g., 2001, ApJ, 553, 321) represent one of the best approaches for studying the physics of illuminated accretion disks around PMS stars. An on-line library of models covering the whole relevant physical parameter space for the central star and the disk has been built and will be used in the analysis.

- Future Data. Despite the wealth of information provided by the IRAS and ISO satellites, the sensitivity of these telescopes did not allow to answer many key questions concerning the evolution of the disks and the possible formation of planets. For that purpose, future telescopes are being planned by ESA, NASA, and ESO to get the observational basis for a complete picture of the evolution of circumstellar disks around young stars and the possible mechanisms of planetary formation.

  The first mission of this series is the Spitzer Space Telescope from NASA. It is already providing an unprecedented improvement in sensitivity allowing major advances in this area of research. To make use of Spitzer data, synthetic magnitudes for the four IRAC (InfraRed Array Camera) photometers, operating at 3.6, 4.5, 5.8 and 8.0μ, have been calculated by convolving the model fluxes with the respective filter passbands. Thus we are able to predict the magnitudes and colours of the young stars harbouring planet-forming disks and interpret their physical state according to the IRAC colours (Allen et al, 2004, ApJS, 154, 363). During its lifetime, Spitzer will observe many young stars with ages spanning a wide range of evolutionary stages, from objects still embedded in their parental clouds to stars with ages around 10^6 yr, the so-called Vega-type stars. This collection of observations will represent a big step towards an unbiased statistical study of the early evolution of stars and disks.

  Major new telescopes under construction include the Photodetector Array Camera and Spectrometer (PACS) on-board Herschel (operating between 60 and 210μ), the Mid InfraRed Instrument (MIRI) planned for the James Webb Space Telescope (operating between 5 and 28μ), and the Atacama Large Millimeter Array (ALMA) (operating between 70 and 900 GHz [$\sim$ 300 and 4,300 μ]). Their targets will need careful selection to take advantage of the high resolution of these instruments. The expertise gained from our studies will contribute to designing observing campaigns as well as testing the theoretical frameworks accessed via the model library.

`VO aspects`

- Need of VO. SEDs building, model fitting and the associated detailed analysis require to devote a tremendous amount of work and time which makes it quite inefficient for large datasets. This was clearly demonstrated in Merín et al. (2004) even if it was focused on only two stars. This situation can be alleviated if the analysis is performed under the VO framework by continuing to develop tools for workflow construction.

- Suitability of the Science case. SEDs building requires accessing to a variety of services (e.g., name resolver at CDS), astronomical catalogues and archives both local and remote. The data transfer protocols already defined in the VO framework (SIAP, ConeSearch, SSA, VOTable) have demonstrated how efficient they are to handle all this heterogeneous and dispersed information.

- Expanding the VO view. The Spanish Virtual Observatory is developing a general-purpose facility that will be implemented as a Web service to build spectral energy distributions. The user's interface is an HTML page where the user enters the object identification or coordinates and chooses the archives where the search has to be made. It is also possible to enter one's own data. After this, the system asks CDS to resolve the name of the object. This request is made via the SOAP protocol to one of the Web services currently running at CDS. The service returns the object information (including its coordinates) in XML format. Having the coordinates of the object, a request is made to each of the data centres (INES, ISO, VIZIER, etc.) using the VO SIAP protocol. Each one of the data centres returns the data in VOTable format. The data will be interpreted by the Web application using its VOTable interpreter module. Once the data retrieved by each of the data centres have been interpreted correctly, it can be combined, processed, or passed to the data mining layer (see "Data mining" section) to perform the model fitting. After this, the user will be able to visualise the result using the visualisation tools that will be implemented in the Web application.

- Implementation of analysis tools. We strongly feel that, in the VO framework, a deeper involvement of the astronomical data centres in the development of analysis tools is required. Accordingly, the system incorporates a set of tools to estimate the stellar physical parameters. In order to fit the best model to the observations it is necessary to have a first estimation of the stellar physical parameters (effective temperature, surface gravity and metallicity) as well as the interstellar reddening. In all cases, the parameter estimation will follow a pre-defined hierarchy in which the values provided by the user are on the top. Significant differences between the input values and the values obtained after the iterative process by the data mining tools may indicate either a non-adequate set of input values (due, for instance, to variable reddening and lack of simultaneity between the Strömgren and 2MASS photometry) or the presence of peculiarities in the spectral energy distribution that might deserve further and more detailed analysis.

- Data mining. Data mining is a key aspect in the VO framework. The system includes a tool to estimate the model parameters following a Bayesian method. The main aim is to help users to quantitatively analyse the data in terms of evidence favouring one or several sets of parameters, what other alternative models can compete with the most a posteriori probable one, and what are the most discriminant observations to discard alternatives. These functionalities will be linked with the capacity of the server to query other archives in search of values for the discriminant observables.

2.2 Intermediate Velocity Clouds

C. Bot (CDS, IAP), L. Cambresy (CDS), F. Boulanger (IAP), D. Egret (Obs. de Paris)

`Science Case`

Intermediate Velocity Clouds (IVC) are clouds of neutral gas that move in unexpected directions and with unusual velocities with respect to the normal rotation of the rest of our Galaxy. They are detected due to their 21cm emission of atomic hydrogen. Some (but not all) also show significant dust emission. The idea is to separate two populations in the IVCs: the dust rich ones, and the dust poor ones, in order to test the underlying idea that their origin is different.

`Science Aspects`

- The study needs to be done on the whole sky to understand the spatial distribution of these two populations.
- The study needs first to identify regions of excess in the ionised or atomic hydrogen maps. This is the equivalent to a source extraction but for extended emission.
- The main core of the study relies on map comparisons at different wavelengths and with different resolutions.
- To detect the IR emission of the IVC, the first step is to remove the foreground emission, using HI data.
- The infrared contribution detected for the IVC will have to be compared to the H_α emission to check that there is no contamination by the warm interstellar medium.
- The intercomparison of the different maps leads to a classification of the IVCs in two groups (dust poor IVC and dust rich IVC).
- The spatial distribution of the two groups is displayed.
- The data have to be all-sky surveys in the far infrared (dust emission), in the 21cm line (atomic hydrogen emission) and in the H_α line (ionised hydrogen emission), like COBE/DIRBE, Leiden-Dwingeloo/IAR HI survey and SHASSA respectively.

`VO aspects`

- Data: The study would be done using the DIRBE far infrared emission all-sky survey to trace the dust abundances, the combined surveys of IAR and Leiden-Dwingeloo plus HIPASS in HI and the SHASSA survey for the H_α emission. All these data are public, except the IAR HI survey, but it is expected to be released on a medium time scale (in the coming years). The data will also be compared to the existing catalogues of IVCs in the literature (through Vizier).
- Tools:

- extended source extractor.
- image manipulation tool (simple operations: +,-,*,/, etc.)
- image convolution tool (to degrade the resolution of maps previously to inter-comparison.)
- plotting tool allowing pixel to pixel maps scatterplots with simple fittings, and the possibility to select regions of interest in the plot. This could be a simple adaptation of the VOPlot tool.
- continue development of spectral datacube manipulator and other spectral manipulation tools.

`VO need`

This SRM proposal is a typical example of what the studies on the extended emission would be with the Virtual Observatory. The gain of the VO is in efficiency, but the main interest of this SRM is that all the requirements for this study are generic tools that can be used in completely different contexts. Furthermore, this study is chosen to be an all-sky study, that is dealing with large amounts of data, and it is therefore an example of the kind of problems will have to be solved and the tools that will be needing with the forthcoming large projects on extended emission (e.g., Planck, Herschel, SKA, ALMA).

2.3 Which star will go Supernova next?

F. Kerber (ST-ECF)

`Science Case`

The late phases of the evolution of massive ($> 8\ M_\odot$) stars and the physical changes of the star leading up to a supernova (SN) explosion are not well understood. The main reason, of course, is that no Galactic supernova has been observed for 400 years. Massive stars, their wind blown circumstellar nebulae, and their supernova explosions are main contributors to the chemical enrichment of the interstellar medium (ISM) and exert an important influence on its dynamics. This makes them main drivers of the chemical evolution of galaxies and the formation of stars. About 400 highly evolved massive stars are known in our Galaxy. Most of these stars are bright sources from the X-ray (hot and fast winds, supernova remnants (SNR)), through the UV (winds and star and circumstellar matter), the visual (star and circumstellar nebulae), IR (star and dusty circumstellar matter) into the radio (circumstellar nebulae and SNR). Because they are bright a wealth of data spanning a large wavelength range is available for them. In a significant number of cases the data cover very long periods of time which is most relevant for this variable and quickly evolving objects. By combining the existing multi-wavelength data the VO promises to provide new insight and can demonstrate its potential as a true discovery tool. This project will make the VO decidedly more powerful by requiring an intensive exchange between observation and stellar modelling based on physical theory and by opening the interesting time domain. One important result will be to establish a list of possible candidates for the next SN, which could then be monitored more closely. Of course, no one can predict when the next Galactic supernova with a massive progenitor will occur but statistically it is clearly overdue. This makes such a project a timely effort which will put us in a position to take maximum advantage of the next Galactic supernova when it actually takes place. Moreover, an improved understanding of the massive stars and their influence on the evolution of galaxies will be fundamentally important for the correct interpretation of observations of the distant and early universe.

The main objectives are:

- combine existing heterogeneous archives to build a multi-wavelength, multi-epoch data cube of the star and its environment and SNR
- extract a status quo of knowledge on evolved massive stars
- search for common features and derive evolutionary sequence
- ranking of SN candidates and list of likely next SN
- study interaction between circumstellar nebula and the ISM

Scientific relevance:

- supernova explosions are of fundamental importance to nucleosynthesis and thus to the chemical enrichment of the universe

- SNe drive chemical evolution of galaxies
- determine dynamics of interstellar medium, trigger star formation
- possible relation to gamma-ray bursts (GRBs).

Practicability:

- huge multi-wavelength, multi-epoch archives exist (bright objects)
- active community (observations and theory) exists in Europe
- solid basis of stellar evolution, nuclear processes exists
- local galaxies within reach of our instrumentation
- modern technology (AO, VLTI) give access to circumstellar environment, mass loss history
- different wavelength domain probe different aspects of the objects: stars, stellar wind, circumstellar nebulae providing the opportunity to develop insight into the full range of phenomena associated with massive stars and supernovae
- Galactic SN explosion is overdue

`VO aspects`

- Data:
  - large body of relevant multi-wavelength and multi-epoch data exists
  - better use of existing data
  - combination of data with different resolution etc., emphasises need for meta-data
  - add time domain to VO
  - include theory (stellar modelling) in VO
- Tools:
  - cross matching tool to combine data from different sources for each individual source
  - tool for comparison of observations with models (spectral cross correlator)
  - tool for extraction of data in time domain
  - visualisation tool for data from different wavelength regions obtained at different resolution etc.

VO Need

Study of individual stars have been done in the past. A comprehensive view of the known most massive stars requires full use of large data volumes distributed over diverse archives. VO tools for analysis and combination of such data will greatly facilitate their interpretation. Inclusion of the time domain and quantitative comparison with theory/models in the VO context will truly enable new and better science.

Relevance to astronomical community

- very fundamental questions addressed
- real service to community (after SN there will be rush to find pre-explosion data and explain things)
- prepare community for next SN, which will be one of the most important events for astronomy to exploit
- enable us to ask better questions, write better proposals
- prepare observatories for SN event, have suitable telescope/instrument

Education & Public Outreach

- large amateur community involved by default (photometry, AAVSO)
- superlatives: most massive, hottest, etc.
- "Top Ten" list of SN candidates should appeal to public
- body of solid and accessible physics: stellar evolution, children of the universe
- nice images of circumstellar matter/nebulae
- possible effect of nearby SN on life on earth, previous mass extinctions
- link to GRBs

2.4 Initial Mass Function within 1 kpc: from planetary to stellar masses

T. Prusti (ESA)

`Science Case`

The initial mass function (IMF) is a fundamental astrophysical quantity linking galactic and extragalactic astronomy together. As an outcome of the star formation process, the IMF constrains different scenarios: no theory of star formation is valid without being able to explain the IMF. On the other hand, the IMF is a key input parameter for our understanding of the chemical evolution of the universe.

Deducing the IMF spanning from planetary to high stellar masses is not easy. Typically scientific articles proposing an IMF over a wide mass range are compilations of several pieces of work published in astronomical literature. Although "The Initial Mass Function" sounds like an appealing title for the science reference mission (SRM) of the Astrophysical Virtual Observatory, it is not suitable as such. The elements needed for the compilation of the IMF, however, are suited for VO work. This proposal will concentrate on deducing the IMF at the lowest masses, as indicated in the project title. The next case deals with the requirements for deducing the IMF at the high mass end.

Science aspects:

- Even the lowest mass stars can be recognised in the youngest regions because sources are more luminous in their youth.
- Going to younger regions means more dust in the surroundings leading to the necessity of extinction correction.
- Close-by regions contain typically few hundreds to few thousands members which means that all members need to be found and distinguished from field stars for statistical studies.
- Starting point has to be a deep survey data-set in optical, NIR and/or MIR wavelengths.
- Known (bright) members have to be identified by using X-ray, H_α emission line, variability, proper motion, IR excess etc. criteria.
- By using the known members one tries to characterise the properties of members in the deep survey data and use these characteristics to find the faint members.
- Construct a sample of members with known completeness levels as a function of some observable
- Calculation of individual luminosities of sources with a possibility to make extinction corrections
- Binning luminosities into an observed luminosity function

- Construction of model luminosity function with input parameters such as pre-main sequence tracks, star forming history, binary fraction, and, above all, a model IMF
- Comparison of observed luminosity function with the modelled one to find the most probable IMF for the sample

`VO aspects`

- Data

  The best starting point would be to look at deep Spitzer data of close-by star forming regions. After finding a suitable public set, one should look for the availability of optical and NIR data of the region. The X-ray, H_α emission line, and other data may be accessible via VO-enabled archives or can be found in the literature using Vizier.

- Tools

  - Point source extractor for images
  - Recipes and tools which allow comparison of data taken with different coverage and resolution in different wavelength regimes to: model sensitivity changes across observed fields; compensate for different effective apertures or resolved objects to align photometry and astrometry; and provide statistical parameters for using upper limits, incomplete data and the probability of identification, etc.
  - Cross matching tool to combine data from different wavelengths to each individual source
  - Table manipulation/plotting tool to examine the properties of members
  - SED analysis tool with possibility to display data, overlay models, correct for extinction, and calculate luminosities

`VO need`

The gain of VO in this IMF work is mainly in efficiency. It is foreseeable that work for a region like Chamaeleon I could be done 'manually' by using available tools, but VO would provide large benefit when the same has to be repeated to other regions, especially Orion with an order of magnitude more members.

2.5 Initial Mass Function for massive stars

M. Kontizas (Athens), A. I. Gomez de Castro (Madrid), A. Richards (Jodrell Bank)

`Science Case`

Massive stars in stellar systems

High mass stars are the most luminous tracers of young stellar populations. O- and B-type stars and rare M-type supergiants only make up a small fraction of each stellar population but the fact that they are very bright makes them easy to observe in our galaxy and other nearby galaxies. The Initial Mass Function (IMF) of the bright end of the mass scale is a crucial parameter for determining the upper limit of mass and its connection to the environment such as metallicity.

To find the IMF we require to know the mass of the observed stars as it is determined from its position on the Colour Magnitude Diagram (CMD) and then we compare with theoretical tracks. We need to know the absolute magnitude of each star, i.e., the exact distance of the star and absorption. The distance in extragalactic systems is not difficult to find, since the overall distance of the parent galaxy is known. The interstellar extinction is not always well determined.

However, an efficient way of overcoming this problem, particularly for our own Galaxy, is to consider the stars in stellar associations, where the distance and extinction is found for all stars in each system. Young associations are also found in the Magellanic Clouds and M31 and a lot of data are available.

Embedded massive stars

The very massive stars at the protostar phase start as small dense ionized gas objects known as Ultra-Compact HII regions detected in the radio domain. The pressure is higher than the surrounding field and the expansion starts quickly so that the region becomes a compact and then a normal HII region. The lifetime of such a protostar is $10^4 - 10^5$ yr. Yet some massive protostars manage to produce bipolar outflows and masers. The types of maser emission associated with young stellar objects (YSOs) is a mass indicator, e.g., methanol with high mass YSOs. It is found that more evolved low-mass YSO, e.g., T Tauri stars can have similar IR colours to high-mass YSO, but T Tauris are more likely to be X-ray emitters and even if the high-mass stars are, the colours are distinctive.

The nature of discs and jets and in some cases stellar activity gives some information on stellar mass and can be used as a good indicator of close binarity/multiplicity.

The extragalactic HII regions are observed like beads along the spiral arms of the parent galaxy. The UV spectrum of the HII regions is the best way to discriminate the number and spectral type (i.e., mass) of the stars. Population synthesis is often done in a self-consistent manner with stellar evolution models for a given metallicity.

For the very massive stars (with $M > 10\ M_{\odot}$) there is no pre-main sequence phase so they begin to emit in the UV.

- Targets

- Massive stars in M31, our own Galaxy (MW), Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) to obtain a galaxy sample with different metallicities from very high to very poor.
- Select the young stellar associations in these galaxies. Find the available CMDs. Extinction and distance. Production of the IMFs
- Select the embedded clusters from IR surveys to have the youngest objects in the galaxies.
- Radio masers with high mass indications.
- Extragalactic Ultra Compact HII regions.

• Data Catalogues - Tools

- Archives in the optical domain of very young star clusters and associations in the nearby galaxies such as the MW, LMC, SMC, M31. The superclusters in M31 have been found with point sources, which provide the high mass end of the IMF.
- Radio archives such as ATCA, MERLIN, Puschino and Nancay (i.e., at least 50% European!) and maybe the VLA will provide the necessary spectra with very accurate positions.
- We need UV spectra of standard massive stars for the population synthesis. This should come from IUE (since HST is too sensitive for them, and IUE low dispersion spectroscopy is well calibrated). We could check the IUE calibration with the new set of standards proposed by HST.
- Tools which use as input the UV spectra (for instance HST/STIS long-slit) and the stellar evolution codes can be used to determine the high mass end of the IMF and its dependence on the metallicity of the region (or galaxy). This tool would be most useful both to determine whether there are significant differences in the high mass end from an area to another and also for the study of the chemical evolution of the Universe (for high redshift galaxies the relevant region of the spectrum moves from the UV to the optical).

2.6 The contribution from low and intermediate mass stars to the interstellar medium

P. Garcia-Lario (ESAC), F. Kerber (ST-ECF), A. Richards (Jodrell Bank)

Science Case

The knowledge of the chemical evolution of stars of low- and intermediate-mass (1 - 8 $M_{\odot}$) is crucial to determine the overall contribution of nuclear-processed material from these stars to the ISM and, subsequently, to develop models which can predict the chemical enrichment of galaxies in general.

Although all stars are born O-rich (C/O ratio < 1) in the local Universe, several dredge-up processes take place during the evolution of low-mass and intermediate-mass stars which may eventually turn their atmospheres into C-rich (C/O ratio > 1) environments. This switch between O-rich and C-rich chemistry has an enormous impact on the nature of the gas and dust returned to the ISM and usually takes place during the latest stages of the evolution of these stars in the so-called 'Asymptotic Giant Branch' phase (AGB). The processes involved are known to be (according to dredge-up and hot bottom burning theoretical models) strongly dependent on the main-sequence mass of the star and on its metallicity. Clues to the C-O transition can be found in a rare class of objects displaying lines indicative of both O- and C-dominated chemistries. One possibility is that these possess a fossil O-rich shell (e.g., with dust showing silicate features, or water masers) and a newly C-enriched CS (central star).

The study of stars on the AGB is hampered from the observational point of view for two main reasons: firstly, this is a short-lived phase ($10^5 - 10^6$ yr), so that there are only a limited number of stars evolving in this phase at any given moment. The transition from 0- to C-rich can happen at any time during this evolutionary stage (in some cases only at the very last moment after a so-called 'late thermal-pulse' which may be triggering the formation of WC-type CSPNe). Secondly, AGB evolution is dominated by heavy mass loss (driven by radiation pressure on grains), which results in the formation of thick circumstellar shells of gas and dust. Mass loss from the more massive stars is so huge (10^{-4} $M_{\odot}$/yr) that the shells become optically thick at visible wavelengths (sometimes even beyond). The CS loses most of its original stellar atmosphere and appears as a heavily obscured source in the optical range being detectable only at infrared (or radio) wavelengths. Such observations also reveal a wealth of molecular and even solid species in the stellar ejecta. It only reappears (usually after a few $\sim 10^3$ years), first in the near infrared and then in the optical, when the mass loss has ceased. By this time, the envelope has expanded sufficiently to become diluted in the ISM and thus becomes optically thinner.

Planetary Nebulae (PNe) result from this heavy mass loss by AGB stars. They are understood in terms of the two-wind model by Kwok et al. (1978, ApJ, 219, L125) as the product of the mass-loss history on the AGB and the CS evolution. After leaving the AGB the CS becomes hotter. A fast collimated wind is often observed in proto-PNe. Upon reaching $\sim 30,000$ K the CS is able to ionize the hydrogen in its vicinity and a PN becomes observable. During the PN phase the temperature of the CS can reach in excess of 100,000 K. The resulting hard UV radiation creates a fully ionized PN plasma which is rich in the recombination lines of H and He and strong forbidden lines from heavier elements (C, N, O, S, Ne etc.). This makes it possible to derive the chemical abundances in the matter ejected on the AGB. The hot CS gives off a fast thin wind which interacts with the much slower but denser AGB ejecta shaping the complex morphologies of PNe.

Over recent years, it has been recognized that the interaction of PNe with the surrounding ISM becomes a very important process in the late stages of PN evolution. Once considered a curiosity only seen in a few odd examples, like A35, recent survey work has shown that it is commonly found in old PNe. These old PNe represent the final observable stages during which the matter, which was processed inside the star and expelled during the AGB phase, is actually returned to the ISM.

When a PN ages the density drops because of the expansion of the nebular shell; at some point the ISM pressure upstream will become comparable to the pressure in the PN shell and interaction with the ISM is initiated. The gas in the shell will be compressed which results in an asymmetric brightness distribution, while the shape remains largely spherical. Later in the evolution the density decreases further and finally the expansion of the PN is significantly slowed upstream, first leading to a deformation of the nebula and later to its disruption. During this process the CS, which is not affected by the slowing, will move out of the geometrical centre of the PN, subsequently leaving its nebula behind, as observed in Sh2-174. The final result of this evolution will be a white dwarf stripped of its nebula, which mixes with the ISM. In a recent survey Kerber et al. have found that in about 50% of interacting PNe the CS is no longer located in the geometric centre of the PN. This is a direct manifestation of the galactic orbital motion of the CS after the interaction process has decoupled the motion of the nebula and its CS.

Recent studies (we have done some work here and references can added) have demonstrated that the relative motion between the CS (plus PN) and the ISM is determining the details of the interaction process. This is a direct manifestation of the Galactic orbital motion of the CS. Namely:

- low-eccentricity, low inclination orbits (typical for thin disk population) result in low relative velocities and mild forms of interaction, the PN basically fossilizes and slowly diffuses into the ISM.
- high-eccentricity, high inclination orbits (typical for thick disk or halo population) results in high relative velocities and violent forms of interaction in which the PN shell is broken apart and significant amounts of matter is stripped off the PN along its path.

The interstellar magnetic field may also play a very important role in the PN shell break up and the mixing of the matter into the ISM. The interaction of the stellar and interstellar fields can be studied using radio lines from magnetisable molecules in (and before) the formation of young PNe, whilst IR polarimetry reveals the magnetic field acting on dust grains from this stage onwards. A more quantitative understanding of these processes will provide valuable insight into the chemical evolution of galaxies.

In this context it is clear that the study of the late stages of stellar evolution of low- and intermediate-mass stars can strongly benefit from the VO ability:

- to perform multi-wavelength analysis of individual sources in order to interpret the rapid changes that these stars experience in such a short period of time, in both continuum and spectral features, especially in the transition phase from the end of the AGB until a new PN is formed;
- to obtain data from different instruments which can cope with objects which are suddenly changing from optically bright to heavily obscured infrared sources and back again;
- to discover new sources in this short transition phase hidden in existing catalogues and archives in order to improve the reliability of diagnostics for the analysis of data on candidate transition sources. At present the literature only contains a very limited number ($\sim$ a hundred) of sources which are known to be in this elusive evolutionary stage;
- to combine proper motion information with radial velocity data in order to derive the kinematic properties, orbits and population membership of these stars.

`VO Tools`

1. for multi-wavelength analysis of individual sources; this can be done by putting together photometric and spectroscopic data taken with various ground-based and space facilities to determine the mass-loss history and the overall chemical properties of the circumstellar gas and dust ejected by the central star and injected in the ISM as a function of the stellar mass/luminosity (e.g., using VOSpec as starting point).
2. for the discovery of new candidate sources in this AGB to PN transition phase; this can be done by determining the characteristic location of transition sources in multi-colour-colour diagrams and/or colour-magnitude diagrams. This would increase the number of sources known to be evolving along this short-lived transition phase in our Galaxy and in other galaxies to study the impact of different metallicity environments on the observed chemistry and on the efficiency of dust formation (e.g., using existing filtering facilities in Aladin and in VOPlot combined with cross-match plugins as the starting point for future developments which are now being extended to cross-match capabilities against SIMBAD to characterize the colours of known sources and to identify new sources not yet identified in the literature in the diagrams).
3. for studying the rapid SED changes that these stars experience; this can be done through the systematic characterization of the SEDs associated to stars known to be evolving in the AGB to PN transition phase. For this purpose, automated classification systems (based on neural networks, knowledge-based systems, or similar) should be developed under the VO environment. These systems must be able to learn by themselves and improve their classification scheme as a consequence of the experience achieved during the classification process (this will require the development of specialized Web services with these goals). (Some partners in VOTech have an interest

in developing such systems.) As an alternative, Web services can certainly be developed which should be able to perform systematic comparisons of a large number of SEDs against a given model (or a set of models) and determine the main physical and chemical properties of the central star and of the circumstellar shell as a result of this comparison (e.g., given a SED, determine the best model fit, or given a model, search for sources that fit a given evolutionary stage predicted by that model, etc.).

4. for studying the physical and chemical evolution of the star and the chemistry of the mass-loss; in some cases the nature of the CS has changed since human records began. Data mining and access to the AAVSO data base will reveal long-term photometric trends, whilst spectra show transitions from, e.g., M to F-type. To date, there has been no attempt at a comprehensive correlation of observed pulsational and chemical changes with the rate and nature of mass loss, although models exist. Line spectroscopy and imaging shows changes in the chemistry of the shell as material moves away from the star, such as the history of the O-C transition. Automatic identification of diagnostic lines (in all wavelength regimes) could be provided by a specialised Web service.

5. for high-resolution studies of the morphology and kinematics of CSE and PNe; this requires optical/IR/radio data (depending on evolutionary stage) with high spatial or spectral resolution or both. Stellar astrometry will reveal evolutionary variability, proper motions and binarity; VOs provide access to data over a long time baseline and need to develop access to tools for astrometric solutions. Space-based or interferometric imaging reveals the nebular morphology and (with good astrometry) the relative motion of the CS with respect to its shell. Surveying and classifying the appearance of a large number of PNe (resolved in the radio or optical) and the properties of their central stars (spectra in the UV/optical/IR) will contribute to settling the heated debated about whether all asymmetric PNe are binary and whether an off-centre CS position is due to a companion or to its relative motion with respect to the nebula as outlined above. The expansion velocity is deduced from multi-epoch imaging. The kinematics of the shell are measured using 1-, 2- or 3-dimensional spectral data (which also yields the distance if the expansion velocity is known). VO tools currently under development can be extended to handle long-slit spectroscopy, data cubes etc. and to handle velocity axes, complemented by increased publication of radio spectral archives.

6. for using polarimetry to measure magnetic fields. Existing tools can display spectra and images in the various Stokes or other polarization parameters but the International Virtual Observatory Alliance (IVOA) needs to establish standards to characterise polarization data before we can provide access to analysis tools. We also need means of displaying and analysing polarization angle vectors. In liason with experts in this field, VOs will eventually provide access to specialised services for comparisons with models of the inferred magnetic fields. As well as their role in the dispersal of PNe, magnetic fields of stellar origin influence the shape of the circumstellar shell. Combined with the studies in 5), systematic comparison of the properties of objects

known to be binary or to lack a significant companion, will enable us to use the nebular morphology and polarization characteristics of obscured systems to deduce whether or not the CS has a companion, and relate this to the chemistry of material returned to the ISM.

All six points above can be generalized to other scientific goals other than the analysis of the AGB to PN transition phase. For example:

(1) the on-the-fly generation of SEDs for multi-wavelength analysis can be applied to any kind of astronomical sources

(2) the search for astronomical sources of a given class based on colour-colour diagrams and/or colour-magnitude diagrams can be extended as well to any type of astronomical source showing characteristic colour properties.

(3) the automated classification of SEDs can be used for spectral type classification of normal or peculiar stars, for the derivation of chemical abundances in stars or nebulae, comparisons with photoionization models can be applied to PNe or HII regions, with circumstellar disk models in the case of YSOs, with galaxy models trying to determine whether the observed SEDs are compatible with the unification scheme, etc.

(4) tools for the manipulation of heterogenous spectral data (for example, comparing radio and IR line systems on the same velocity scale) will assist the interpretation of present and near-future observations nearby galaxies on scales small enough to distinguish the influence of interactions, star-forming regions of varying ferocity, etc., on chemical gradients.

(5) and (6) contain important tools for studying star formation, such as the kinematics of protostellar discs, in the transition from accretion to planet formation, and the role of magnetic support in cloud collapse. Star forming regions are also heavily obscured and have a complex 3D structure. VO methods can be used to investigate fully a 'learning set' of well-understood objects with multi-wavelength multi-resolution observations and establish diagnostics for the age, mass etc. of other protostars for which only partial information is available.

`Why is the VO approach unique?`

The VO approach will be a new, powerful tool for:

1. multi-wavelength analysis of individual sources. For this purpose the VO should be able to put together in a plot data (photometry and/or spectroscopy) taken from various archives/catalogues in order to have a quick look to the overall SED.

   The tool should be intelligent enough to cope with:

   - different physical units
   - different beams/apertures - especially for extended sources
   - astrometric information and its uncertainties - again needed for extended sources
   - different dates when data were taken (for variability analysis; the combination of multi-wavelength data taken at different epochs can in principle be used to derive light curves at different wavelengths)

- display on a velocity axis if suitable metadata are provided
- polarized spectra
- different photometric systems (ideally a Web service should exist providing a database of filter profiles, central wavelengths, zero points, colour corrections, etc.)
- different quality of the input data (data quality information should be always attached to any data plotted in a SED to facilitate reliability analysis and evaluate the need of retrieving original files from the contributing archive/catalogue if needed). The required information must be found either attached to any input data ingested in the VO as metadata information or must be retrievable from specialized Web services providing this information (e.g., the filter database information) automatically during the ingestion process.

VO tools such as SpecView and VOSpec can already perform some of these tasks, such as the construction of SEDs from multiple input tables in heterogenous formats by VOSpec. These tools can perform basic analysis, e.g., black-body (BB) or line fitting. However, VOs should encourage data and model providers to supply user-friendly interfaces to more specialised tools for, e.g., multiple BB fitting, comparison with existing models, automated classification of SEDs under a given scheme, line identification linked to chemical databases. The job of the VO is to develop standards which will allow data to be transferred between such services and provide the interface for easy access.

2. for the discovery of new candidate sources in this AGB to PN transition phase. VO should provide:

   - Filtering facilities applicable:
     - before loading the catalogue from Vizier or any other data server so that only a subset of a given catalogue/data set is ingested for further manipulation with the VO (according to more or less complex user selected criteria).
     - at any moment during the VO session: as a real-time decision of the user.
   - Manipulation of columns
     - especially interesting is the creation of new columns from existing ones (already exists in current VO prototype)
   - Cross-matching tools:
     - between multiple catalogues containing a very large number of sources (at least $10^5 - 10^6$ sources each or even more in the future) all them previously ingested as VOtables in the VO session
     - between one catalogue (or a subset of a catalogue which may have resulted from the application of filter(s) within the VO session) and an external catalogue in Vizier when the latter is too big to be ingested as a VOtable during a VO session ('query by list' option which has just been implemented in the latest version of the VO prototype).

  - between one catalogue (or a sub-catalogue) and the SIMBAD database (this is very important to distinguish between sources which are well known in the literature and those which are not yet identified in SIMBAD for discovery purposes). The result of this cross-match could be one or more new columns added to the original VOtable incorporating for instance the classification code assigned to that source in SIMBAD (best match) and the number of references in the bibliography (also recently implemented in the AVO prototype although still the output is an independent VOtable which still needs to be cross-matched against the original list)
  - the cross-matching tools should be able to deal with the astrometric uncertainties of the sources associated to individual entries in the catalogues to be matched and to return an estimate of the reliability of each cross-identification (potentially incorporating statistical methods using parameters such as sample completeness).
- Easy comparison of spectra and images (superposition of slit direction, aperture size etc.) to identify transition sources where molecular emission is associated with the distinctive morphology of a fossil shell.

3. for studying the rapid SED changes that these stars experience. VO role here should be to provide a link to specialized Web services as explained above. Both automated classification of SEDs and comparison with models should be considered under this area of development. Eventually, it may be possible to schedule robotic telescope observations via VOs (e.g., the ESTAR project) for objects which are passing through the sort of rapid changes seen in Sakurai's Object or IRC+10420.

4. for study of PNe

   - comparison between observed and model spectra of the CS
   - comparison between observed and model spectra of the PN
   - combination of proper motion information with radial velocity data in order to derive the kinematic properties (this will benefit tremendously from GAIA)
   - morphological classification using pattern recognition or other automated image analysis of extended objects.

The role of VOs is not to write tools, but to provide access to tools as well as data. To this end, the VO establishes protocols and standards in consultation with the astronomical community. This project will need strengthening of these standards, in particular:

- Metadata standards for describing polarization
- Implementation of the metadata velocity standards so that spectral tools can be used to investigate multi-wavelength molecular transitions. In the longer term, the VO may provide interfaces to chemical modelling databases for line identification.

In return, VOs need providers of data, tools, and models to liaise with VOs to enable access to data in suitable formats and to their resources via interfaces which also communicate with all other relevant resources.

2.7 Galaxy Formation and Evolution from $z = 10$ to $z = 0.1$

A. Cimatti (Arcetri), D. de Young (NOAO)

`Science Case`

The recent detection of objects at redshifts near 6.5 (and perhaps even as high as $z \sim 7 - 10$), together with the impressive success of the Cold Dark Matter (CDM) scenario to account for the Cosmic Microwave Background power spectrum, have led to statements that Cosmology is now a "solved" problem. However, one of the most crucial and profound astrophysical problems of our time remains unresolved. This is the issue of the formation and evolution of galaxies from the earliest epochs of the Universe to the present day. An understanding of this phenomenon will provide insight into the formation of the first stars, the construction of galaxies as a function of cosmic epoch and the origins of metals in the intergalactic medium (IGM). It will also provide knowledge of how galaxy morphology changes with epoch and how star formation is enhanced or reduced as a function of dynamical, chemical, and environmental factors as they change within an evolving universe. It is the objective of this proposal to outline a means by which this problem can be addressed, using techniques, capabilities, and resources that are only now becoming available through Virtual Observatory efforts worldwide.

More specifically, the current hierarchical Λ-CDM model predicts that the most massive galaxies are formed at late times via a long process of the merging of smaller galaxies. The early work of Madau and subsequent efforts that suggest a peak in star formation and the occurrence of active galaxies around $z \sim 1 - 3$ with little subsequent evolution has provided impetus to this view. However, the data, though sparse, do not provide strong support for this simple model. In particular, the high redshift passive elliptical galaxies and the large number of lower mass systems required at large redshift have not yet been found. In addition there is no evidence for the very high redshift "first light" population of objects that are needed in very large numbers to support this model. It seems clear that much more data at rather small redshift intervals from a few tenths to $z \sim 8$ are needed to examine these questions. Unraveling this mystery of the earliest formation of stars and galaxies and their subsequent evolution is a problem of the highest priority, and it is clear that very large datasets, organized in a manner that make them accessible to a wide segment of the worldwide astronomical community, are essential to developing a solution to this problem. In particular, an organized effort in the form of an AVO Science Reference Mission could address the following key scientific questions:

1. When did the first objects form?
2. What are the progenitors of the present-day massive ellipticals?
3. What type of galaxies populate the Universe at $z > 1$, 2, and 4?
4. How many massive galaxies are already in place at $z > 1$, 2 and 4?
5. How do the star formation and galaxy stellar mass densities evolve?
6. What is the history of metallicity of the IGM at $z > 1$, 2, and 4?

`Required Data`

The ideal dataset for this kind of studies is represented by a deep multi-wavelength coverage of a significant area such as the GOODS and the COSMOS fields. The existing database of the GOODS-South (CDFS) field can be taken as an ideal laboratory for training and to optimize the use of AVO. The GOODS-South (150 arcmin^2) public database includes:

- HST+ACS images in bviz bands (deeper in the Ultra Deep Field)
- VLT and ESO 2.2m images in UBVRIZJHK bands
- optical spectroscopy taken in a variety of surveys
- MERLIN, GMRT, VLA, ATCA radio data[1]
- Chandra and XMM-Newton X-ray data
- Spitzer Mid-IR imaging (3.6-24μ)
- Future submm imaging (APEX? SCUBA?)
- GALEX UV imaging

Another valuable dataset that could be used is the Sloan Digital Sky Survey (SDSS). Most images and catalogues are already available at STScI and ESO Web sites.

`Unique capabilities provided by the VO`

Until very recently the joint use of such multi-wavelength datasets required very specialized expertise in both the knowledge of data characteristics from X-ray to radio and in the reduction of that data. Thus combining such datasets required the coordinated effort of many different people and consumed a great deal of time and funds. The net result was that such efforts have been very rare, and the astrophysical insights made possible via multiwavelength studies have not been realized. The VO completely revolutionizes this situation because its framework will enable easy access to datasets at all wavelengths for all astronomers. The tools and capabilities to do this are already in place or under development at VO sites around the world. This unique software capability, when coupled with the extremely large datasets now becoming available, provides an astrophysical problem solving ability that has never been possible in the history of astronomy. Only now, and only through the VO, are the datasets large enough and the software and network tools mature enough that the Formation and Evolution of Galaxies can be examined in a meaningful way. The VO now makes it possible to determine the History of Light.

`Some VO possibilities`

[1]There is an urgent need for radio imaging datasets with good resolution. The current "archives" available at NRAO and other radio facilities are in the form of visibility data virtually unusable to the vast majority of astronomers. VO organizations need to urge radioastronomy facilities to rectify this situation.

The key aim of the VO should be to provide output results **scientifically usable and reliable**, and not only to act as a simple "cross-correlator" of datasets. Obviously, some human intervention will be needed to develop VO tools and to test their results.

The main steps required by the AVO tools can be (from a purely scientific point of view) summarized as follows:

- Select the band where the sample is to be extracted
- Extract the sample from the whole image or from a sub-section of it, and do photometry with appropriate tools (e.g., the usual SExtractor for optical/near-IR, or simply selecting the objects from an already existing catalog). This step will require inputs from the users to optimize the extraction and photometry parameters.
- Cross-correlate the extracted sample with the existing images, catalogues and spectra. This is one of the crucial steps. In order to build reliable and scientifically usable multi-band catalogues and SEDs, this step should include: matching of the point spread functions (PSFs) of the used images to the PSF of the first image where the sample was originally extracted, use of consistent photometric apertures, correct treatment of the noise and estimate of upper limits for undetected objects, sanity checks such as comparing the observed and predicted colours of stars. These steps are critical for optical/near-IR/mid-IR images, but less for data characterized by a larger PSF (e.g., radio, submm, X-ray data). Correction for Galactic extinction should also be included here.
- The output of the last step should be in the form of: (i) a multi-band catalog with positions, fluxes, flux errors, upper limits, and available spectra (if any) (ii) colour-colour diagrams defined by the user. We strongly suggest to make the colour-colour diagrams **clickable** on individual plotted objects or to have the possibility to mouse-select regions of the plots.
- Photometric redshifts: if the SEDs are built accurately, the user should have the possibility to run photo-z code(s) (e.g., Hyper-z) on all sample objects or only on specific objects selected from the colour-colour diagrams. This step will require some inputs and iterations from the users to optimize the code parameters, to choose the SED template library, the preferred dust extinction curve, etc.
- Physical parameters: photo-z codes can also provide as outputs some physical quantities which characterize the properties of the selected galaxies such as luminosity, E(B-V), SFR, M/L ratio and hence stellar mass.
- It would be great to have on-line libraries of synthetic spectra (e.g., Bruzual & Charlot, Jimenez, PEGASE, etc.) with a wide range of ages, metallicities and star formation histories and IMFs in order to allow the user also to perform a comparison/fitting between the available spectra for the objects in the sample and the synthetic spectra. (This is now possible and was demonstrated at the last AVO demo in January 2005).

- It would be great also to have the chance to ask for morphological information when HST imaging is available and to have software which computes the concentration, asymmetry, clumpiness (CAS) parameters (Abraham et al., Conselice et al.) for faint galaxies, and public software (e.g., GALFIT, GIM2D, GASPHOT) which fits the surface brightness profiles for objects bright enough.

- Another extremely useful performance would be the possibility to produce multi-band thumbnail images to visualize and check how the selected objects look like in all selected images. (This is also already possible and was demonstrated at the last AVO demo in January 2005).

- The possibility to "stack" images at the same or at different wavelengths, or spectra at different redshifts would be extremely valuable in order to search for ultra-faint objects (e.g., visible only in J+H+K bands), to confirm their absence of flux in some bands (e.g., to confirm no flux in U+B+V+R+I bands for I-band dropouts), or even to build "average" spectra for specific classes of objects.

- It shouldn't be impossible also to implement some routines which perform an analysis of the angular clustering of the selected sources in the selected field (e.g., the $w(\theta)$ estimator of Landy & Szalay 1993).

- The availability of theoretical simulations from models of galaxy formations, such as "mock catalogues", would make this game even more promising as the user will have a chance to quickly compare the observations with the model predictions. Most theory groups are already used to make the simulations publicly available on the Web (VIRGO, GALICS, etc,). Thus it is just a question to make them readable and usable in a VO context.

2.8 Build-up of supermassive black holes

P. Padovani (ESO)

`Science Case`

Census of black holes as a function of redshift.

The determination of the time-history of accretion is crucial to our understanding of how supermassive black holes form and evolve. The first step towards this goal is to identify them. However, it is now well established that much of the accretion power in the Universe is absorbed, making it impossible to measure at optical wavelengths. The Chandra and XMM-Newton observatories have revolutionized this field by making it possible to map the history of the Active Galactic Nuclei (AGN) population using hard ($2-10$ keV) X-ray surveys. This is because AGN are ideal tracers of supermassive black holes, as their features make them easily visible and recognizable and the hard X-rays can directly probe AGN activity and are uncontaminated by star formation processes at the X-ray luminosities of interest. Compton-thick sources, that is objects with column density in excess of 1.5×10^{24} cm^{-2}, will however still be missed by hard X-ray surveys. The recently launched Spitzer satellite, with its near- and far-infrared coverage, provides an even more valuable probe of AGN demographics. The obscuring dust which hides AGN from ultraviolet, optical, and soft X-ray surveys should be a largely isotropic emitter at wavelengths $\gtrsim 30\mu$. Furthermore, even Compton-thick sources will have to re-emit the photons absorbed at shorter wavelengths in the far-infrared.

The main aims of this project are the following:

1. expand the work already done in the X-ray band on small areas of the sky (e.g., the GOODS fields) to larger areas to improve the statistics; include sources with enough X-ray counts ($\gtrsim 100$) to derive an X-ray spectrum, which is vital to estimate the absorbing column (most previous works had to rely on hardness ratios). This will also allow us to recognize high-redshift absorbed sources, which can be mistakenly classified as soft sources by using hardness ratios (e.g., at $z = 3$ the rest-frame $2-8$ keV band shifts to $0.5-2$ keV). The XMM serendipitous catalogue will combine a large area with the depth and counts we need;

2. utilize the infrared Spitzer data to complete the black hole census and include also Compton-thick sources; quantify the fraction of such sources so that Chandra and XMM surveys can be properly "corrected" for missing them.

`Required Data`

- XMM serendipitous catalogue
- Spitzer Legacy data: IRAC at 3.6, 4.5, 5.8 and 8.0 μ, MIPS at 24, 70 and possibly 160μ
- ancillary data in various bands: X-ray (Chandra), UV (Galex), optical (ground based).

Strategy

1.
   - estimate photometric redshifts for X-ray sources when no spectra are available
   - correlate the X-ray catalogue with optical/near-IR catalogues and catalogues in other bands; build a Spectral Energy Distribution (SED) to classify the source using all available data
   - estimate the X-ray power by alternative means for the really faint sources, e.g., via the Fiore et al. (2003) relationship between X-ray-to-R-band flux ratio and X-ray power; calibrate first this relationship at faint magnitudes, as there might be a magnitude dependency
   - derive absorbing columns and spectral indices from X-ray spectra to estimate the relative number of absorbed and non-absorbed sources.

2.
   - select AGN using mid-IR (IRAC) colours (see Stern et al. 2004)
   - improve on selection by using far-IR (MIPS) data
   - use available spectra and multi-wavelength information to classify sources
   - estimate photometric redshifts when no spectra are available
   - examine AGN undetected in the X-ray band to look for Compton-thick sources

Why is the VO approach unique?

VO tools will allow to carry out the project in a much faster and reliable way. Required tools include:

- multi-wavelength tool to be able to cross-correlate and overlay images in different bands taking into account the different Point Spread Function (PSF)
- generation of photometric redshifts from catalogues/images by running appropriate codes
- colour-colour tools to facilitate object selection
- data-mining tools to look for outliers
- classifier tools based on the properties of previously known sources
- SED builder and comparison with synthetic spectral libraries
- availability of relevant models, e.g., X-ray spectrum for various absorbing columns.

2.9 The Formation and Evolution of Galaxy Clusters

D. de Young (NOAO)

`Science Case`

Clusters of galaxies are the largest entities in the Universe that are in gravitational equilibrium. Their formation and evolution hold keys to fundamental questions is cosmology, large scale structure formation, the formation and evolution of galaxies, the reionization of the Universe, and the origins of metallicities in the intergalactic medium. Clusters of galaxies and their intracluster media also act as crucial test beds for examining the evolution of AGNs and collimated outflows as well as for examining the details of galaxy encounters and galaxy stripping by dense gas.

Some fundamental questions that could be addressed are:

- What is the necessary environment for the formation of rich clusters?
- Are they already in place when the first density fluctuations begin to appear? Or do they form at later epochs?
- Do clusters only occur at the intersections of cosmic filaments?
- Do clusters always form from subassemblies of small groups?
- Is there a distribution in the epoch of formation of clusters? Why?
- When does the Intra-Cluster Medium (ICM) form? When does it first reach virial equilibrium?
- What is the origin of the ICM?
- What is the frequency of cluster-cluster interactions as a function of epoch?
- What is the role of AGN in the evolution of clusters? As a function of epoch?
- What is the role of AGN in the evolution of the ICM?
- When does the presence of dark matter first become significant in the evolution of clusters?
- How does the galaxy population in clusters evolve as a function of epoch?

`Required Data and Strategy`

A multiwavelength, multi-institutional approach will be the most effective in addressing this very large topic. In addition, extensive comparisons between large-scale databases and very detailed theoretical simulation datasets will be needed.

X-ray data from all archives will be essential in looking for the first signatures of the ICM and its evolution, including its metallicity, as a function of redshift. Moreover, X-ray data may be of some use in characterizing AGN in clusters that are at lower redshift.

Radio data are invaluable in detecting the first onset of AGN activity out to large redshifts. The presence of AGN indicates that at least one galaxy has formed a massive central black hole at that epoch, and finding this at large redshift is extremely important. Moreover, the increasing awareness that radio sources may have a significant role in influencing the evolution of the ICM and perhaps other cluster features makes it imperative to have much more complete radio data. In addition, the superposition of radio, optical and x-ray data immediately gives a wealth of information about the AGN-ICM interaction. Hence image archives at these wavelengths, together with superposition tools such as Aladin will be essential.

Optical and IR data archives will of course be central in determining redshifts, morphologies of clusters and galaxies as a function of redshift, metallicities, emission line signatures, indicators of interactions, etc.

Theoretical simulations are now able to address the entire range of physics involved, from the first formation of large scale structure through the dynamical interactions of galaxies and clusters, the evolution of the ICM, cluster mergers with ICM, the injection of energy and mass from AGN, galaxy stripping, etc. Hence, extensive use of present and future theoretical simulation databases will be essential to the success of this project.

2.10 Correlation of Cosmic Microwave Background, radio/mm and optical/NIR Galaxy Surveys

P. Lehay (Jodrell Bank), T. Banday (MPA)

`Science Case`

The study of full-sky maps of emission in the microwave regime provides considerable opportunities for the VO world of federated data archives to enable the study and disentangling of various cosmological and astrophysical effects. Whilst the Cosmic Microwave Background (CMB) anisotropy is the main scientific driver of such observations, there is overlap and complex interplay with other branches of astronomy over a variety of angular scales. The large-scale emission from our own Galaxy is an important contaminant of the CMB observations which must be modelled and corrected to high accuracy making use of multi-wavelength observations and inferred frequency dependencies. On smaller scales, individual bright objects much be excised from the maps, whilst the integrated contribution of fainter objects to the power spectrum must be corrected for. However, emission from galaxies and clusters of galaxies create interesting secondary anisotropies in their own right, which can give insight into the physical nature of the objects or universe over a range of redshifts. Below we summarise a few more detailed examples of this interplay.

One important aspect for all of the full-sky diffuse emission survey work, however, is the need for a common pixelisation scheme. The current standard in CMB science is HEALPix (`http://www.eso.org/science/healpix/`). It should also be noted that such partitions of the sphere can also be used as indexation schemes, which allow the fast selection of catalogued objects on the sky.

1. Integrated Sachs-Wolfe Effect

   The majority of the observed CMB fluctuations originate at or near the surface of last scattering. However, a significant contribution to these fluctuations in those Universes with either spatial curvature or an effective cosmological constant can arise from the passage of the CMB photons through a time-varying gravitational potential. This so-called integrated Sachs-Wolfe (ISW) effect is generated after the Universe ceases being matter dominated and can contribute significantly to the CMB fluctuations on large angular scales.

   Since the phases can be correlated with the matter distribution, one may attempt to disentangle the ISW effect from other sources of CMB anisotropy by cross-correlating the full-sky CMB emission maps with probes which trace the large-scale matter distribution and are thus sensitive to the evolution of the gravitational potential at late times. Since such effects are expected in both cosmological constant dominated cosmologies and the so-called 'quintessence' models in which the dark-energy component itself undergoes time-variation, the change in the magnitude of the ISW effect with redshift interval affords a useful discriminant between these scenarios.

   The two key aspects of the project relate to: i) generating a map of the CMB sky cleaned of Galactic foreground emission and nearby point sources; ii) the production of a homogeneous catalog of objects to trace the large-scale gravitational potential at the moderate redshifts at which the ISW effect imposes observable structure on the CMB.

A good tracer of this low-redshift distribution should satisfy several criteria:

(a) the tracer objects must be numerous enough that the Poisson error term in any cross-correlation is small

(b) the tracer must probe as large a volume of redshift space as possible, to maximise the cross-correlation signal

(c) the tracer survey area should cover a large angular fraction of the sky to minimise the "sample variance" effect due to incomplete sky coverage in the CMB map.

`Project Outline`

- combine WMAP, Planck etc. CMB sky maps at various frequencies to minimise Galactic foreground signals or utilise surveys of Galactic emission at wavelengths where particular components of the emission (synchrotron, free-free, dust) are dominant as tracers of the foreground emission. Multiwavelength measurements may allow the construction of templates for subtraction from the CMB anisotropy data.
- from radio/IR surveys determine local sources to be removed from the CMB sky maps
- identify potential tracer objects of the large-scale matter density distribution from multi-wavelength catalogues as a function of redshift interval
- produce a homogeneous catalogue of tracer objects, accounting for the selection functions and other catalogue specific details apply reconstruction methods to the tracer objects to recover the underlying gravitational potential at the optimal redshift(s) to probe the ISW cross-correlate the template potential with the CMB sky map

`VO aspects`

- Federation of relevant datasets including interchange/merging of meta-data
- Identification of suitable objects and redshift intervals by appropriate query applications
- Generation of 2D binned object data set for cross-correlation purposes
- Visualisation of data-sets
- Statistical analysis tools to compute cross-correlation, comparison to theoretical models (possibly represented by an ensemble of simulated data-sets which must also be derived from the archives)

2. Sunyaev-Zel'dovich Effect

The Sunyaev-Zel'dovich (SZ) effect arises from the inverse compton scattering of photons by plasma in the hot intra-cluster medium. The SZ effect is in fact made up of two separate effects: One is due to the bulk velocity of the cluster, and the other due to the thermal velocities of the electrons in the cluster gas. These are called the kinematic and thermal SZ effects respectively. The kinematic effect measures the cluster peculiar velocity, whereas the thermal effect can be used, in conjunction with images and spectra of its X-ray emission, to study the cluster gas. The two components have different frequency dependencies and can therefore be separated by multi-wavelength observations. The thermal effect is null at a frequency of 210 GHz.

In the context of full-sky CMB surveys, the SZ effect can be viewed in two ways. Firstly, as a foreground to the primary CMB anisotropies, which must therefore be modelled and removed from the data. Secondly, as a tracer of galaxy clusters on the sky. It is expected that the Planck Early Compact Source Catalogue (ECSC) and Deep Early Compact Source Survey (DECS) will contain approximately 5000 and 240 clusters of galaxies, and indeed significant numbers of very distant clusters. Such an unbiassed survey of the sky will provide a catalogue of clusters to be followed-up in the optical, X-ray, and indeed by dedicated SZ telescopes such as AMI and APEX.

These latter high sensitivity, high angular resolution instruments will allow high-resolution imaging of the SZ effect in cluster sources thus allowing us to understand the dynamical processes within the cluster. The capacity to resolve embedded sources (radio cocoons, other point sources, lensed images of background galaxies) will improve the understanding of the SZ measurements and their implications for cosmology.

`Project Outline`

- extract candidate clusters from full-sky survey using a matched filter technique and frequency information
- verify sample of galaxy clusters using X-ray/optical data
- utilise optical multi-colour images to derive photometric redshifts
- search for IR correlation and quantify galaxy evolution in clusters
- determine correlation with radio surveys to identify the frequency of radio galaxies in clusters, search for radio halos
- follow-up observations with dedicated SZ telescopes for detailed imaging

`VO aspects`

- Federation of relevant datasets including interchange/merging of meta-data
- Verification of candidate cluster members by appropriate query applications to optical/X-ray catalogues
- Acquire multi-color information to determine photometric redshifts

- Identification of candidate radio galaxy cluster members by querying radio catalogues with search criteria (e.g., location) tailored to the derived cluster sample
- Visualisation of multi-wavelength cluster data
- Deprojection algorithms to allow study of morphology in survey data.