Implementing Astronomical Image Analysis Pipelines using VO standards

François Bonnarel, André Schaaff, Jean-Julien Claudon, Cyril Pestel,
CDS, Observatoire de Strasbourg

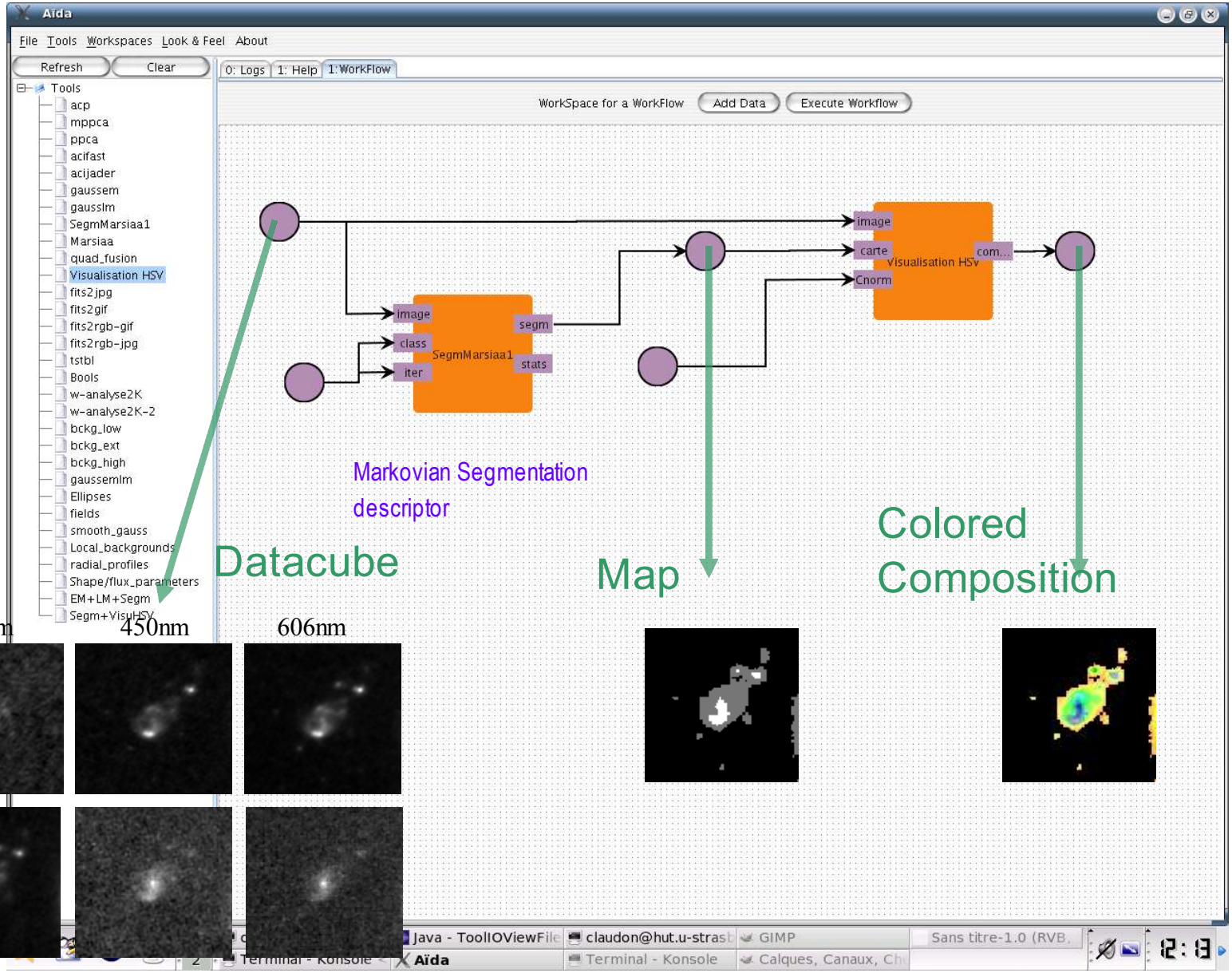
Mireille Louys **LSIIT, Université Louis Pasteur, Strasbourg,**

Goals and requirements

- Describe and publish image processing software suites in order to keep and propagate expertise
- Distribute specific image analysis tools to astronomers in the VO framework
- Describe not only tool boxes but also templates of specific analysis procedures corresponding to some particular problem
- Allow users to reproduce analysis results using published data and published procedures

The AIIDA prototype

- AÏDA (**A**stronomical **I**mage processing **D**istribution **A**rchitecture) allows to encapsulate image processing programs developped in any language such as C, C++, FORTRAN, MATLAB, ….
  - A testbed, funded by the MDA project (Massive Data in Astronomy - French Research Ministry)
- Features
  - It allows to sketch out a chain of processing steps as a graph (JLOW library)
  - The workflow engine interprets the language and orchestrates the execution of the workflow
  - The server part executes the WF via CGI and WebServices interfaces



AIIDA use case

Lessons learned

- A first experiment for tools descriptions via metadata descriptors and chaining
- A valid approach for the gathering and dissemination of image processing tools
  - As an internal collaborative tool for interaction between astronomers and signal processing people and for collaborators in astronomy
  - Need for more metadata:

    classes of tools, algorithm description (including relevance domain) and parameters, image metadata …
- Uses standards formats : FITS for images and data cubes, VOTable for tabular data

What is to be described

- Describe each step → Tools and data
  - The scientific purpose of each tool/program
  - The input and output parameters
- Describe the content of the data consumed by a processing tool
- Describe the execution → Workflows
  - The sequence of steps as a graph
  - The data flow within the graph
  - How the steps are distributed for the execution (local programs, cluster , grid)
  - The execution status of each step (execution log)

Tools and Data

- A scientific tool description is needed to propagate knowledge
  - VOApplication Model , (Registry WG), currently based on:
    - Resource Metadata structure
    - CEA Application Model by Astrogrid for the parameters description

    http://ivoa.net/twiki/bin/view/IVOA/RegDMApplications#Application_model

  - More elaborate descriptions for parameters:
    Hierarchical and dynamical description of parameters for numerical simulation codes (OV France Workflow effort)

- Observational data :
  - Describes axis types , coordinates, coverage field, and resolution
  - Allows for validation of data inputs before launching the execution

  - Use VO Data Models: Characterisation and Spectrum
    http://ivoa.net/internal/IVOA/IvoaDataModel/CharacterisationDraft-06May15.pdf
    http://ivoa.net/internal/IVOA/IvoaDataModel/spec97d.pdf

    Just add the characterisation of the input and output data to each processing block

Workflows description

- Large Workflow effort conducted by the Astrogrid project
  - Provides a workflow scripting language (Groovy), a workflow engine and an interface
  - Fully integrated within the Astrogrid Workflow System with
    - Interfaces to VO applications via the Common Execution Architecture
    - Distributed storage (MySpace)

http://www.ivoa.net/Documents/Notes/AstrogridWorkflow/AstrogridWorkflow-20060227.pdf
http://www.ivoa.net/Documents/Notes/CEA/CEADesignIVOANote-20050513.html

  - Question: How can I navigate from my specialized workspace to the Astrogrid workspace and vice versa?

- Distributed Computation (INAF, ESAC, Grid Community)
  - Local clusters
  - Submission to Grid Services

Conclusion

- Workflows reference implementations will help to propagate data analysis experience
- They should support VO standards for reproducing procedures
- WF descriptions can now benefit from existing Data Models
  - Helpful for the users to define the steps
  - Useful for consistency checking before job submissions
- A wish for an homogeneous WF description for the VO