Mem. S.A.It. Vol. 80, 574

**Memorie** della

# AstroGrid: Taverna in the Virtual Observatory

K.M. Benson[1] and N.A. Walton[2]

[1] Mullard Space Science Laboratory, University College London – Holmbury St. Mary, RH5 6NT, UK e-mail: kmb@mssl.ucl.ac.uk
[2] Institute of Astronomy, University of Cambridge Cambridge, CB3 0HA, UK. e-mail: naw@ast.cam.ac.uk

**Abstract.** This paper reports on the implementation of the Taverna workbench by AstroGrid, a tool for designing and executing workflows of tasks in the Virtual Observatory. The workflow approach helps astronomers perform complex task sequences with little technical effort. Visual approach to workflow construction streamlines highly complex analysis over public and private data and uses computational resources as minimal as a desktop computer. Some integration issues and future work are discussed in this article.

**Key words.** Astrogrid – Virtual Observatory – Taverna – Workflows – Astro Runtime

## 1. Introduction

Execution of Virtual Observatory (VO)-related data retrievals commonly requires construction of workflows. A workflow is a sequence of tasks that can be stored in a standard format and then sent for execution immediately or saved for future use. For instance, querying a specific image service (in the VO this is a service which conforms to the Simple Image Access Protocol (SIAP) standard[1] to find an image at a particular position on the sky; sending the image found to an invokeable copy of SExtractor (an application which detects objects in an image[2], putting in suitable parameters, and deriving an object catalogue; saving the resulting table to your VOSpace (a distributed storage system developed for the VO); and then also loading the catalogue into Topcat[3], will require construction of a workflow sequence[4].

The Taverna workflow approach, commonly used in Bioinformatics (Oinn 2005), brings flexibility to VO-related tasks. One can change the query parameters on a stored workflow before its execution. Besides the parameters, workflow allows changing the tasks, their sequence, and other options. In the example above, one may need to alter the target SIAP service, or using a different sky position. The workflow approach helps astronomers perform complex task sequences with minimal technical effort. Graphical workflow editors have been widely adopted for building workflows.

AstroGrid, the UKs Virtual Observatory (VO) System[5], developed earlier its own work-

[1] http://www.ivoa.net/Documents/latest/SIA.html
[2] http://terapix.iap.fr/rubrique.php?id_rubrique=91
[3] http://www.star.bris.ac.uk/$\sim$mbt/topcat/)
[4] http://www.astrogrid.org/wiki/Help/IntroScripting
[5] http://www.astrogrid.org

flow solution, named Job Execution Service (JES)[6]. However, the JES suffered from scalability issues, thus as AstroGrid moved to full VO operations, it had been deprecated. In order to provide a fully capable workflow system, AstroGrid made an analysis of alternative workflow solutions. The Taverna workbench , a tool for designing and executing workflows, has been chosen for its robustness and scalability, and use in other astronomical applications. It provides a straightforward visualisation mechanism for users with limited computing background and moderate technical resources.

AstroGrid has been developing a version of Taverna with VO plugins (Walton 2008), which added a number of signicant capabilities. AstroGrid's implementation of Taverna helps astronomers search, retrieve and work with a wide range of astrophysical data. One element of the AstroGrid system is the ability to create workflows using visualisation. The advantage of visual interface streamlines construction of highly complex analysis over public and private data and computational resources, all from a standard PC, UNIX box or Apple computer. AstorGrid workflows are saved as documents in a standard XML format; the Taverna plugin allows loading such workflow documents into the visual flow diagram and modify it according to user needs (see Fig. 1.).

When a workflow is executed the user is prompted for input parameters. A list of values can be accepted instead of a single parameter as it is a dataflow model. Implicit parallelism in the workflow is exploited, and the user can specify error- handling tactics - like retrying a service, or alternatives. Taverna contains an interactive execution monitor which shows the progress of the workflow - the nodes in the flowchart change colour depending on the status of the tasks such as whether they are running, completed or pending.

[6] `http://software.astrogrid.org/components/jes/index.html`

## 2. Pre-built workflows and astronomical communities

Further facilitating sharing of workflows among the astronomical community, myExperiment, a social web site for scientists, supports search and open access to already constructed workflows.[7]

The plug-in architecture of Taverna is being extended to provide the capability to connect to the VO community. ESO uses a tool called Reflex in implementing data reduction pipelines via Taverna. JVO will have a plugin connected to Taverna in the near future. AstroGrid have developed a plug-in adaptor for Taverna, thereby exposing the standard range of IVOA standard services, thus those for images (SIAP), spectra (SSAP), tables (TAP), and applications (CEA-Common Execution Architecture) to Taverna/.

## 3. Astro RunTime

The AstroGrid implementation of Taverna relies on the Astro Runtime, a client-side library of functions to access VO. As seen in Fig. 2 applications, such as Taverna, leverage RMI to connect to the Astro RunTime. The scope of the library is to integrate all VO standards, popular ad-hoc services, and suitable help functions.

This exposes all the functions of the AstroGrid Astro Runtime (AR) as services in Taverna. The plugin connects to a running AstroGrid Astro Runtime instance via Java RMI. Implementing the plugin was simplied because the AR provides functions to return lists of functions, along with the associated documentation and type information. As the AR is self-describing, all the plugin code has to do is massage this data into the model expected by Taverna. Taverna is utilised in the AstroGrid context, both as the graphical workflow builder, and as the workflow enactment engine.

Astro RunTime is exposed as a desktop service accessible from almost all programming languages via a range of connections including

[7] `http://www.myexperiment.org/`

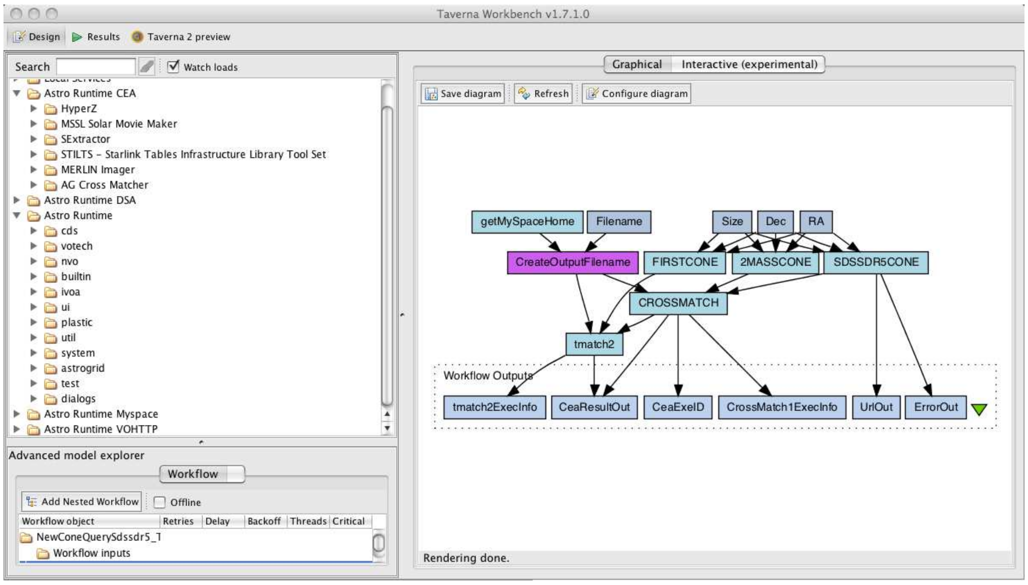


**Fig. 1.** This shows an exampe Taverna Workflow using Astrogrid. This crossmatches the results from a positional query of the all sky optical (SDSS) and infrared (2MASS) image surveys, placing the resulting VOTable in a Virtual storage space known as VOSpace.

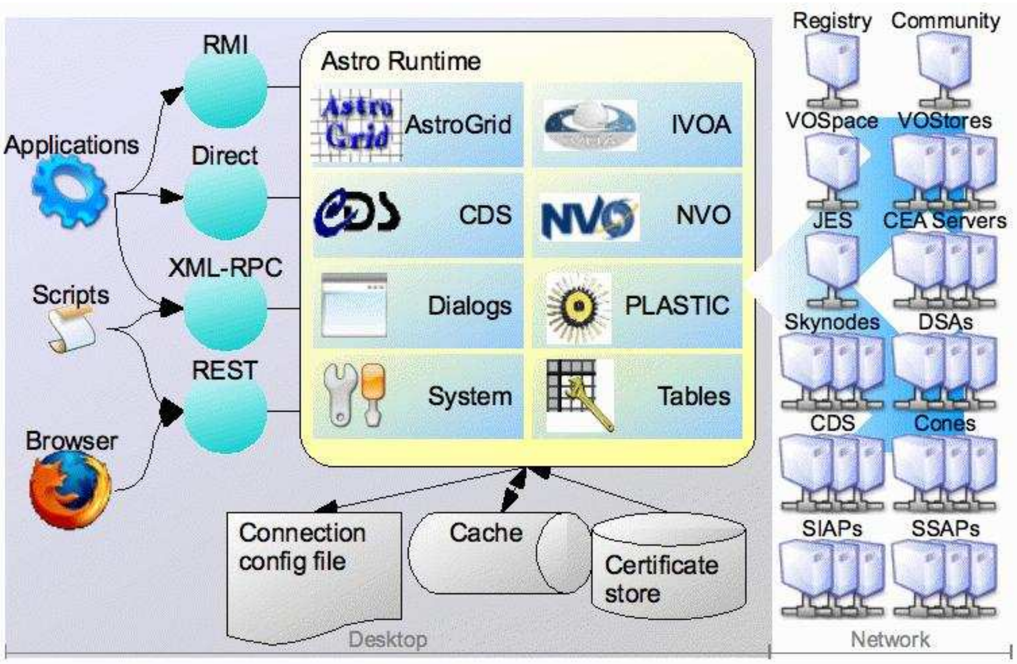


**Fig. 2.** Astro Runtime

XMLRPC, HTTP, RMI, C-binding, etc. The Astro RunTime API is designed to use consistent abstractions and types to minimise learning curve and insulate from change.

Via the Astro Runtime the AstroGrid Taverna implementation is now integrated into the new AstroGrid VODesktop (Walton 2008). Thus astronomers can search relevant data services, and compose workflows connecting queries of these resources, integrated with applications to carry manipulations of the data. All of these services are connected with other elements of the AstroGrid infrastructure, enabling for instance staging of data via storage in the VOSpace distributed storage areas. The asynchronous nature of the entire system allows to support running and logging of resource intensive workflows.

## 4. Future work

Astrogrid will be placing Taverna wrapped as a asynchronous service allowing very long workflows to be run as a service and not just on a clients computer. This will also allow faster computation of workflows that use intensive memory and cpu. The official release of the Astrogrid plugin of Taverna will be autumn 2008. It will be possible to save and share user 'Workflows' utilising the myExperiment community space. This will encourage great distribution by research teams of not only the results of their scientific investigations, but also the processing steps that they took in those investigations. This has the potential to allow for richer community use of high value datasets as exposed by the astronomical Virtual Observatory.

*Acknowledgements.* AstroGrid is funded by the UK Science and Technology Facilities Council through its eScience Programme.

## References

Hook, R., et al. 2008 in ASP Conf. Ser. XXX, ADASS XVII, ed. J. Lewis, R. Argyle

P. Bunclark, D. Evans, & E. Gonzalez-Solares (San Francisco: ASP), [P8.7]

Oinn, et al. 2005, Concurrency and Computation Practice and Experience, 18, Issue 10,1067

Walton, N. A., et al. 2008 in ASP Conf. Ser. XXX, ADASS XVII, ed. J. Lewis, R. Argyle,

Skrutskie, R. M., et al. 2006, AJ, 131, 1163

Tedds, J. A., et al. 2008 in ASP Conf. Ser. XXX, ADASS XVII, ed. J. Lewis, R. Argyle,

P. Bunclark, D. Evans, & E. Gonzalez-Solares (San Francisco: ASP), [O8.2]

Walton, N. A. 2005, Astronomy & Geophysics, 46, 23

Walton, N. A., et al. 2005 in ASP Conf. Ser. 347, ADASS XIV, ed. P. Shopbell, M.

Britton, & R. Ebert (San Francisco: ASP), 347, 273

P. Bunclark, D. Evans, & E. Gonzalez-Solares (San Francisco: ASP), [P1.21]

Bunclark, D. Evans, & E. Gonzalez-Solares (San Francisco: ASP), [D14]