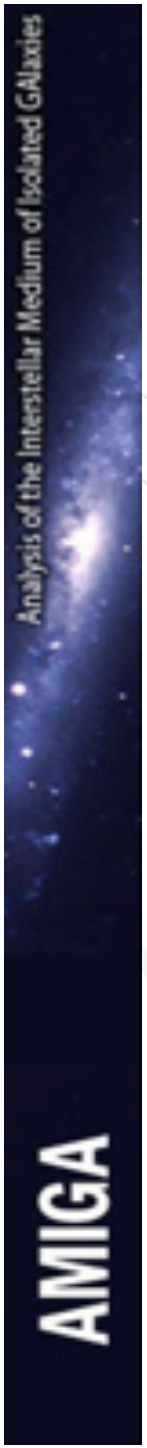


Workflows
Access and Massage VO Data

José Enrique Ruiz
on behalf of the Wf4Ever Team

IVOA INTEROP SPRING MEETING 2013
HEIDELBERG, MAY16th 2013

Wf4Ever
Advanced Workflow Preservation Technologies for Enhanced Science
2011 - 2013

1. Intelligent Software Components (ISOCO, Spain)
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (P
5. University of Oxford and OeRC (OXF, UK)

6. Instituto Astrofísica Andalucía (IAA-CSI
7. Leiden University Medical Centre (L

Reproducible Science

The University of Manchester

MANCHESTER 1824

INSTITUTO de ASTROFISICA de ANDALUCIA CSIC IAA

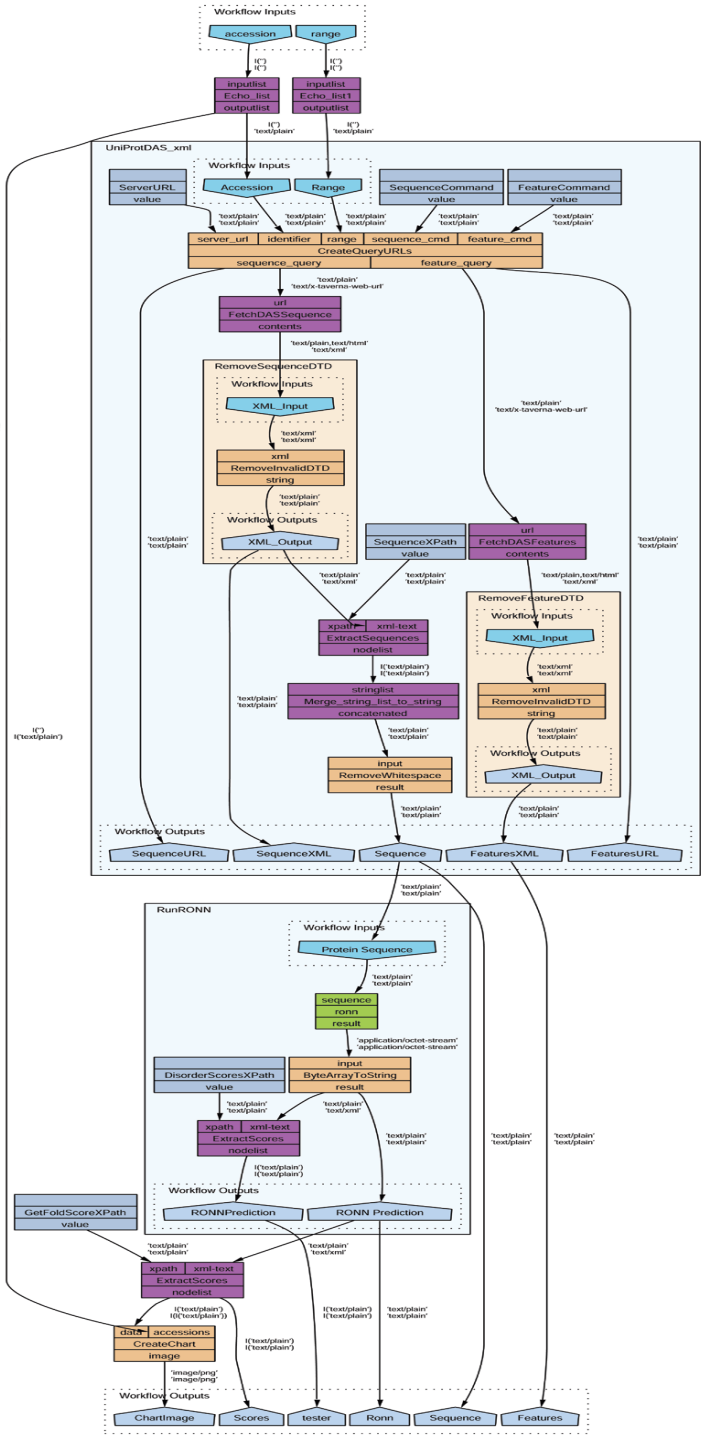
POZNAN SUPERCOMPUTING NETWORKING

iSOCO enabling the networked economy

LUMC

SEVENTH FRAMEWORK PROGRAMME

European Commission Information Society and Media



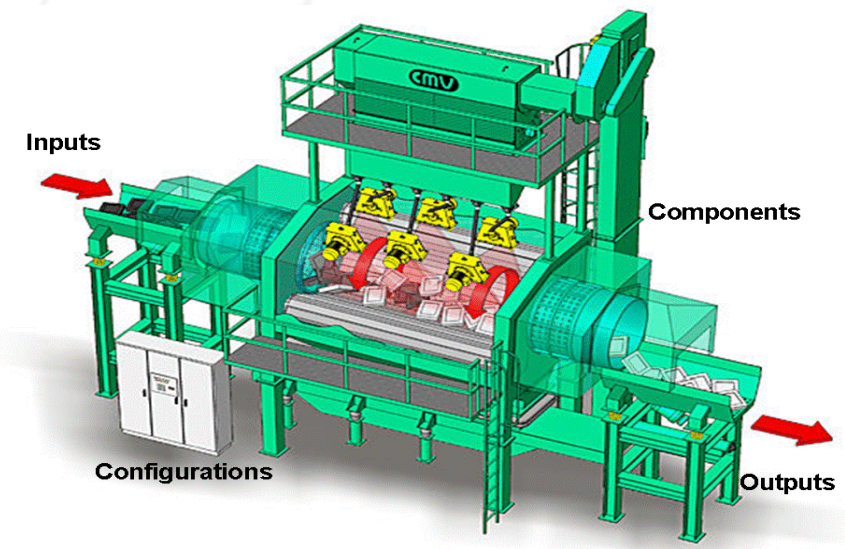
Workflows to Access and Massage VO Data

What is a Scientific Workflow?

» A mechanism for coordinating the execution of services and codes, and linking together resources.

» The combination of data and processes into a configurable, modular, structured set of steps that implement semi-automated computational solutions in scientific problem-solving.

» The implementation of a scientific method.



State of the art in Astronomy

» IVOA Note Definition

These are networks of analytical steps that may involve, e.g., database access and querying steps, data analysis and mining steps, and many other steps including computationally intensive jobs on high performance cluster computers.

» Wf Software

- Taverna
- Kepler
- Pegasus
- Triana
- ESO Reflex

Related Initiatives

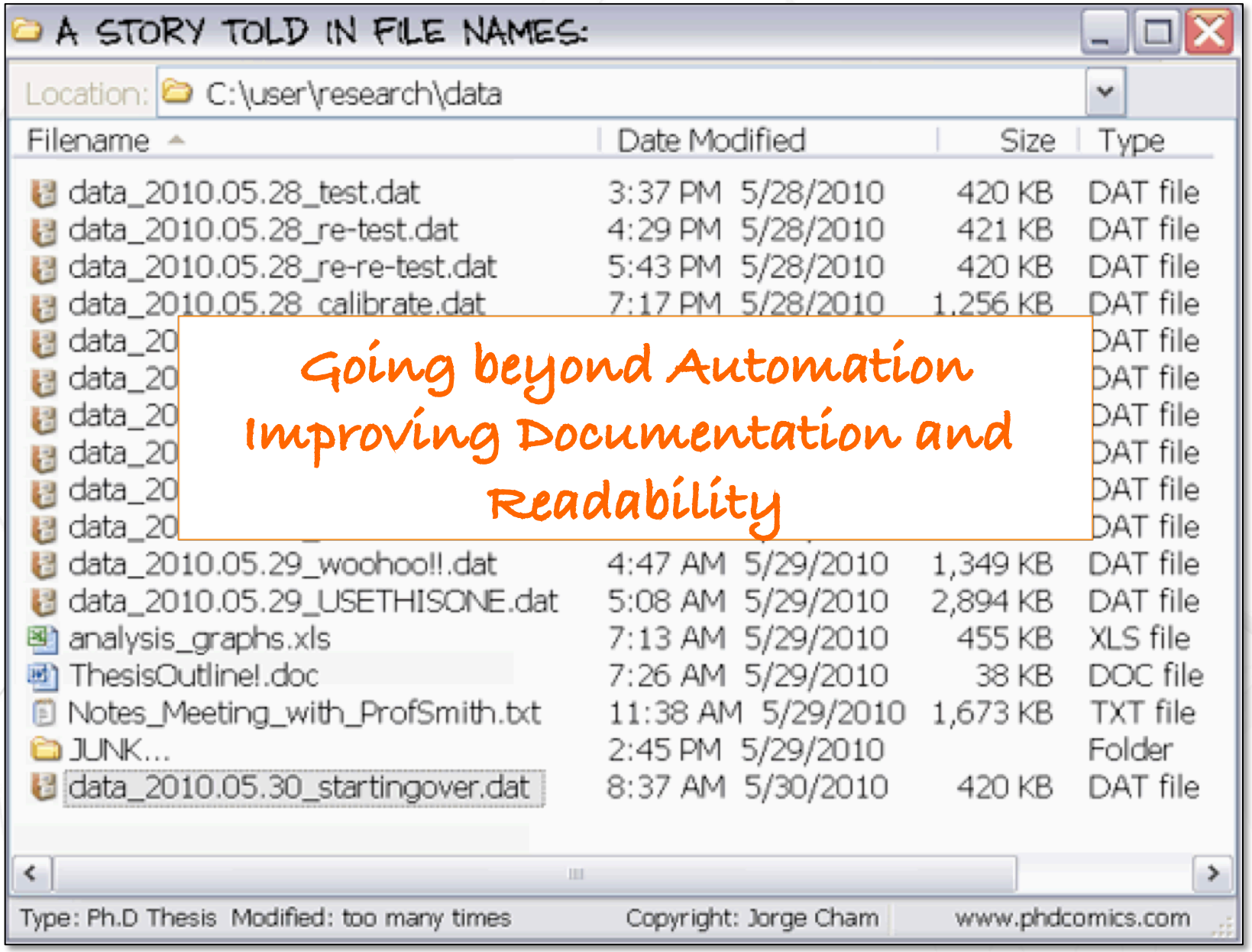
- ER-Flow
- VAMDC
- Helio-VO
- Cyber-SKA
- IceCore
- Montage
- Astro-WISE
- AstroGrid

In the VO

- GWS WG
- VO France WF WG
- VAMDC
- AstroGrid

Workflows to Access and Massage VO Data

Digital Astronomy



Going beyond Automation
Improving Documentation and Readability

AstroTaverna Workflows

Retrieving and Manipulating VO Data **+ Catalog … TML Pages**

- ConeSearch
- SIA
- SSA
- TAP coming soon…

… Web Services

- Tabular Data (VOTables)
- Images, but not yet Spe…

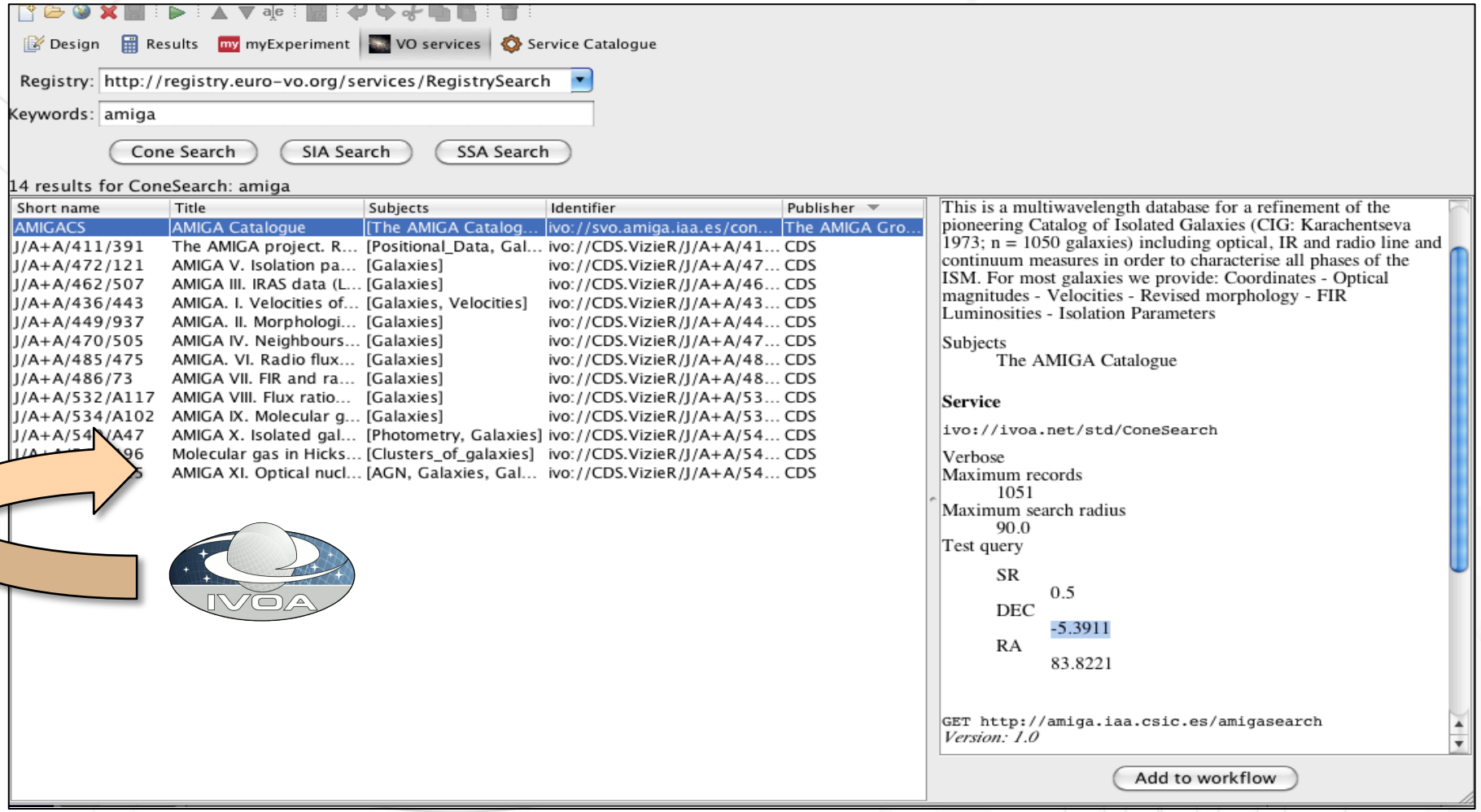
…ccess to JDBC databases

- Crossmatching, Fi… …solving, Coordinates and reference system transfor… …ssage.. (**STILTS**)
- Overplotting sc… …s on Images and filtering, overplot circles, ellipses, etc. as a… …of physical magnitude. Resampling, crops, blinks, mosaics, mov… …blinks, RGBs, fusion, diff.. (**ALADIN**)
- **SAMP** for final inspection

No interactive actions and decisions based on visual inspection

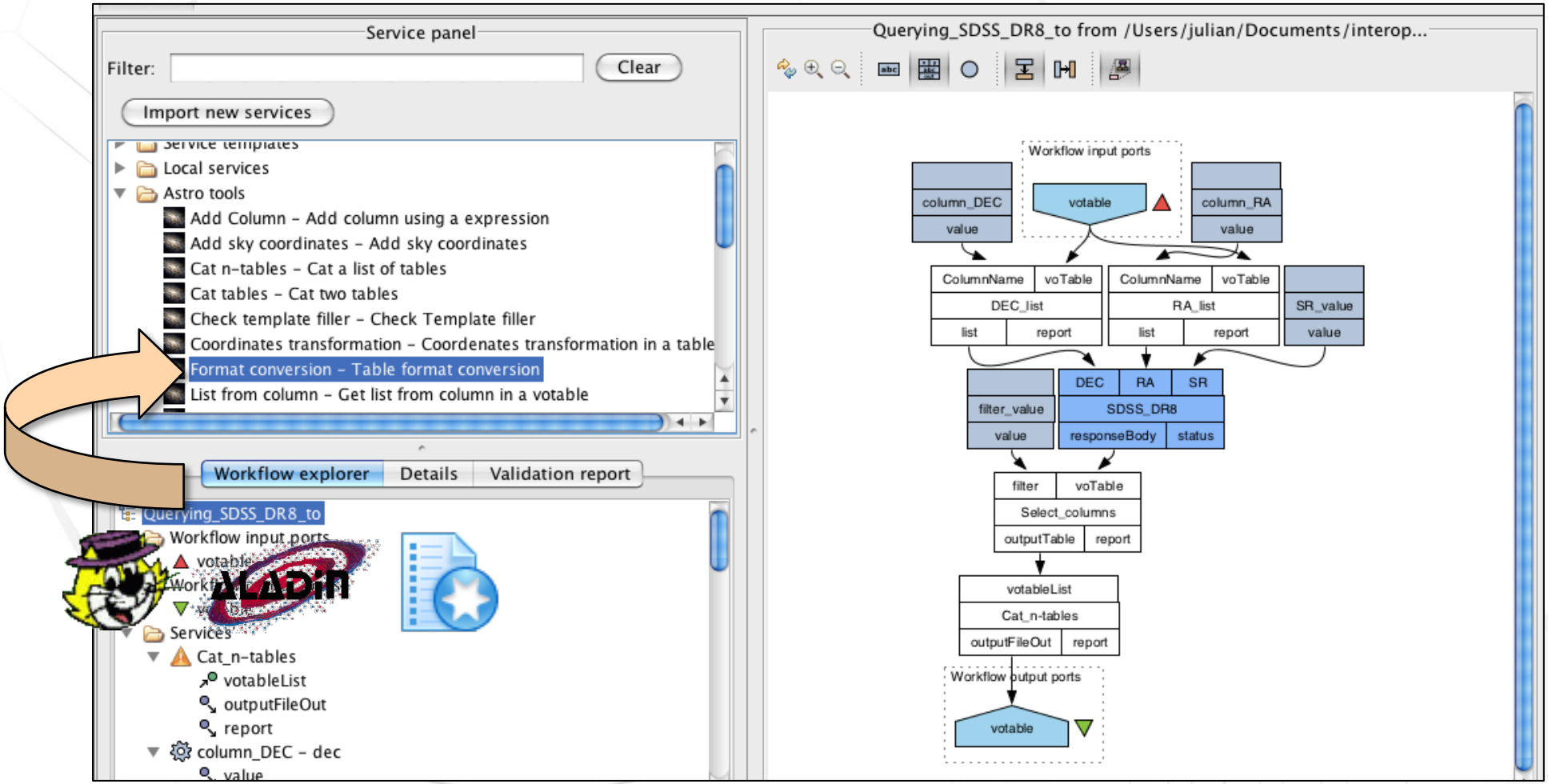
+ Advanced Analysis using Scripts

VOData Access: VO Services Discovery



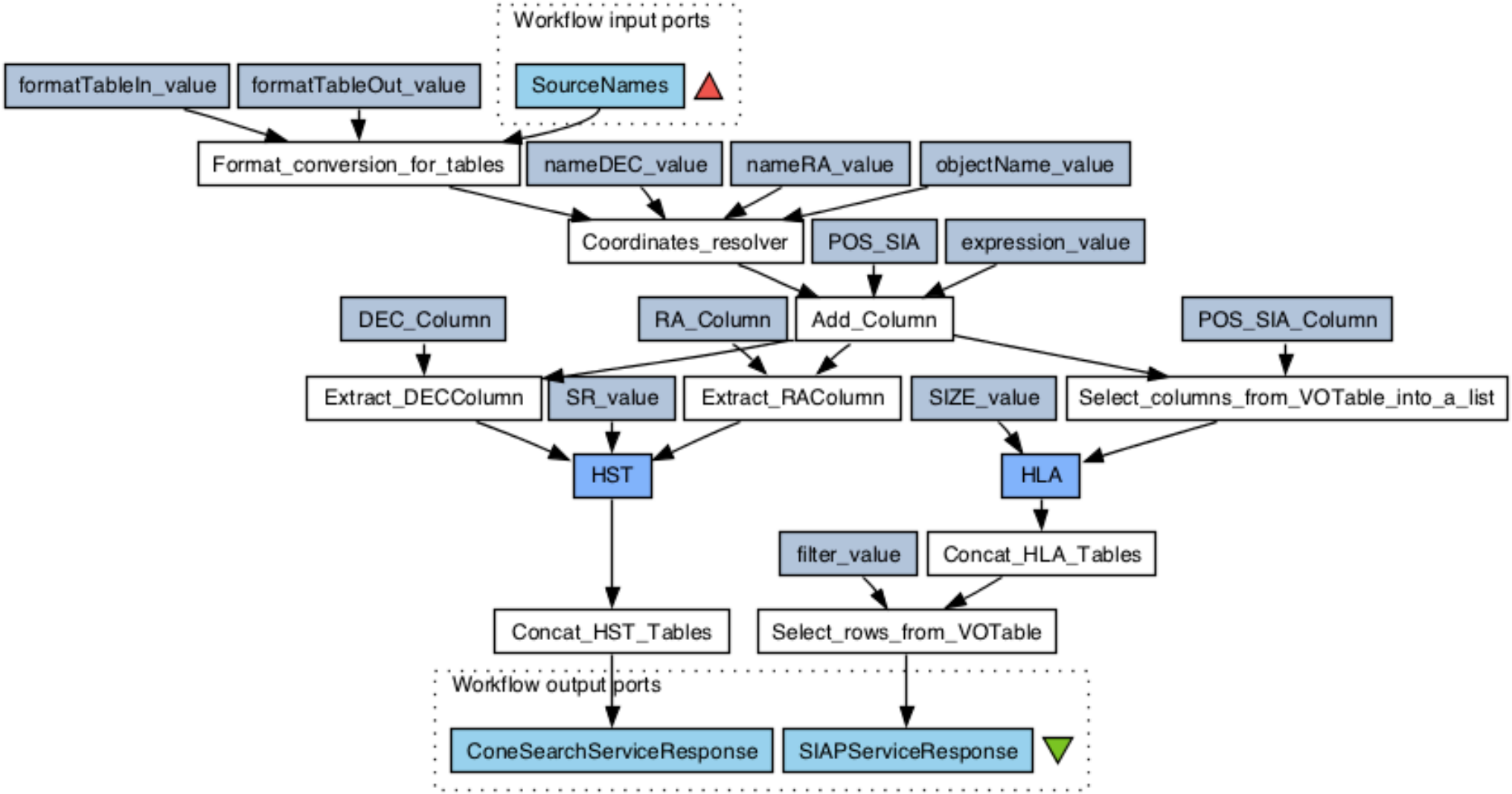
http://amiga.iaa.es/p/290-astrotaverna.htm

VOData Massage: VOTables, STILTS, Aladin, TerminalSim



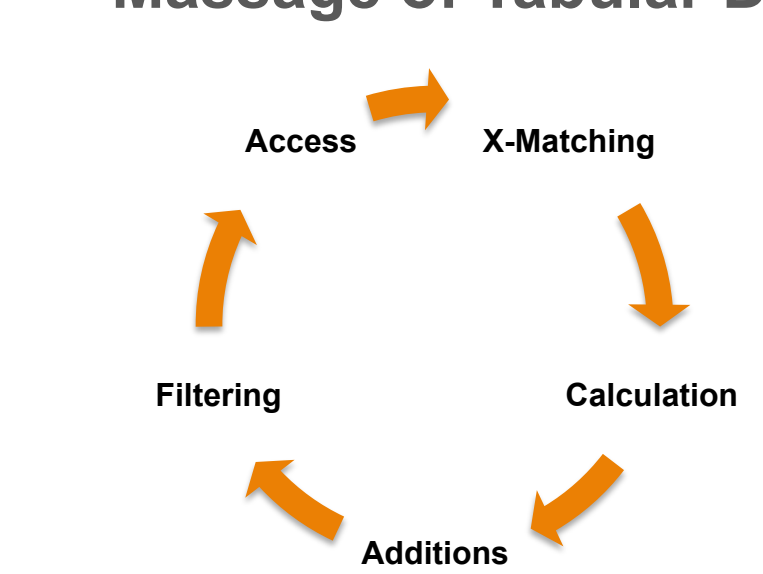
http://amiga.iaa.es/p/290-astrotaverna.htm

VOData Consumers

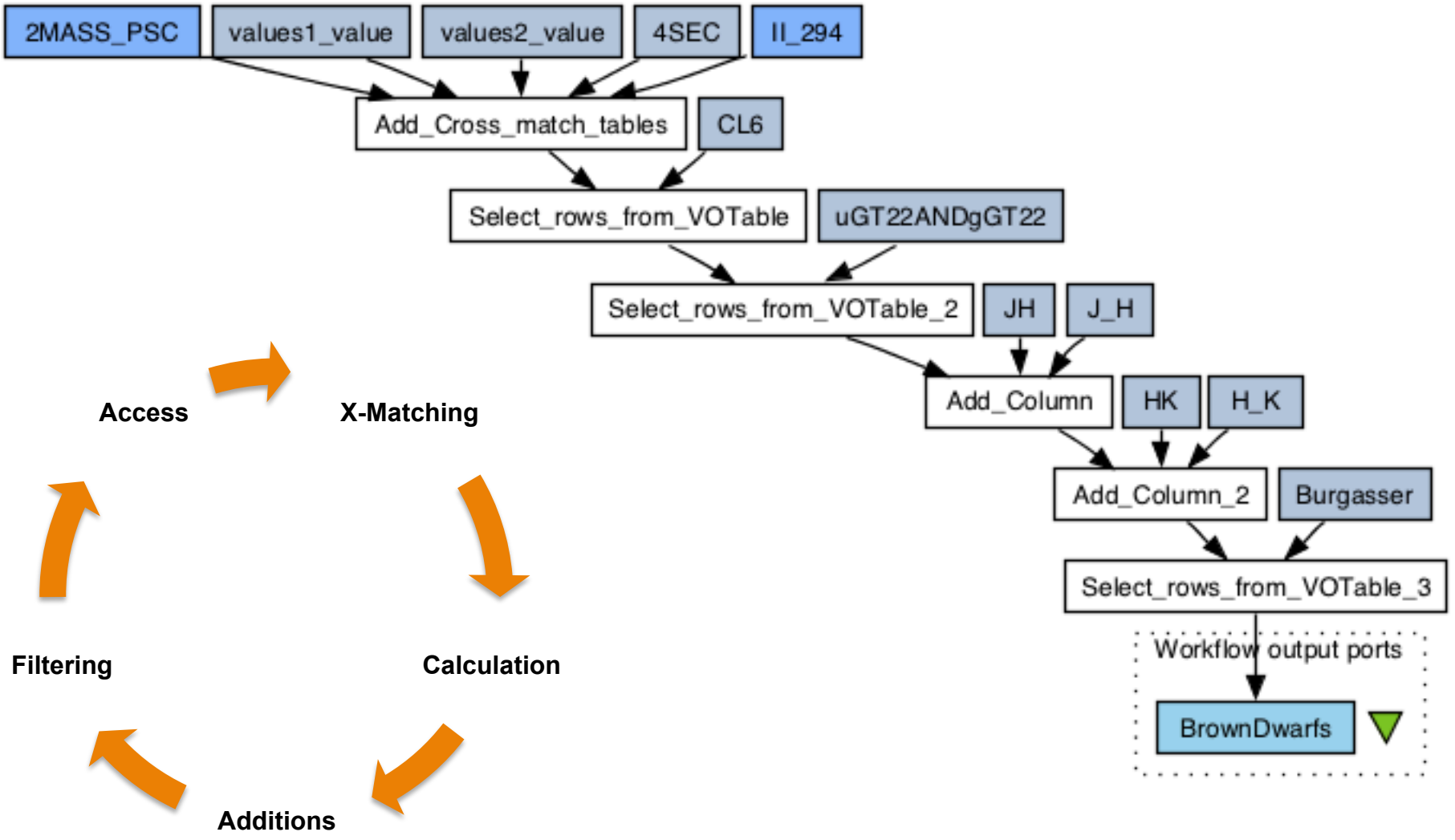


VOData Manipulation

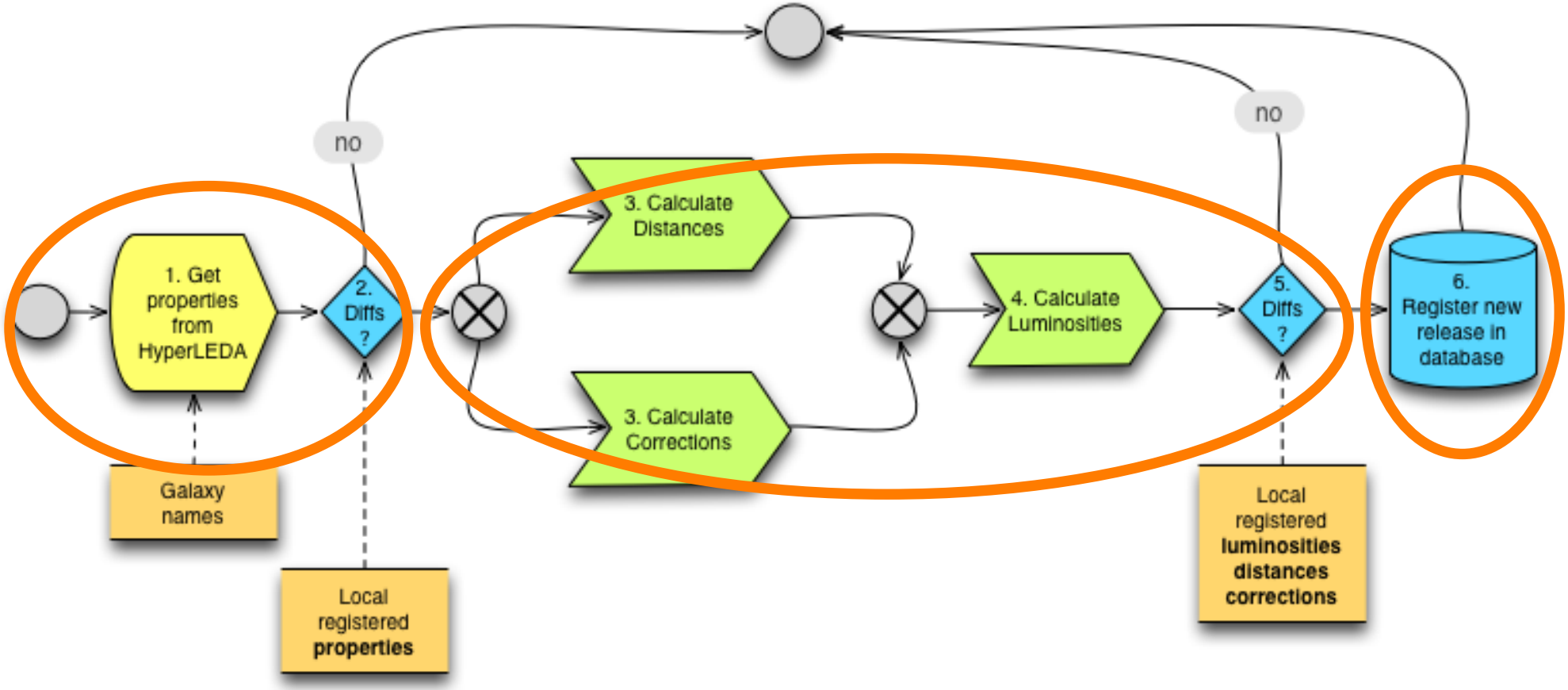
Massage of Tabular Data



VOData Manipulation

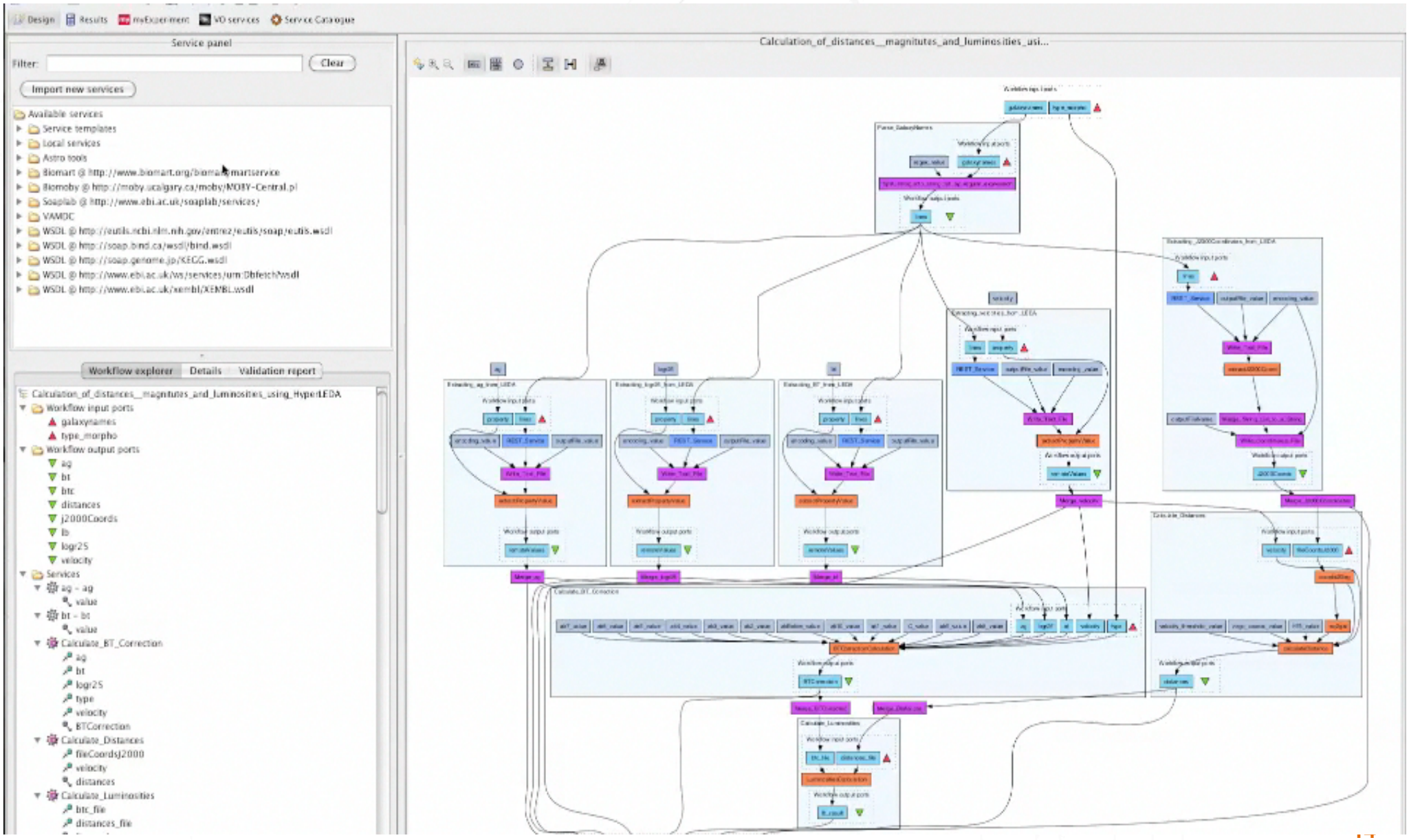


VOData Curation



Workflows to Access and Massage VO Data

VOData Curation

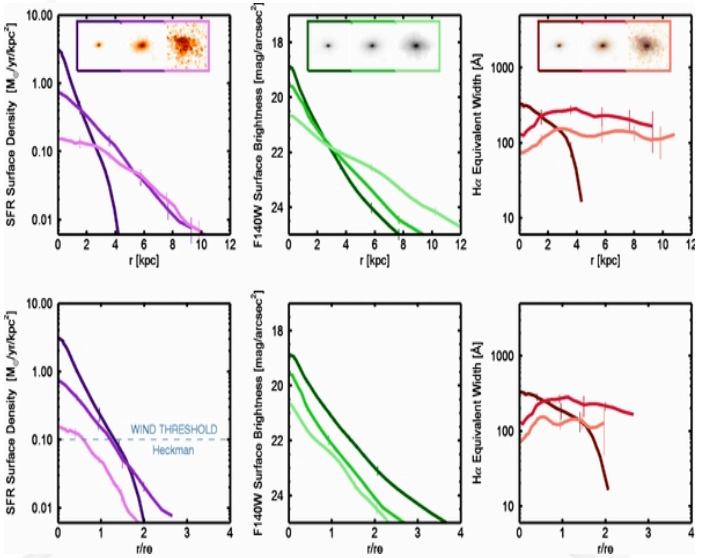


Calculation of Luminosity Profiles for a Sample of Galaxies extracted from SDSS DR8

IRAF
Image Reduction and Analysis Facility

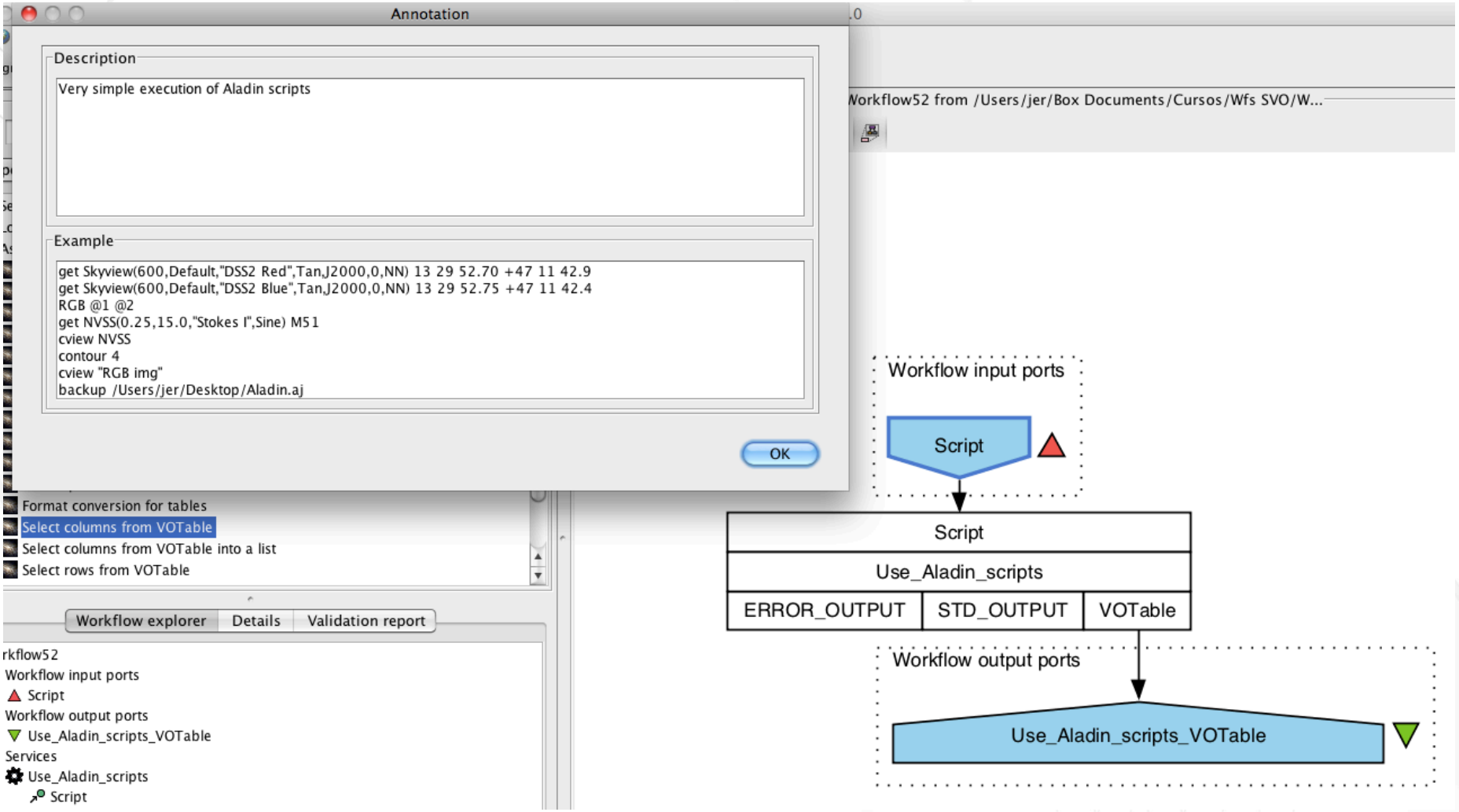
SExtractor

GALFIT



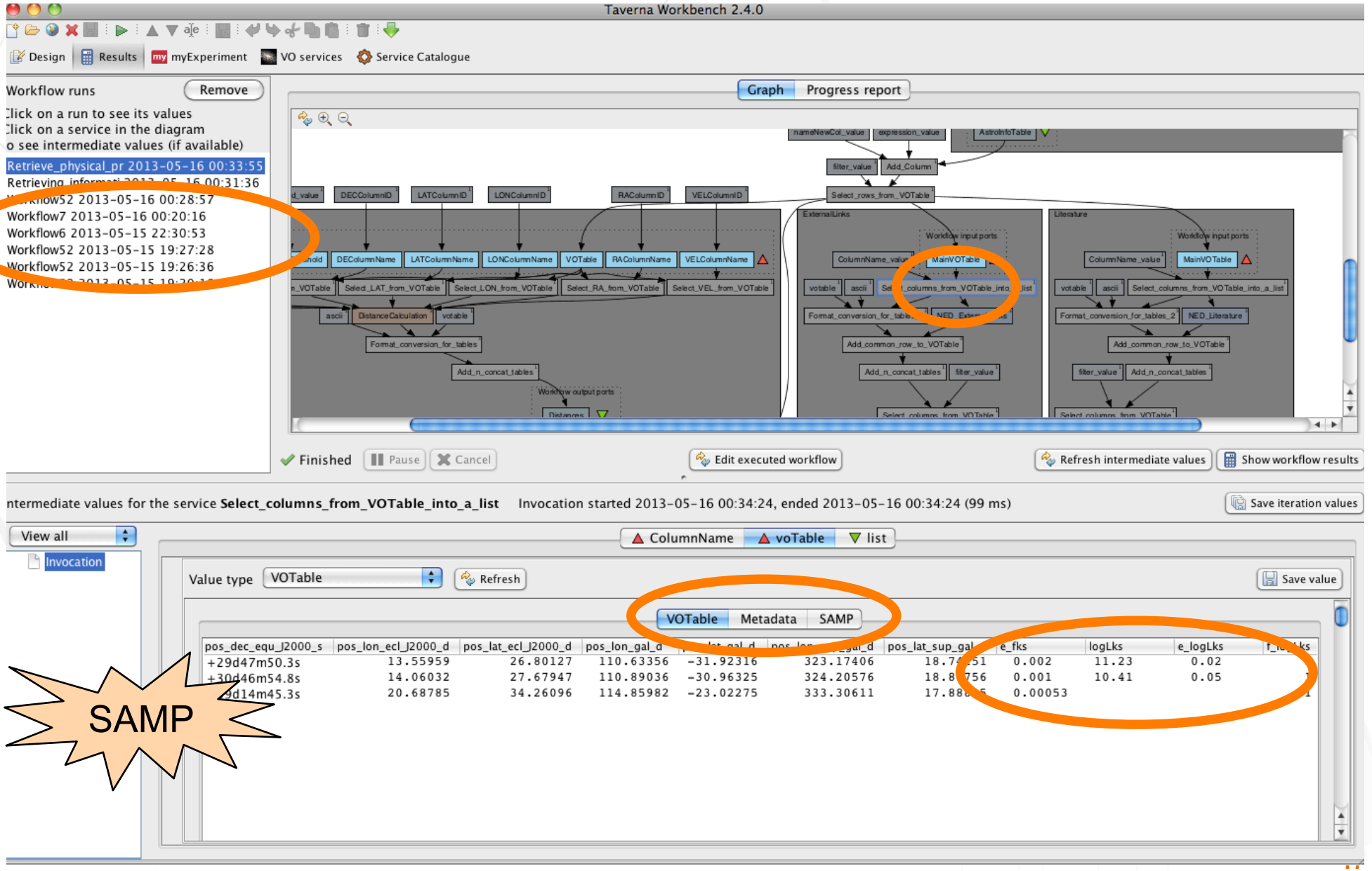
90 galaxies observed in 3 bands

Aladin Scripts and Macro executing in GUI/noGUI mode



Workflows to Access and Massage VO Data

VOData Inspection



Workflows to Access and Massage VO Data
Learning by the example

Pack: AstroTaverna Starter Pack

Update Upload Open Import

Pack: AstroTaverna Starter Pack

Creator: Jose Enrique Ruiz
Created at: Fri Apr 12 18:10:31 CEST 2013
Last updated at: Tue Apr 23 19:41:01 CEST 2013

This is a set of small snippets using AstroTaverna plugin, developed in the Wf4Eve
project. The purpose of this pack is to make the design and development of astrono
Virtual Observatory workflows easier, learning by the example.

AstroTaverna may be installed on Taverna 2.4 Workbench

http://wf4ever.github.io/astrotaverna/

Pack Items (28) | Tags (5) | Comments (0)

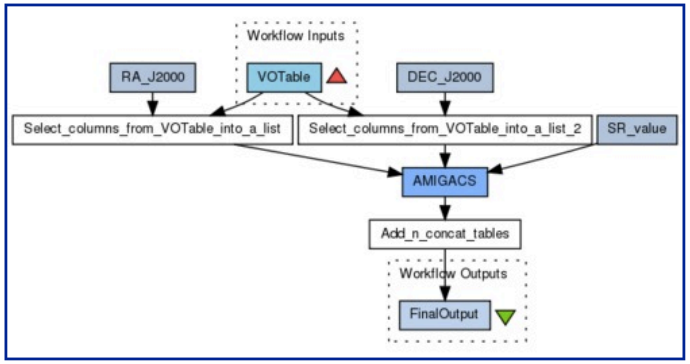
- Add a common row to a VOTable
  Added by Jose Enrique Ruiz [2013-04-22 15:15:42 +0000]
- Adding a column to a VOTable using a mathematical expression
  Added by Jose Enrique Ruiz [2013-04-12 19:08:32 +0000]
- Cocatenates several VOTables into one
  Added by Jose Enrique Ruiz [2013-04-12 20:04:14 +0000]
- Concatenates two VOTables with the same numer of columns into a single one
  Added by Jose Enrique Ruiz [2013-04-12 19:22:47 +0000]
- Coordinate units conversion
  Added by Jose Enrique Ruiz [2013-04-12 18:15:12 +0000]
- Crossmatching VOTables
  Added by Jose Enrique Ruiz [2013-04-12 17:21:39 +0000]
- Discovery of Brown Dwarfs mining the 2MASS and SDSS databases
  Added by Jose Enrique Ruiz [2013-04-12 16:38:32 +0000]
- Executes Python script
  Added by Jose Enrique Ruiz [2013-04-22 14:55:27 +0000]
- Extract a column from a VOTable into a List
  Added by Jose Enrique Ruiz [2013-04-12 19:41:57 +0000]
- Joins two VOTables with the same numer of rows into a single one
  Added by Jose Enrique Ruiz [2013-04-12 19:23:00 +0000]

Update Upload Open Import

Workflow Entry: Perform Multi-ConeSearch queries to a VO Service (version 1)

Type: Taverna 2

Uploader: Jose Enrique Ruiz
Created at: Mon Apr 22 21:43:39 CEST 2013
License: Creative Commons Attribution-Share Alike 3.0 License



Snippet showing how to use AstroTaverna "VO service perspective" and other tools for performing Multi-ConeSearch queries to a VO Services. The input is a VOTable with a list of source names and coordinates to perform the multi-query. Please *note how the AMIGACS block list handling is cofigured* with right click -> Configure running -> List Handling. Add_n_concat_tables tool is used to concatenate the list of responses issued from the multiquery. The result is a VOTable that may be renderized properly in the perspective Results, coosing Value Type as VOTable.

VO compliant data from pipelines

Traditional data processing pipelines, e.g., instrumental or survey data processing pipelines, which produce higher, level data products. At present there are many variants of these and they have little or no direct connection to VO, aside from possibly producing VO-compliant data or being optionally driven from VO.

It is not clear how much VO mechanisms are needed at this level (VO compliant data and metadata, modelling provenance, etc.)

Driving Data Processing Pipelines from the VO

In this case we have a traditional data processing pipeline and the remote user or client software invokes a job to do some pipeline reprocessing, e.g., to custom reprocess an instrumental dataset to produce a new image, cube, etc. The "workflow" in this case runs at a single site, and VO is used to drive the job remotely (SSO, UWS) and manage the results (VOSpace, VO data services).

We could think on integrating the traditional data processing pipelines we already have with VO, to allow VO users to do on-the-fly reprocessing to generate data products which can be analysed with VO (custom reprocessing of observatory data for example)

Some attempts to integrate general processing applications have been made with CEA and UWS.

Distributed Data Analysis Workflows

In this case a user or a client defines and executes a distributed workflow, which invokes services on multiple remote sites via the VO infrastructure. The workflow would be entirely in VO-space, driving simpler services at the individual sites.

The AstroTaverna developments provide a graphical tool for the composition and design of workflows based on VO services and data from different archives and facilities.

Self Descriptive Web Services: S3, SimDAL, PDL, DataLink