VO Science Goals, Tools and services

Mark Allen

Centre de Données astronomiques de Strasbourg

Data Centre Alliance – Project Scientist

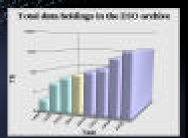
IVOA – Applications Interest Group Chair

Outline

- Introduction to Virtual Observatories
- Science with the VO
  - Goals
  - Science Rerence Mission
  - Example: Search for type 2 Quasars
- Tools and Services available now
- Practical – trying some VO tools & services

Introduction to VO

Why do we need VO?



- Imminent Technical Challenges
  - Data volumes and rates
  - Major archives growing: Tbytes/yr
    - Doubling time: 6-12 months
    - Faster than Moore's Law (18 months)
    - Device access rates & last mile of thin wire are the real bottle necks

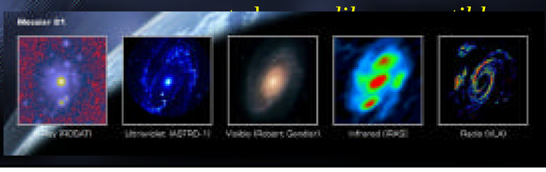
➔ *Take the computation to the data*

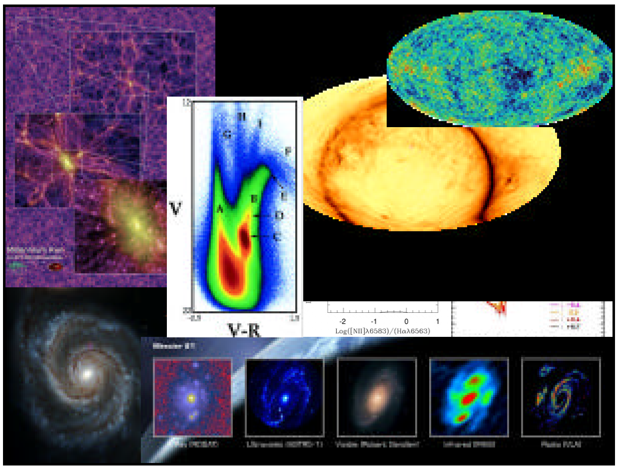
➔ *Ship the results, not the data*

Scientific Motivation

Multi-● science requires

- Data from different telescopes
- Analysis tools
- on-line services
- archived information





Virtual Observatories

- VO = framework for interoperable systems
- VO Vision: *All Astronomy resources as if they were on your desktop*
- *"Observing the digital sky"*

Interoperability

- Common query language
- Uniform interfaces for diverse data
- Analysed by the same tools
- Astronomical interoperability
  - Coordinates, Units, photmetric systems, ...
  - Describing content in standard way
- The key to interoperability is the use of Standards

Core Components

- Finding available data, services
  - Registries
- Accessing the data and services
  - Data Access layer protocols: SIA, SSA
  - Common query language: VOQL
  - Contents Description: Metadata, formats
  - Workflow: Linking services together

Standards

- Registry standards
  - Describing resources/services
- Simple Image/Specta Access Protocols
- Format standard: VOTable, (FITS)
- VO Query Language
- UCD: semantics
- Data Model
- Web Services

Status

- ***IVOA Standards***

Status

- Milestone: Phase-A studies finalised
  - AVO, AstroGrid, NVO
  - Moving on to building real services
- Prototype VO tools - Scientific Results
  - ***Many projects/archives adopting standards***
  - ***Tools developing and maturing rapidly***
- ***Scien***tific Content
  - ***Data & Service providers motivated to 'publish' to the VO***

Science with the VO

Example science drivers

- Find all info on a given set of objects
  - Defined by positions, colours, morphology,...
- Build SEDs from multi-archive data
  - Accounting for instrumental, sensitivity, aperture effects
- *"Outlier Science"* Multi-d parameter searches
- Compare LSS with *'virtually observed'* N-body simulations
- Re-analyse the SLOAN, MACHO
- Multi-● census of AGN
- Build a survey to search for Cosmic Shear

AVO Science Reference Mission

- AVO Science Advisory committee (2005)
- Define the key scientific results that a European VO should be able to achieve when fully implemented
- Consists of science cases over broad range of astronomy and the related requirements

Science Cases

- Circumstellar disks: from pre-Main Sequence stars to stars harbouring planets
- Intermediate Velocity Clouds
- Which Star will go Supernova next?
- Initial Mass Function within 1kpc: Planetary to Stellar Masses
- Initial Mass Function for Massive Stars
- Contributions of Low and Intermediate Mass Stars to the ISM
- Galaxy Formation and Evolution from z=10 to 0.1
- Build-up of Supermassive Black Holes
- Formation and Evolution of Galaxy Clusters
- Correlation of CMB, radio/mm and optical/NIR Galaxy Surveys

Galaxy Formation and Evolution from z=10 to 0.1

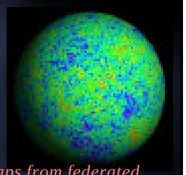
- When did the 1st objects form?
- What are the progenitors of present day massive ellipticals?
- How many massive galaxies at z>1,2,4?
- How do SF and galaxy stellar mass densities evolve?
- ***Required data***
  - ***Deep Multi-wave surveys (GOODS, COSMOS)***
  - ***HST+ACS bviz imaging***
  - ***SLOAN***
  - ***Optical spectroscopy***
  - ***MERLIN, GMRT, VLA, ATCA radio***
  - ***Chandra and XMM-Newton X-ray***
  - ***Spitzer mid-IR***
  - ***Future sub-mm***

"Only now, and only with through the VO, are the datasets large enough, and the tools mature enough that Galaxy Formation and Evolution can be examined in a meaningful way."

VO steps

- *Extract sample from data*
  - Perform Sextractor type photometry
    - Cross correlate with images, catalogues, spectra. *Crucial that output results are scientifically useable and reliable* Matching of PSF, consisitent photometric apertures, treatment of noise, and upper limits
      - Sanity checks like stellar colours
        - Output multi-band catalogue, and colour-colour diagrams
      - Visualize output colour-colour space
    - Photometric z from SEDs (Template SED libraries, extinction curves etc.)
    - Physical Parameters – L, E(B-V), SFR, M/L ∈ stellar mass
  - Comparison with star formation scenarios and synthetic spectra
  - Morphological analysis
    - Stack images at same wavelength, or spectra at different redshifts
      - Build average spectra for specific object classes
        - Angular clustering analysis
  - Comparison with mock catalogues from theoretical simulations

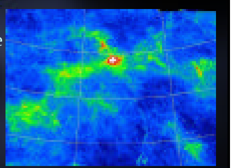
Correlation of CMB, radio/mm and optical//NIR Galaxy Surveys



- Integrated Sachs-Wolfe Effect
  - *CMB fluctuations from passage through time varying gravitataional potential*
- Sunyaev-Zel'dovich Effect
  - *Inverse compton scattering of photons by plasma in the hot intra-cluster medium*
- ***Required data***
  - ***WMAP***
  - ***Planck***
  - ***radio/IR surveys***
  - ***X-ray/optical cluster data***

Study of full-sky maps from federated archives to disentagle various cosmological and astrophysical effects

Intermediate Velocity Clouds



- What is the origin of neutral gas clouds moving with unexpected velocities in the galaxy ?
- ***Required data***
  - ***All sky far-IR surveys DIRBE***
  - ***IAR, HIPASS 21cm***
  - ***SHASSA H•***

- VO Steps

  Identify regions of excess HI or H
  To detect IR IVC, remove foreground using HI
  Check contamination by warm ISM
  Classify dust rich/poor IVCs
  Analyse spatial distribution

Common requirements

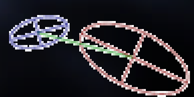
- ***Browsing/searching of data and distributed information***
- ***Manage large amounts of distributed heterogeneous data***
  - ***Combining Multi-wavelength data taking into account different:***
    - ***Units***
    - ***coverage***
    - ***resolutions/PSF, observing technique***

Common requirements (cont.)

- Multi-wave cutouts of individual sources
- Generate and visualize SEDs from image, and spectral and catalogue data
  - Taking into account different
    - Beams/apertures (extended sources)
    - Backgrounds
    - Photometric systems
- Time axis:
  - Light-curves
  - Multi-epoch imaging

Common requirements (cont.)

- Cross-Matching
  - Catalogues: Large-Small, Sparse-Dense
  - Taking into account:
    - positional uncertainties, resolution, completeness
    - extra constraints: e.g. colour, object type, environment
  - Statistical confidence of results

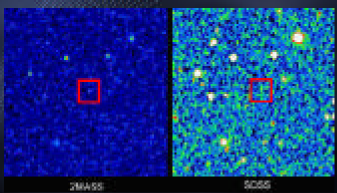


Common requirements (cont.)

- Compare observations with models
  - Virtual observations of models
  - Projection of models to observed parameter space
  - ***Spectral fitting/classification***
  - ***Colour-colour*** **visualization** ***Tool***
  - ***Astronomy functionality alongside visualization***
    - ***Reproject data, correct for extinction, calulate luminosities etc.***
    - Visualization requirements ∈ Analysis requirements

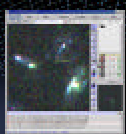
Science Demos and early VO Science

- Discovery of Brown Dwarf -

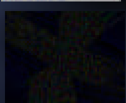


- R&D on scientific requirements and technology for building a VO
- 6 European organisations
- 3 yr phase-A, 2001-2004/5
- Driven by strategy of scientific VO demonstrations

Demo 2003: First Light
VO Visualisation and Data Discovery



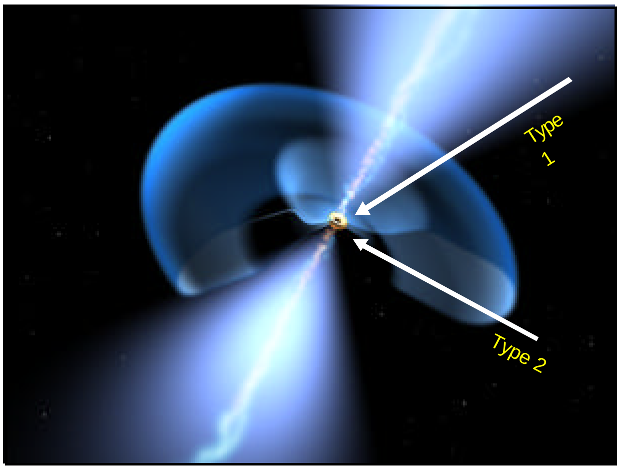
Demo 2004: First Science
Searching for type 2 Quasars using Virtual Observatory tools



Demo 2005
AGB to Plantary Nebulae transition
Workflow: SEDs to Models

AGN Unified Model

- Different types of AGN explained by the physics of the
  - Black hole
  - Accretion disk
  - Jet and
  - Torus
- Convolved with the geometry of the viewing angle
- Urry & Padovani (1995), Jaffe et al. (2004)

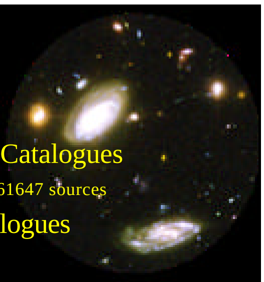


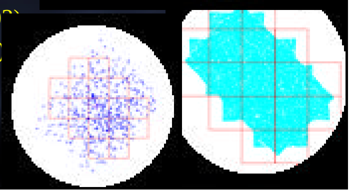
Type 2 AGN

- Local *low power* type 2s : Seyfert 2s
- High power counterparts: QSO 2 *(type 2 quasars), difficult to find.*
- *Heavily reddened and fall through optical/UV selection*
- *However…*
  - *Hard X-rays can penetrate the torus*

Data

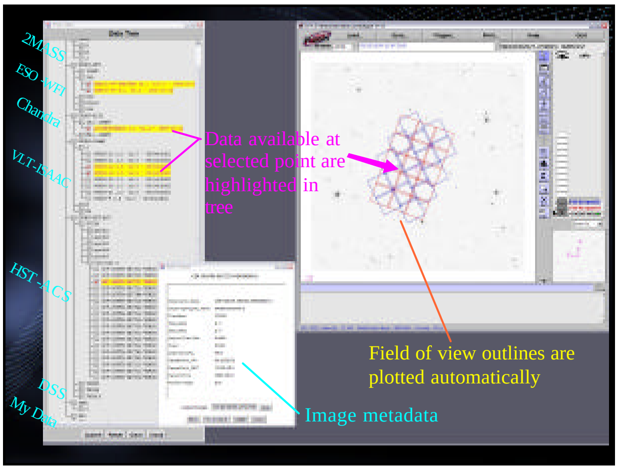
- GOODS: *Bviz* Images & Catalogues
  - CDF-S (+UDF), HDF-N : 61647 sources
- Deep X-ray Chandra Catalogues
  - Alexander et al. 2003
  - HDF-N 2Ms (50
  - CDF-S 1Ms (326)





AVO prototype

- Registry of services (GLU)
- CDS Aladin interface
  - Interactive manipulation of image and catalogue data
  - "Portal" for access to services/data
- Cross-matching service for catalogues
- Conventions for accessing remote data
- Remote calculations
- Interoperable with other VO tools

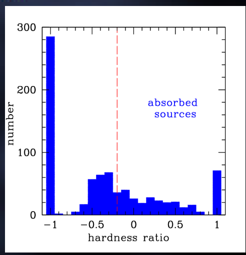


Method

- Select candidates based on two key physical properties of Type 2 AGN:
  - Obscured – absorbed
    - Chandra X-ray catalogues
  - High power – to be classed as AGN
    - Empirical relation for Lx using flux(2-10 keV) / flux(R-band)

Absorbed sources

- HardnessRatio
  HR=(H-S)/(H+S)
  S = 0.5 – 2.0 keV
  H = 2-8 keV
- Type 2 AGN have HR > -0.2
- Increasing z makes sources appear softer therefore discard some high-z type 2s
- 294 absorbed sources (CDF-S: 104, HDF-N:190)

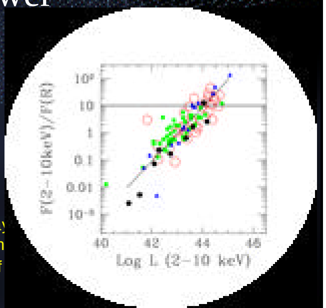


Optical counterparts

- X-match absorbed sources with GOODS *z*-band catalogs
- Detect and correct for systematic shift
- Take positional uncertainty into account
  - (Match distance)/error < 1
  - Most match distances < 1.25”
  - False match estimate: 8 – 15%
- Almost all (HDF-N, CDF-S) X-ray sources have optical counterparts

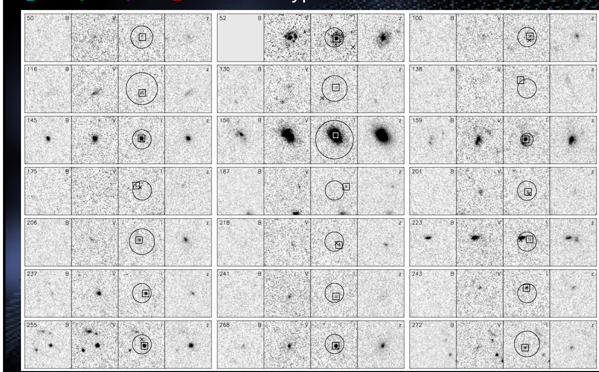
Estimate X-ray Power

- Fiore et al. (2003) relation for type 2 objects
- flux(2-10 keV) / flux(R) ~ L(nucleus) / L(host galaxy R)
- Since R luminosity (unlike X-ray power) show modest scatter, th flux ratio is a good estimator of ray power



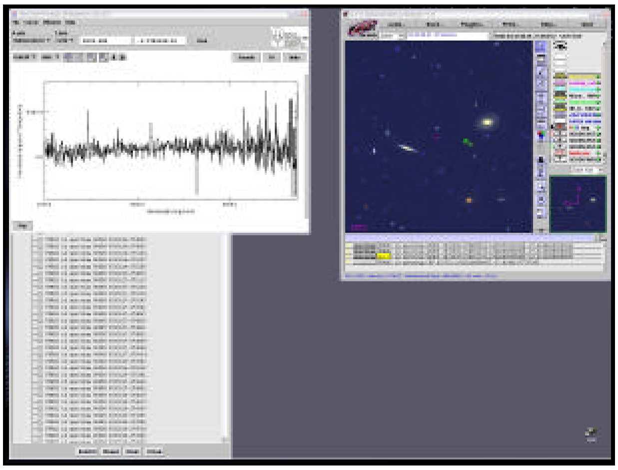
Log L(2-10 keV) = Log f(2-10 keV)/f(R) + 43.05 (erg s^{-1})

B V *i* *z* Type 2 AGN candidates: CDF-S



Results

- 147 type 2 AGN
- 27% of the 546 X-ray sources
- But some of these sources are already known…
  - Spectroscopically classified as type 2 AGN
    - Szokoly et al. (2004)
    - Barger et al. (2003)



Known sources

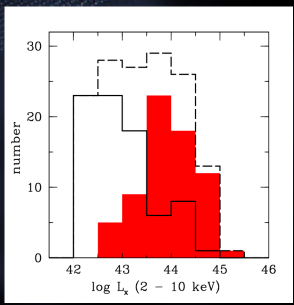
- Remove known sources from candidates, and use as a check on the L_x estimation method
- Identified by X-match with Szokoly et al. (2004) CDF-S, and Barger (2003) HDF-N.
- Check of estimated Lx
  ‹log $L_{x,est}$› = 42.57±0.08 compared to ‹log L_x› = 42.49±0.09
  ...consistent

Results : NEW Type 2 AGN

- 68 new type 2 AGN candidates
- 31 have $L_x > 10^{44}$ erg s^{-2} : QSO 2
  - Only 9 previously known in GOODS fields
  - Now 40 QSO 2s: Quadrupled the QSO 2s in the GOODS fields !

Luminosity distribution

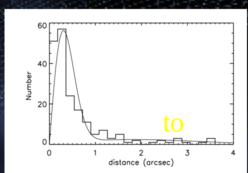
- New type 2 AGN
- Fills a gap luminosity distrib.
- AGN 2 $<z_{est}>$~2.9
- QSO 2 $<z_{est}>$~3.7



QSO 2 surface density

- >330 QSO2 deg^{-2} (down to 10^{-15} erg $cm^{-2}s^{-1}$)
- Consistent with Perola et al (2004) predictions for $>10^{-14}$ erg $cm^{-2}s^{-1}$
- ~5x higher than predictions for faint flux limits
  - Gandhi et al. (2004) predicts 19 deg^2
  - We find ~ 100 deg^2
- Resolved X-ray background : 10±2% down to f(2-8keV)=10^{-15} erg $cm^{-2}s^{-1}$

False Match Rate



- Fit X-match distances model
- Expected match rate D

$$N_1\Phi(d,\lambda) + N\Phi(d,\lambda)\left[\frac{1+2\pi\lambda\sigma^2}{2\pi\lambda\sigma^2}\exp\left(-\frac{d^2}{2\sigma^2}\right) - 1\right] \quad (A.1)$$

2-d Poisson distribution: $\Phi(d,\lambda) = 2\pi\lambda d \exp(-\pi\lambda d^2)$

True matches: $\alpha = \frac{N}{2\pi\lambda\sigma^2}\Phi(d,\lambda)\exp\left(-\frac{d^2}{2\sigma^2}\right)$

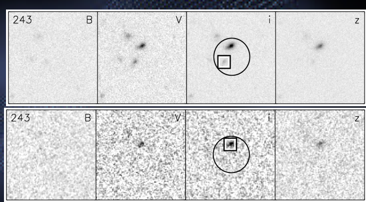
True partner, but correctly assigned: $\beta = N\Phi(d,\lambda)\exp\left(-\frac{d^2}{2\sigma^2}\right)$

No counterpart, but have been matched: $\psi = (N_1 - N)\Phi(d,\lambda)$

Ultra Deep Field

UDF

GOODS



- 3 AGN-2 candidates fall within UDF
- 1 has closer, fainter optical counterpart in UDF

Science Conclusions

- Using
  - the deepest Chandra X-ray, and HST imaging
  - & Empirical estimator for Lx we find
- 68 New type 2 AGN
- 31 Qualify as QSO 2, z~4
- Many more QSOs than predicted

Published: Padovani, Allen, Rosati, Walton A&A 2004

'First Science' (Jan 2004)

- Prototype VO tools for science
  - *Aladin & CDS services, AstroGrid*
  - *Using distributed information*
- Enabled by real gains in standards for:
  - Data access
  - Manipulating image and catalogue data
  - Remote calculations

- Difficulties
  - *Minimal interoperability between tools – required local saving*
  - *Getting info from one tool to the next*
  - *Data size limitations for client tools*
  - *X-match limited by local memory*
  - *Standards changing*
- *Now possible to do this in more streamlined way*
  - *next lecture – Tools & Services*