# Workflows in the VO

## Grid and Web Services Session

Jose Enrique Ruiz
IAA-CSIC

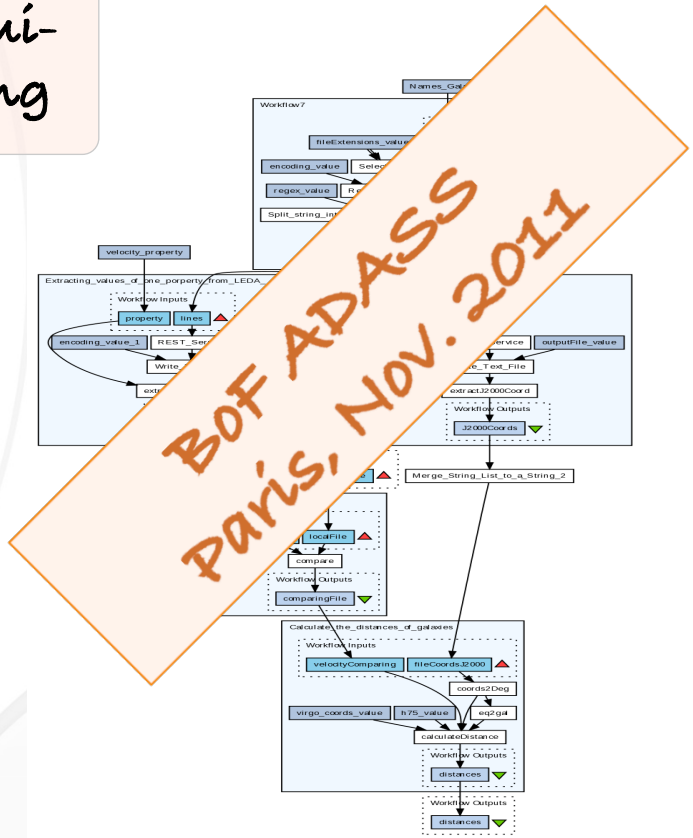
October 19th 2011
2011 IVOA Fall Interop Meeting - Pune

# What are workflows ?

Combination of data and processes into a configurable and structured set of steps that implement semi-automated computational solutions in problem solving

## Types of workflows in Astronomy

- Personal script-based recipes
  Python, IDL, Software..
- Multi-archive VO recipes
- Internal group developments
  GRID, Clusters..
- Processing pipelines
  Provide Data, Computing Infrastructure, Tools..




Scientifically exploitable results vs. scientific insight
Easily accessible and reproducible (Shared)

Wfs on steroids !

## Taverna

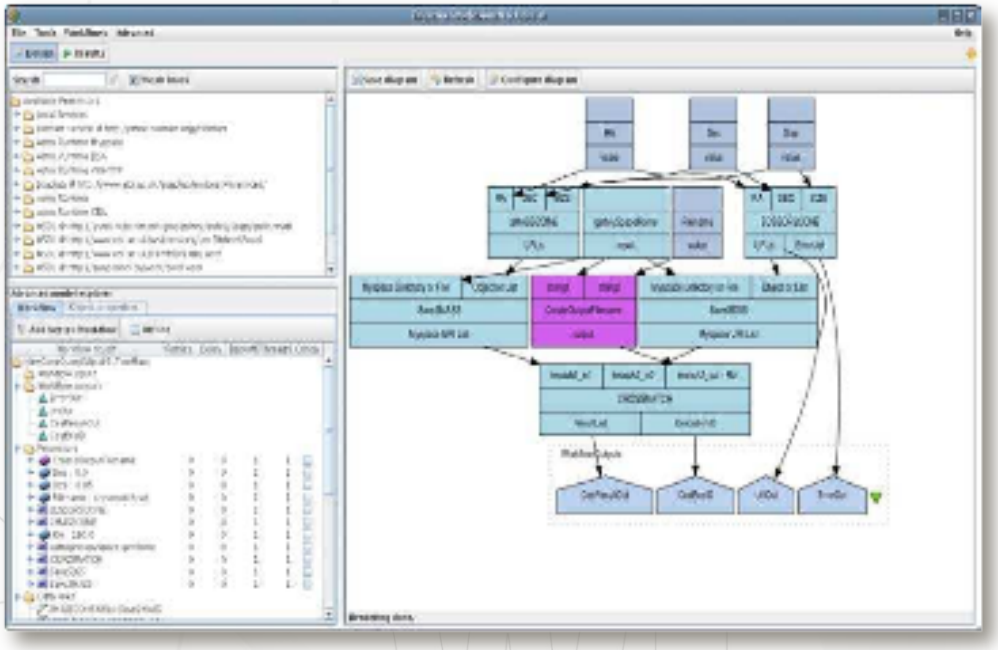
- Strongly typed bioinformatics
- Taverna Engine
- Taverna Server
- Taverna Workbench

## Kepler

- Generic Science
- Workflow System

## Triana

- Local execution
- Clusters RMI
- GRID
- Web Services



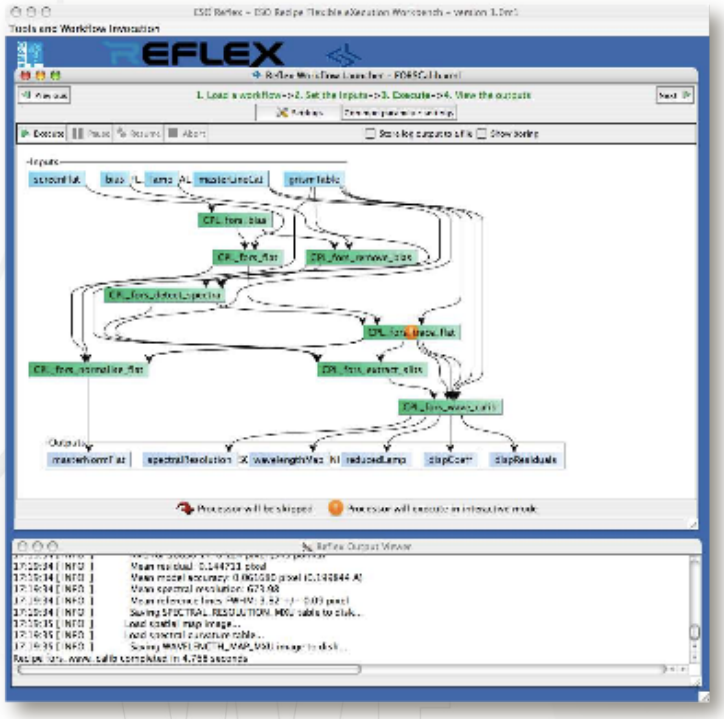
# ESO Reflex

Finland's in-kind contribution to ESO

- Prototype/feasibility study
- Initially based on Taverna 1

Current implementation based on Kepler

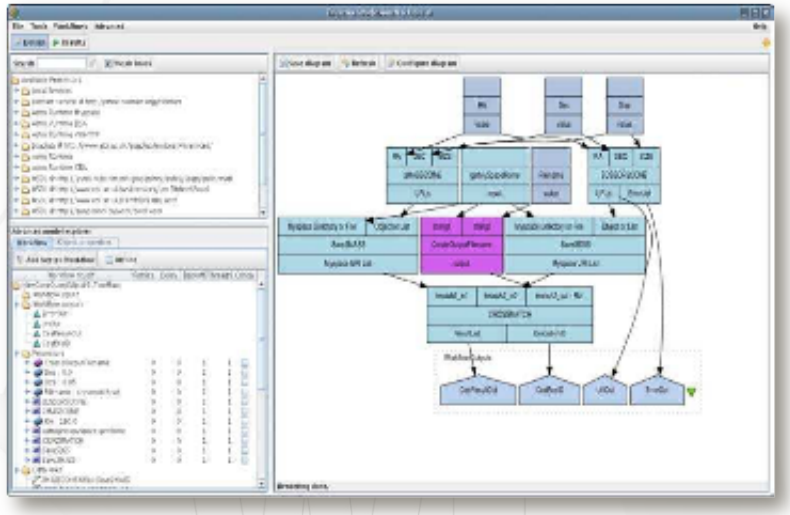
- Main intent: replace CLI (pipeline orchestrator)



# AstroTaverna

AstroGrid Development
Prototype, marrying of VO Desktop & Taverna 1
Library of Taverna functions to access VO Desktop's API

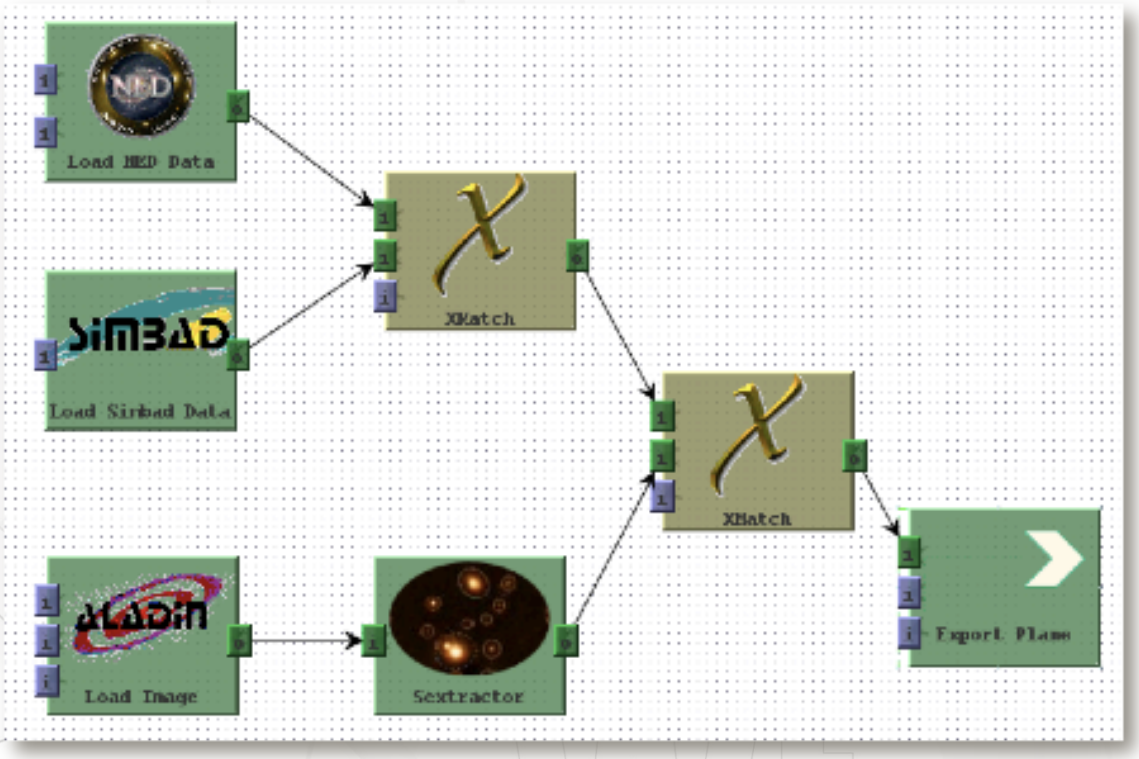
Status
Wrapper libraries only for Taverna 1



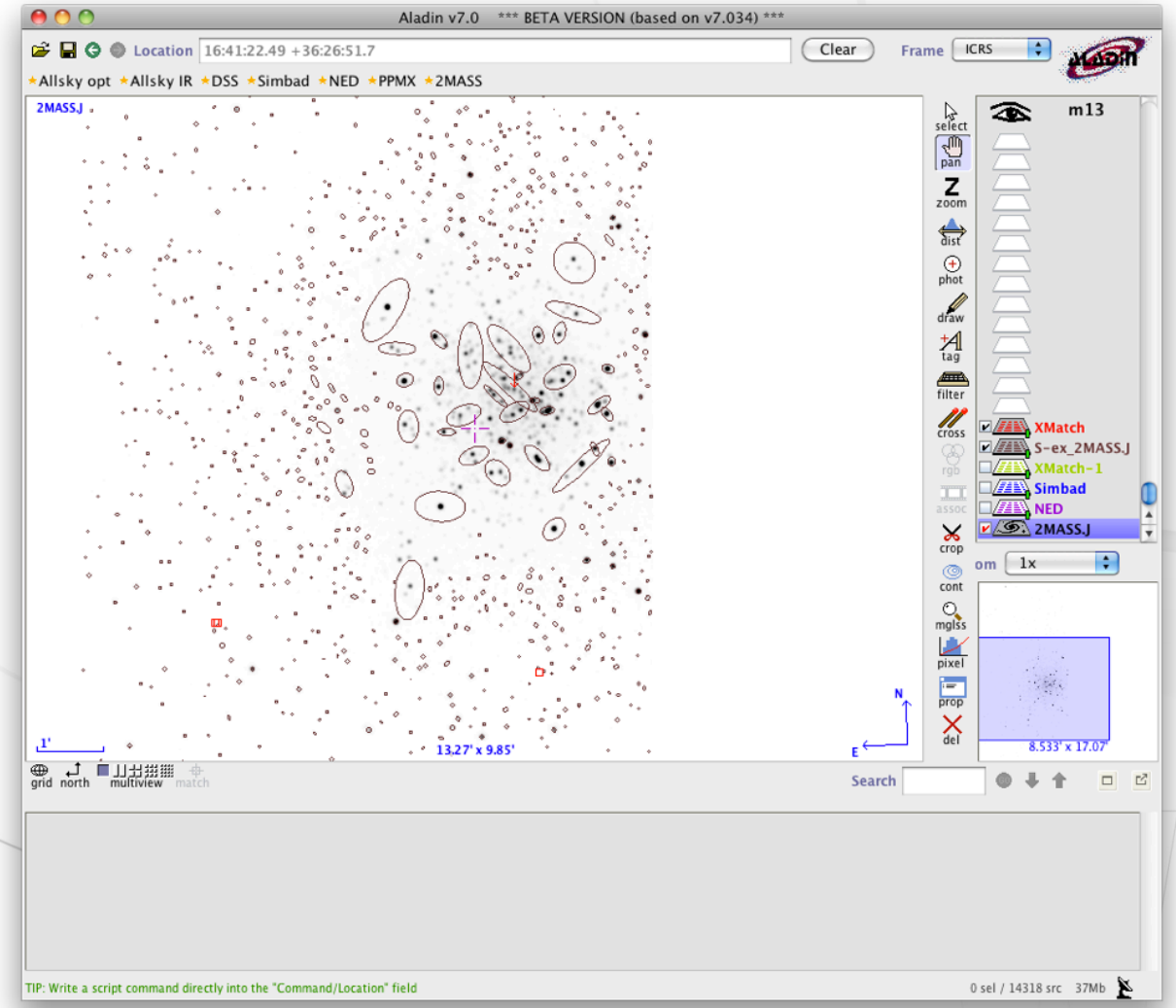
# Aladin JLOW Plugin

Aladin plugin API permits graphical replacement of Aladin tools

```
#AJS
get NED m13 14'
get Simbad m13 14'
sync
xmatch NED Simbad 4
sync
get Aladin(2MASS,J) m13
sync
get SExtractor(2MASS.J)
sync
xmatch "XMatch*1" "SExtractor*"
sync
export -votable "XMatch*2"
```



# Aladin JLOW Plugin



# Wf4Ever

2011 – 2013 EU funded FP7 STREP Project

Preservation of Workflows

- Preservation of workflows and associated material
- Archival, classification, indexing in semantic repositories
- Provide advanced access and recommendation abilities
- Collaborative working platform
- Sharing, re-use, re-purpose
- Digital Libraries

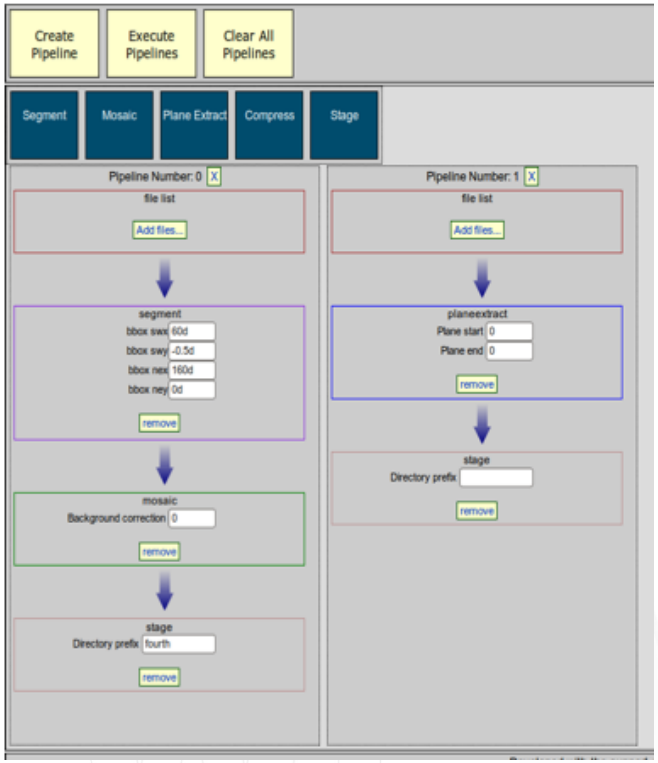
Jose Enrique Ruiz
Tuesday DCP Session

# Cyber-SKA

Provide infrastructure that will be required to address the needs of future radio telescopes such as the Square Kilometre Array

Web based workflow builder

- Image segmentation
- Image mosaicking (Montage)
- Spatial reprojection
- Plane extraction from data cubes



# Related Initiatives

## Montage

- FITS Image Mosaicking
- Toolkit for Desktops, Clusters and Grids

## Astro-WISE

- Distributed data storage and computing infrastructure
- Track process provenance of final data products
- Calibration and analysis of images

## Helio-VO

- Solar physics Virtual Observatory
- Enable workflow execution via Taverna Server

## Workflows VO France

- Provide use cases mainly oriented VO
- AÏDA Workflow System implements FITS validation with CharDM

# The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~100 TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as virtual data generated on-the-fly

Automated surveys

- Huge amount of tabular data
- Services for Knowledge Discovery in Databases

# The upcoming context

We are moving into a world where

- computing and storage are cheap
- data movement is death

Archives should evolve from data providers into virtual data and services providers, where web services may help to solve bandwidth issues.

Archives speaking web services

- Smaller virtual data subproducts
- Distributed, multi-archive, multi-wavelength astronomy

## Web services based workflows as a disruptive working methodology

- Reproducible
- Repeatable results
- Encourage best practices
- Modular nature allows
  - Re-use
  - Re-purpose
- Expose
  - Provenance
  - Scientific method
- Formative
- Foster collaborative work

myexperiment

## Distributed data analysis in the VO

- Panchromatic, multi-archive, multi-facility
- Executes in the VO Infrastructure
- Orchestration of simple services

## Present processing pipelines

- Produce exploitable data
- Provenance modelling
- VO compliant data

**Workflows VO Characterization**

- Inputs
- Outputs
- Processes
- Descriptions
- Metadata
- Etc..

## Data processing from the VO

- Provide custom re-processing to VO users
- Virtual data generation through UWS in VOSpace

# Related activities in the VO

## IVOA Working Groups

- **Data Modeling**
  Characterization, Provenance..
- **Semantics**
  Ontologies, Vocabularies, Annotat
- **Data Access Layer**
  TAP, self-descriptive Protoco
- **Grid and Web Servi**
  UWS, VOSpace, SSO
- **Applications**
  SAMP
- **IG . KDD**
  Knowledge Discovery . Mining
- **IG . Data Curat d Preservation**
  Persistent Identif ration of VO Resources..
  Wf4Ever Project


Carlo Maria Zwölf
Thursday GWS Session

Fabio Pasian
Liason with Open Grid Forum

### IVOA Note

Scientific Workflows in the VO

**workflow@ivoa.net**