AstroGrid: Virtual Observatory Service

Nicholas Walton
AstroGrid and Euro-VO Technology Centre
Project Scientist
(Institute of Astronomy,
University of Cambridge)

Outline: Today

- Data and Analysis challenges for astronomy
  - science
  - technical
- Building a solution: The Virtual Observatory
  - Alliances and standards
  - Global Projects
- AstroGrid: In detail
- AstroGrid: Architecture
- AstroGrid: Science Usage Examples
  - IPHAS – a Galactic Plane Example
- Check: AstroGrid workbench works on YOUR machine

Outline: Friday

- Run through of 'live' AstroGrid
  - demonstration of key capabilities
  - discussion of Python scripting access to AstroGrid
- Practical Session
  - you use AstroGrid
  - Workthrough examples from http://www2.astrogrid.org/science
  - discover data using AstroScope
    - visualisation
  - database access – IPHAS, SDSS, 2MASS
    - colour-colour diagrammes
- Feedback and Summary Session

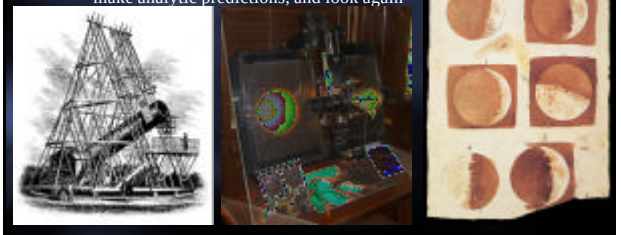
Why Virtual Observatories?

Traditional Observatories: e.g. ESO, HST

Key VO Outcome: enable access, to the community, to the best astronomy resources (data, compute, algorithms)

The Evolving Scientific Process: 1

- Astronomy is an observational science
  - observe the sky & analyse the observations
  - compare and contrast with models
  - make analytic predictions, and look again

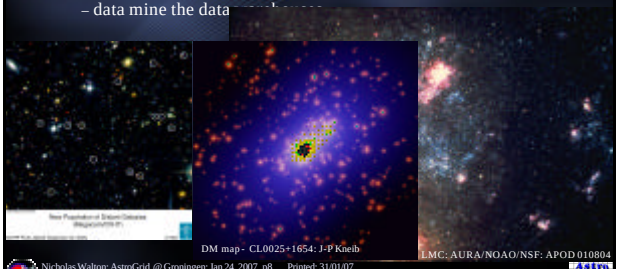


The Evolving Scientific Process: 2

- Astronomy becomes a computational science
  - observe all the sky in many colours
  - generate large scale simulations
  - compare and contrast with observations

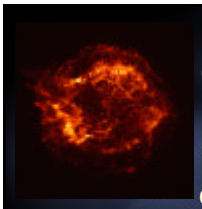
The Evolving Scientific Process: 3

- Astronomy moves to the extreme
  - observe all the sky, in many colours, many times
  - compare & contrast with sophisticated multi-dimensional models
  - data mine the data

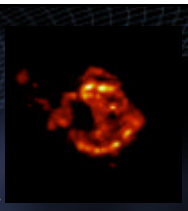


Supernova Remnant Cassiopeia-A – a 300 year old Supernova

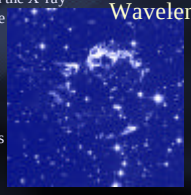
The Challenge and Opportunity of multi-Wavelength data:



Shocks seen in the X-ray Chandra image



Dust shows in the IR

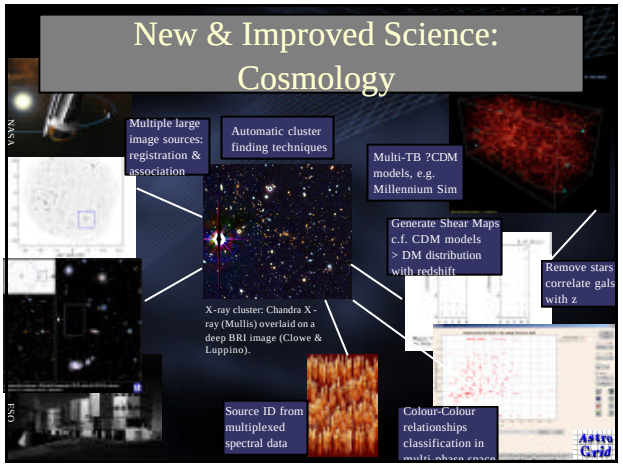


Heavy elements seen in the optical

Mapping e⁻s in the magnetic field as revealed by Radio data

Images from Chandra Science Centre

New & Improved Science: Cosmology



New & Improved Science from VO's: Space Weather

What happens to the Earth's magnetosphere during a coronal mass ejection ?

Event imaged by space based solar observatory

Effect detected later by satellites and ground radar

SOHO/EIT – EUV

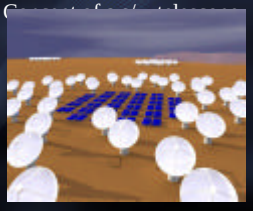
Yohkoh – Xray

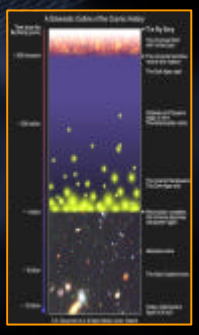
The Need for Virtual Observatories: Technological Advances

- The massive Growth of Data
  - Number + size of telescopes
    - Optical: ESO's 4x8m VLT, 2x8m Gemini
    - X-ray: XMM-Newton
    - sub-mm: ALMA
  - Increase in size and multiplex capabilities of instrumentation:
    - Infra-Red: VISTA > 100 GB/nights
    - Radio: e-Merlin > data rates ~320 Gbps

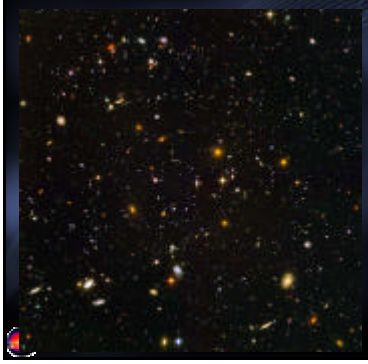
SKA: future radio challenge

- Huge processing challenge
  - data rates at the >100 TB/s
  - local processing
  - GB/s data flows of science product
- C[illegible]



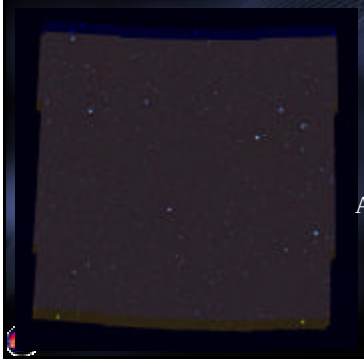


The need for a VO



- Hubble UDF
- Million second exposure
- 6000x6000 pix
- 11.5 sq. arcmin
- 10,000 galaxies

The need for a VO



- SWIRE ELAIS N1
- 9 sq. degrees (~3000 UDF)
- (moon ~0.2 sq. deg.)
- ~ 600,000 objects

All Sky – 40,000 sq. deg!

Solutions: a Flexible Framework

- Create a system that recognises:
  - no one data provider or repository: thus data interoperability
  - application provision
- Requires
  - a system built upon agreed interoperability standards
- Exploits
  - wider IT developments: Grid and WS technologies
    - power of XML/ SOAP etc
  - access to high speed networks
    - but note: backbones ~10Tb/s, desktops ~100Mb/s
  - reduced costs of h/w: all data now on spinning disks

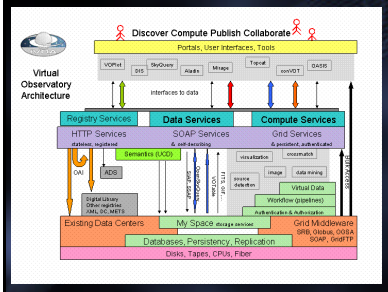
Building the Virtual Observatory:

Global scope
International partnerships
Agreeing interoperability standards
Building regional implementations
Based on new computational technologies
Deployed on the fastest networks

IVOA: Stds Enabling Interoperability

- The International Virtual Observatory Alliance http://www.ivoa.net
- A global partnership
- Projects represent global astronomy data providers
- IVOA a forum for interoperability standards
- VO projects build on these standards
- Global reach

IVOA Architecture Analysis



Analysis of a VO:

- Multi-layer
- Complex
- User interfaces thru a portal
- Astro-apps interface to a VO abstraction layer
- Lower level middleware provided by the 'grid' world
  - e.g. SRB
- Hardware at bottom layer

Ref: IVOA Architecture Overview: Williams et al, 2004 IVOA Note 2004-06-14: http://www.ivoa.net/Documents/Notes/IVOAArch/IVOAArch-20040615.html

IVOA has working groups to address 'astro' specific issues

IVOA Working Groups: http://www.ivoa.net/forum

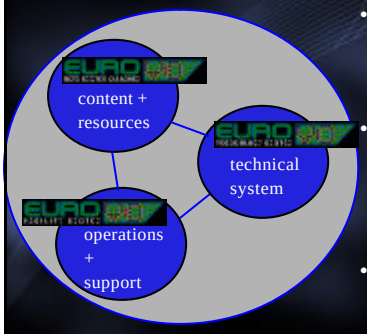
- Registry:
  - how to 'register' resources: concept of VOResources
- VOEvent: alert notifications – supporting e.g. GRB studies
- Data Access Layer
  - Standards for remote data access: e.g. SIAP, SSA
- Data Model
  - Standards for the actual data: e.g. XML'ing of FITS
- VO Query Language
  - Standards for 'astro' database access: e.g. Openskyquery, 'circle'
- Unified Content Descriptors
  - Standards for common ways of describing data: metadata
- VOTable
  - XML representation of tabular data

IVOA: Interest Groups

Aim: issues of importance across working groups

- Applications
  - Interfacing new and legacy apps (e.g. Iraf) to the VO
- RadioVO
  - Issues relevant to 'Radio': e.g. The UV plane, interferometry
- Theory
  - Simulations, mass scale compute
- Data Curation & Preservation
  - Linkages to the Digital Libraries world, a-ph, ADS
- Architecture
  - Fitting it all together
- Networks
- Semantics

AstroGrid and the Euro-VO



- Euro-VO: balancing
  - Technical System with
  - Content + Resources and
  - Operations + Support
- All elements science driven
  - VO Technology Centre
  - VO Facility Centre
  - VO Data Centre Alliance
- Euro-VO Science Balance via Euro-VO Science Advisory Group

Euro-VO: VOTC Technology Centre

Infrastructure, new tools, resource discovery, data mining+visualisation

Euro-VO : Technology Centre

- Work programme organised into thematic areas
  - Aim to produce robust overall design for the Euro-VO
  - Eventual implementation across Europe from 2007 onwards
- Infrastructure: the VO middleware
  - Workflows, job execution, security, transport layer etc
- New Tools: applications for the VO
  - Footprint, best fitting, SED builder, etc
- Resource Discovery: finding the needle in the haystack
  - Building ontology's, dictionaries, resource browsers, etc
- Data Mining and Visualisation: mass scale analysis
  - Large scale compute, multi dimensional visualisation, etc

AstroGrid in Detail

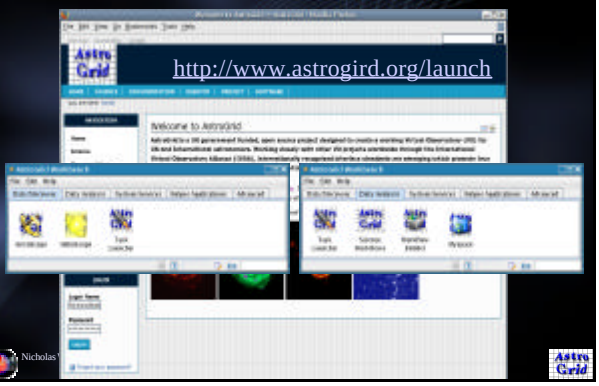
AstroGrid: UK's Virtual Observatory

Empowerment of scientists

- Improve the quality, ease, speed and cost effectiveness of on-line astronomy
- Make comparison and integration of data seamless
- Removing barriers to multi-wavelength as
- Enable acces to very large data sets

University of Leicester, University of Bristol, University of Cambridge, CCLRC Rutherford Appleton Laboratory, Jodrell Bank Observatory, University of Portsmouth, Queen's University Belfast, UCL MSSL

AstroGrid 2006.3 Release: Aug 2006

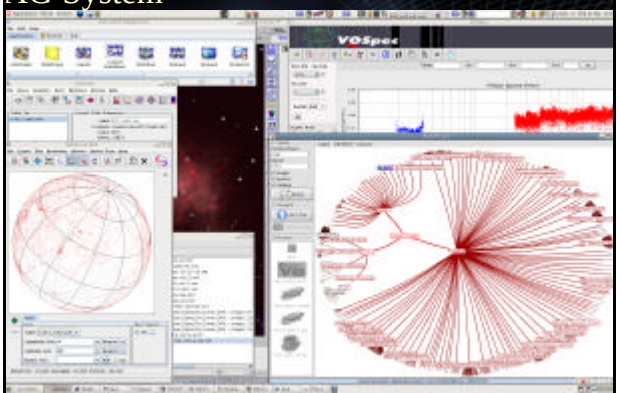


AstroGrid: A place for science

- Simple workbench access to VO services
- Concept of 'MySpace': virtual user space for data, workflows, results
  - ability to share with research teams, collaborators
  - longer term – support publication of user data: analagous to ebay, c.f. astro-ph vs ADS
- Powerful, yet simple tools to enable data discovery
  - astroscope
- System to support creation of user defined workflows
  - find data, process data, integrate data
  - save workflows
- Range of applications on server and client side

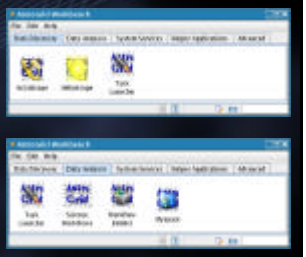
S2-147 a SNR
5° x 5° Rana
Credit: A Zilistra, Irwin

AG System



AstroGrid Workbench

- User Interface to VO services
- Delivery via Java Webstart technology
- Components
  - Registry
  - Find Data
  - Work with Apps
  - Workflows
  - Client Visualisation
- Enables Science



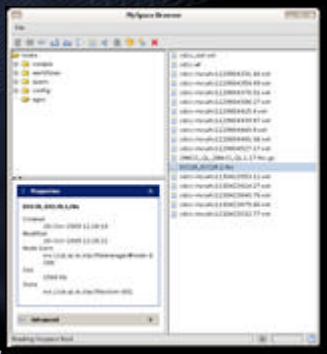
Registries – our 'AstroGoogle'

- As an astronomer – how do you find the data that you require?
- VO solution: Registries, used to discover and locate resource
- A list of resource descriptions, described by structured metadata: enables automated searching and processing
- Types of resource
  - Generic services, web services, applications, …
  - Data collections
  - VO-specific resources (e.g. MySpace servers)
- Collects all global information

US VO, EuroVO, AstroGrid, JVO, etc

AstroGrid: MySpace

- Virtual disk space where you can store results, temporary files, and new things like query files and workflow files, so you can adjust and re-run jobs on a later day.
- Visible from any computer.



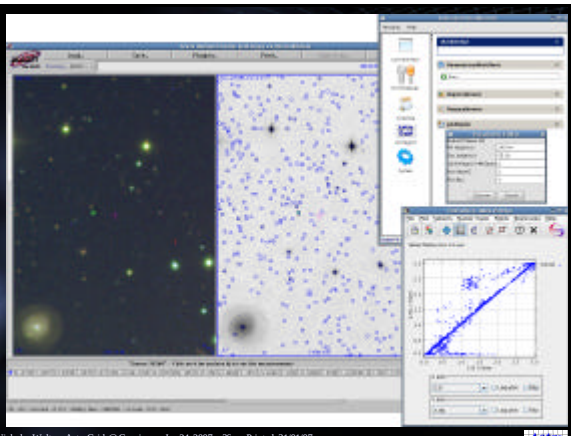
Server Side Workflows and Science Services

AstroGrid Redshift Science Service

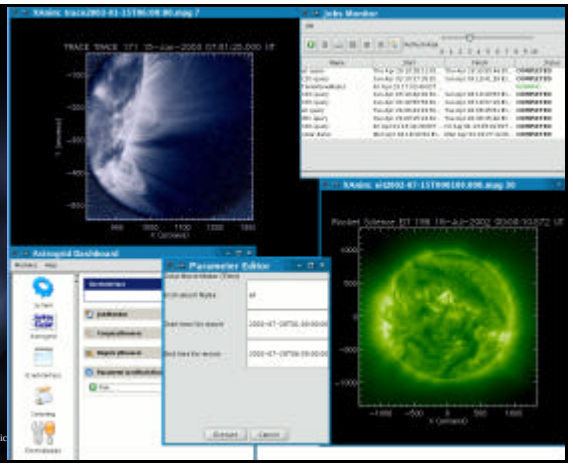
- Packaged workflow
  - Enter RA, Dec, radius
  - System returns objects and redshifts
- User Options
  - Input data (INT-WFS, SDSS)
  - Redshift apps (hyperz, bpz)
- Defaults
  - Source extraction double image mode (r' image ref)
  - Plus lots of other sensible default configs for the cross match, the redshift apps etc

See http://wiki.astrogrid.org/bin/view/Astrogrid/RedshiftMaker

AstroGrid Redshift Maker



Solar Movie Maker

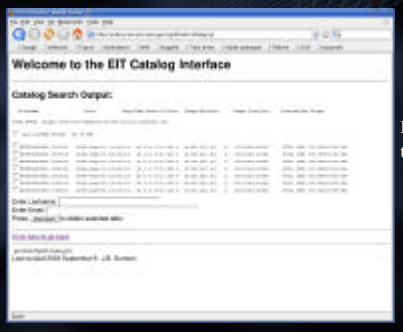


Comparison of the workflow with and without AstroGrid

Making a solar movie – non VO

Web interface to *database* of SOHO/EIT observations

Making a solar movie – non VO



Download images to local machine

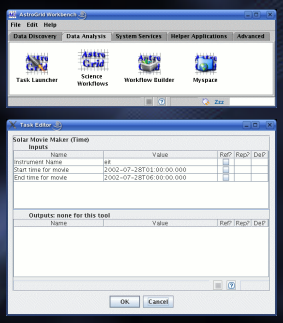
Making a solar movie – non VO

IDL SolarSoft required to calibrate images and make a movie. Routines available, however each user rewrites code calling them to produce a movie.

Making a solar movie – non VO: weaknesses

- If user wishes to make a movie for a different time period, the above steps need to be repeated – by hand
- Similarly for astronomy data if one is interested in many objects
- The entire archive is not 'visible' to the user – only the downloaded subset can be processed
- Scripting eg with Python, Perl etc is possible though code is different for different archives – not easy for the 'general' user / for many datasets. This limits feasibility of multi-wavelength, multi-instrument work

Making a solar movie – with AG



Solar Movie Maker science workflow.

Other science workflows available:
Redshift Maker, Colour Cutter, Cone Search, SWIRE images

Making a solar movie – with AG: under the hood

- Access to database of observations – via AG DSA (DataSet Access) software
- Requests are sent using ADQL (Astronomy Data Query Language), similar to SQL

AG Workbench/ workflow engine ⟷ DSA ⟷ Database of observations (at archive)

Input: ADQL query

Output: table of observations satisfying the query, in VOTable format

Making a solar movie – with AG: under the hood: CEA

- Capability to call a set of processing routines that do operations on data, or a model (apps in any programming language can be wrapped) – CEA Application

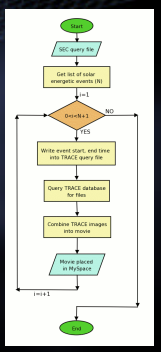
AG workbench/ workflow engine ⟷ CEA ⟷ Application. Eg software that processes data, model etc (on apps server)

Input: CEA application input parameters

Output: whatever the output of the application is, delivered to user's Myspace

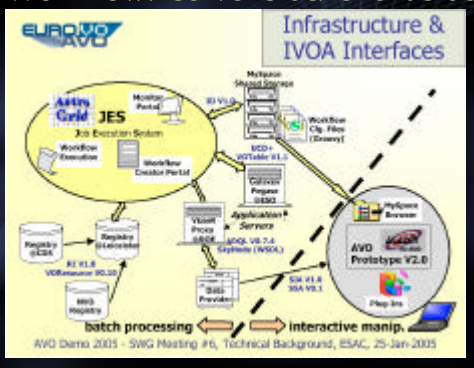
Making a solar movie – with AG: under the hood: Workflow

- Workflow capability so that queries to datasets and calls to applications can be managed
- Jobs are run remotely and asynchronously
- Queries and workflows can be re-used and shared
- AstroGrid is currently the only VO project with a workflow capability

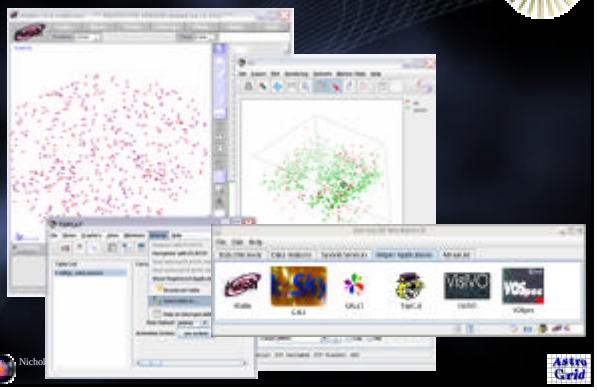


Client Side Applications

Workflow: server side/ client side ...

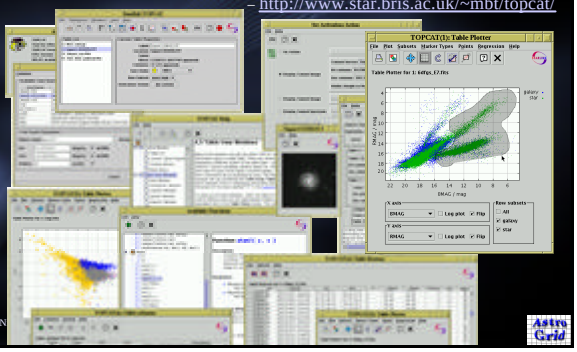


Plastic – VO tools on the desktop



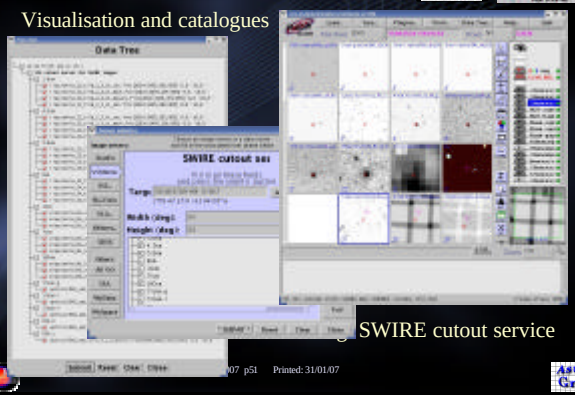
TopCat

- TOPCAT is an interactive graphical viewer and editor for tabular data
- Available from:
  - http://www.star.bris.ac.uk/~mbt/topcat/



Aladin

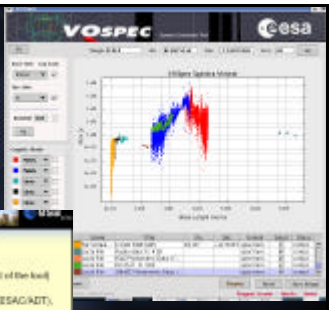
Visualisation and catalogues

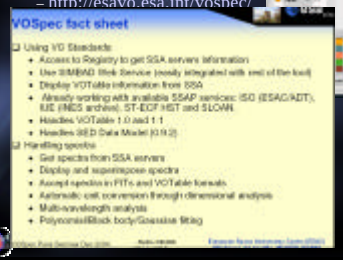


SWIRE cutout service

VOSpec

- VO spectral access tool
- Developed at ESAC
- Startup from:
  - http://esavo.esa.int/vospec/



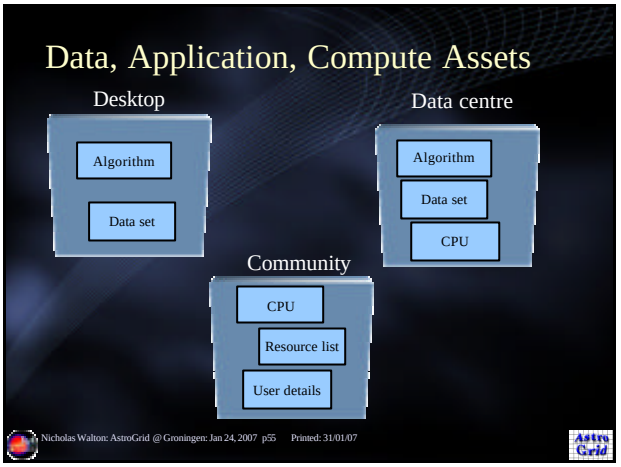


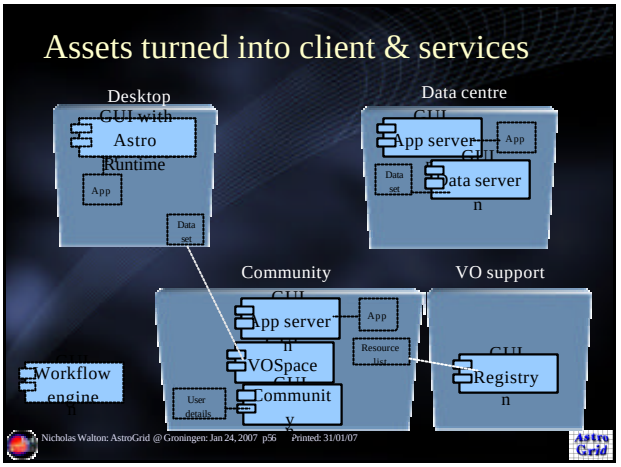
Technology:

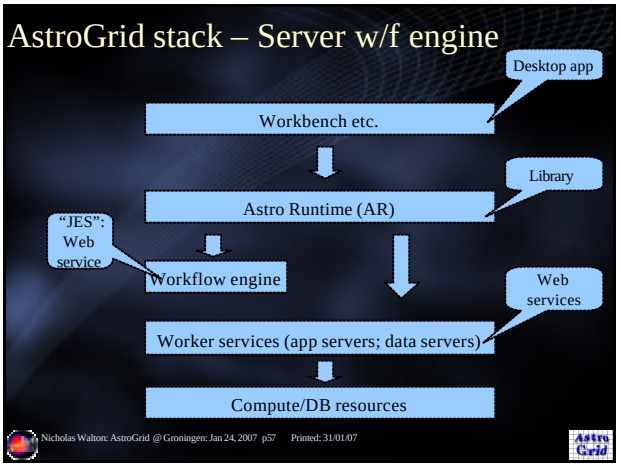
Powers the VO – transparent to Astronomers

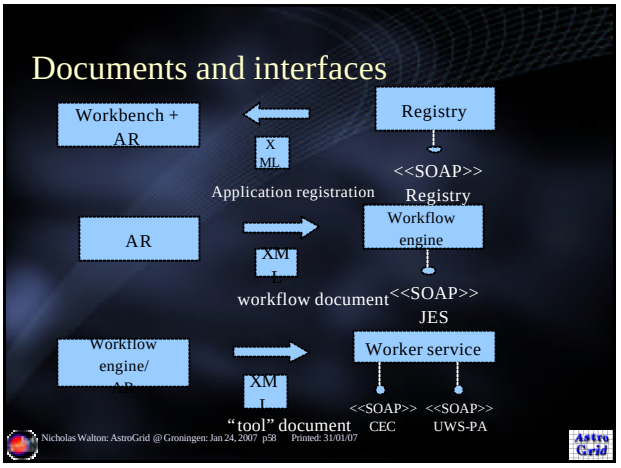
AstroGrid – an eScience Project

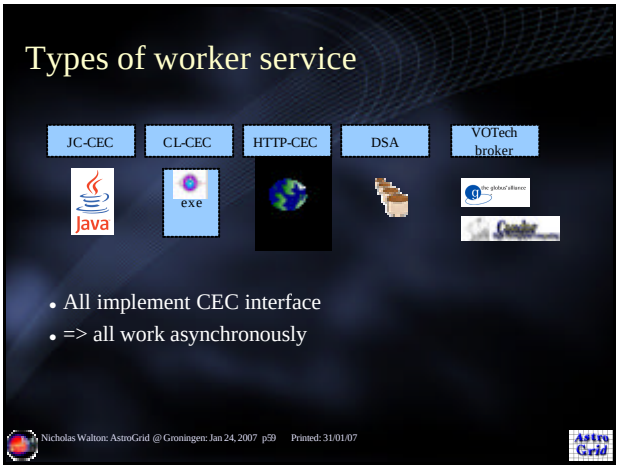
- Open source – see http://software.astrogrid.org
- Web service technologies
  - developed in Java
  - utilises the Apache toolset
  - Eclipse IDE for development
  - Maven/Ant for building
  - JUnit for testing
  - CVS for code control
- Distributed project development
  - Plone and Wiki
  - Bugzilla
  - Jabber
- Professional development team

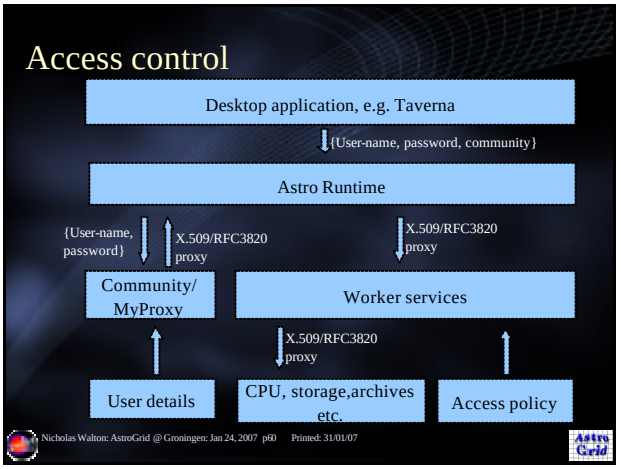

Data, Application, Compute Assets
Desktop
Algorithm
Data set
Data centre
Algorithm
Data set
CPU
Community
CPU
Resource list
User details
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p55 Printed: 31/01/07


Assets turned into client & services
Desktop
Astro Runtime
App
Data set
Data centre
App server
App
Data set
Data server
Community
App server
App
Resource list
VOSpace
User details
Community
VO support
Registry
Workflow engine
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p56 Printed: 31/01/07


AstroGrid stack – Server w/f engine
Desktop app
Workbench etc.
Library
Astro Runtime (AR)
"JES": Web service
Workflow engine
Web services
Worker services (app servers; data servers)
Compute/DB resources
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p57 Printed: 31/01/07

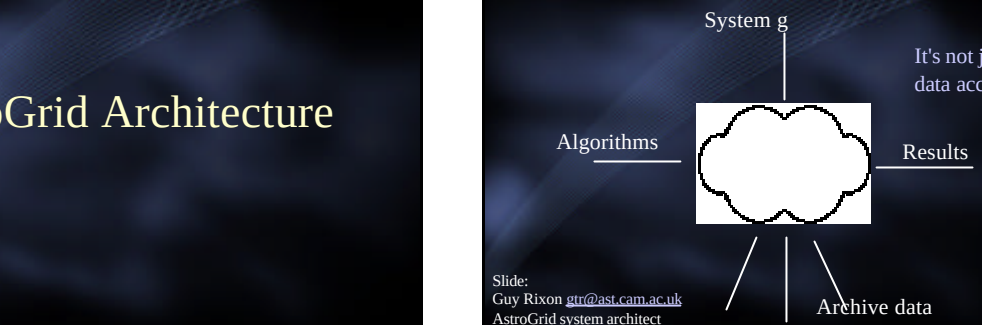

Documents and interfaces
Workbench + AR
XML
Registry
<<SOAP>>
Application registration
Registry
AR
XML
Workflow engine
workflow document
<<SOAP>>
JES
Workflow engine/
XML
Worker service
<<SOAP>>
CEC
<<SOAP>>
UWS-PA
"tool" document
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p58 Printed: 31/01/07


Types of worker service
JC-CEC
CL-CEC
HTTP-CEC
DSA
VOTech broker
Java
exe
• All implement CEC interface
• => all work asynchronously
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p59 Printed: 31/01/07


Access control
Desktop application, e.g. Taverna
{User-name, password, community}
Astro Runtime
{User-name, password}
X.509/RFC3820 proxy
X.509/RFC3820 proxy
Community/ MyProxy
Worker services
X.509/RFC3820 proxy
User details
CPU, storage,archives etc.
Access policy
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p60 Printed: 31/01/07

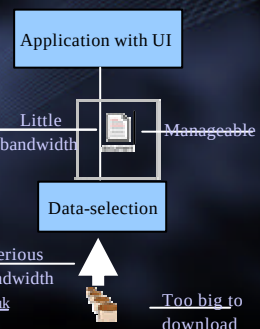
AstroGrid Architecture

The VO as a virtual laboratory



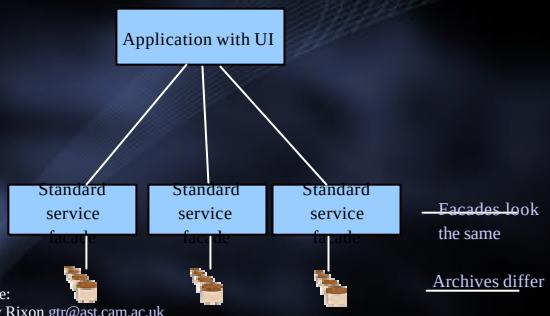
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

Pattern: data selection



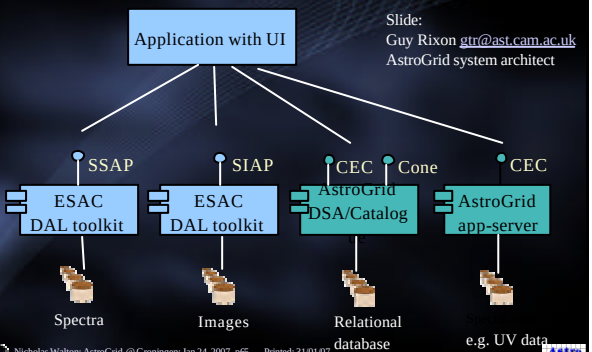
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

Pattern: facades



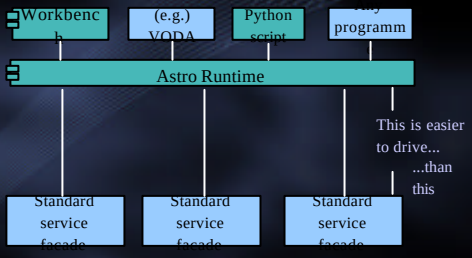
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

Implementation: standard interfaces

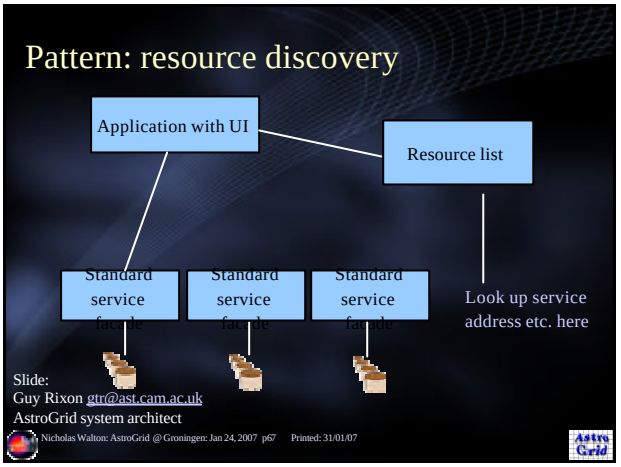


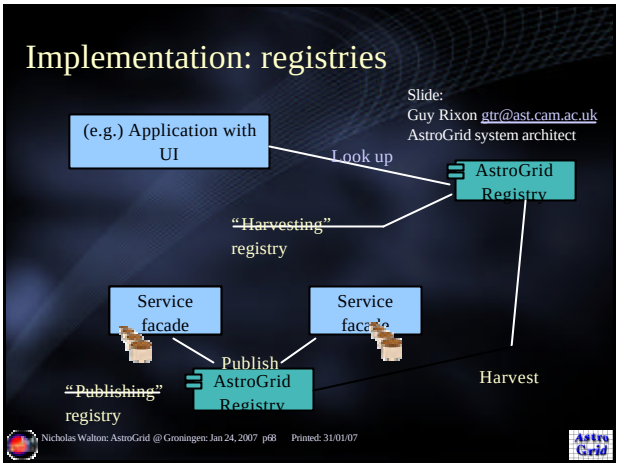
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

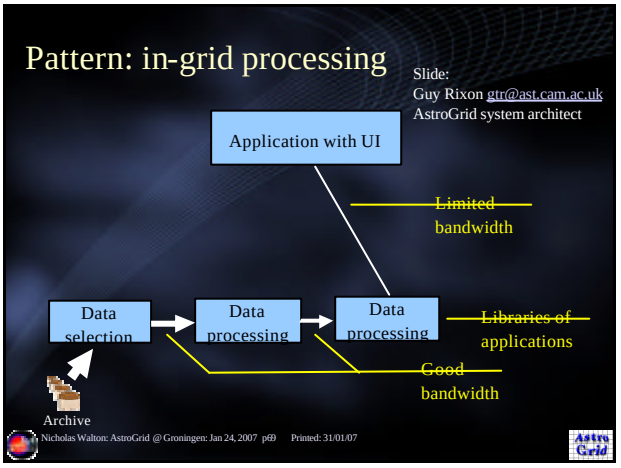
Implementation: Astro Runtime

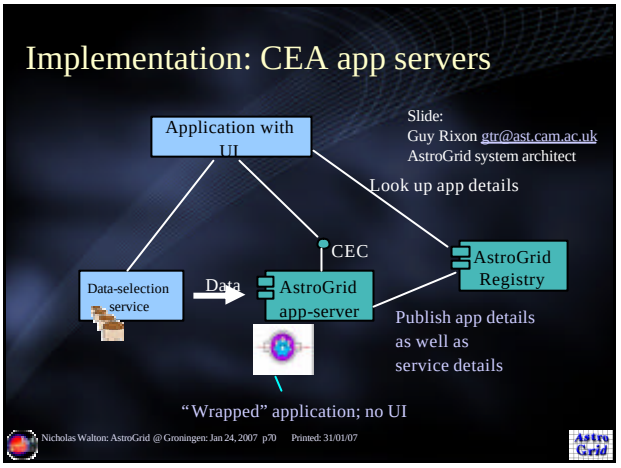


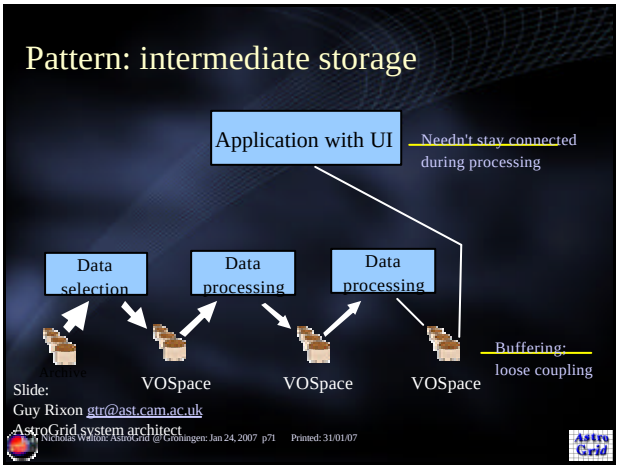
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

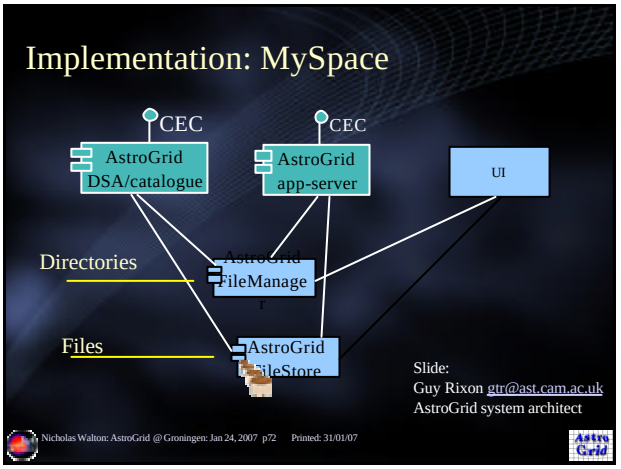

Pattern: resource discovery
Application with UI
Resource list
Standard service facade
Standard service facade
Standard service facade
Look up service address etc. here
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p67 Printed: 31/01/07


Implementation: registries
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
(e.g.) Application with UI
Look up
AstroGrid Registry
"Harvesting" registry
Service facade
Service facade
Publish
AstroGrid Registry
Harvest
"Publishing" registry
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p68 Printed: 31/01/07

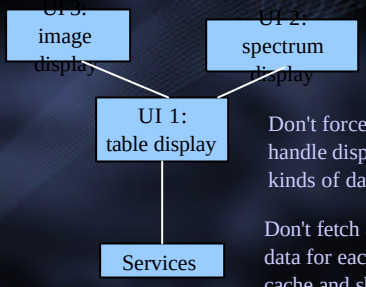

Pattern: in-grid processing
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
Application with UI
Limited bandwidth
Data selection
Data processing
Data processing
Libraries of applications
Good bandwidth
Archive
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p69 Printed: 31/01/07


Implementation: CEA app servers
Application with UI
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
Look up app details
CEC
AstroGrid Registry
Data-selection service
Data
AstroGrid app-server
Publish app details as well as service details
"Wrapped" application; no UI
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p70 Printed: 31/01/07


Pattern: intermediate storage
Application with UI
Needn't stay connected during processing
Data selection
Data processing
Data processing
Archive
VOSpace
VOSpace
VOSpace
Buffering: loose coupling
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p71 Printed: 31/01/07


Implementation: MySpace
CEC
CEC
AstroGrid DSA/catalogue
AstroGrid app-server
UI
Directories
AstroGrid FileManager
Files
AstroGrid FileStore
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect
Nicholas Walton: AstroGrid @ Groningen: Jan 24, 2007 p72 Printed: 31/01/07

Pattern: specialization of UIs

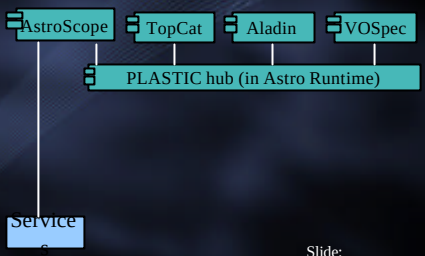


Don't force each UI to handle display of all kinds of data: specialize

Don't fetch a new copy of data for each display: cache and share it

Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

Implementation: PLASTIC



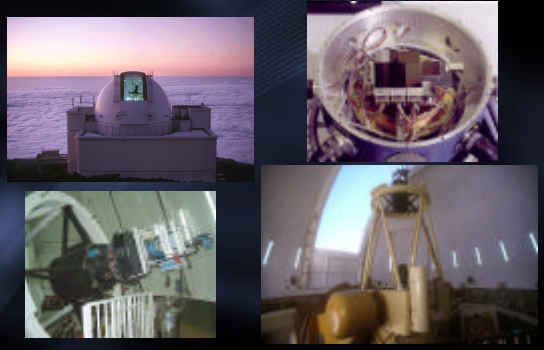
Slide:
Guy Rixon gtr@ast.cam.ac.uk
AstroGrid system architect

AstroGrid in Use: Now

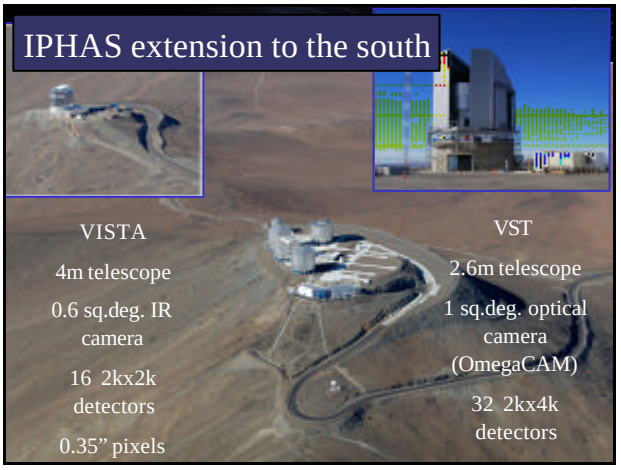
Example Large Scale Survey: the IPHAS Survey of the Galactic Plane

- IPHAS: The Isaac Newton Telescope/ Wide Field Camera Photometric H-alpha Survey of the Norther Galactic Plane (PI: Janet Drew: Imperial, UK)
- Element of the wider IPHAS/VPHAS+/UVEX consortium: forming EGAPS (European Galactic Plane Surveys) – see http://www.egaps.org
  - large collaboration of scienctists from ~10 countries
- Key Goals: Large scale Milky Way structure and study of early and late type populations (preferentially selected via H-alpha emission line properties)

IPHAS data source: 2.5-m INT + WFS



IPHAS extension to the south



VISTA

4m telescope

0.6 sq.deg. IR camera

16 2kx2k detectors

0.35” pixels

VST

2.6m telescope

1 sq.deg. optical camera (OmegaCAM)

32 2kx4k detectors

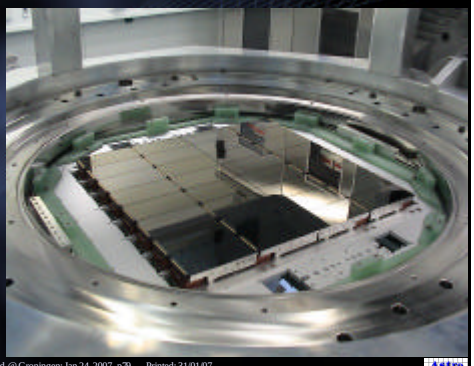
OmegaCam – a large camera

Focal plane geometry of OmegaCAM on the 2.6-m VST

(typical of large CCD mosaics)

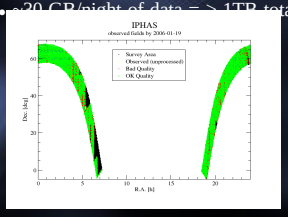
32 4kx2k CCDs

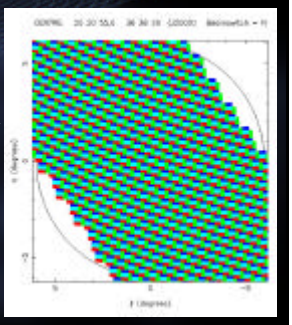
one image ~0.5GB



IPHAS Survey

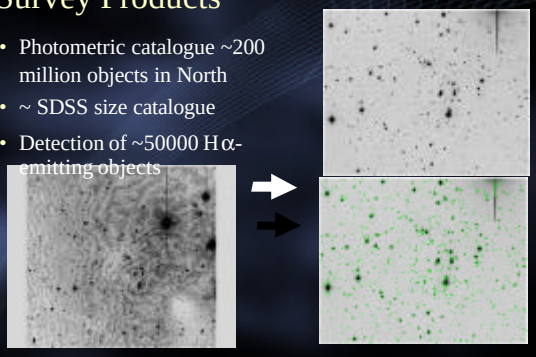
- 7635 Pointings ~2000 sq deg
- x 2 (overlap, 5'x5' offset)
- thus 15270 Pointing in total
- r, Ha, and i
- ~30 GB/night of data => ~1TB total





Survey Products

- Photometric catalogue ~200 million objects in North
- ~ SDSS size catalogue
- Detection of ~50000 Hα-emitting objects

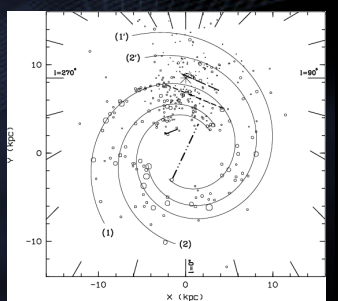


IPHAS: Images

IC 5070 – the pelican nebula. Credit Nick Wright (UCL)

Models for Galactic Structure

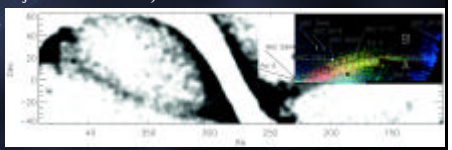
- Young populations good tracers of galactic structure
- e.g. Russeil (2003) study based on positions of star forming complexes (from HI, HII, CO observations)
- Vallee (2005) statistical study of recent work – points also to 4 arm model
  - uncertainties over interarm distances
  - incompleteness in certain directions e.g. 180<l<270, anticentre direction



Russeil 4-arm model - 1) Sagittarius-Carina, 2) Scutum-Crux, 1') Norma-Cygnus, 2') Perseus arms – sun marked by X

Aside: streams in the Milky Way

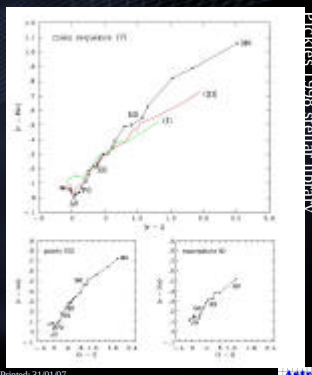
- Stellar streams are a good indication of galaxy hierarchical merger models (accretion of galaxies)
- Various (old) tracer populations used to map galactic streams
  - blue A coloured stars (inc BHBs) (e.g. Yanny et al 2000, Ibata et al, 2001)
  - M giants (e.g. Majewski et al 2003)
- IPHAS to trace young (for galactic spiral structure) and old (for streams)



Panoramic view of the Sgr stream, obtained by combining the 2MASS M giants of Majewski et al. (2003) with the SDSS stars (2x10e7) of Belokurov et al. 2006). Marked on the figure are branches A and B of the stream, together with some of the globular clusters

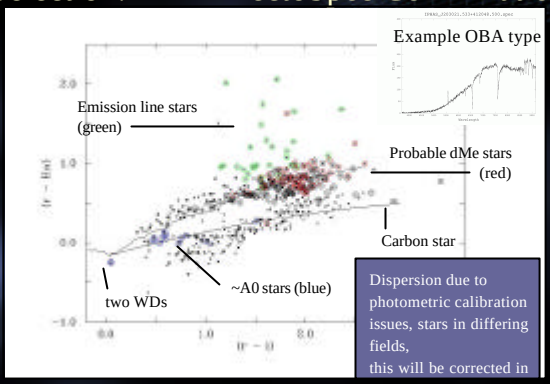
Structure: use of the IPHAS survey

- IPHAS is deep enough to sample most of the plane
  - r~20 = unreddened A0 dwarfs at 20kpc
- A stars are luminous to allow for the study of distant clusters
- A0V reddening line → a population of easily-modelled 'standard candles'
- Thus select early type (A, B stars) from their position in the colour-



Pickles 1998 stellar library

Selection: MMT HectoSpec Confirmations



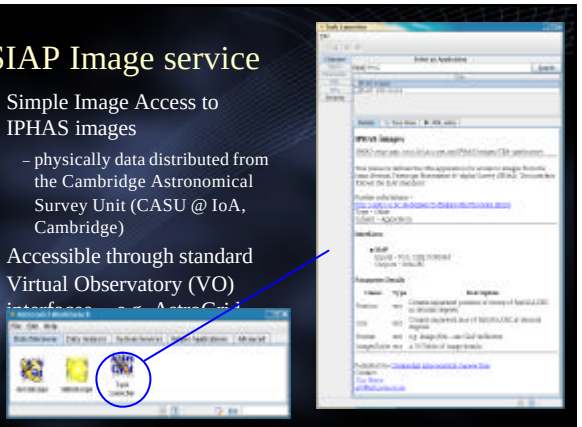
Dispersion due to photometric calibration issues, stars in differing fields, this will be corrected in upcoming data release

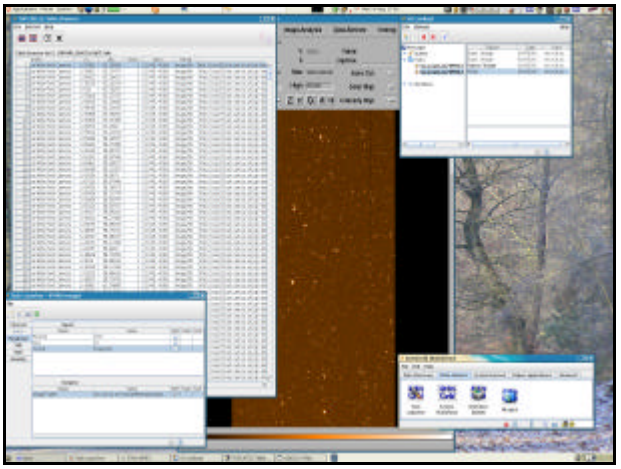
Integrating IPHAS Data into the VO

- Data Products from Cambridge survey pipeline include:
  - FITS images (per image pointing)
  - FITS table catalogues (per image pointing)
  - Single unified merged object source catalogues
    - Sybase IQ DMS system
- Access to these products expedited by use of AstroGrid interface mechanisms:
  - Images: via Simple Image Access protocol service
  - Catalogues (FITS tables):
  - Merged catalogues: Data Set Access component to database

SIAP Image service

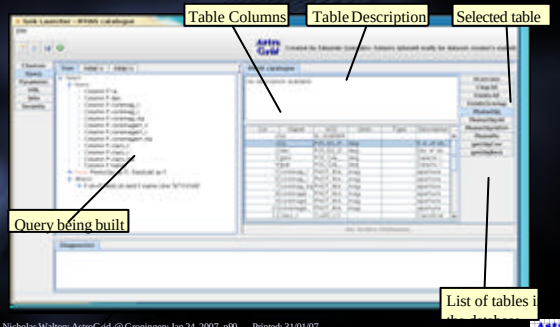
- Simple Image Access to IPHAS images
  - physically data distributed from the Cambridge Astronomical Survey Unit (CASU @ IoA, Cambridge)
- Accessible through standard Virtual Observatory (VO)

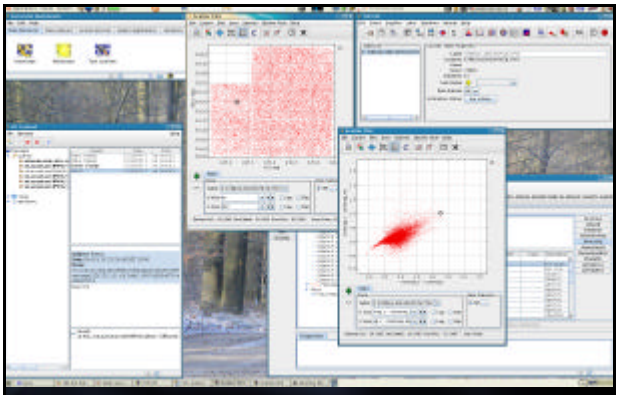




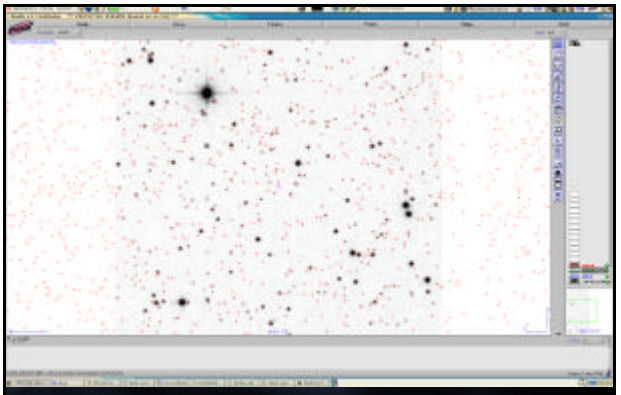
Catalogue Access

- Catalogue data available both as FITS tables on a per pointing basis
- In release – unified object catalogues



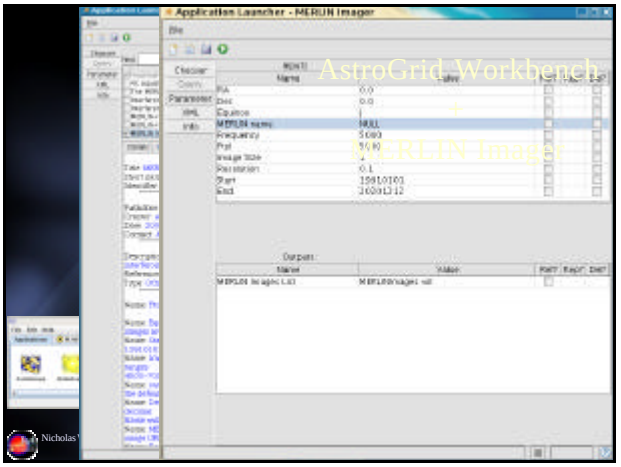


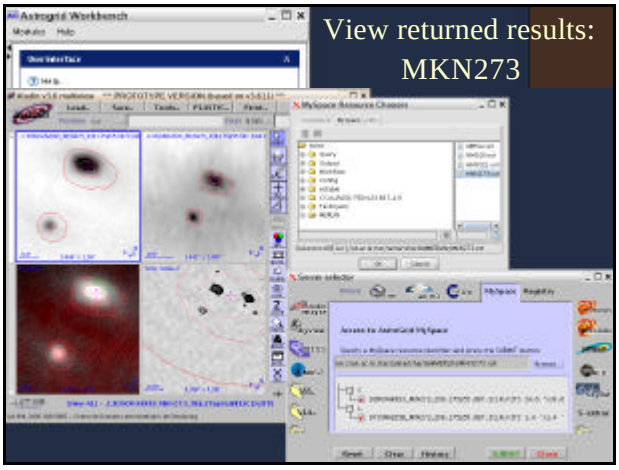
SQL query selecting IPHAS field 7300o objects = 24000 sources



Cross match of IPHAS and 2MASS for element of Field 7300o

AstroGrid: Radio Usage



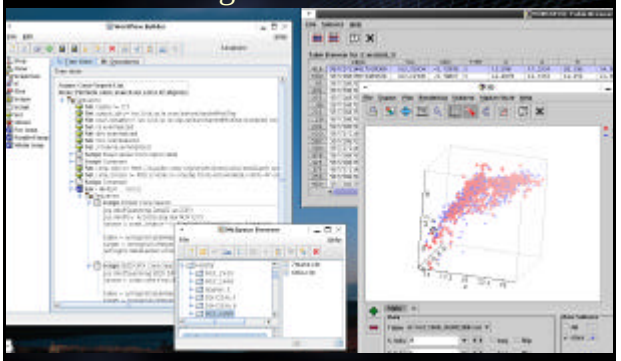


AstroGrid: Usage Examples

Hot Star Discovery: Use of AstroGrid

- Science workflow
  - search optical and infrared catalogues around user supplied list of star clusters
  - cross match the results
  - select optically classified stars
  - display u, g and K (optical and IR colours)
  - select massive stars based on colours (U > G and faint in K)

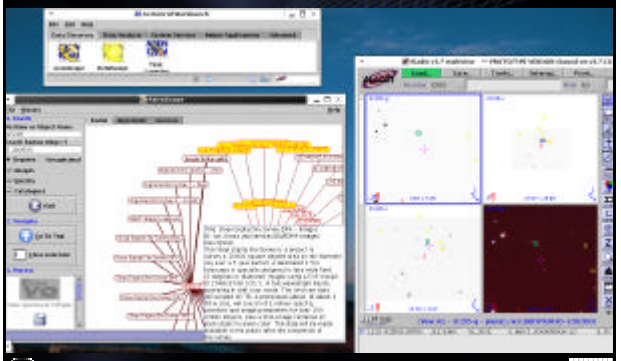
AstroGrid usage to discover hot stars



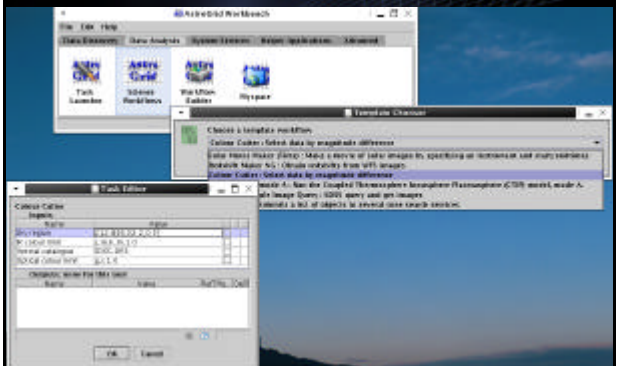
Cluster of Galaxies: VO use example

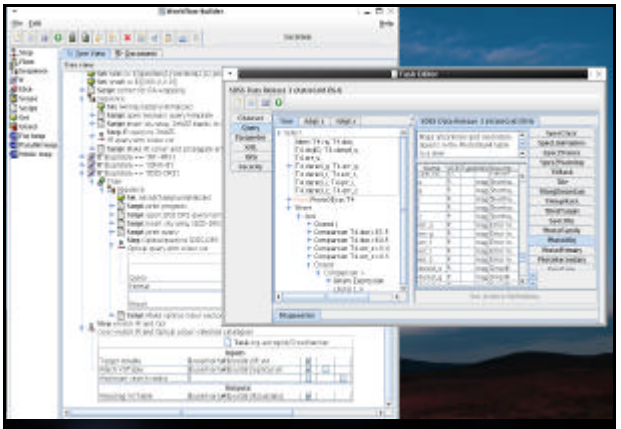
- Investigation of 3C295 – a large radio galaxy
  - interactions with cluster galaxies
  - heating intergalactic medium
- Interesting object to study at multiple wavelengths and scales
  - X-ray emmision for the cluster
  - radio jets
  - optical core
  - environment – nearby galaxies in the cluster

AstroScope around 3C295: red sources

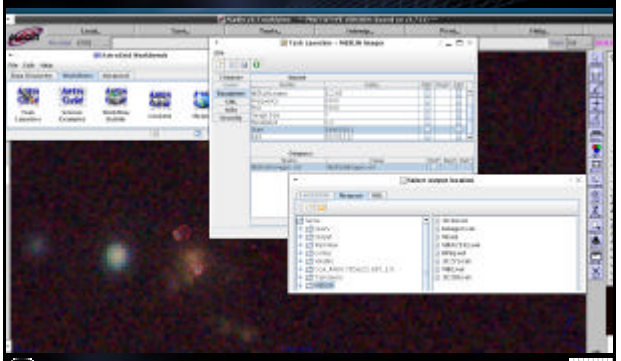


Colour Cutter: select red objects

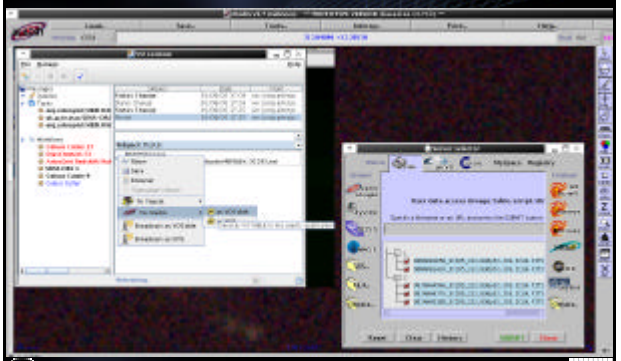


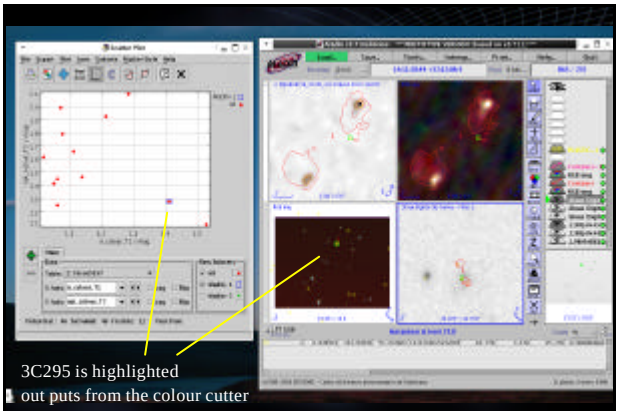


3C295 Radio Images from Merlin



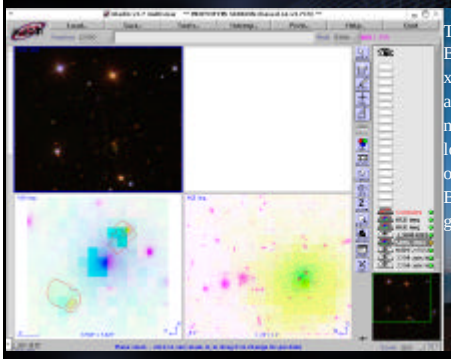
Visualise Merlin Images





3C295 is highlighted
out puts from the colour cutter

3C295: optical and X-ray data



TL: SDSS optical
BL: Blue is hard x-rays, violet optical and contours radio – note non thermal lobes in radio/xray only
BR: soft x-ray cluster gas in yellow

Closing

- Key Links
  - AstroGrid: http://www.astrogrid.org
  - AstroGrid Science: http://www2.astrogrid.org/science
  - AstroGrid software : http://software.astrogrid.org/
  - Euro-VO: http://www.euro-vo.org

Friday: Practical session ...