Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Workflows
Access and Massage VO Data

José Enrique Ruiz
on behalf of the Wf4Ever Team

IVOA INTEROP SPRING MEETING 2013
HEIDELBERG, MAY16th 2013

INSTITUTO de ASTROFISICA de ANDALUCIA CSIC I·A·A

CSIC

Workflow 4Ever

SEVENTH FRAMEWORK PROGRAMME

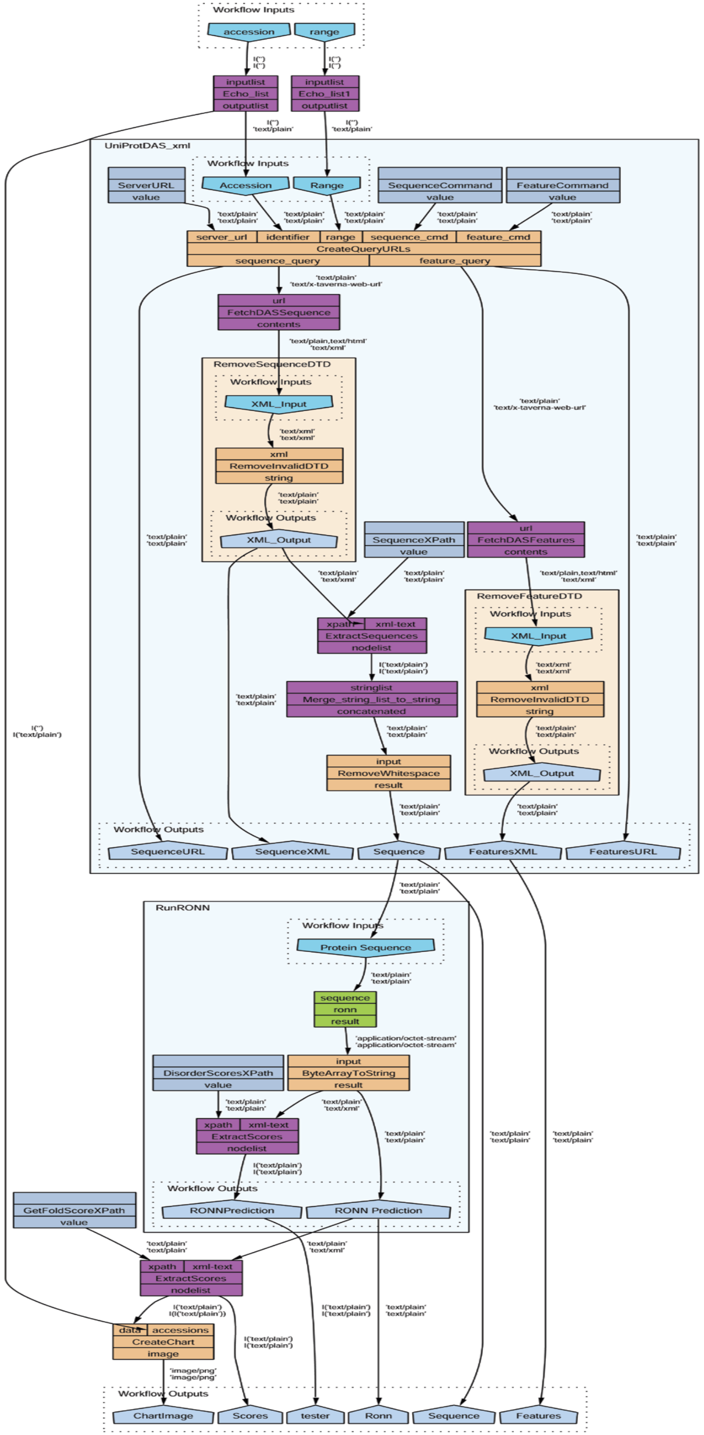
European Commission
Information Society and Media

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 (Po
5. University of Oxford and OeRC (OXF, UK)

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

Reproducible Science



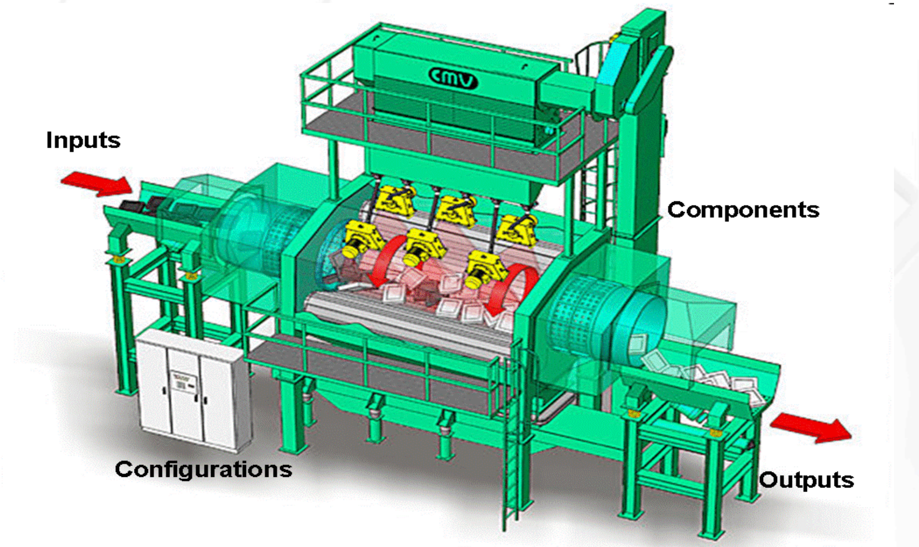
Workflows to Access and Massage VO Data

What is a Scientific Workflow?

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

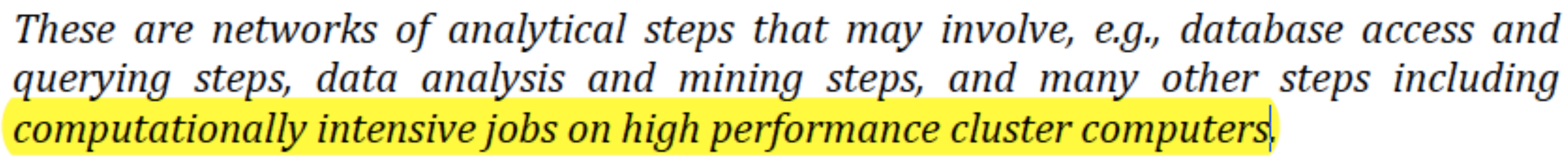
» The combination of data and processes into a configurable, modular, structured set of steps that implement semi-automated computational solutions in scientific problem-solving.

» 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

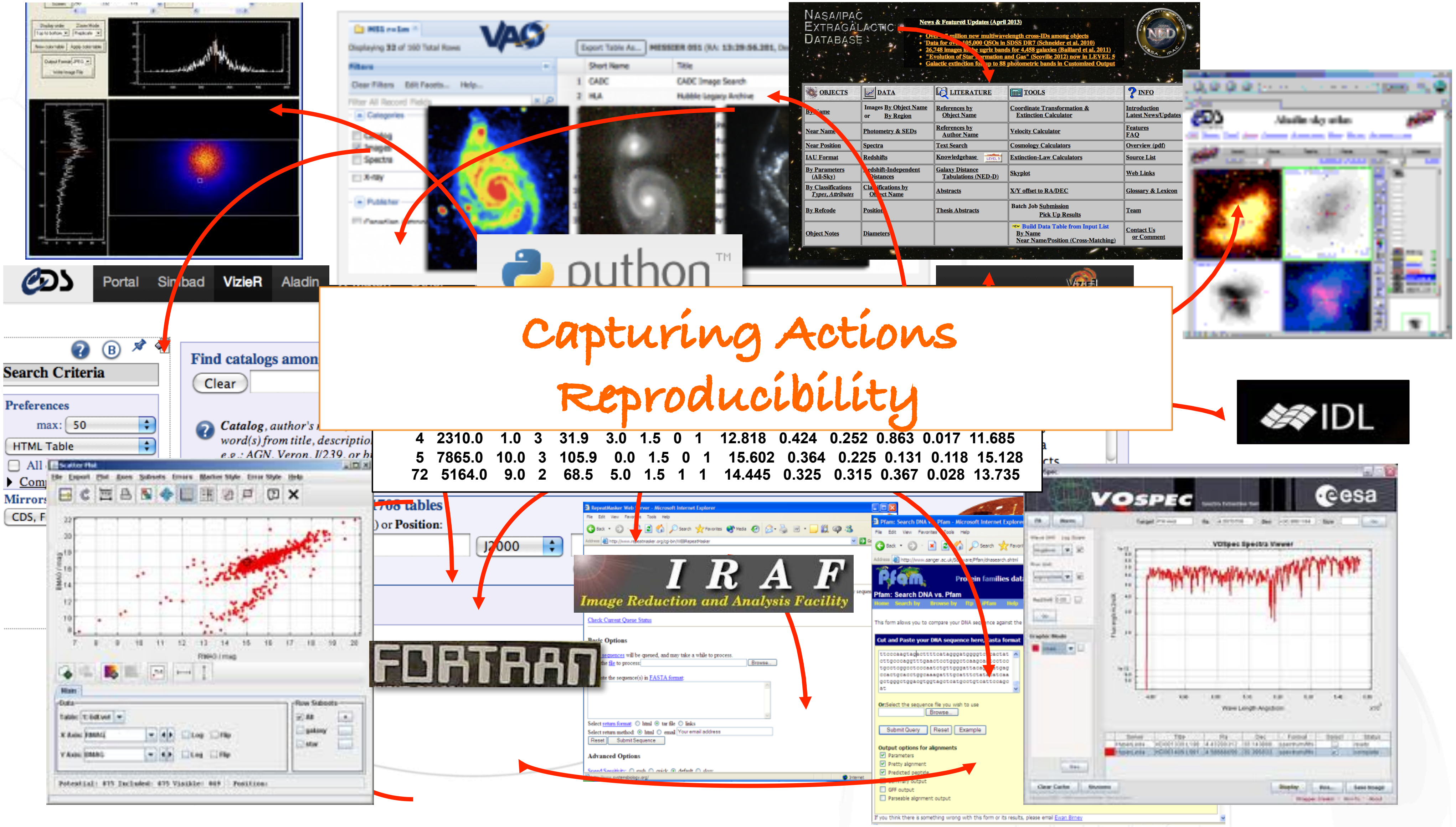
- ER-Flow
- VAMDC
- Helio-VO
- Cyber-SKA
- IceCore
- Montage
- Astro-WISE
- AstroGrid

In the VO

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

Workflows to Access and Massage VO Data

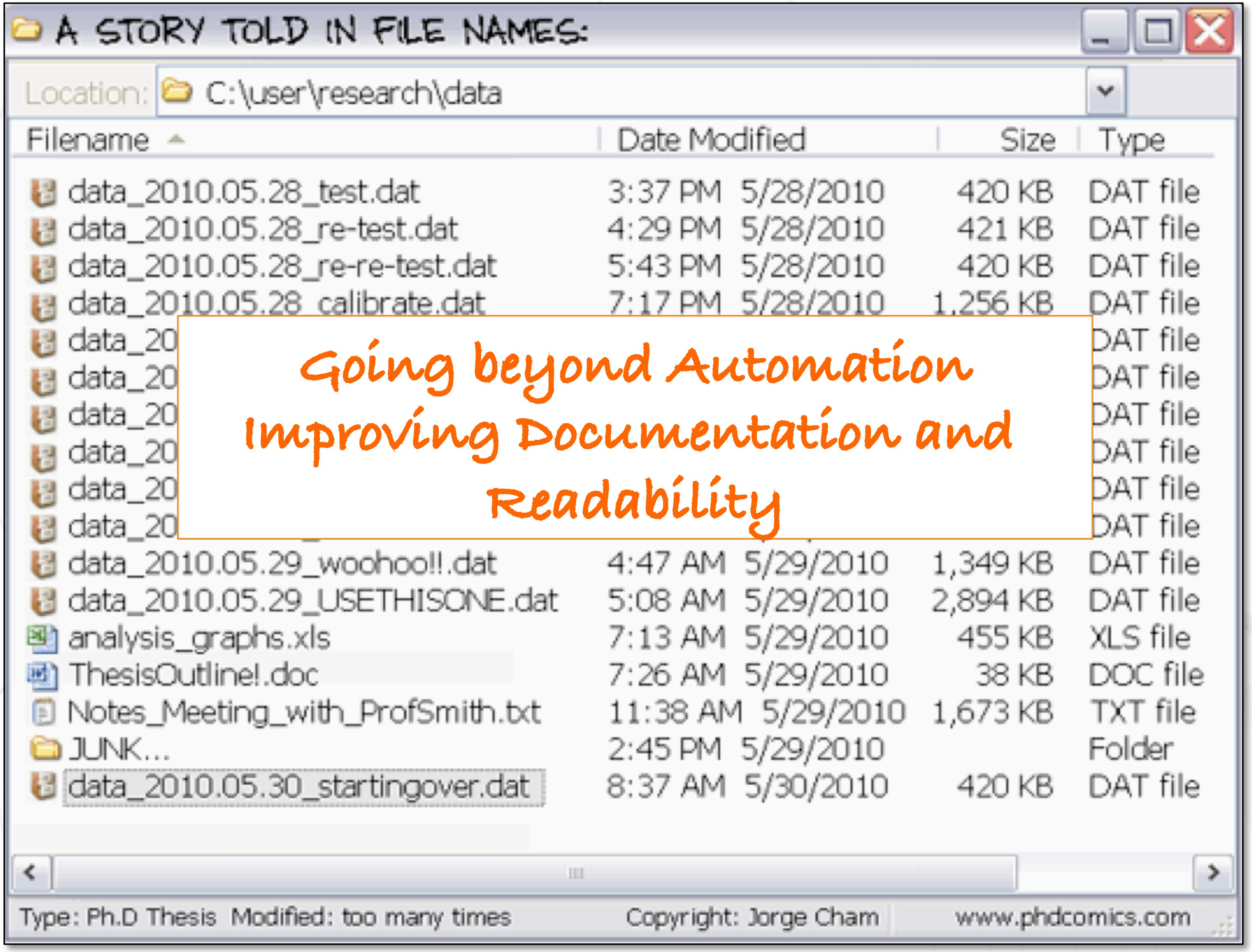
Digital Astronomy



Capturing Actions

Reproducibility

```
 4  2310.0   1.0  3   31.9   3.0  1.5  0  1   12.818  0.424  0.252  0.863  0.017  11.685
 5  7865.0  10.0  3  105.9   0.0  1.5  0  1   15.602  0.364  0.225  0.131  0.118  15.128
72  5164.0   9.0  2   68.5   5.0  1.5  1  1   14.445  0.325  0.315  0.367  0.028  13.735
```



A STORY TOLD IN FILE NAMES:

Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_20			DAT file
data_20			DAT file
data_20			DAT file
data_20			DAT file
data_20			DAT file
data_20			DAT file
data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

Type: Ph.D Thesis Modified: too many times Copyright: Jorge Cham www.phdcomics.com

Going beyond Automation
Improving Documentation and Readability

AstroTaverna Workflows

Retrieving and Manipulating VO Data **+ Catalog ML Pages**

- ConeSearch
- SIA
- SSA
- TAP coming soon…

Web Services

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

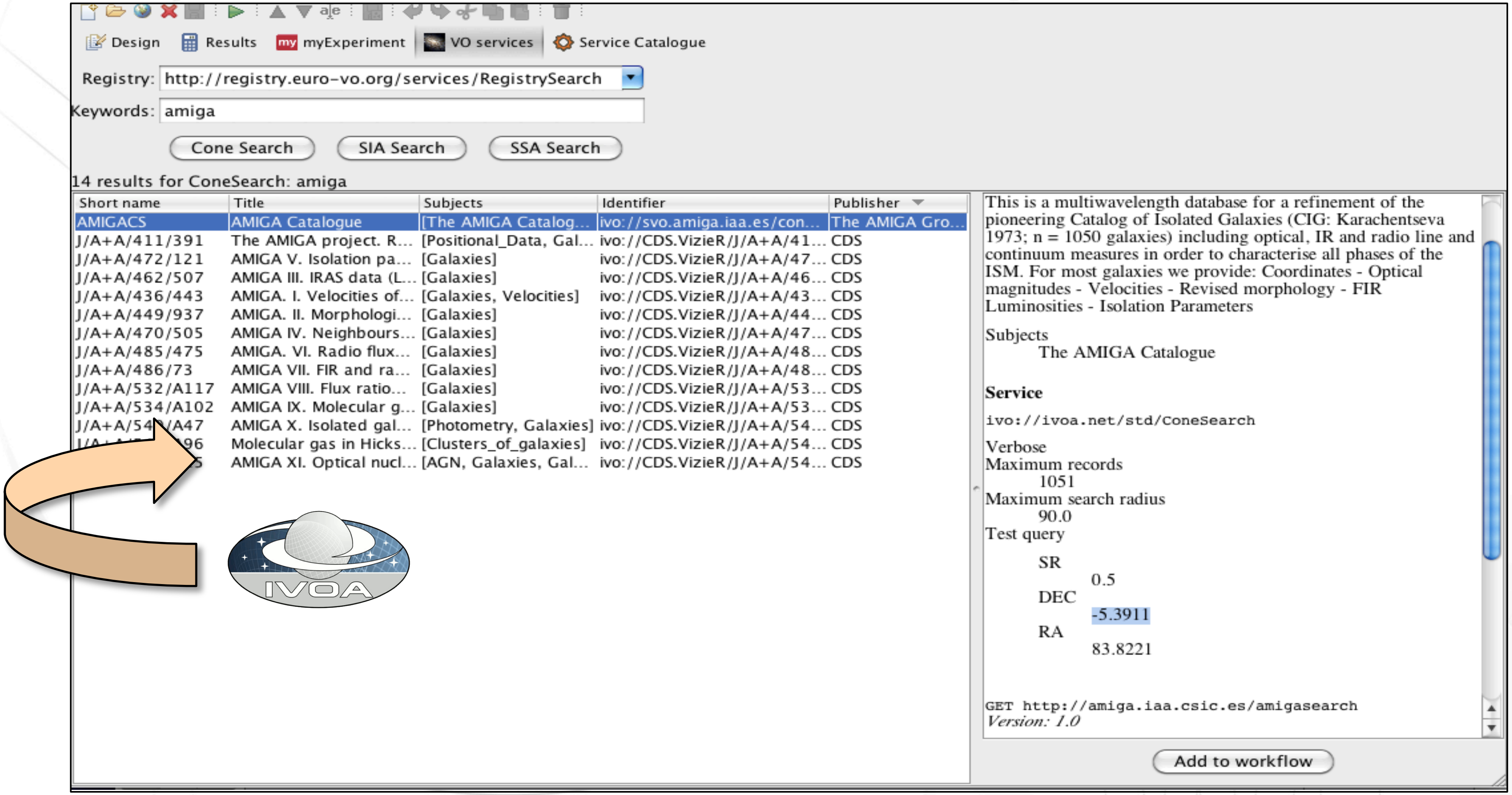
ccess to JDBC databases

- Crossmatching, Fi esolving, Coordinates and reference system transfor ssage.. (**STILTS**)
- Overplotting sc 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

+ Advanced Analysis using Scripts

No interactive actions and decisions based on visual inspection

VOData Access: VO Services Discovery



Registry: http://registry.euro-vo.org/services/RegistrySearch

Keywords: amiga

14 results for ConeSearch: amiga

Short name	Title	Subjects	Identifier	Publisher
AMIGACS	AMIGA Catalogue	[The AMIGA Catalog…	ivo://svo.amiga.iaa.es/con…	The AMIGA Gro…
J/A+A/411/391	The AMIGA project. R…	[Positional_Data, Gal…	ivo://CDS.VizieR/J/A+A/41…	CDS
J/A+A/472/121	AMIGA V. Isolation pa…	[Galaxies]	ivo://CDS.VizieR/J/A+A/47…	CDS
J/A+A/462/507	AMIGA III. IRAS data (L…	[Galaxies]	ivo://CDS.VizieR/J/A+A/46…	CDS
J/A+A/436/443	AMIGA. I. Velocities of…	[Galaxies, Velocities]	ivo://CDS.VizieR/J/A+A/43…	CDS
J/A+A/449/937	AMIGA. II. Morphologi…	[Galaxies]	ivo://CDS.VizieR/J/A+A/44…	CDS
J/A+A/470/505	AMIGA IV. Neighbours…	[Galaxies]	ivo://CDS.VizieR/J/A+A/47…	CDS
J/A+A/485/475	AMIGA. VI. Radio flux…	[Galaxies]	ivo://CDS.VizieR/J/A+A/48…	CDS
J/A+A/486/73	AMIGA VII. FIR and ra…	[Galaxies]	ivo://CDS.VizieR/J/A+A/48…	CDS
J/A+A/532/A117	AMIGA VIII. Flux ratio…	[Galaxies]	ivo://CDS.VizieR/J/A+A/53…	CDS
J/A+A/534/A102	AMIGA IX. Molecular g…	[Galaxies]	ivo://CDS.VizieR/J/A+A/53…	CDS
J/A+A/54[illegible]/A47	AMIGA X. Isolated gal…	[Photometry, Galaxies]	ivo://CDS.VizieR/J/A+A/54…	CDS
[illegible]96	Molecular gas in Hicks…	[Clusters_of_galaxies]	ivo://CDS.VizieR/J/A+A/54…	CDS
[illegible]5	AMIGA XI. Optical nucl…	[AGN, Galaxies, Gal…	ivo://CDS.VizieR/J/A+A/54…	CDS

This is a multiwavelength database for a refinement of the pioneering Catalog of Isolated Galaxies (CIG: Karachentseva 1973; n = 1050 galaxies) including optical, IR and radio line and continuum measures in order to characterise all phases of the ISM. For most galaxies we provide: Coordinates - Optical magnitudes - Velocities - Revised morphology - FIR Luminosities - Isolation Parameters

Subjects
The AMIGA Catalogue

Service

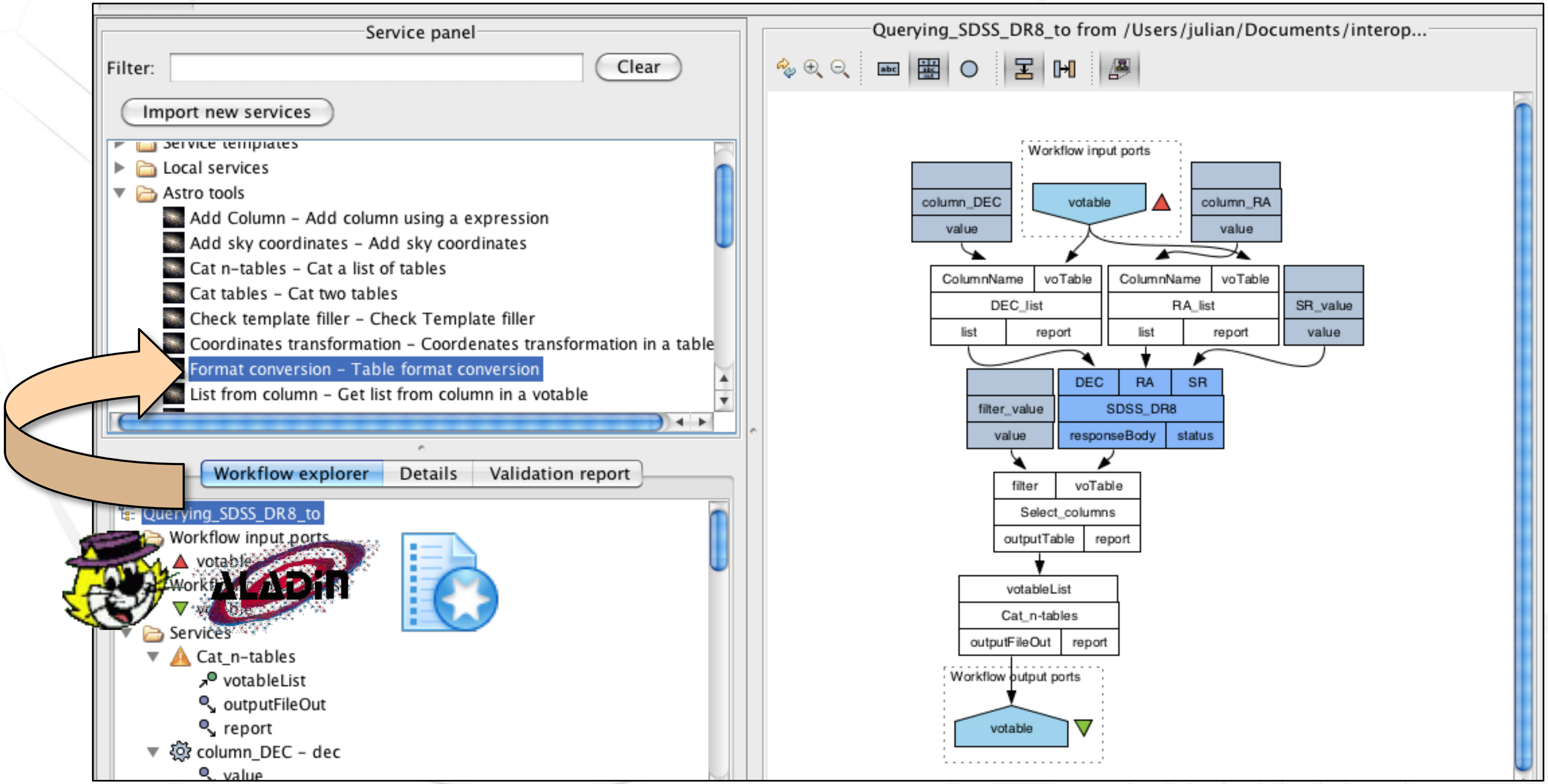
`ivo://ivoa.net/std/ConeSearch`

Verbose
Maximum records
1051
Maximum search radius
90.0
Test query
SR
0.5
DEC
-5.3911
RA
83.8221

`GET http://amiga.iaa.csic.es/amigasearch`
Version: 1.0

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

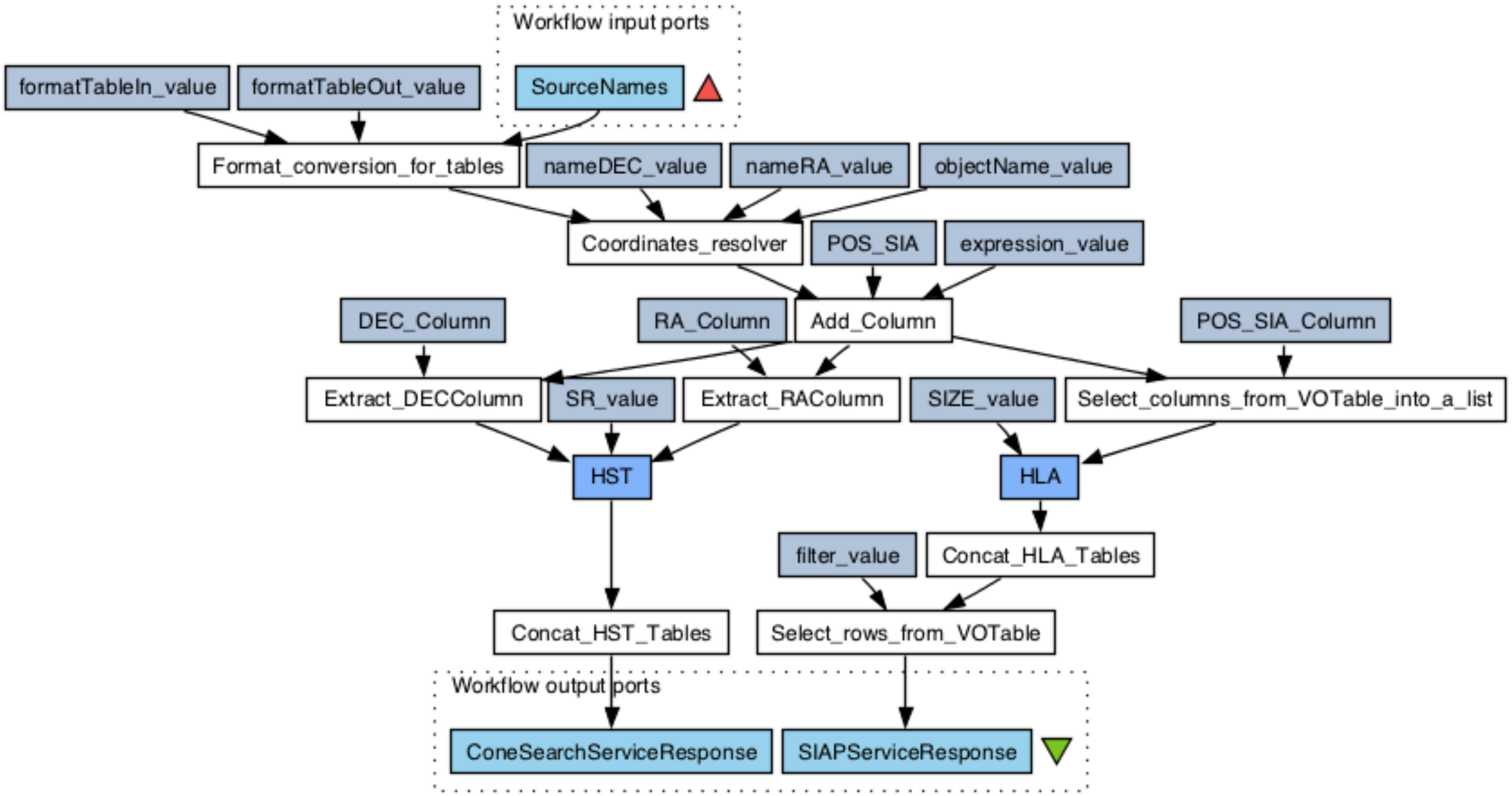
VOData Massage: VOTables, STILTS, Aladin, TerminalSim



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

Workflows to Access and Massage VO Data

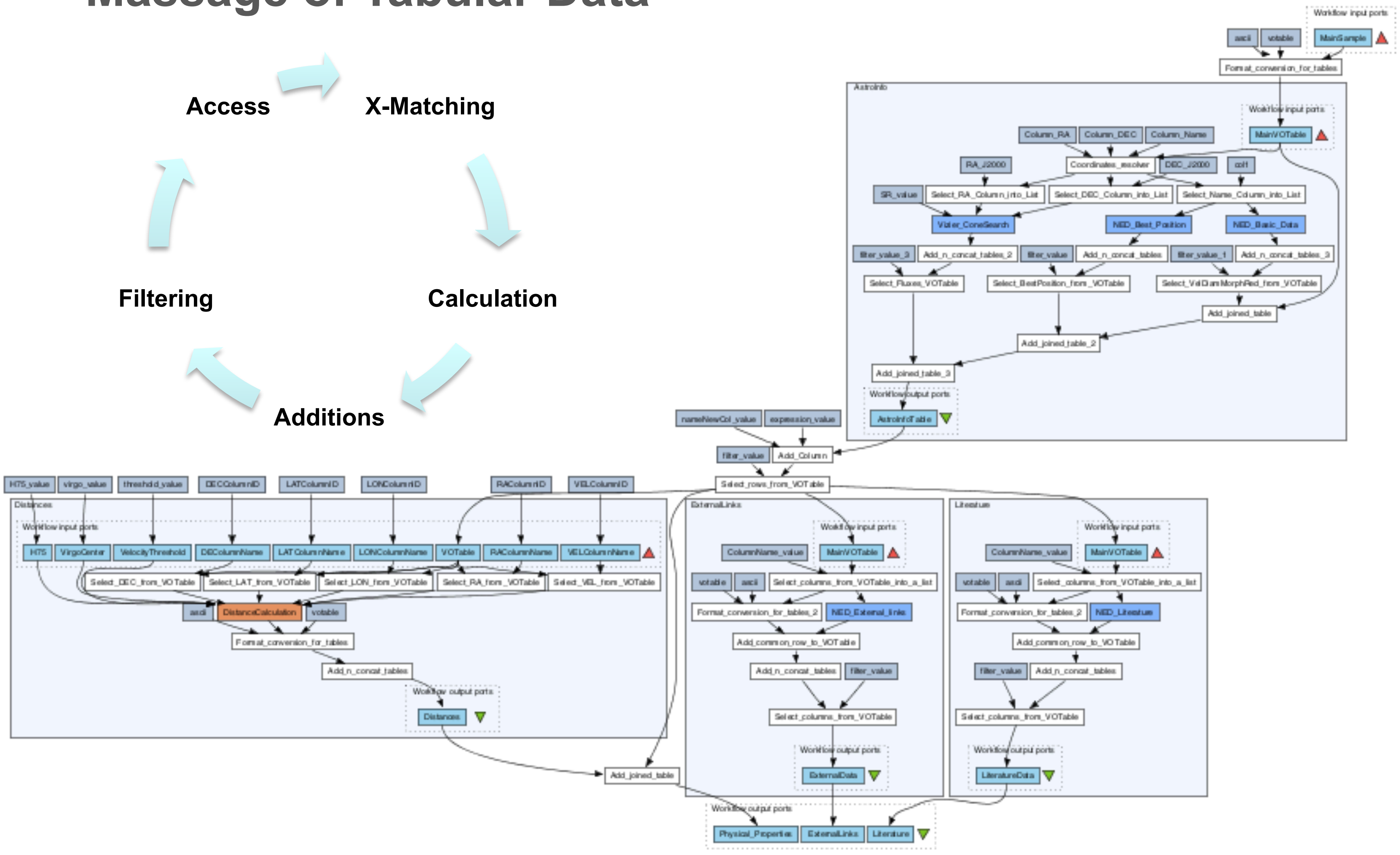
VOData Consumers



Workflows to Access and Massage VO Data

VOData Manipulation

Massage of Tabular Data



Access

X-Matching

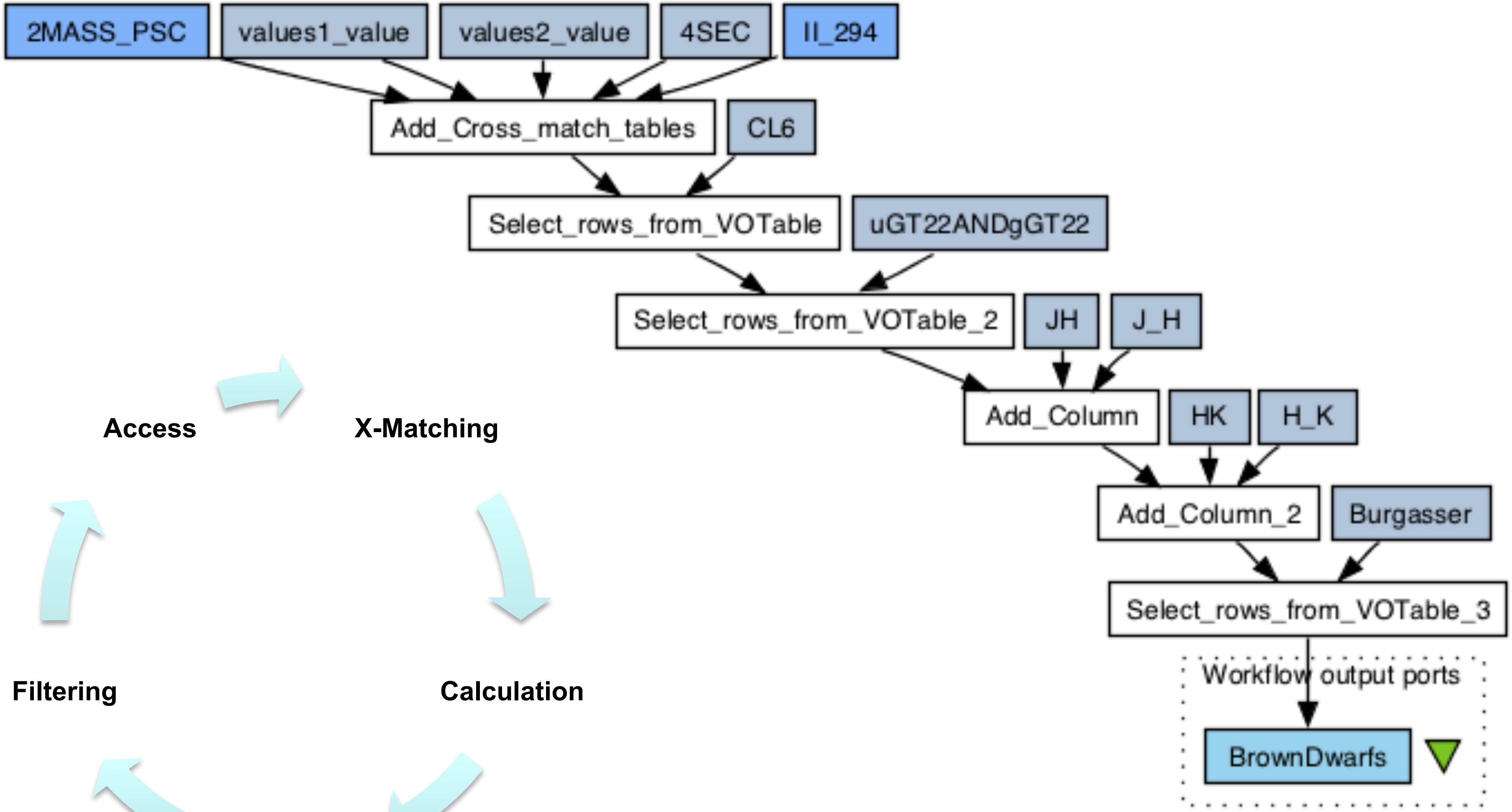
Calculation

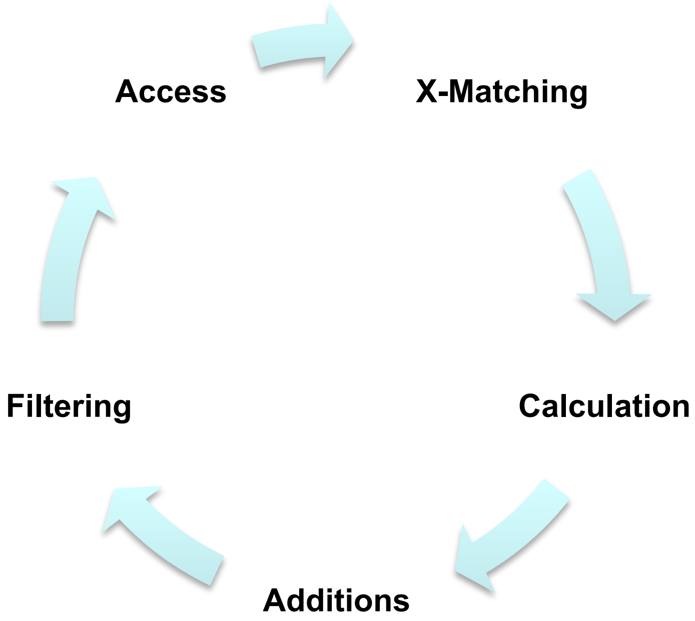
Additions

Filtering

Workflows to Access and Massage VO Data

VOData Manipulation





Access

X-Matching

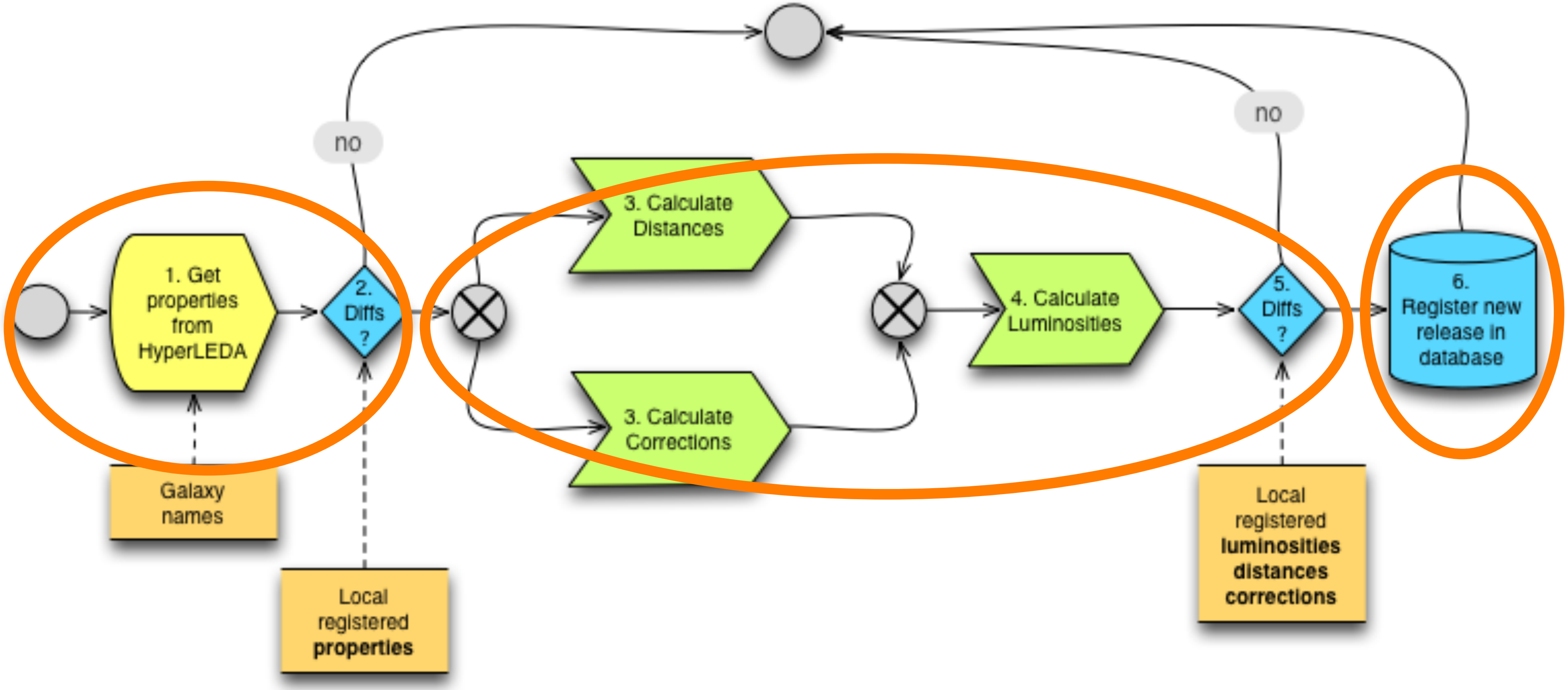
Calculation

Additions

Filtering

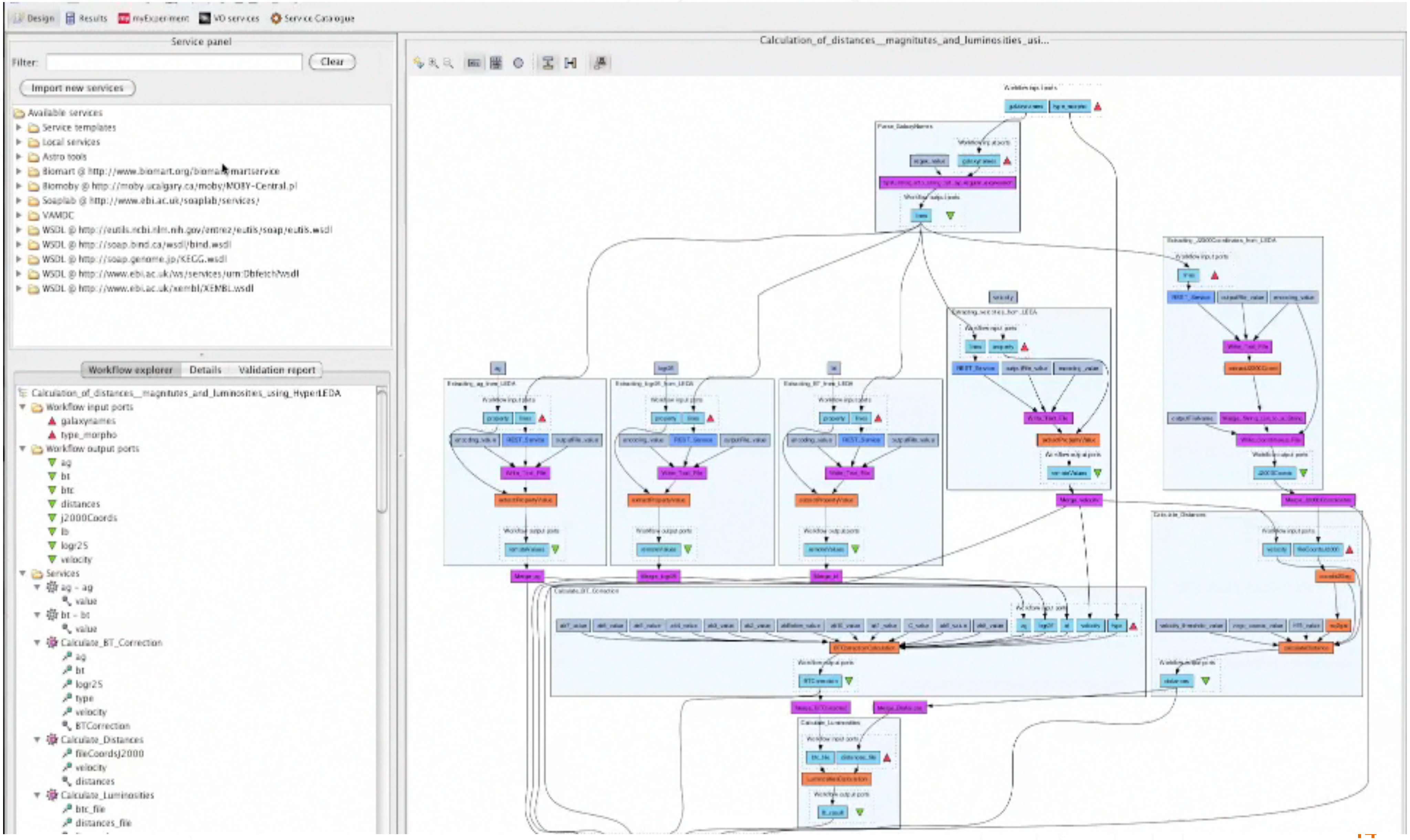
Workflows to Access and Massage VO Data

VOData Curation


no
no
1. Get properties from HyperLEDA
2. Diffs ?
3. Calculate Distances
3. Calculate Corrections
4. Calculate Luminosities
5. Diffs ?
6. Register new release in database
Galaxy names
Local registered properties
Local registered luminosities distances corrections

Workflows to Access and Massage VO Data

VOData Curation



- Available services
  - ▶ Service templates
  - ▶ Local services
  - ▶ Astro tools
  - ▶ Biomart @ http://www.biomart.org/biomart/martservice
  - ▶ Biomoby @ http://moby.ucalgary.ca/moby/MOBY-Central.pl
  - ▶ Soaplab @ http://www.ebi.ac.uk/soaplab/services/
  - ▶ VAMDC
  - ▶ WSDL @ http://eutils.ncbi.nlm.nih.gov/entrez/eutils/soap/eutils.wsdl
  - ▶ WSDL @ http://soap.bind.ca/wsdl/bind.wsdl
  - ▶ WSDL @ http://soap.genome.jp/KEGG.wsdl
  - ▶ WSDL @ http://www.ebi.ac.uk/ws/services/urn:Dbfetch?wsdl
  - ▶ WSDL @ http://www.ebi.ac.uk/xembl/XEMBL.wsdl

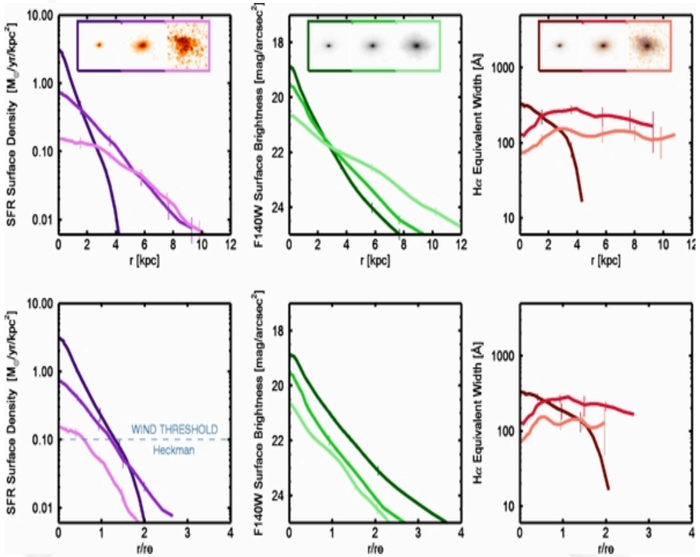
- Calculation_of_distances__magnitutes_and_luminosities_using_HyperLEDA
  - ▼ Workflow input ports
    - galaxynames
    - type_morpho
  - ▼ Workflow output ports
    - ag
    - bt
    - btc
    - distances
    - j2000Coords
    - lb
    - logr25
    - velocity
  - ▼ Services
    - ▼ ag - ag
      - value
    - ▼ bt - bt
      - value
    - ▼ Calculate_BT_Correction
      - ag
      - bt
      - logr25
      - type
      - velocity
      - BTCorrection
    - ▼ Calculate_Distances
      - fileCoordsJ2000
      - velocity
      - distances
    - ▼ Calculate_Luminosities
      - btc_file
      - distances_file

VOTable Format Interoperability

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

I R A F
Image Reduction and Analysis Facility

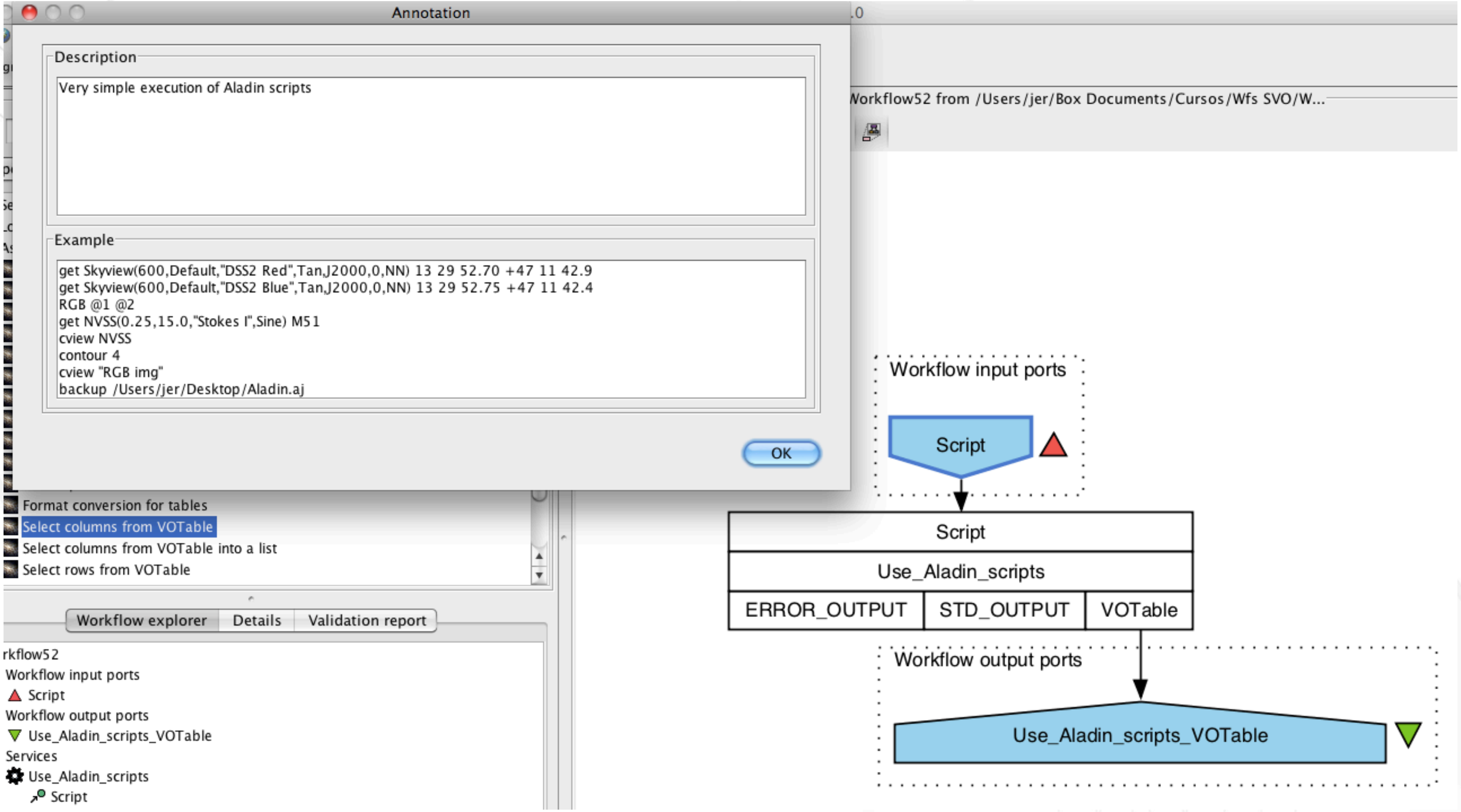
GALFIT



90 galaxies observed in 3 bands

SExtractor

Aladin Scripts and Macro executing in GUI/noGUI mode



Description: Very simple execution of Aladin scripts

Example:

```
get Skyview(600,Default,"DSS2 Red",Tan,J2000,0,NN) 13 29 52.70 +47 11 42.9
get Skyview(600,Default,"DSS2 Blue",Tan,J2000,0,NN) 13 29 52.75 +47 11 42.4
RGB @1 @2
get NVSS(0.25,15.0,"Stokes I",Sine) M51
cview NVSS
contour 4
cview "RGB img"
backup /Users/jer/Desktop/Aladin.aj
```

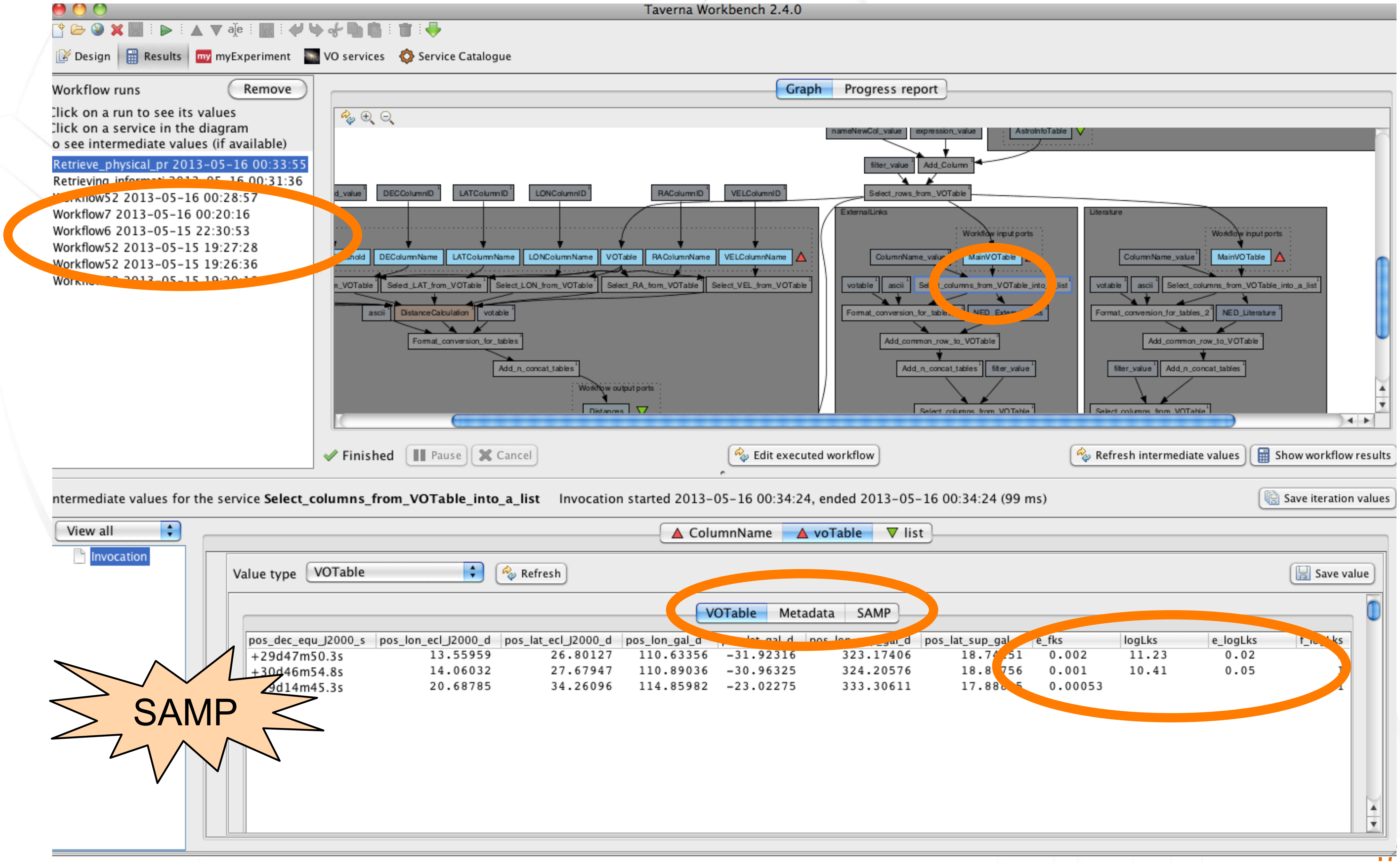
Workflow52 from /Users/jer/Box Documents/Cursos/Wfs SVO/W...

- Format conversion for tables
- Select columns from VOTable
- Select columns from VOTable into a list
- Select rows from VOTable

rkflow52
- Workflow input ports
  - ▲ Script
- Workflow output ports
  - ▼ Use_Aladin_scripts_VOTable
- Services
  - Use_Aladin_scripts
    - Script

Workflows to Access and Massage VO Data

VOData Inspection



Taverna Workbench 2.4.0

Workflow runs

Click on a run to see its values
Click on a service in the diagram
to see intermediate values (if available)

Retrieve_physical_pr 2013-05-16 00:33:55
Retrieving_informati 2013-05-16 00:31:36
Workflow52 2013-05-16 00:28:57
Workflow7 2013-05-16 00:20:16
Workflow6 2013-05-15 22:30:53
Workflow52 2013-05-15 19:27:28
Workflow52 2013-05-15 19:26:36
Workflow52 2013-05-15 19:20:1[illegible]

Intermediate values for the service **Select_columns_from_VOTable_into_a_list** Invocation started 2013-05-16 00:34:24, ended 2013-05-16 00:34:24 (99 ms)

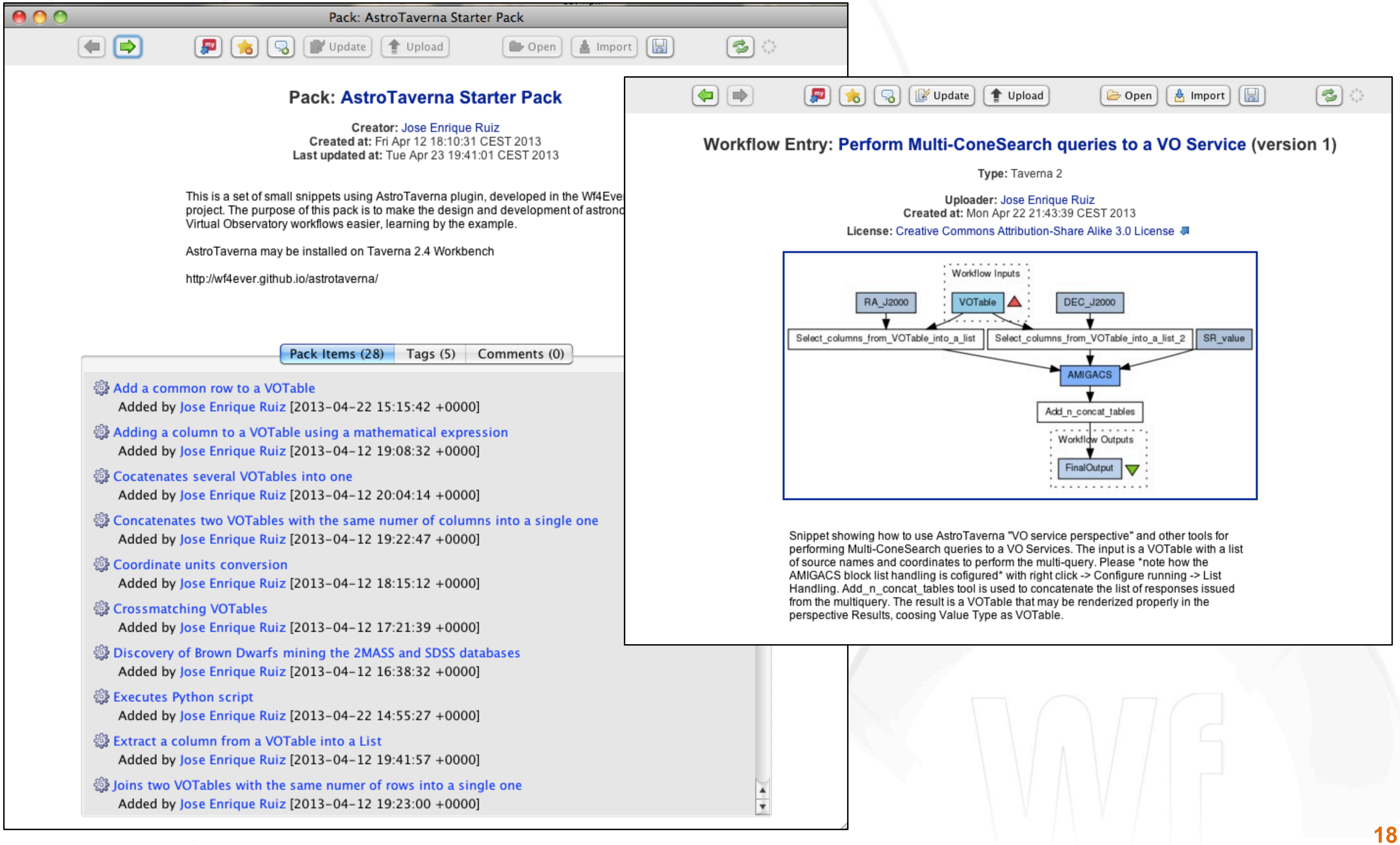
Value type VOTable

pos_dec_equ_J2000_s	pos_lon_ecl_J2000_d	pos_lat_ecl_J2000_d	pos_lon_gal_d	[illegible]lat_gal_d	pos_lon[illegible]_gal_d	pos_lat_sup_gal	e_fks	logLks	e_logLks	f_lo[illegible]ks
+29d47m50.3s	13.55959	26.80127	110.63356	-31.92316	323.17406	18.7[illegible]51	0.002	11.23	0.02	
+30d46m54.8s	14.06032	27.67947	110.89036	-30.96325	324.20576	18.8[illegible]756	0.001	10.41	0.05	[illegible]
[illegible]9d14m45.3s	20.68785	34.26096	114.85982	-23.02275	333.30611	17.88[illegible]5	0.00053			[illegible]

SAMP

Workflows to Access and Massage VO Data

Learning by the example



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/

- 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]

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.

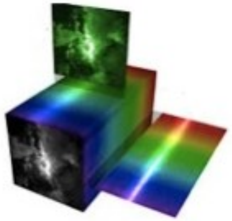
The AstroTaverna developments provide a graphical tool for the composition and design of workflows based on VO services and data from different archives and facilities.

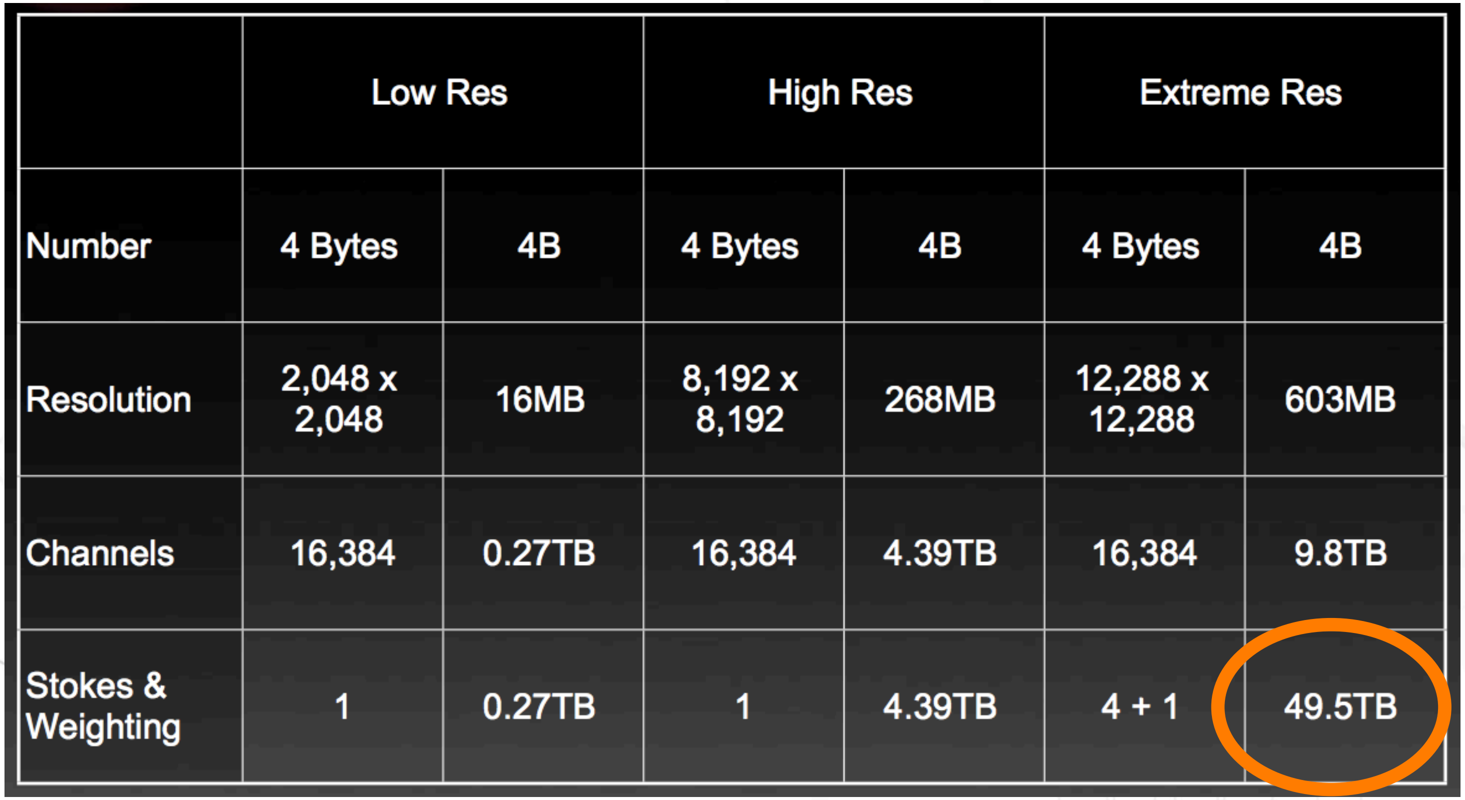
Self Descriptive Web Services: S3, SimDAL, PDL, DataLink

Much wider FoV and spectral coverage

» Large volumes for an observed datacube

» Subproducts are **Virtual Data** generated on-the-fly





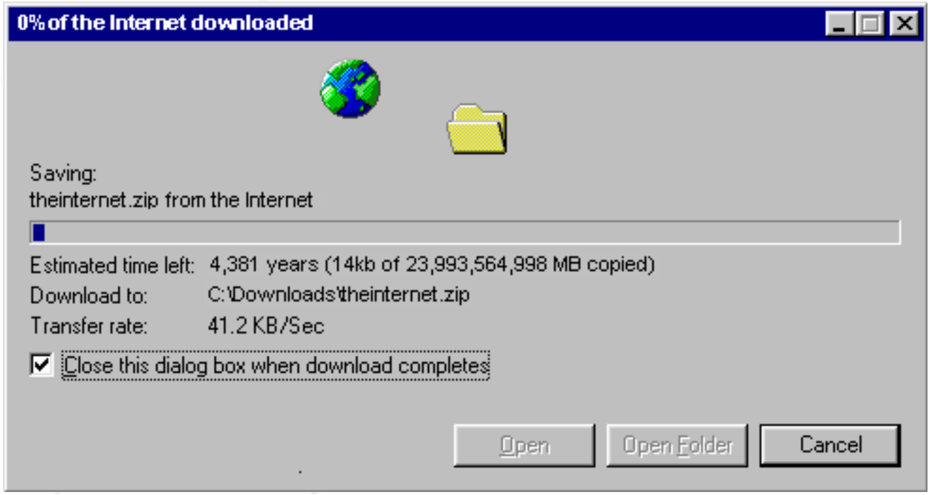
	Low Res		High Res		Extreme Res	
Number	4 Bytes	4B	4 Bytes	4B	4 Bytes	4B
Resolution	2,048 x 2,048	16MB	8,192 x 8,192	268MB	12,288 x 12,288	603MB
Channels	16,384	0.27TB	16,384	4.39TB	16,384	9.8TB
Stokes & Weighting	1	0.27TB	1	4.39TB	4 + 1	49.5TB

ASKAP Cubes
Prof. Kevin Vinsen

The next generation of archives

Automated surveys

- » Huge amounts of tabular data
- » Services for KDD

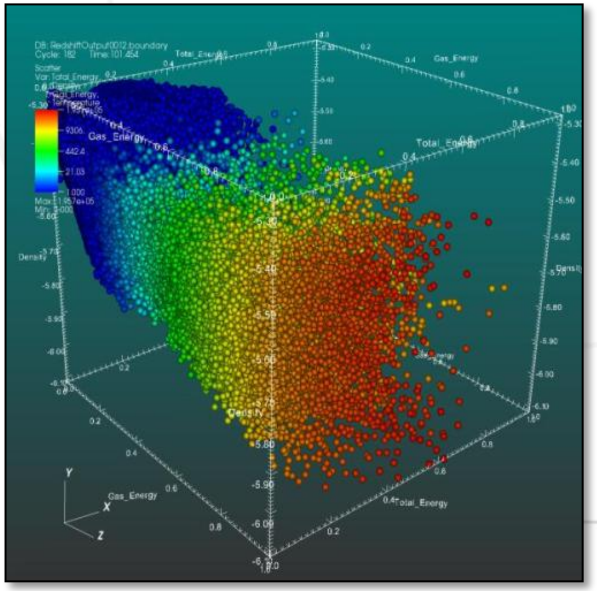


0% of the Internet downloaded

Saving:
theinternet.zip from the Internet

Estimated time left: 4,381 years (14kb of 23,993,564,998 MB copied)
Download to: C:\Downloads\theinternet.zip
Transfer rate: 41.2 KB/Sec

☑ Close this dialog box when download completes



Extraction of scientifically relevant information from a multidimensional parameter space

- » Exploration services
- » Anomaly detection
- » Cross-matching data
- » Dimensionality reduction

The next generation of archives

- A cloud of Web Services

Archives should evolve from data providers into

- Virtual data providers
- Software tasks providers

- Archives speaking Web Services

Astronomy of multi archives/facilities/wavelength

Interconnected and interoperable archives

- Data -> Virtual Observatory
- Software Tasks

Preservation

Process should benefit of the same privileges acquired by data

Preserving the method ensures replication of final results at any moment