

Implementing astronomical image analysis pipelines using VO standards

Mireille Y. Louys¹, François Bonnarel², André Schaaff²,
Jean-Julien Claudon² and C. Pestel²

¹LSIT, Université Louis Pasteur, 4 rue Blaise Pascal, F-67070 Strasbourg, France
email: louys@astro.u-strasbg.fr

²CDS, Observatoire de Strasbourg, 11 rue de l'Université, F-67000 Strasbourg, France

We consider here image analysis pipelines and examine how data and processes could be described in the context of the VO. The tasks chain is considered as a workflow, not only in terms of computing and resource allocation as in the Grid community, but in terms of data analysis know-how. Such pipelines may be published as coarse grain tool boxes and provide reference examples to the user.

We have designed a prototype named AIIDA (Astronomical Image processing Distribution Architecture) which allows encapsulating image processing programs developed in any language such as C, C++, FORTRAN, and MATLAB. The AIIDA client allows to sketch out a chain of processing steps using a graphical tool (JLOW library) encodes it into an XML description and passes it to a workflow engine which in turn interprets the language and orchestrates the execution of the workflow. The server part executes the workflow via CGI and Web Services interfaces. From this simple project has been an interesting collaborative tool between astronomers and signal processing developers inside and outside our laboratory. It helped as a test bed to understand how data analysis workflows could be represented and documented.

What is to be described, lays into two categories. On one hand: *tools and data*, i.e, the scientific purpose of each tool or program, with input and output parameters, and the physical content of the data file consumed by each processing tool. On the other hand, *the workflow execution*, i.e, the sequence of steps as a graph or a list, the data flow within the graph, the allocation of computing resources and the execution status of each step.

There are emerging VO standards matching these requirements: Processing blocks can be described, using the VOApplication Model, with parameters described using the Common Execution Architecture (CEA) data model or via a hierarchical structure as proposed for numerical simulation codes. Many parameters in data analysis workflow are related to observations files whose physical content could be checked before launching a complex workflow. The Spectrum data model for 1D spectra as well the Characterisation data model for the higher dimensions allow to describe for input files the physical axes and the data properties such as coverage or resolution, and check for the compliance with the signal expected in the processing block.

The requirements for workflow description are partly covered by the Astrogrid Workflow System, which provides a workflow scripting language (Groovy), a workflow engine and a user interface for scripting the task chain. It supports interfaces to VO applications via CEA and relies on a distributed storage (MySpace). The description of task allocation and execution on GRID installations is more in the hands of the grid community. Some further work is needed to bridge the gap, between VO workflow descriptions and grid execution logs in order to give feed back to the user, in terms of VO procedures.