Exploiting Virtual Observatory and Information Technology: Techniques for Astronomy

Nicholas Walton
AstroGrid Project Scientist
Institute of Astronomy,
The University of Cambridge

Lecture #3 Goal:

Applications and Theory,

Workflows

Summary: Lecture #3

- Introduction
- Standards for Data Access
- Applications in a VO
  - Common Execution Architecture
  - Workflows
- Theory in a VO
- Science Example
  - Photometric redshifts
  - ... feeding into a observation/model SED case

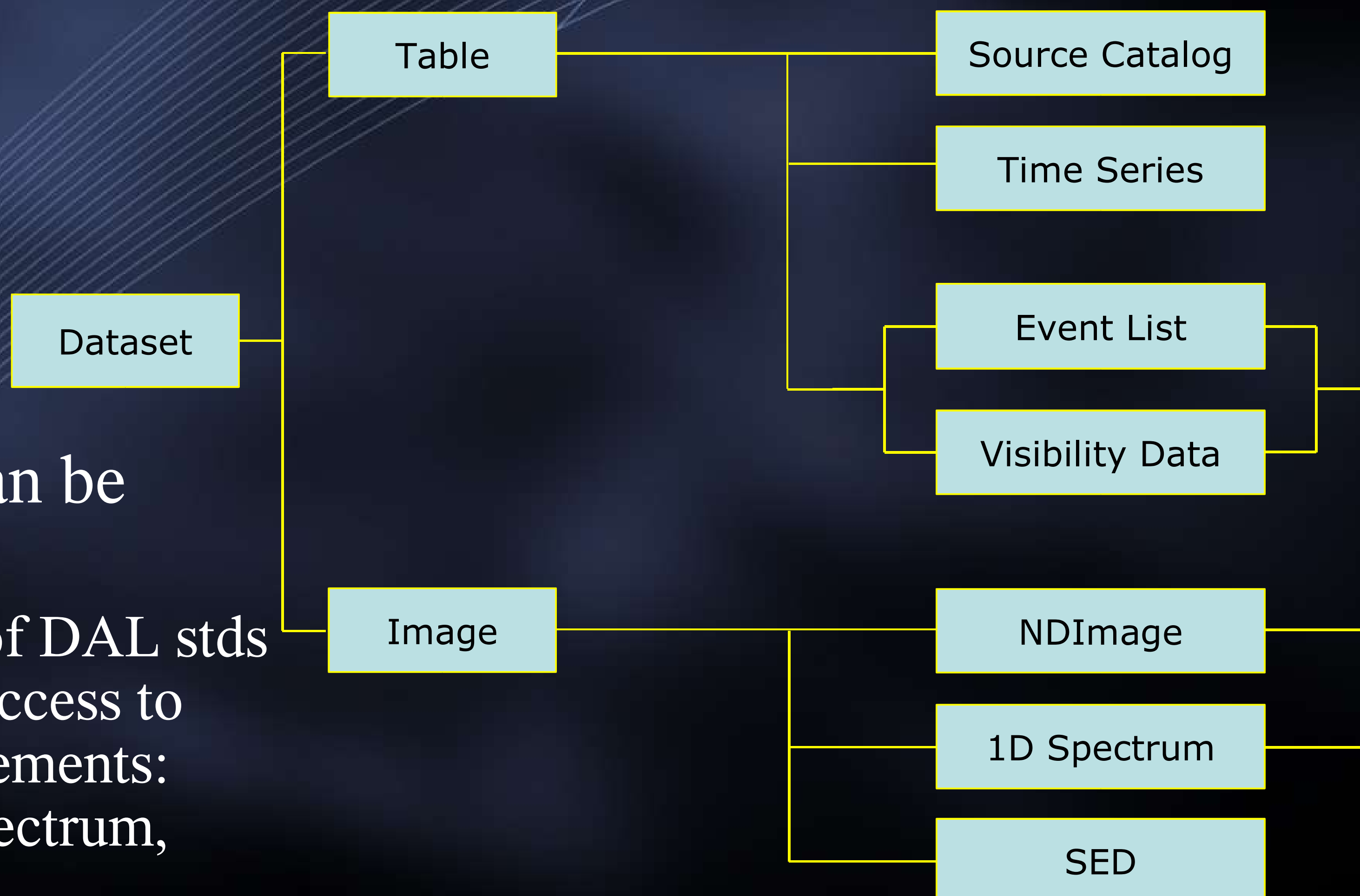
Introduction: Image Access

- Data exists in catalogue form and image form
  - Access to catalogues discussed in Lecture 2
  - Access to images (includes theory data, spectra and so forth)
- Virtual Observatory standards to address these
  - Data Access Layer (SIA, SSA)
  - Data Models
  - Common Execution Architecture
  - Workflow

Data Access Layer: DAL

http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDAL

- Provide standards to enable access to data sets



- A dataset can be complex
  - Concept of DAL stds to allow access to dataset elements: image, spectrum, SED, etc

Simple Image Access: SIA

http://www.ivoa.net/Documents/latest/SIA.html

- Protocol for retrieving image data from a variety of astronomical image repositories through a uniform interface
- Provides access to 'image' data
  - regularly sampled (pixelated) data
  - (instead of spectrum, catalog, etc.)
  - usually an image of the sky, with a World Coordinate System (WCS)
- Service-oriented data discovery
  - query service to discover data
- Access to image *metadata*
  - can get image metadata without retrieving the actual image
  - uniform description based on standard data models
- Access to image *datasets*
  - data may be *virtual* or computed on demand
  - uniform interface to any type of image data

Simple Spectra Access: SSA

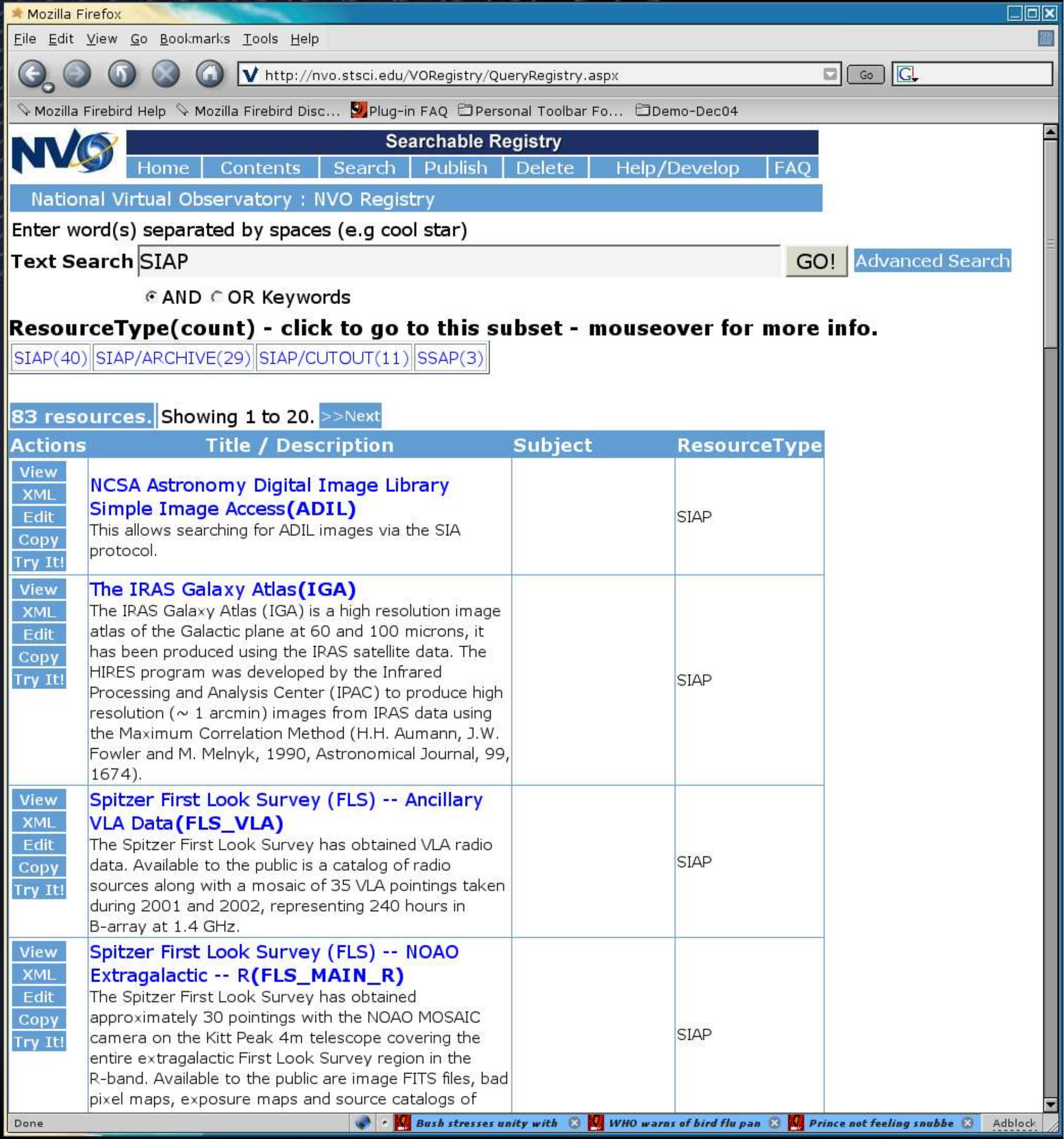
- Provides access to 'spectral' data
  - similar to SIA but deals with tabular spectrophotometric data
- Service-oriented data discovery
  - query service to discover data
- Access to dataset metadata
  - can get dataset metadata without retrieving actual dataset
  - uniform interface based on standard data models
- Access to actual dataset
  - data may be *virtual*, i.e., computed on demand
  - uniform interface to any type of spectral data
  - hides details of how data is stored or represented externally

Virtual Data

- Key VO concept: most data that is worked on is Virtual Data, i.e. Data that is created 'on the fly' during the VO process
  - Creation of virtual data may involve additional processing, e.g. Calibration of raw data from a request to an archive, subsetting of images etc.
  - Access conforms to 'Data Model' standards
- Basic SIAP usage:
  - Position and Size of search region sent via an HTTP GET call
  - Query response is a VOTable describing the images
    - Also gives an access reference to the actual data
  - Get the data via a fetch – use the returned file refs to actual data:
    - e.g http://cass38.ast.cam.ac.uk/cgi-bin/wfs-siap/getImage?run=347497&ccd=2
    - The reference URL may point to a service which returns a processed image, e.g. A cutout, a mosaic.

SIA and SSA Services

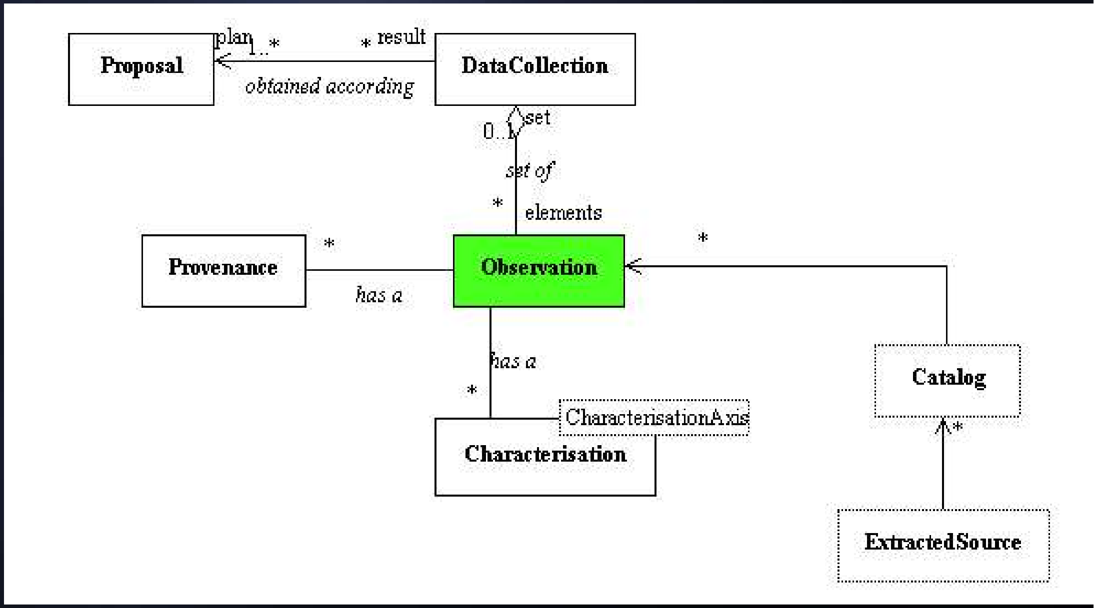
- A wide range of SIAP (the 'P' is for protocol) and SSAP services now provided to give access to image data
  - Cutout services
  - Mosaic services
  - Pointed observations
- Services discoverable through the 'Registry'
- NVO 'datascope' actions multiple SIAP/SSAP calls
  - Demonstrated in Lect 5
  - Science case later ...



Data Models:

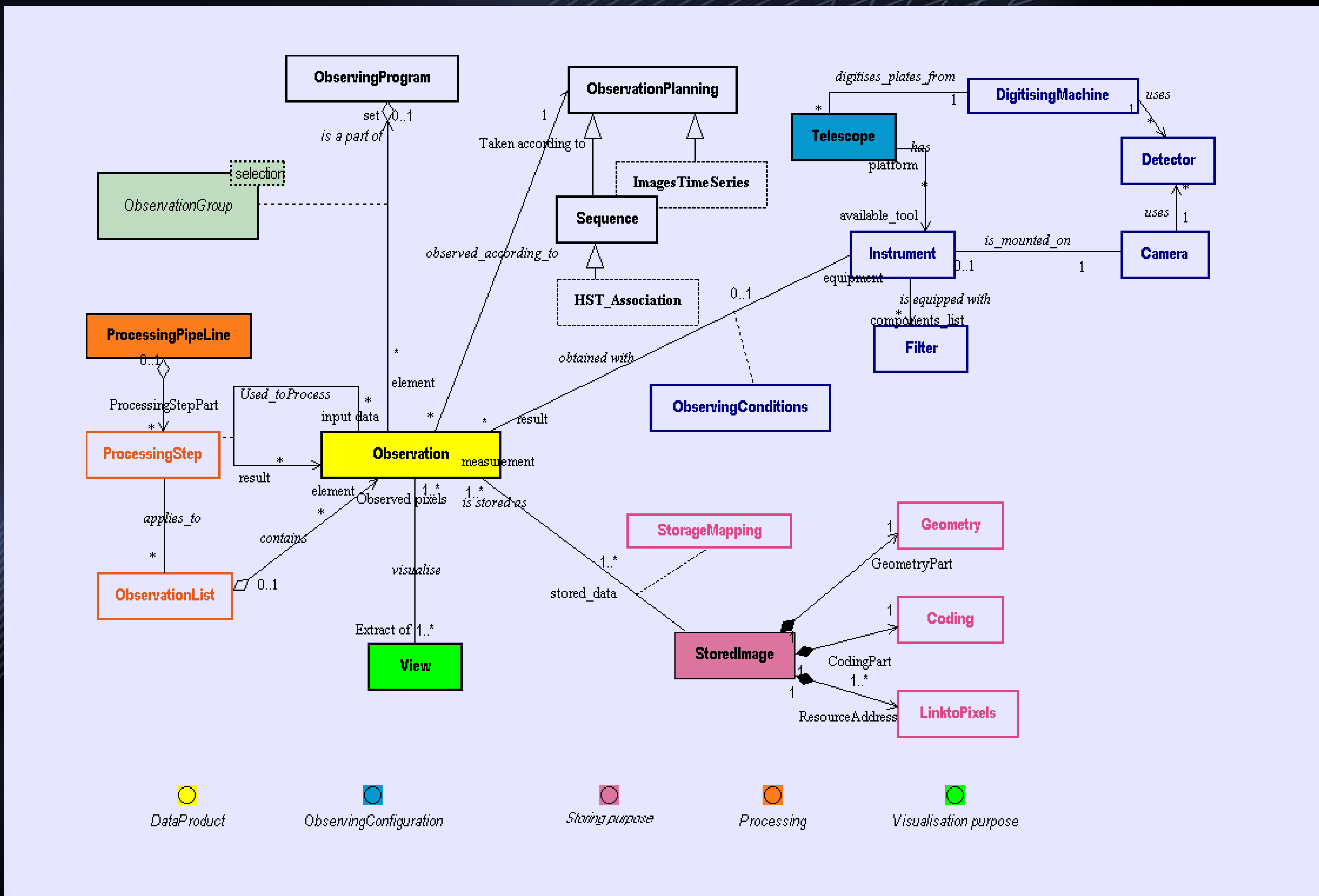
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDataModel

- Define standards to describe the structure and semantics of astronomical data
  - Generic data model broken down into key elements:
    - High level: image, spectrum, time, catalog
    - Low level: quantity, resolution
    - Provenance
  - IDHA model used by CDS
  - Specific SED example
  - DM defines other standard VO interfaces



IDHA Data Model Example:

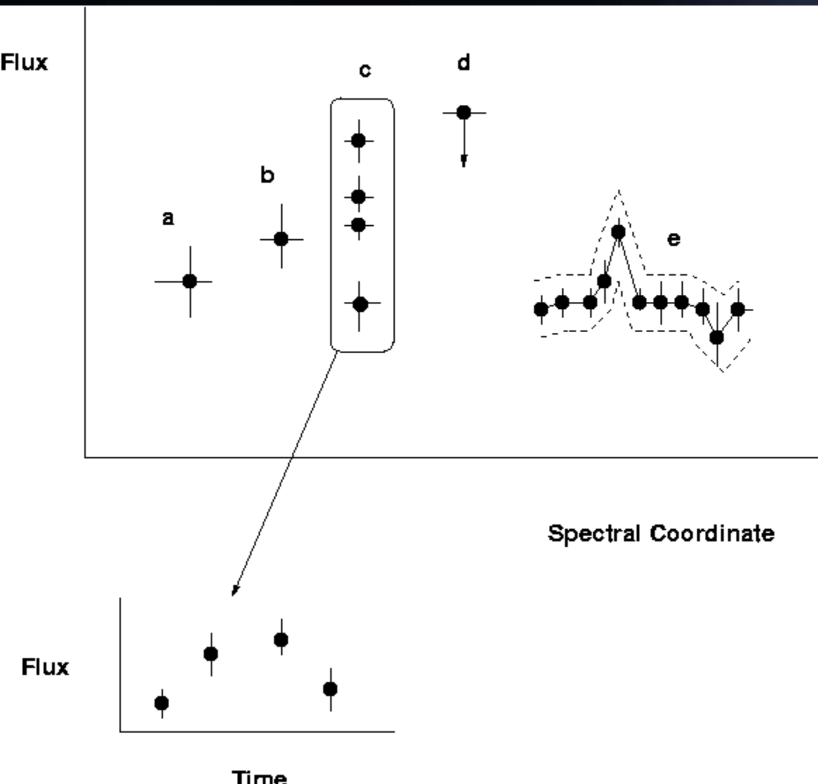
http://alinda.u-strasbg.fr/IDHA/lastmodel/

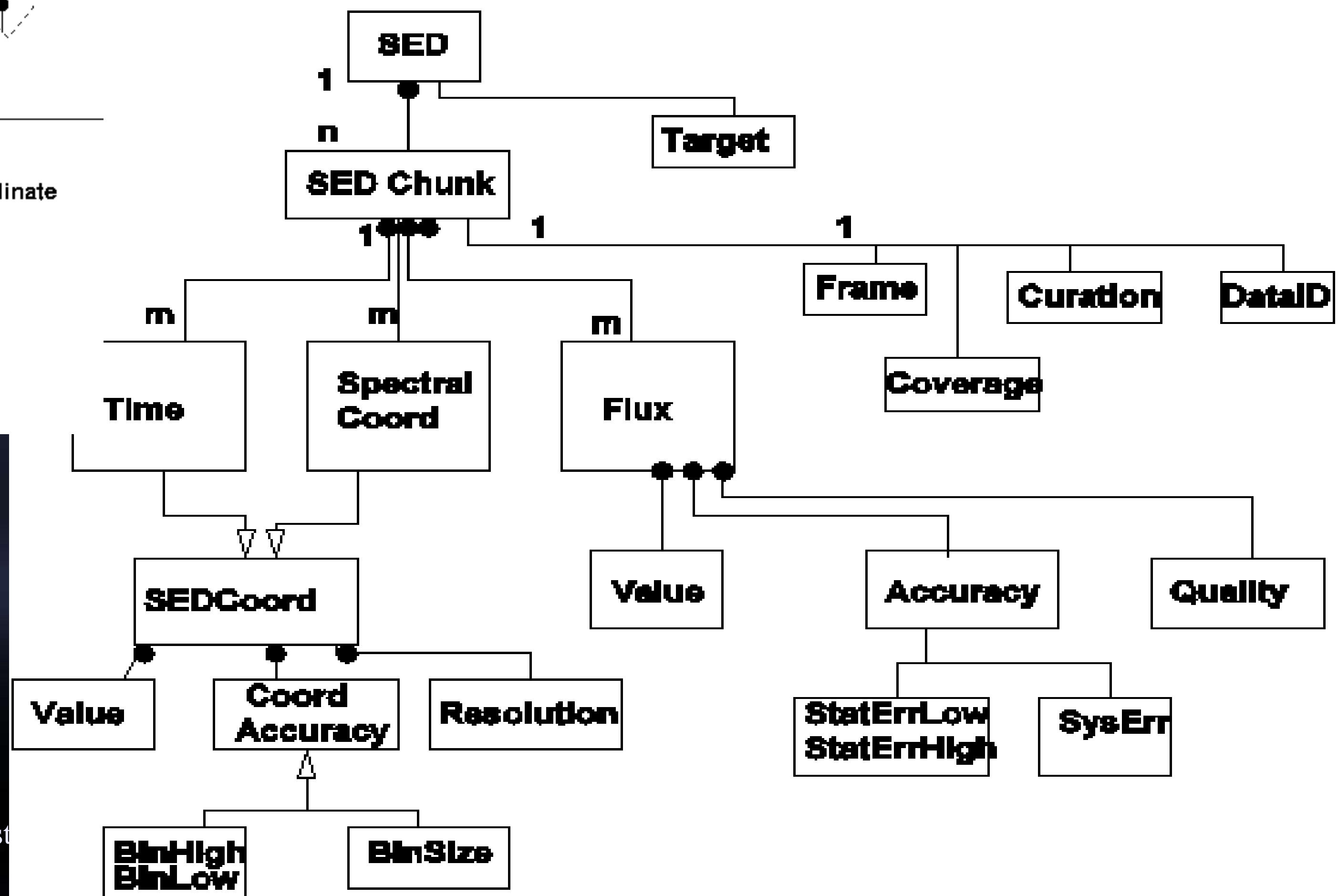


SED: Spectral Energy Distribution:

http://hea-www.harvard.edu/~jcm/vo/docs/spec0.92.html

Implementation available via VOSpec (Lecture 5/6)



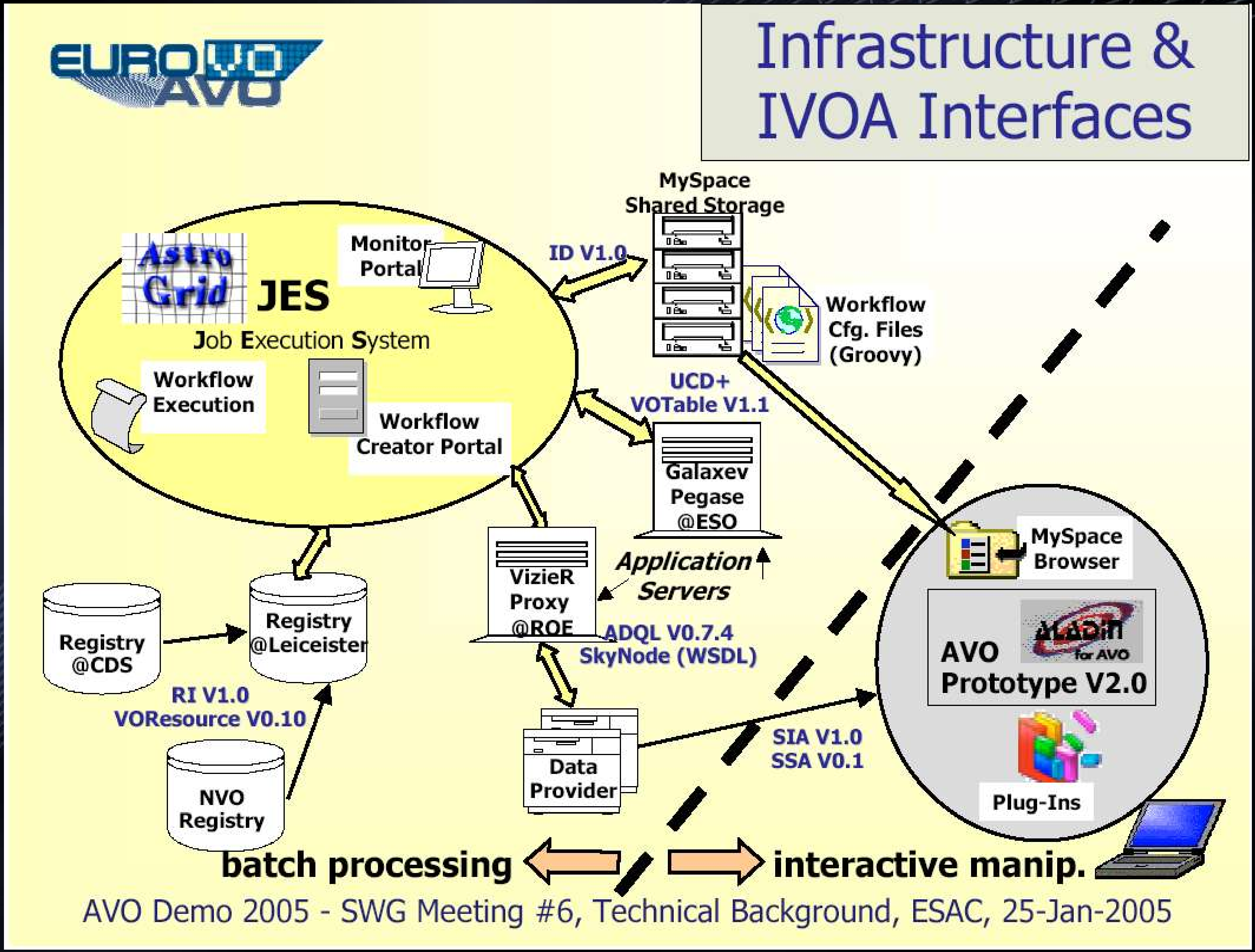


Model to enable construction of a multiwavelegth SED from spectral AND photometric data points

Applications @ IVOA:
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications

- Applications to manipulate and process data
  - Client side tools: conforming VO standards: e.g. Aladin
  - Server side tools: e.g. 'hyperZ' running on 1000's of images
- Applications now being developed to exploit standard interfaces
  - Those employing VOTable for data exchange
    - VOPlot, TopCat, Mirage, Aladin
  - Large scale service applications, e.g. Montage
- For the server side systems, concept of a framework to allow the use of a wide variety of applications in user configurable workflows.
- More on specific applications in Lect 5 + 6.

Workflow: server side/ client side ...



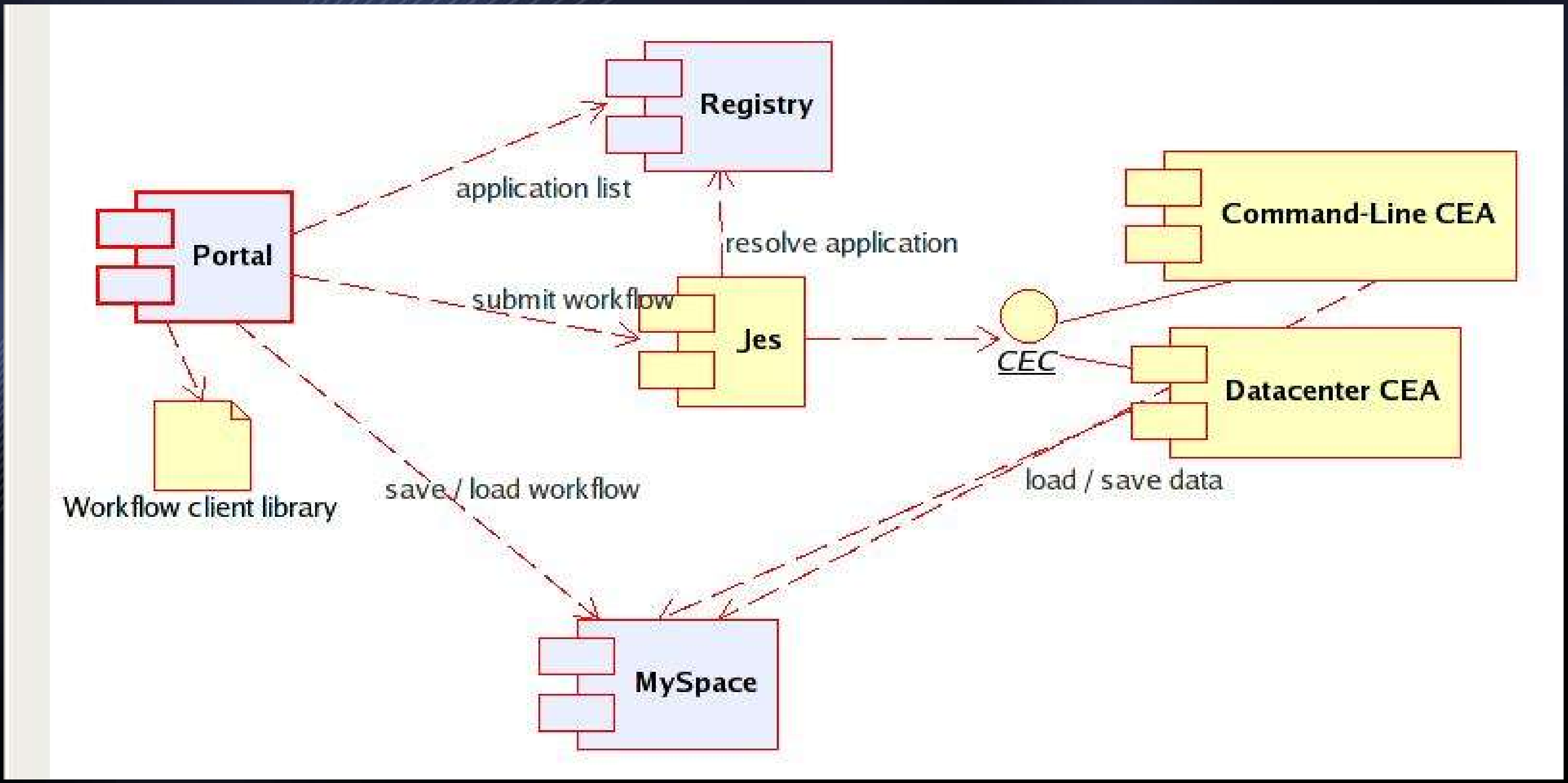
Common Execution Architecture

- Models how an application is run in the Virtual Observatory
  - An application is any process that consumes or produces data
- Set of interface definitions and schema
  - Defines the tool (application) and its parameters
  - How to execute the tool
    - Initialise
    - Registers listeners for logging and results
    - Gather remote data
    - Actually runs the application
- See http://www.astrogrid.org/maven/docs/HEAD/applications/

AstroGrid CEA (applicns)

Workflow

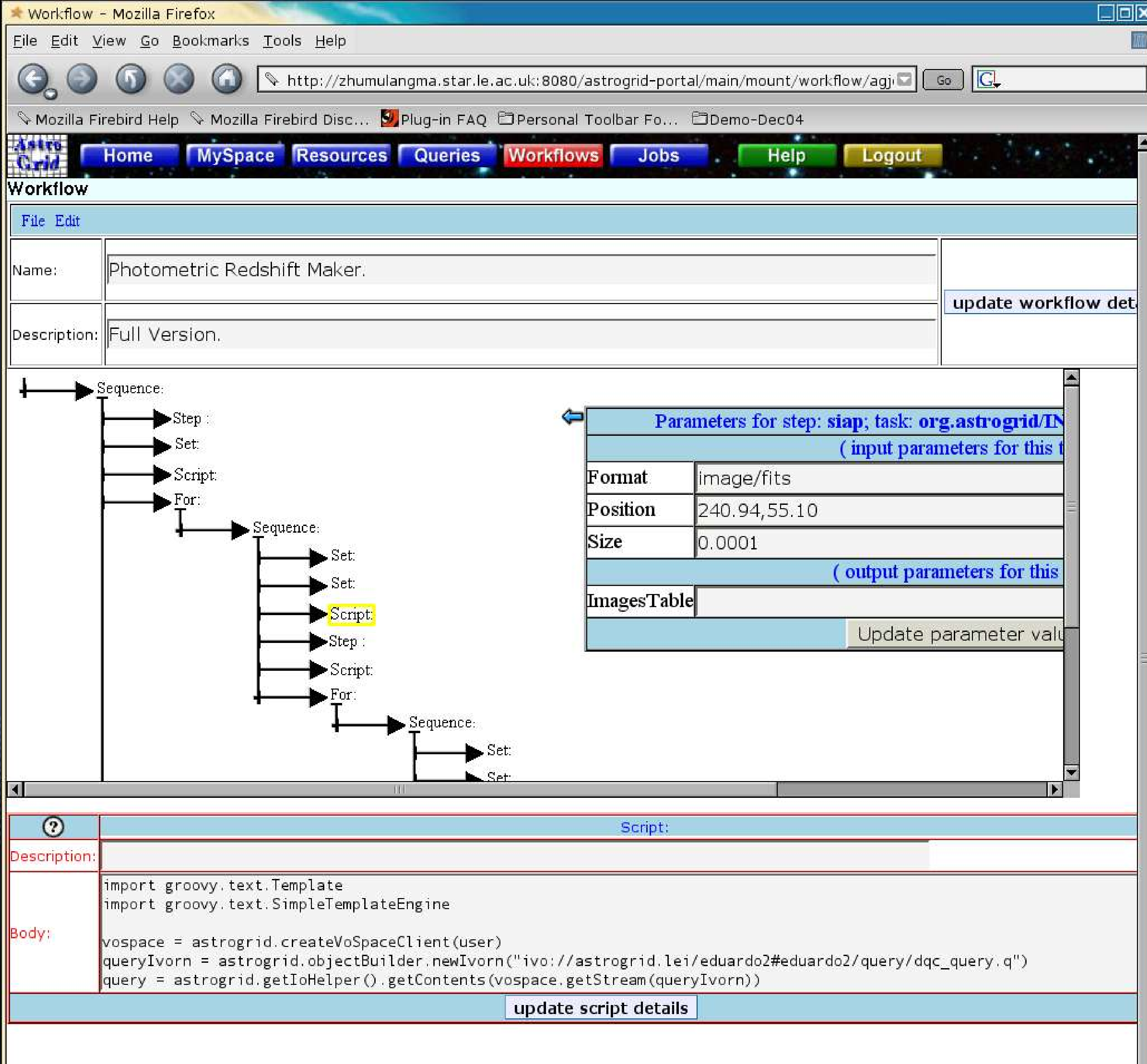
- Work is run remotely and asynchronously
- Archives searched and results manipulated
- Results are stored in a virtual file system
- Queries and workflows can be re-used and shared



Workflow: continued

- Workflow enables complex operations to be carried out, with the exact details of where the operations are occurring being hidden from the user
  - Gives access to CPU away from the desktop
- Workflows can be shared, ammended – enables community sharing of processes in addition to results
- AstroGrid is currently the only VO project with a workflow workbench where scientific workflows can be created and run.
  - This workflow engine is being integrated into the Euro-VO
- Workflows are constructed via discovery of relevant data and applications from the Registry
  - Applications are provided through the CEA

AstroGrid Workflow Builder

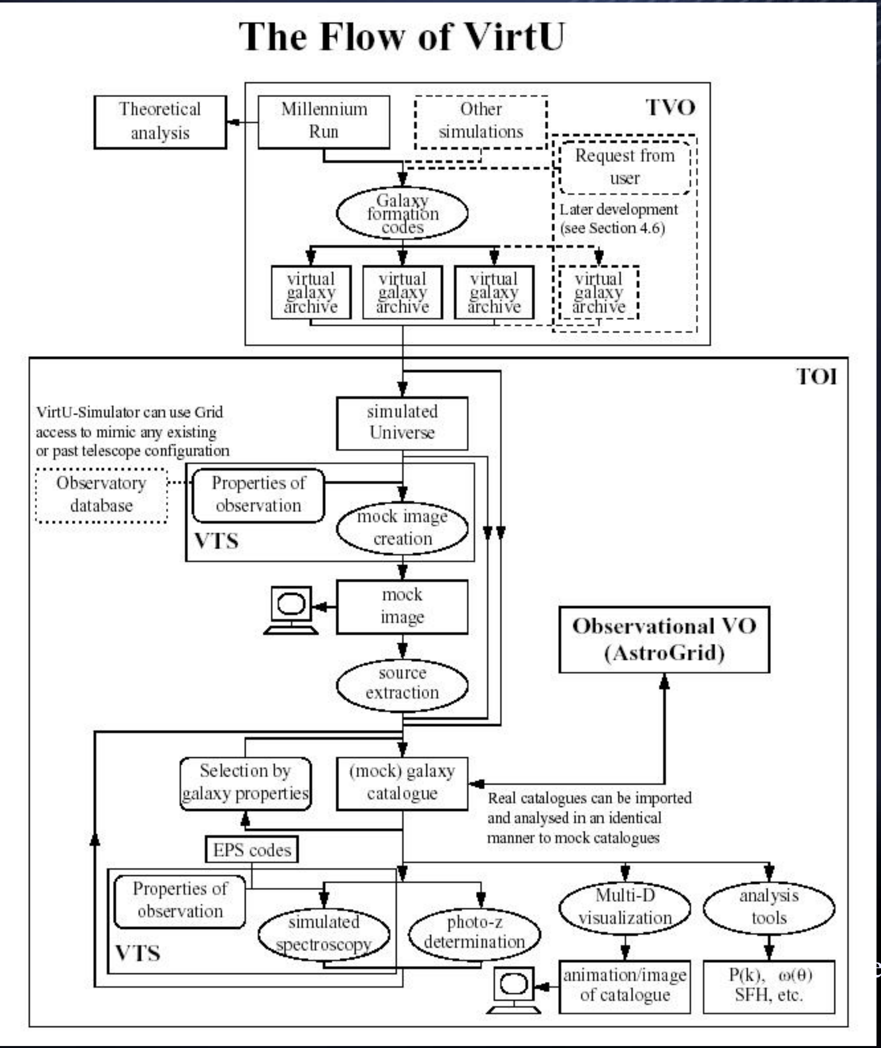


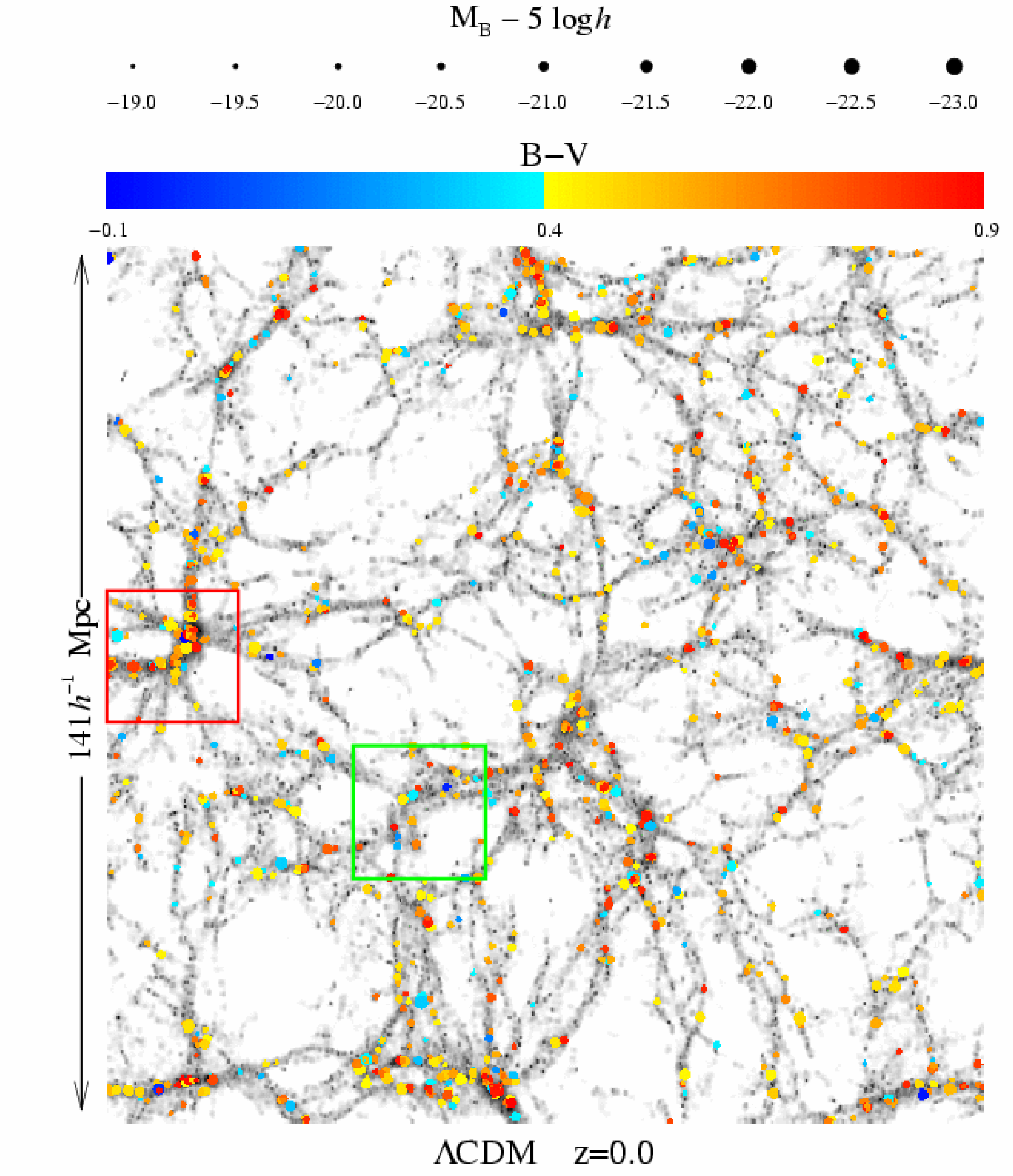
Theory in the VO

- A wide problem:
  - Large scale: e.g. Stellar hydro-codes, N-body simulations
  - Small scale: e.g. Spectral synthesis codes
  - Fundamental data: e.g. Excitation rate coefficients
- In principle treat simulation data as observational data
  - Enable comparison of simulations with observations
  - Enable comparison of simulations with simulations
- Early effort in areas of Data Models, Metadata

Virgo simulations

- Accessing observable properties computed from the simulations





Benson, Frenk, Baugh, Cole & Lacey (2001)

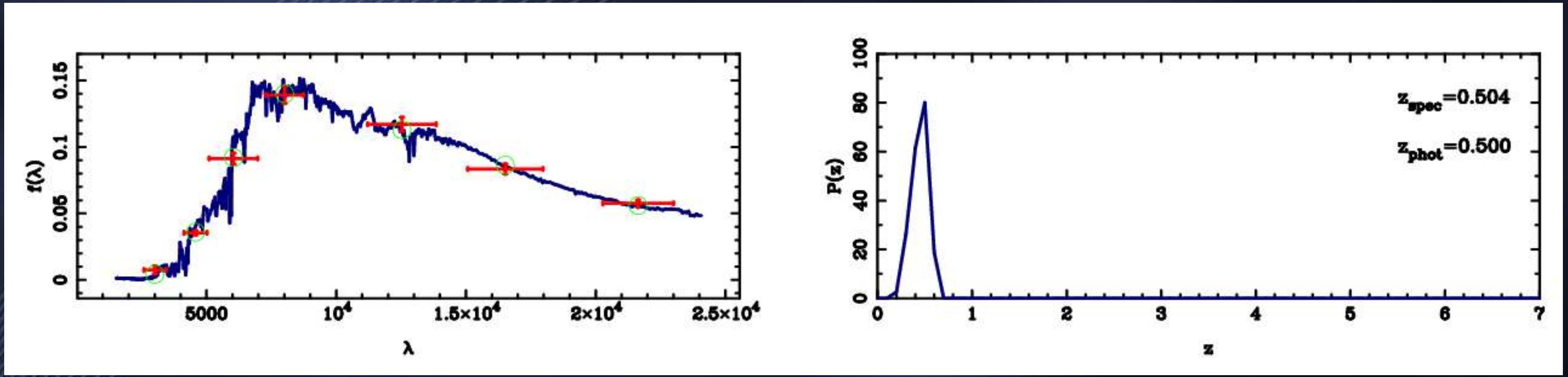
Science Example

Putting the technology to use ...

Science Case: Galaxy Distances

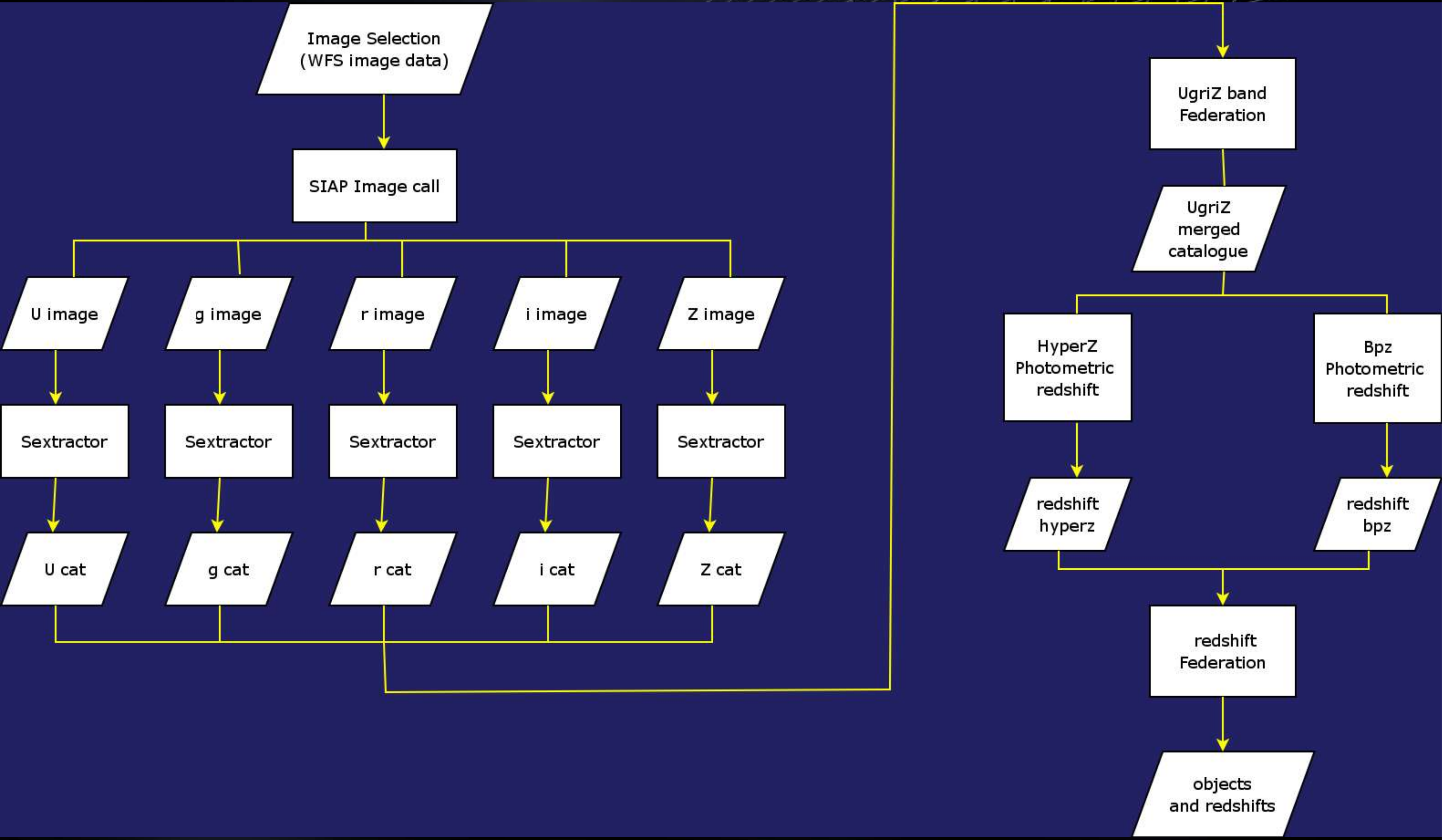
http://wiki.astrogrid.org/bin/view/Astrogrid/AgDemoDec2004Extragalactic

- Determine the distance to galaxies
  - Use of broad band photometry is efficient for large samples
  - Relies on the identification of spectral breaks in galaxies spectral energy distribution



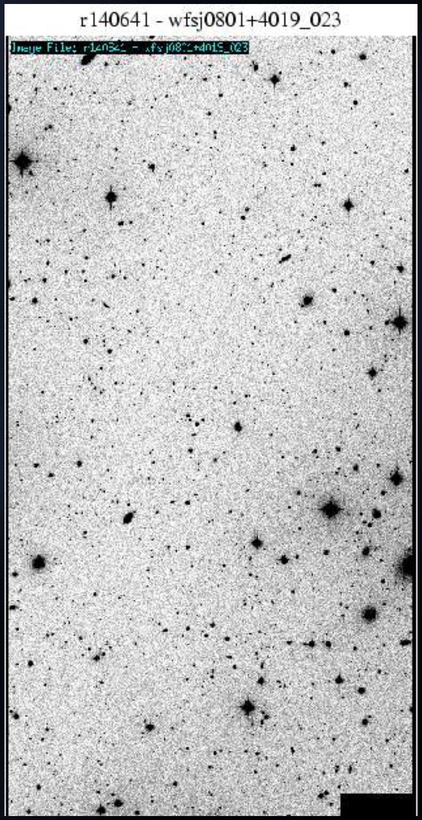
- Aim here is to automate the generation of the galaxy fluxes from survey image data, and feeding these fluxes into a number of specialist applications which return statistical estimates of galaxy redshifts (and thus distance)

Extragalactic Case Workflow



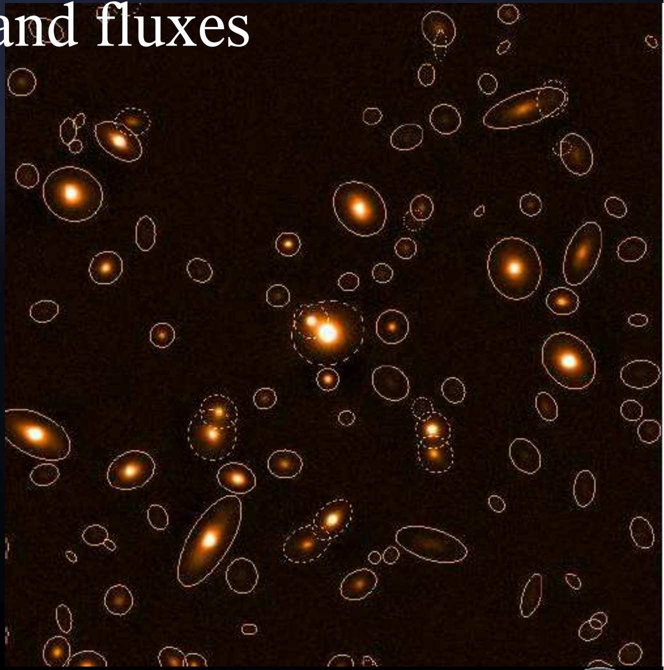
Accessing the image data

- INT Wide Field Survey data
- Data obtained using the Wide Field Camera on the INT
  - Accessed through an IVOA std SIAP call
  - Returns lists of files
  - These uploaded to MySpace via a 'workflow script'
- The data is held in Cambridge at CASU
- The MySpace server is in Leicester

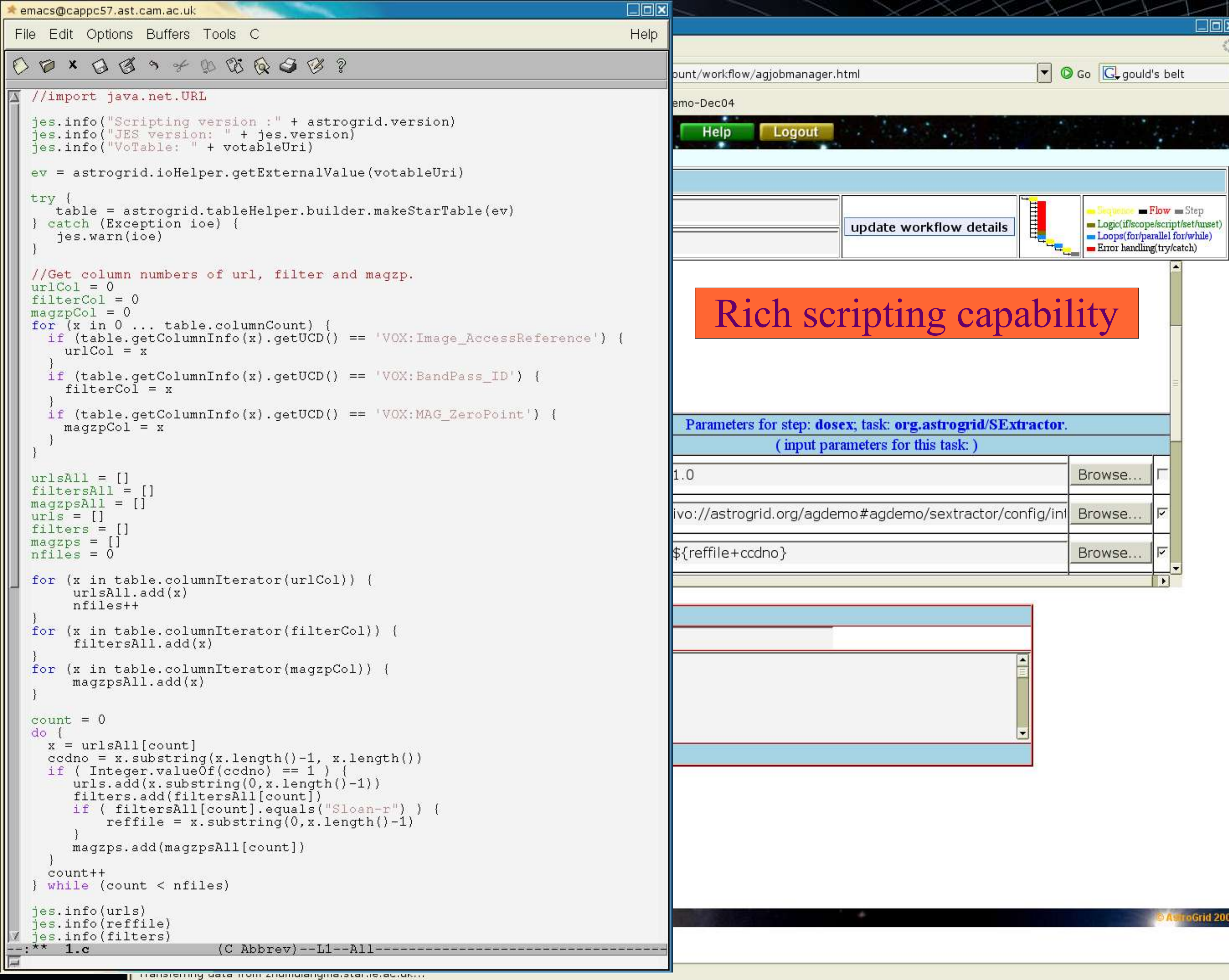


Generating the image catalogues

- For each image field pointing, the mosaiced image data is returned on the basis of individual CCD images
  - One INT WFC pointing is a mosaic image with 4 CCD images
  - Each CCD image 16 MB: one pointing in 5 colours > 1/3 GB
- Each field pointing observed in 5 colours
- Each image file processed with Sextractor
  - This application returns positions of objects and fluxes
- Outputs for each colour federated
  - Metadata added to output file
  - This uses a VO federator application
- Sextractor runs at Jodrell
- The Federator runs in Leicester



The Workflow: Sextractor

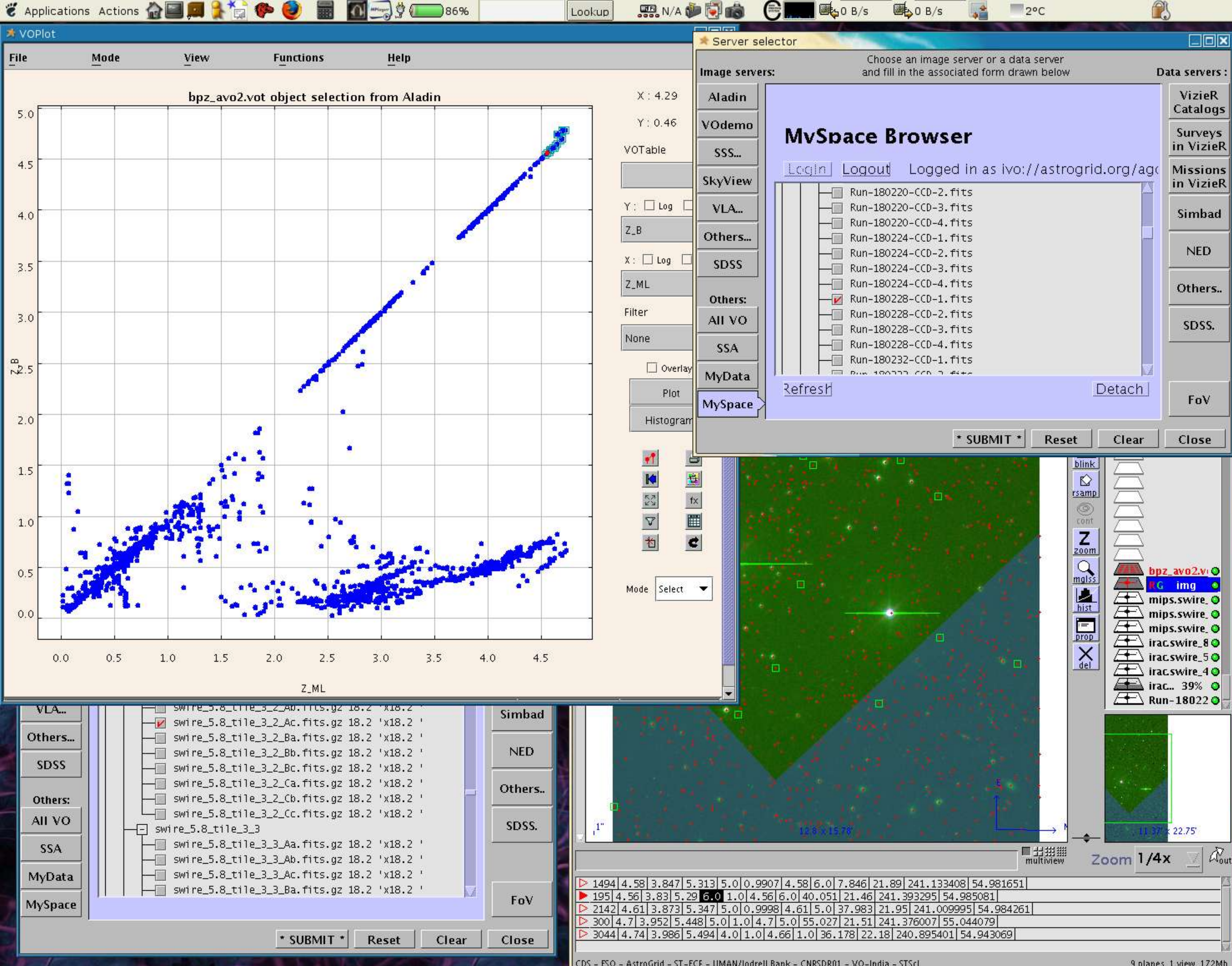


IoA UNIVERSITY OF CAMBRIDGE

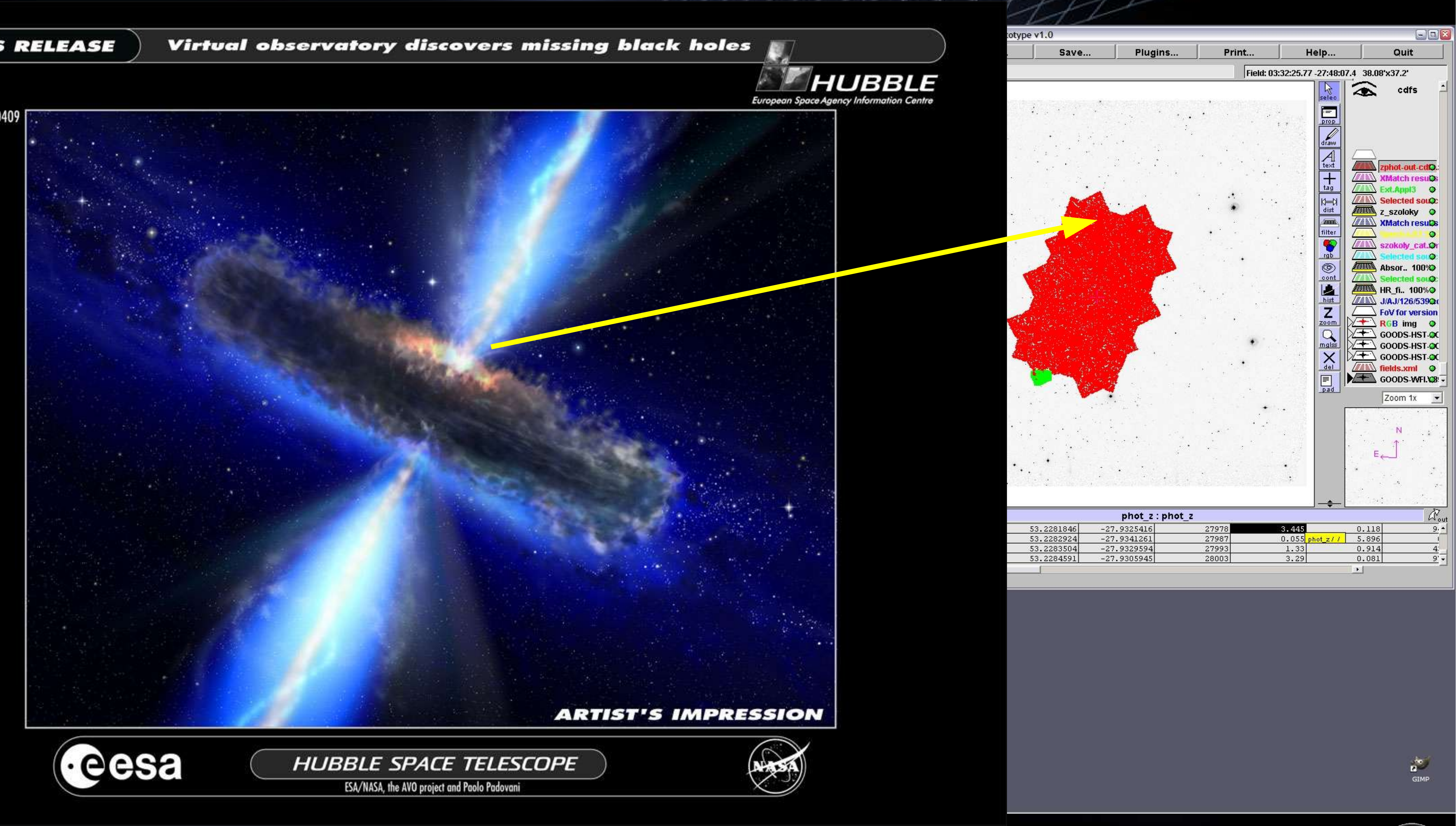
Hyperz and Bpz: The redshift apps

- The object photometry catalogues are fed into two apps
  - Hyperz: determines redshifts by template fitting to the SEDs
  - Bpz: baysian technique, similar to hyperz but includes weighting to reduce degenerate fits at different redshifts
- Applications run at Jodrell
  - Both compute intensive, generating large hyper cubes
- End step is the creation of two output files > then merged
- Final output catalogue:
  - Per field contains objects, photometry and redshifts
  - Integration with external viewers such as AVO/Aladin

Optical – IR – Redshifts



Example Usage:
1st Science from the AVO



Science Case Extension

Extending this science case ...
... this brings many elements of VO technologies lectures 1, 2, and 3 together ...

Observation/ Model Comparisons

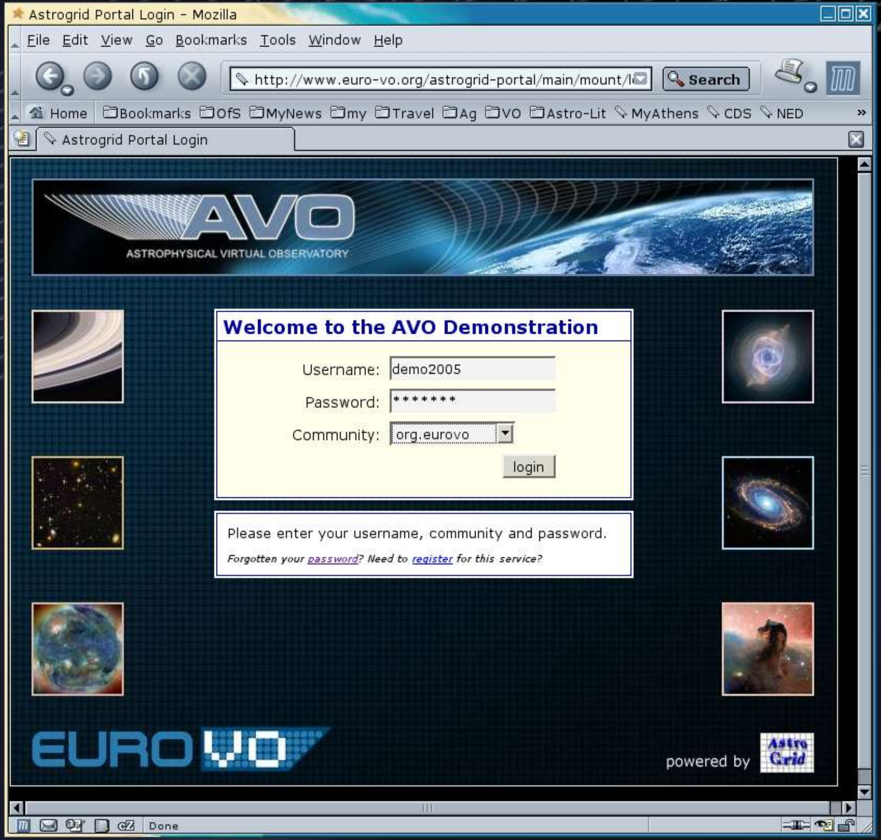
http://www.euro-vo.org/twiki/bin/view/Avo/AvoDemo2005Gal

- Observational Data:
  - At a position, return broadband photometric catalogues
  - Manipulate to generate uniform Spectral Energy Distributions
  - Plot and study the SEDs
  - Select samples: e.g. based on colours
- Theoretical Data:
  - For multiple spectral synthesis models
  - Generate multiple theoretical SEDs (e.g. For range of ages)
- Comparison
  - Access and compare/fit observational and model SEDs
- Iterate
  - Home in on parameter space to generate 'best fit' models

Workflow Overview

Euro-VO Portal

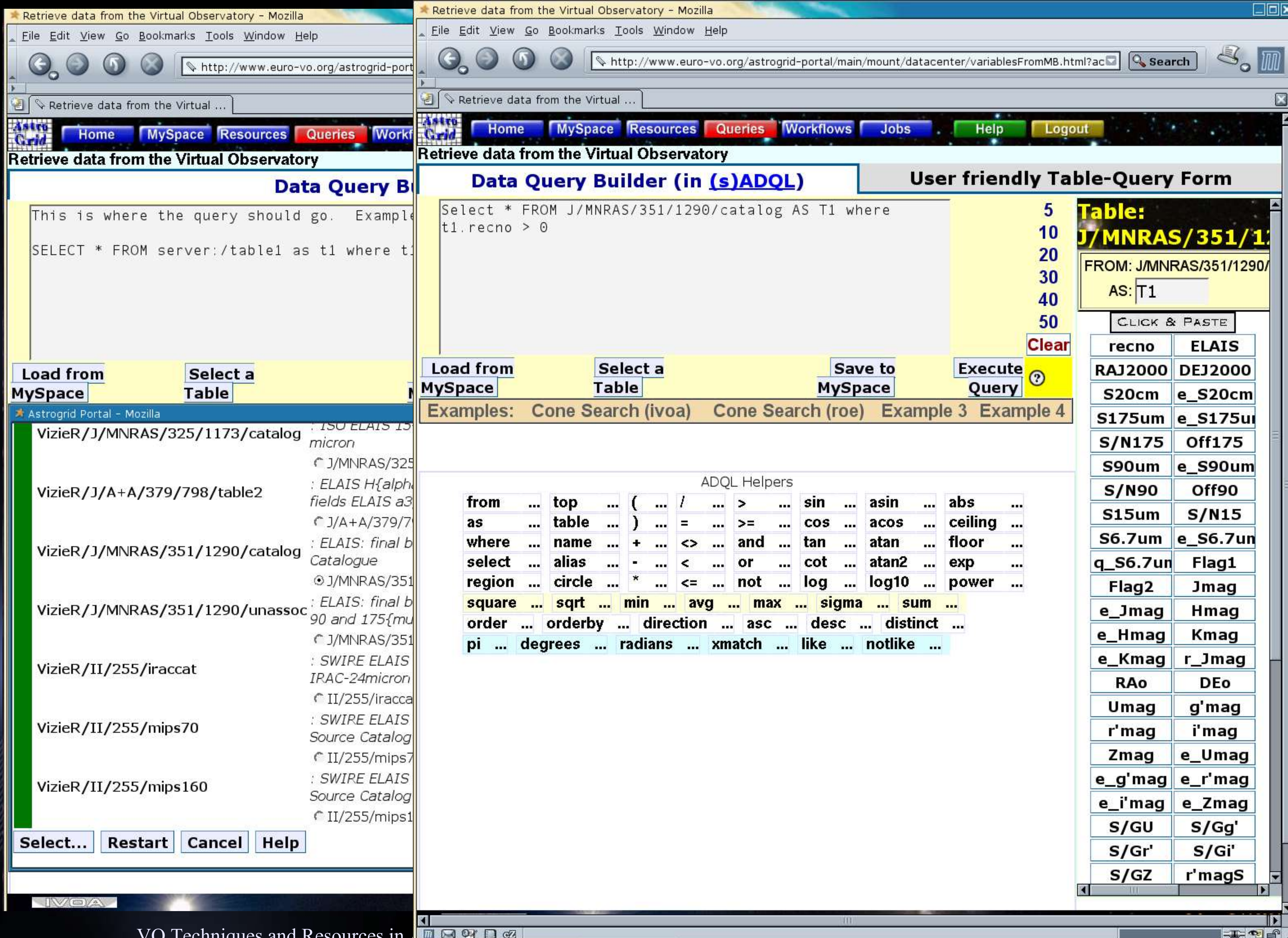
- User login
  - Authentication
  - Authorisation
  - Communities
- Credentials in Registry
- Central Access
  - All further components
- Portal runs in Garching
  - Registry in Leicester
  - Data Access in Cambridge, Edinburgh, Manchester, Strasbour
  - Applications in Garching, Cambridge
  - Thus one user runs processes over a distributed network.



AstroGrid Portal

Querying the catalogue data

- ELIAS photometric catalogue
  - Held in queriable database
- Query constructed with IVOA std ADQL
  - Table metadata uploaded from the registry call
  - Query saved to MySpace
- Workflow element
  - Send data query to the database (CDS)
- Workflow execution
  - Query sent through a standard VOQL call to the database
  - Results of query returned to MySpace in a std VOTable file



Retrieve data from the Virtual Observatory - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://www.euro-vo.org/astrogrid-portal/main/mount/datacenter/variablesFromMB.html?ac
Search
Retrieve data from the Virtual ...
Home MySpace Resources Queries Workflows Jobs Help Logout
Retrieve data from the Virtual Observatory
Data Query Builder (in (s)ADQL)
User friendly Table-Query Form
Select * FROM J/MNRAS/351/1290/catalog AS T1 where
t1.recno > 0
5
10
20
30
40
50
Clear
Load from MySpace
Select a Table
Save to MySpace
Execute Query
Examples: Cone Search (ivoa) Cone Search (roe) Example 3 Example 4
ADQL Helpers
from top (/ > sin asin abs
as table) = >= cos acos ceiling
where name + <> and tan atan floor
select alias - < or cot atan2 exp
region circle * <= not log log10 power
square sqrt min avg max sigma sum
order orderby direction asc desc distinct
pi degrees radians xmatch like notlike
Table: J/MNRAS/351/1
FROM: J/MNRAS/351/1290/
AS: T1
CLICK & PASTE
recno ELAIS
RAJ2000 DEJ2000
S20cm e_S20cm
S175um
S/N175 Off175
S90um e_S90um
S/N90 Off90
S15um S/N15
S6.7um
Flag1
Flag2 Jmag
e_Jmag Hmag
e_Hmag Kmag
e_Kmag r_Jmag
RAo DEo
Umag g'mag
r'mag i'mag
Zmag e_Umag
e_g'mag e_r'mag
e_i'mag e_Zmag
S/GU S/Gg'
S/Gr' S/Gi'
S/GZ r'magS
This is where the query should go. Example
SELECT * FROM server:/table1 as t1 where t
Astrogrid Portal - Mozilla
VizieR/J/MNRAS/325/1173/catalog
VizieR/J/A+A/379/798/table2
VizieR/J/MNRAS/351/1290/catalog
: ELAIS: final b
Catalogue
VizieR/J/MNRAS/351/1290/unassoc
VizieR/II/255/iraccat
: SWIRE ELAIS
IRAC-24micron
VizieR/II/255/mips70
VizieR/II/255/mips160
Select... Restart Cancel Help

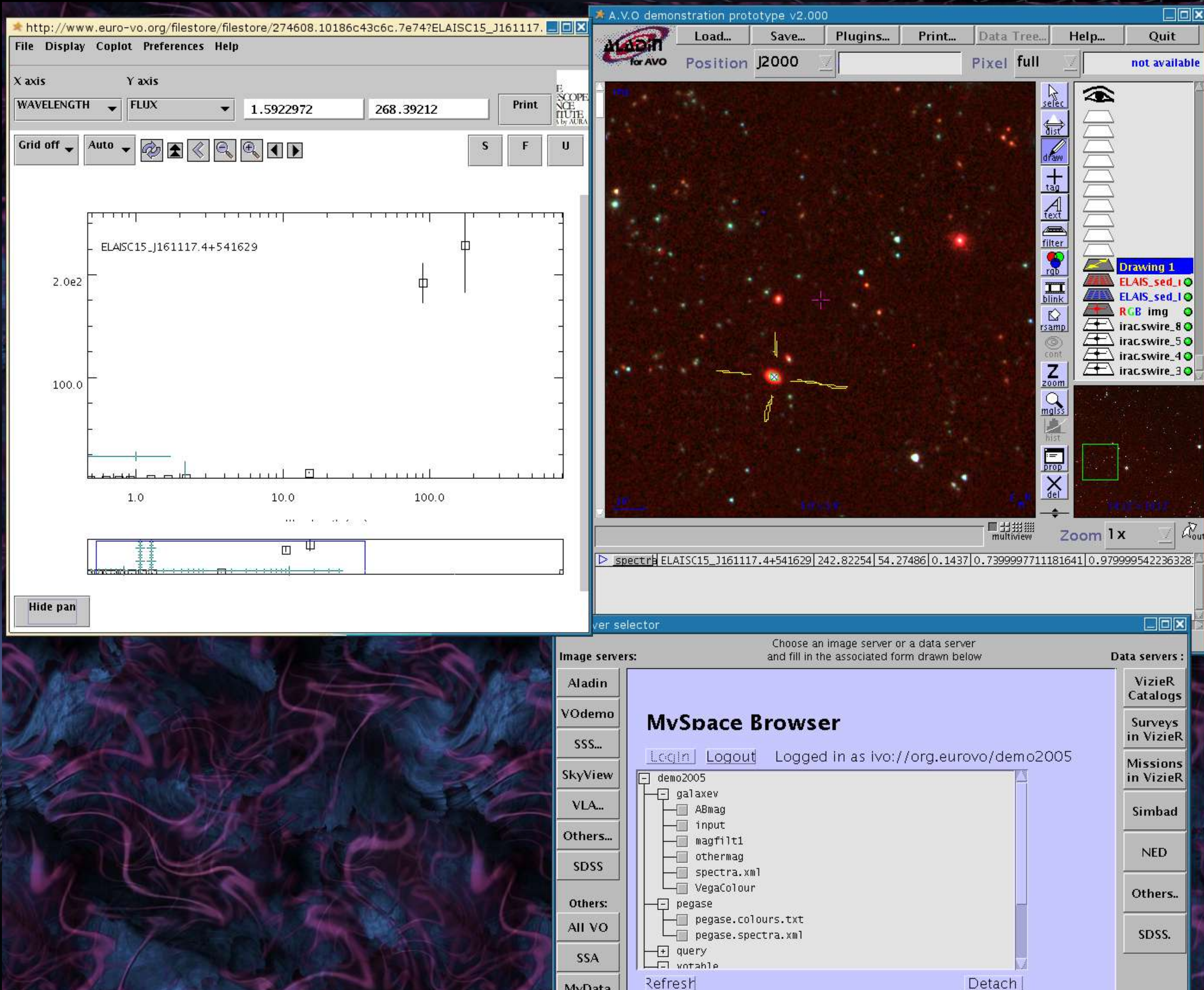
Workflow: Generating the SEDs

- The ELAIS catalogue contains 3500 objects
  - IR + opt photometry
- Problem:
  - IR fluxes, but optical magnitudes
- Solution:
  - Convert magnitudes to fluxes based on magnitude zero points
- Script
  - Takes input catalogue
  - Converts magnitudes to fluxes, and creates per object SED files
  - Each SED file in a standard format
  - Script implemented in Groovy (http://www.groovy.org)
- Workflow:
  - Find catalogue – retrieve catalogue – generate SEDs

Visualising the SEDs

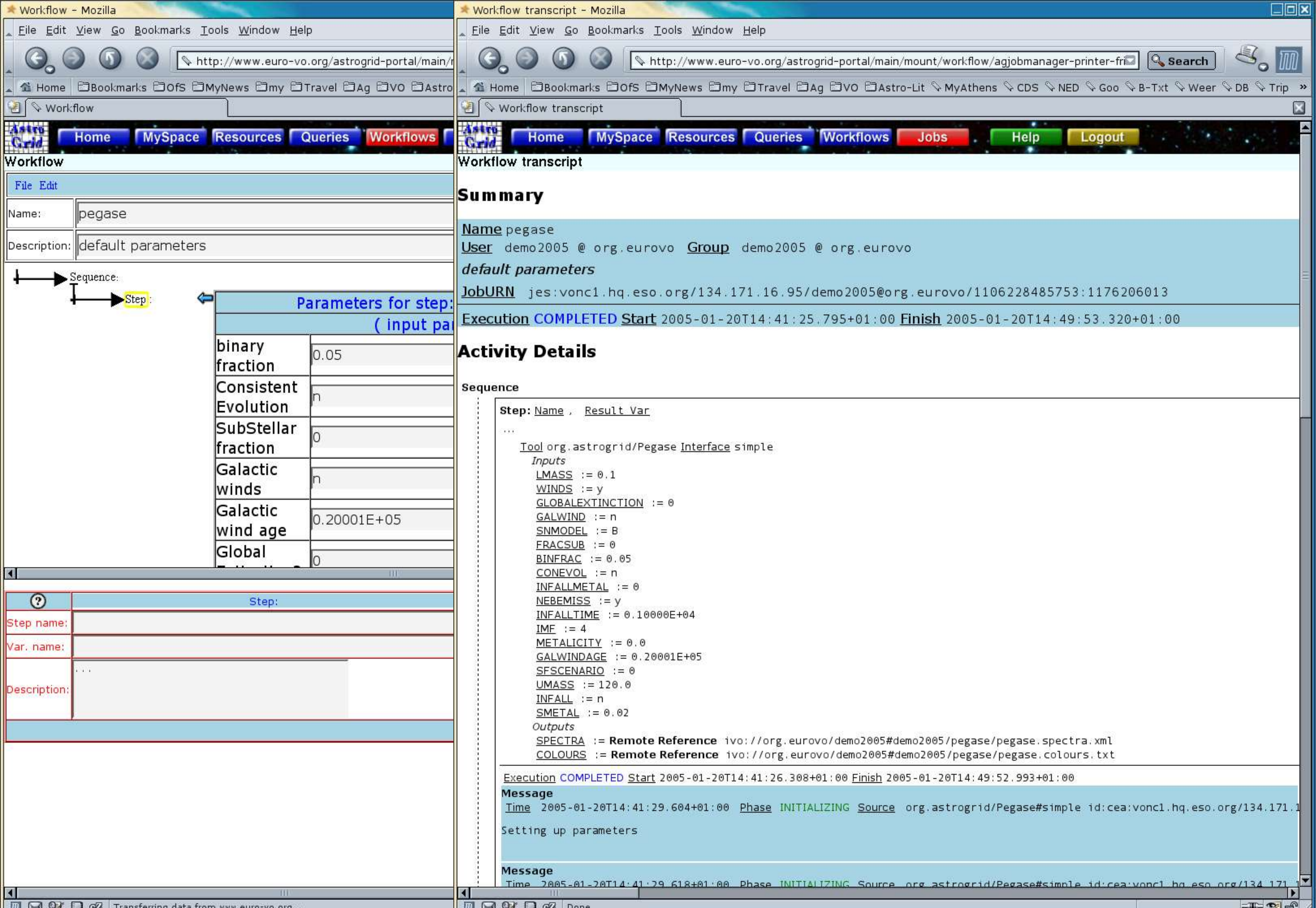
- SEDs stored in MySpace
  - SED Index: Simple Spectrum Access standard
- Accessible:
  - The portal
  - External applications: e.g. Topcat, Treeview
  - In this case Aladin and Specview
- Aladin:
  - Myspace browser
  - Select SED
  - Each SED displayed in an Aladin plane
  - In the metadata browser, click 'view' spectrum
    - SED transferred to Specview for analysis

SED generate & visualise



Pegase and GALEXV: The spectral synthesis model applications

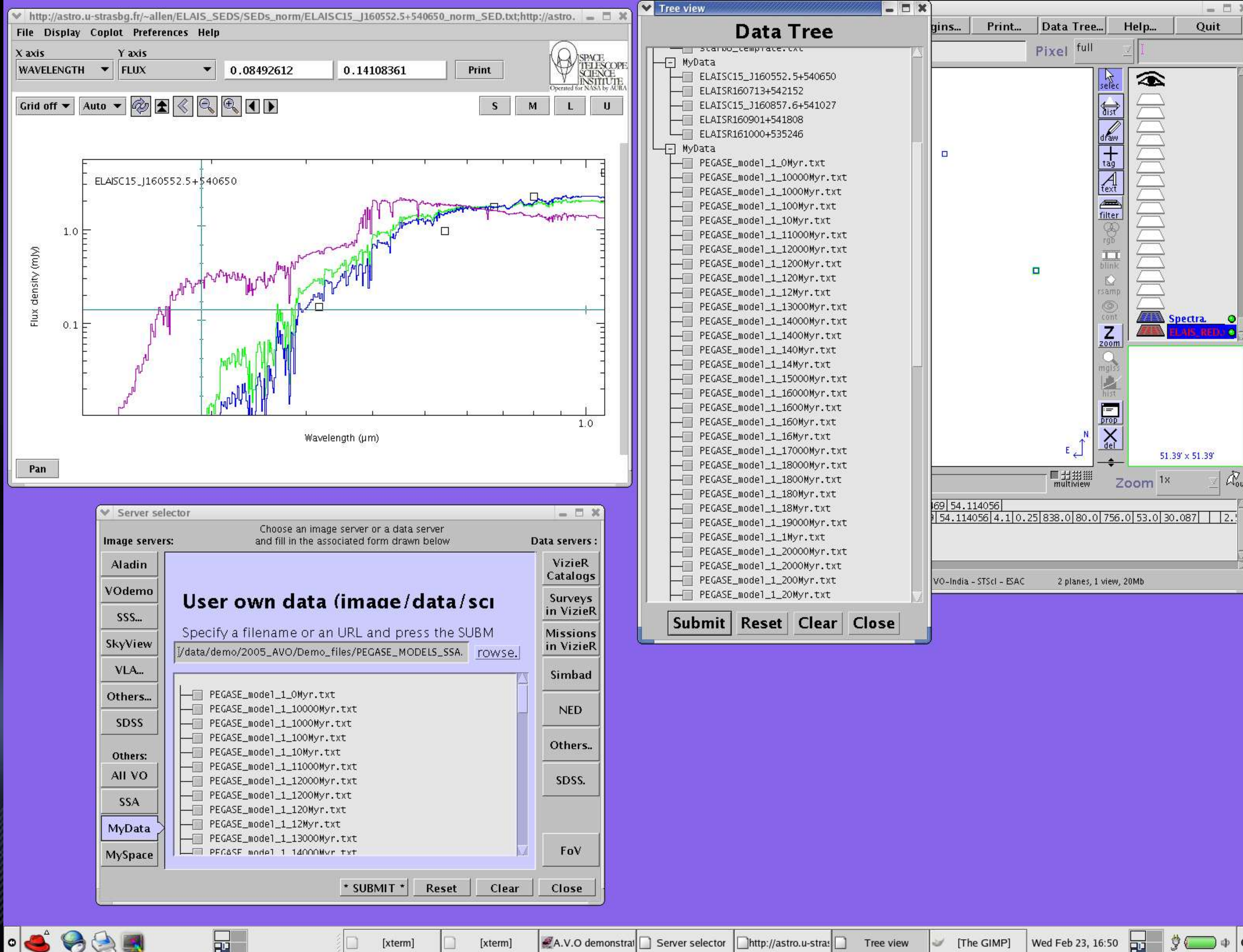
- Two techniques to generate theoretical galaxy spectral energy distributions.
- Search 'registry' to 'discover' relevant applications
- For each, include in workflow:
  - User ability to alter input parameters
  - Each run generates sets of output spectra
  - Output formats conform the VO interoperability standards
- Applications run within the AstroGrid CEA
  - Common Execution Architecture
    - Standard framework for applications
    - Metadata describes application, discovererable through 'Registry'


Workflow - Mozilla
Workflow transcript - Mozilla
http://www.euro-vo.org/astrogrid-portal/main/mount/workflow/agjobmanager-printer-fri
Home MySpace Resources Queries Workflows Jobs Help Logout
Workflow
File Edit
Name: pegase
Description: default parameters
Sequence:
Step:
Parameters for step:
(input pa
binary fraction 0.05
Consistent Evolution n
SubStellar fraction 0
Galactic winds n
Galactic wind age 0.20001E+05
Global 0
Step:
Step name:
Var. name:
Description:
Transferring data from www.euro-vo.org...
Workflow transcript
Summary
Name pegase
User demo2005 @ org.eurovo Group demo2005 @ org.eurovo
default parameters
JobURN jes:voncl.hq.eso.org/134.171.16.95/demo2005@org.eurovo/1106228485753:1176206013
Execution COMPLETED Start 2005-01-20T14:41:25.795+01:00 Finish 2005-01-20T14:49:53.320+01:00
Activity Details
Sequence
Step: Name , Result Var
Tool org.astrogrid/Pegase Interface simple
Inputs
LMASS := 0.1
WINDS := y
GLOBALEXTINCTION := 0
GALWIND := n
SNMODEL := B
FRACSUB := 0
BINFRAC := 0.05
CONEVOL := n
INFALLMETAL := 0
NEBEMISS := y
INFALLTIME := 0.10000E+04
IMF := 4
METALICITY := 0.0
GALWINDAGE := 0.20001E+05
SFSCENARIO := 0
UMASS := 120.0
INFALL := n
SMETAL := 0.02
Outputs
SPECTRA := Remote Reference ivo://org.eurovo/demo2005#demo2005/pegase/pegase.spectra.xml
COLOURS := Remote Reference ivo://org.eurovo/demo2005#demo2005/pegase/pegase.colours.txt
Execution COMPLETED Start 2005-01-20T14:41:26.308+01:00 Finish 2005-01-20T14:49:52.993+01:00
Message
Time 2005-01-20T14:41:29.604+01:00 Phase INITIALIZING Source org.astrogrid/Pegase#simple id:cea:voncl.hq.eso.org/134.171.1
Setting up parameters
Message
Done

Observation/ Model Comparisons

- Observational and Theory data stored in MySpace
  - 'one click' access through Aladin
  - Read SEDs into Specview
  - Read models into Specview
- For any observed SED
  - Select and 'double left click'
  - Allows overplotting of model
  - Select best model
    - Rerun models to tune input parameters for best match
- Automatic model/observed SED fitting
  - This tool required – possible development through VOTECH.

SED fitting …



http://astro.u-strasbg.fr/~allen/ELAIS_SEDS/SEDs_norm/ELAISC15_J160552.5+540650_norm_SED.txt;http://astro.
File Display Coplot Preferences Help
X axis
Y axis
WAVELENGTH
FLUX
0.08492612
0.14108361
Print
Grid off
Auto
S M L U
ELAISC15_J160552.5+540650
Flux density (mJy)
Wavelength (µm)
Pan
Tree view
Data Tree
MyData
ELAISC15_J160552.5+540650
ELAISR160713+542152
ELAISC15_J160857.6+541027
ELAISR160901+541808
ELAISR161000+535246
PEGASE_model_1_0Myr.txt
PEGASE_model_1_10000Myr.txt
PEGASE_model_1_1000Myr.txt
PEGASE_model_1_100Myr.txt
PEGASE_model_1_10Myr.txt
Submit Reset Clear Close
Server selector
Choose an image server or a data server
and fill in the associated form drawn below
Image servers:
Aladin
VOdemo
SSS...
SkyView
VLA...
Others...
SDSS
Others:
All VO
SSA
MyData
MySpace
Data servers:
VizieR Catalogs
Surveys in VizieR
Missions in VizieR
Simbad
NED
Others..
SDSS.
FoV
User own data (image/data/sci
Specify a filename or an URL and press the SUBM
/data/demo/2005_AVO/Demo_files/PEGASE_MODELS_SSA.
* SUBMIT * Reset Clear Close
Print... Data Tree... Help... Quit
Pixel
Zoom
Spectra
VO-India - STScI - ESAC
2 planes, 1 view, 20Mb
[xterm]
A.V.O demonstra
Server selector
Tree view
[The GIMP]
Wed Feb 23, 16:50

Lecture 3: Acknowledgements + Refs

- SSA slides 5 to 7 adapted from Doug Tody: see http://www.us-vo.org/summer-school/proceedings/presentations/dal-nvoss.ppt
- IVOA standards – see http://www.ivoa.net/forum/
- Pegase: see http://www2.iap.fr/users/fioc/PEGASE.html
- Galexv: see http://www.cida.ve/~bruzual/bc2003

Next Lecture:
Mining the Sloan Digital Sky Survey