AstroTaverna - Building workflows with Virtual Observatory services

Ruiz, J.E.[1], Garrido, J.[1], Soiland-Reyes, S.[2], Santander-Vela, J.D.[1], Sánchez-Expósito, S.[1], and Verdes-M, J.D.[1]

[1] Instituto de Astrofísica de Andalucía - CSIC, Glorieta de la Astronomía s/n, 18008 Granada, Spain e-mail: `jer@iaa.es`
[2] School of Computer Science, University of Manchester, Kilburne Building, Manchester, UK

ABSTRACT

Despite the long tradition of publishing digital datasets in Astronomy, and the existence of a rich grid of services providing astronomical datasets in standardized formats through the Virtual Observatory (VO), there has been little use of scientific workflow technologies **in this field**. In this paper we present AstroTaverna, a plugin that we have developed for the Taverna scientific workflow management system. It integrates existing VO web services as first-class building blocks in Taverna workflows, allowing the digital capture of otherwise lost procedural steps manually performed in e.g. GUI tools, providing reproducibility and re-use. It improves the readability of digital VO recipes with a comprehensive view of the entire automated execution process, complementing the scarce narratives produced in the classic documentation practices, transforming them into *living tutorials* for an efficient use of the VO infrastructure. The plugin also adds astronomical data manipulation and transformation tools based on the STIL Tool Set and the integration of Aladin VO software, as well as interactive connectivity with SAMP-compliant astronomy tools.

Key words. Workflows, Virtual Observatory, e-Science, Interoperability, , look at list of keywords in the Journal

1. Introduction

In recent decades Astronomy has witnessed the development of instrumentation performing observations of the sky across the full range of the electromagnetic spectrum. The advent of automated survey facilities has dramatically increased the number of archival datasets and transformed the approach to deal with key science topics, now potentially benefitting from seamless access to interoperable multi-wavelength data. The International Virtual Observatory Alliance[1] (IVOA) was created with the goal of developing standards and their implementations for exchange of astronomical data, and maintaining registries of data repositories and services. The Virtual Observatory (VO) consists of a grid of web-distributed data services which provides free access to astronomical catalogues, image and spectrum data, combined with local visualization and analytical tools (VO Apps) where astronomers graphically explore the different data sets, all following IVOA recommendations. The VO provides a way to transparently access and extract scientific knowledge from multiple archives of astronomical data, enabling standardised services discovery and access to interoperable data.

VO Apps are any kind of software able to access the elements of the VO data grid, following IVOA protocols, and sometimes permitting discovery of services available in the VO Registry (Benson & Walton 2009) in order to provide the users with the datasets of their interest. Current VO Apps are mainly of two-kinds: general purpose tools for accessing a particular kind of data (catalogues, images, or spectra); and custom-built programs and scripts created by astronomers in order to perform their scientific research. Both kinds of tools are typically used together for an experiment, with the user performing each step with different tools. This means that, in order to document the fully procedural protocol, the user needs to manually capture the dependencies between steps in an **Electronic Lab Notebook (Slominski 2010)**, which is currently accomplished by writing scripts driving the execution of other tools or scripts.

Digital scientific workflows (Gil et al. 2007; Gil & Rey 2008) allow capturing the actions performed in distributed data access, aggregation and analysis using third-party services, as they identify and record the provenance leading to the final obtained results. They facilitate the means to come up with "in silico" **reproducible** research, that can be shared, executed or even modified by other scientists in the community. **Workflows** provide a comprehensive explicit exploratory view of the scientific protocol, while being at the same time the precise and well-codified description needed by the underlying engine to execute the whole multi-step process. Unlike traditional pipelines, which tend to produce scientifically exploitable data, most of the scientific workflows are aimed at producing scientific insight.

Taverna[2] **(Wolstencroft et al. 2013)** is an open source, domain-independent scientific workflow management system that has grown popular in fields like bioinformatics, chemistry, text mining, and image analysis. It is a suite of tools including Taverna Workbench, a desktop client application used to graphically design, create, edit and execute scientific workflows. **Taverna workflows can also be executed on the command line or on an installation of the Taverna Server.** The Taverna Workbench is specially designed to combine distributed web services and local tools into complex workflows, the basics building blocks of VO-services-based scientific **workflows**. It also integrates workflow discovery and publishing through dedicated UI for access to MyExperiment[3] (Goble et al. 2010) public digital library of workflows. Moreover, its plugin architecture allows

---

[1] http://www.ivoa.net

[2] http://www.taverna.org.uk
[3] http://www.myexperiment.org/

loose integration and open-source developments of plugins, decoupled from core developments of Taverna Workbench.

There have been several attempts of bringing workflow technologies to the astronomy field. The ESO Sampo[4] project performed a feasibility study for the integration of the ESO pipeline processing tools with the Taverna 1 workflow engine. The study gave birth to the ESO Reflex [5] subproject (Hook et al. 2009), using the Kepler engine (Altintas et al. 2004) instead. The HELIO-VO[6] project **(Bentley et al. 2013)** has successfully used the Taverna 2 workflow engine, by creating heliophysics-specific services that provide support for their scientific use cases in a Service-Oriented Architecture. In the CyberSKA[7] project (Kiddle et al. 2011) a web-based workflow builder is being developed to address the evolving data-intensive science needs of future radio telescopes such as the Square Kilometre Array. Other related initiatives explored in Astronomy are the Triana[8] (Taylor et al. 2007) and Pegasus[9] **(Deelman et al. 2005)**workflow environments, workflow implementations of Montage[10] image mosaicing software, the VO-France Workflow working group (Schaaff et al. 2008), the Aladin (Bonnarel et al. 2000) JLOW plugin, as well as projects like VAMDC[11] (Walton et al. 2011), Astro-WISE[12], ER-Flow[13] or IceCore (Maisala et al. 2012). The Astro-Grid[14] project has been, to our knowledge, the sole inititative to deliberately undertake the development of tools to design and execute VO services based scientific workflows. AstroGrid created a modified version of the Taverna 1 workflow engine (Benson & Walton 2009; Walton & González-Solares 2009) which integrated VO Apps in the source code of the tool. Unfortunately, the final result was not publicly relased.

In this paper we present AstroTaverna, a plugin that simplifies the use of VO services from within the Taverna Workbench 2.4, by providing VO-service discovery facilities, and building blocks for VO data access and manipulation. AstroTaverna provides astronomers with the means to compose scientific workflows of standardized VO-services, together with existing local and distributed tools. In Sect. 2 we describe the functional drivers that led the developments of AstroTaverna as well as its main features, illustrating them all with an example use case. In Sect. 3 we provide a discussion on the impact of the use of workflows in the community in the context of the results issued from AstroTaverna developments, as well as future plans for its dissemination and improvements. Finally, in Sect. 4 the main conclusions of the exposed work are presented.

2. Opening Taverna to the Virtual Observatory

2.1. The Taverna Workbench

The Taverna Workbench allows scientists to assemble data and analytical services in a graphical **workflow, connecting tools by specifying the desired flow of data (inputs and outputs) rather than a fixed execution order.** It provides the scientist with a tool for designing and executing workflows, playfully

---

[4] http://www.eso.org/sci/software/sampo
[5] http://www.eso.org/sci/software/sampo/reflex/
[6] http://www.helio-vo.eu
[7] http://www.cyberska.org
[8] http://www.trianacode.org
[9] http://pegasus.isi.edu
[10] http://montage.ipac.caltech.edu
[11] http://www.vamdc.eu
[12] http://www.astro-wise.org
[13] http://www.erflow.eu
[14] http://www.astrogrid.org

combining local and external processes with data, as well as with support for **service** and workflow discovery. Such workflows typically combine a wide range of third-party tools, ranging from RESTful and SOAP web services, SQL databases and grid infrastructures, to local command line tools and embedded scripts. They can then be processed on the Taverna Workbench (desktop application), **using the Taverna Command Line,** or on installations of the Taverna Server (remote location).

Taverna Workbench additionally offers tools for browsing provenance, and inspect workflow runs, allowing the exploration of intermediate results from past invocations. Execution of workflows is optimized because of its intra-processing data parallelism, users may fine-tune the behavior of processes as well as define error-handling tactics, and follow the progress of the execution of the workflow through its dependencies in an interactive monitor, **browsing outputs and intermediate values as they are produced**.

The *executable graph* provided by a workflow does **not guarantee on its own** a full understanding of the scientific method, neither reusability **nor** reproducibility **(Belhajjame et al. 2012)**. This is specially needed in collaborative science, where sharing executable recipes is key in order to avoid duplication of effort and reinvention, as well as to assess soundness of the scientific methodology. To overcome these issues, additional information may be needed to better document the process. Taverna Workbench allows providing free-text descriptions on the elements of the workflow, which can be seen as complementary information about the science involved, assumptions made, hypothesis to be proven, the configuration of the execution environment, caveats and still not well solved issues, etc.

Although these are already major advantages over manual analysis methods and scripting, Taverna Workbench also supports the integration of local scripts and tools as workflow components. These processes are executed as if they were launched at the terminal command line, allowing the full orchestration of local and external resources in a single workflow, where the entire procedural protocol is captured and possibly documented.

2.2. AstroTaverna design drivers

The AstroTaverna plugin targets users with no special technical knowledge of the underlying applications involved in the workflow execution process. It should be easy to install and allow users to design, create and fully understand workflows without commissioning specialists or hiring software engineers.

The AstroTaverna plugin aims to provide universal access to VO data services recorded in the VO Registry. It supports open science and open data access. In the VO context, standardized workflows could be helpful to gather and aggregate data from distributed datasets, in order to engage multi-epoch and multiband comparative astrophysics. The vision of a workflow as the orchestration of tools and tasks running either locally or externally may be greatly improved if we consider the VO as a rich infrastructure of web services and data, where VO services may be used as components for web-services-based workflows.

The main functional drivers to be addressed by the AstroTaverna plugin are:

- *Data discovery, gathering and aggregation.*
  AstroTaverna should be an open window to public astronomical data and services. We take advantage of the infrastructure of interoperable data provided by the VO in order to build, for a sample of objects, tabular representations of their properties extracted from different archives. These ta-

bles may be rebuilt at any moment from the re-execution of the workflow, hence providing up-to-date information when refreshed.

– *Data manipulation, filtering and cross matching.*
The interoperable VOTable (Ochsenbein et al. 2009) data format provided by the VO allows efficient manipulation and combination of datasets coming from different archives. Tabular data should be handled along the workflow with no need to address major format conversion issues, but focusing on basic operations for data management (e.g. cross matching, filtering, table merging and concatenation, row/column extraction and addition, etc.) in order to shape and rebuild information based on very specific needs.

– *Data transformation.*
We should consider allowing actions on datasets going beyond data massage and manipulation, hence altering their values. Operations like transformation of sky coordinates and among reference systems, resolution of source names into coordinates, addition of new data based on existing ones, etc. are commonly used when working with tabular data of astronomical objects. More complex data transformations could be possible by adding small code snippets of Python scripting or local command line software.

– *VO software integration and data inspection.*
In order to benefit from widely used and well known functionalities of mature VO Apps, the integration of some of their specific tasks should be possible. Moreover, in order to achieve a full understanding of final and intermediate data a proper rendering of the VOTable format is needed. Sharing these data with VO Apps in a seamless way would allow the user to perform subsequent more sophisticated data analysis and data inspection.

2.3. Plugin design, implementation and functionalities

Users consume transparently the VO-technologies by using astronomy VO-compliant software (VO Apps), as seen in Fig. 1. A user interacts with the VO via a variety of VO-aware, locally run applications, and accesses either locally archived files, or remote files, either via VO protocols or using a web browser to access a VO-enabled web portal. Users must only be aware of the different VO applications –VOApp in the figure– and/or VO portals of their interest, and that there exist interoperable image, spectra, and catalogue servers, transparently accessed from their toolset. They must also be aware that some VO applications can send messages and data between them, sending for instance data to be analysed to some applications, and sending the received results to other applications for plotting. All the VO systems in the Internet –dotted cloud in Fig. 1– are indispensable for the operation of the VO, but are completely transparent to the user. It is very common that VO App users deal simultaneously with local and remote data by invoking VO services encapsulated in the VO Apps. These services are known by the applications since they are indexed in registry servers (VO Registry).

VO services are either SOAP (for the access to the VO Registry), or RESTful (for data access protocols) web services. Taverna Workbench provides tools for directly accessing this kind of services, but that means that discovery of VO services is left to the user (while service discovery is integral for VO Apps), together with a certain knowledge of web services and the parameters required by each data access. AstroTaverna should provide the means for easy discovery and query of VO data access services. In addition, functionality for the manipulation and vi-
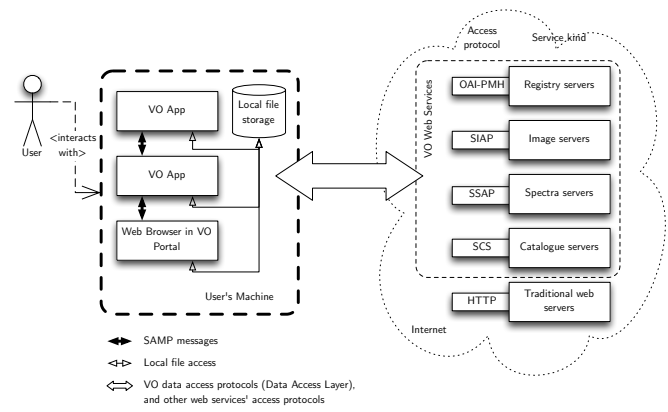

Fig. 1. Virtual Observatory usage and architecture.

sualization of VOTables should be also added in order to make VO services first-class building blocks for Taverna workflows.

The main VO architectural elements that have been implemented in AstroTaverna are:

– The VO Registry: based on the Open Archives Initiative resource metadata (Lagoze et al. 2002), it provides services registered in the form of VO Resources (Hanisch 2007). Specializations of the VO Resource exist for identifying image data services (Simple Image Access Protocol; Tody et al. 2009), spectral data services (Simple Spectral Access Protocol; Tody et al. 2012), positional-search table services (Cone-Search Protocol; Williams et al. 2008), and tabular complex searches (Tabular Access Protocol; Dowler et al. 2010). Each entry contains metadata identifying the curators and publishers of the data, service type, service-type specific descriptions, or the URL for the services' entry points.

– VO Services: the VO data access services implemented in AstroTaverna are those compliant with IVOA standards for *simple data access protocols*[15], with parameterized inputs encoded in one URI that is used as entry point to query the services. Data returned are always in VOTable format. Science data are embedded in the VOTable, or linked from data access fields.

– VO Apps: there are a number of mature VO Apps widely used in the VO community providing enhanced capabilities for tabular data management and inspection, as well as simple tasks for multi-wavelength image visualization, analysis and comparison. Some may be fully integrated in Taverna Workbench as local services performing very specific tasks based on STIL API and STILTS Java framework (Taylor 2006), or as command line calls in the case of Aladin scripts and macros. Integrating STILTS in AstroTaverna also grants direct access to most common functionalities provided by this library. Extended functionalities may be accessed making use of scripts through Taverna Beanshells. Finally, the SAMP (Simple Application Messaging Protocol; Taylor et al. 2012) messaging protocol enables data exchange and communication among local tools, allowing new born tools to benefit from all the functionalities coming from existing VO Apps.

2.3.1. Data discovery, gathering and aggregation

AstroTaverna provides support for discovery of VO data access services, with the addition of a new "VO Services" tab to the
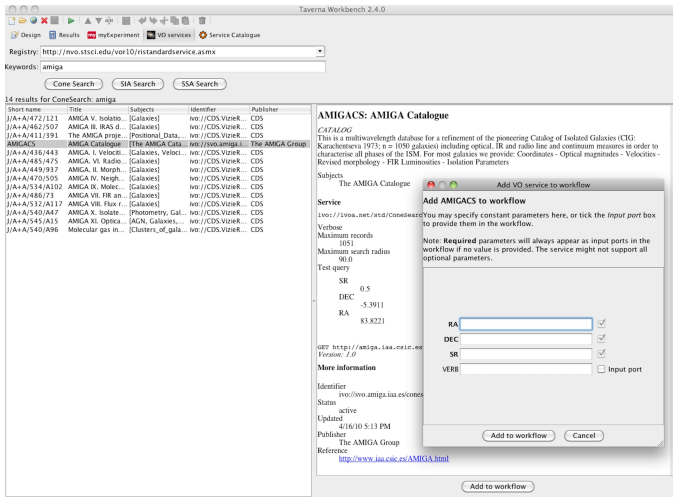

15

Fig. 2. AstroTaverna user interface for discovery of VO data access services

perspectives menu located on the top of Taverna Workbench window. This perspective, as seen in Fig. 2, allows users to discover, browse, inspect and add VO services to workflows in the process of design. Users may select the VO Registry URL endpoint where they would like to perform the search of VO services. This search is based on keywords entered on a free-text box as well as on the specific service type chosen: Cone Search (positionally-indexed tabular data), SSA (positionally-indexed spectra) and SIA (positionally-indexed images). The list of VO services found is then presented as a tabular list with columns for the service *Short name, Title, Subjects, Identifier* and *Publisher*. The query interface is very similar to the one proposed in the widely used TOPCAT VO application (Taylor 2011).

When one of the services is selected among the list, a more detailed description is displayed on the right panel, with metadata extracted from the VO Registry. The user may, at this point, decide to use the service in the design process of a workflow pressing the "Add to workflow" button. This action opens a window allowing a more detailed definition of the specific artifact that will be created associated to this service on the ***Design*** perspective of Taverna Workbench. The AstroTaverna plugin provides the user with the possibility to create these artifacts on the workflow design with **fixed parameter values** entered on this form. **Optional, service-specific query parameters can similarly be enabled as inputs or given a fixed value.** VO services may also be added directly into the *Design* perspective in drag-and-drop actions from the service panel. In this case the user needs to know the exact URL of the added RESTful web service, **as the VO Registry discovery process is not involved. This mechanism is useful while services are under development or not yet registered in the registry.**

2.3.2. Data manipulation, filtering and cross matching

Given that the data returned by VO services are interoperable VOTables, there is a need for tools allowing actions for efficiently manipulating tabular data, as well as tools providing format conversion between VOTable format and others widely used tabular formats in astronomy research (e.g. FITS, CSV, TST, ASCII, HTML, etc.). Furthermore, spatial table cross matching (finding rows in other tables which might be near a given position, with a prescribed tolerance) and table joins on common

keys are also common actions that need to be covered in data manipulation basic capabilities.

AstroTaverna provides an AstroTools **collection** of **built-in** data manipulation services.

- – *Format conversion for tables*
- – *Concatenate two VOTables*
- – *Concatenate a list of VOTables*
- – *Join two VOTables*
- – *Add a common field to a VOTable*
- – *Crossmatch two VOTables*
- – *Select columns from a VOTable*
- – *Extract a column from a VOTable as a list*
- – *Select rows from a VOTable*

They may be dragged and added to the workflow diagram from the **services panel** of Taverna Workbench into the central design panel. They include table row filtering by conditional expressions on field values and column subsets. **The expressions for row filtering and column selection follow the STILTS syntax.**

2.3.3. Data transformation

On the same basis, AstroTaverna also provides in the AstroTools **collection of services** for transformation of VOTable data. These include the addition of columns based on calculations and/or basic combination of existing ones, following algebraic expressions, transformation of sky coordinates and between different reference systems (e.g. equatorial, ecliptic, galactic, and others), resolution of astronomical object names into equatorial (Right Ascension, Declination) coordinates using the Sesame service, as well as the conversion of coordinates between different units. These are the following:

- – *Add a column to a VOTable*
- – *Resolve object name into coordinates*
- – *Coordinates reference system transformation*
- – *Coordinate units conversion*

Other two services have been added related with the creation and validation of text files using templates. These templates are provided in the workflow, and are used together with information extracted from VOTables in order to create local files *on-the-fly* (e.g. software configuration files, input data files in a given specific format, etc.).

- – *Fill template file from VOTable*
- – *Validate template file against VOTable*

More complex data transformations may be possible using the Service Template Tool service provided by Taverna Workbench to locally execute specific command line software or OS commands and small script files. This is especially useful when adding existing snippets of Python scripting, commonly used in astronomy research.

2.3.4. VO software integration and data inspection

Proper rendering of VOTables is needed in order to allow inspection of intermediate and final values issued from the data workflow. AstroTaverna provides the possibility to display the XML-formatted VOTables in the *Results* perspective, together with their field's metadata, in a spreadsheet-like form. It also implements the SAMP message exchange protocol, allowing the user

to send VOTable data to local SAMP-enabled software. SAMP connects AstroTaverna workflows with mature VO Apps, giving the user the possibility to perform post-workflow data inspection and analysis using more specific local software.

AstroTaverna also integrates with the Aladin VO App, allowing the user to execute macros and scripts using data extracted from VOTables provided by VO services. These services may be configured with the GUI option enabled, which triggers the visual automated execution of actions on Aladin during the process of workflow execution. The specific output of these Aladin services is a VOTable containing the path and filenames of the final images produced during the executions of macros and/or scripts.

2.4. Example use case

In this section we illustrate the use of the AstroTaverna plugin with a specific use case consisting on a workflow needing functionalities for VO data gathering, manipulation, transformation and inspection. From an initial sample of galaxy names, we retrieve some of their properties from VO archives and other data services, calculate others based on the previous ones, and complement them with associated bibliography data and links to others related external catalogs.

In this use case we will assumme that this information will be used as the basis for preparing lists of targets for different observational proposals, performing e.g. filtering by positions of galaxies on the sky, size, brightness, etc. The workflow would be re-executed in further occassions in order to provide up-to-date information when needed to prepare a new target list, and possibily slightly modified according to specific needs for different proposals. The whole workflow is composed of four different blocks (nested workflows), according to their distinctive roles and actions performed, as seen in Fig. 3, and it is publicly accessible[16] through MyExperiment portal. Starting from a list of object names in one ASCII file, the workflow builds up three different VOTables: *Object Properties, Bibliography* and *External Links*.

In a first step it queries, using the name of each galaxy as input, two different data access web services provided by NASA/IPAC Extragalactic Database (NED)[17] archive (Mazzarella & Team 2008).

The following web services provide information for the *Object Properties* VOTable.

Service Short Name: NED_basic_posn
Service Id: `ivo://NED/Basic_Position_Data_For_Object`
Service Output VOTable Fields:

– Right Ascension in degrees. (Equatorial J2000.0)
– Declination in degrees. (Equatorial J2000.0)
– Right Ascension in sexagesimal degrees. (Equatorial J2000.0)
– Declination in sexagesimal units. (Equatorial J2000.0)
– Longitude in decimal degrees. (Ecliptic J2000.0)
– Latitude in degrees. (Ecliptic J2000.0)
– Longitude in decimal degrees. (Galactic)
– Latitude in degrees. (Galactic)
– Longitude in decimal degrees. (Super Galactic)
– Latitude in degrees. (Super Galactic)

Service Short Name: NED_basic
Service Id: `ivo://NED/Basic_Data_For_Object`
Service Output VOTable Fields:

– The heliocentric radial velocity in km/sec.
– Quality flag for heliocentric radial velocity.
– Velocity's mean error when known.
– Major diameter in arcminutes.
– Minor diameter in arcminutes.
– Optical magnitude.
– The morphological type.
– The Galactic reddening E(B-V), in magnitudes.

The workflow makes use of the *Format conversion for tables* AstroTaverna service for making ASCII to VOTable conversion, as well as format conversion for datavalues using STILTS expressions on the *Add a column to a VOTables* services. It also queries a specific Vizier ConeSearch service to retrieve data from a published catalog (J/A+A/545/A15). The output VOTable fields retrieved from this catalog are:

– Flux density in Ks-band

– Error in the flux density in Ks-band

– Log of the luminosity in Ks-band

– Error in the log of luminosity in Ks-band

All the previously enumerated fields conform the *Object Properties* VOTable, which is then filtered according to observational and instrumental criteria: e.g. selecting those galaxies with declination greater than twenty degrees and those with observed angular diameter lower than two arc minutes.

The next step is the calculation of the distance to the selected galaxies, based on their heliocentric radial velocities and sky positions related to the center of Virgo cluster. The workflow executes a Python script as from the command line using the aforementioned parameters as input values. The script is registered and configured in the free-text box of Service Template Tool service that is provided by Taverna Workbench. The calculated distances are added to the *Object Properties* VOTable.

Finally, there are other data access web services provided by NED archive that are queried, in order to complement the sample with related information from bibliography and associated external catalogs.

The information for the *Bibliography* VOTable is provided by the following web service:

Service Short Name: NED_search_notes
Service Id: `ivo://ned.ipac/Notes_By_Object_Name`
Service Output VOTable Fields:

– Object Name
– The NED 19-digit Bibliographic Reference code (year, journal, volume number, page number, and the initial of the first author's last name).
– A Note transcribed as faithfully as possible from the catalog or paper.

The following web service provides the information for the *External Links* VOTable.

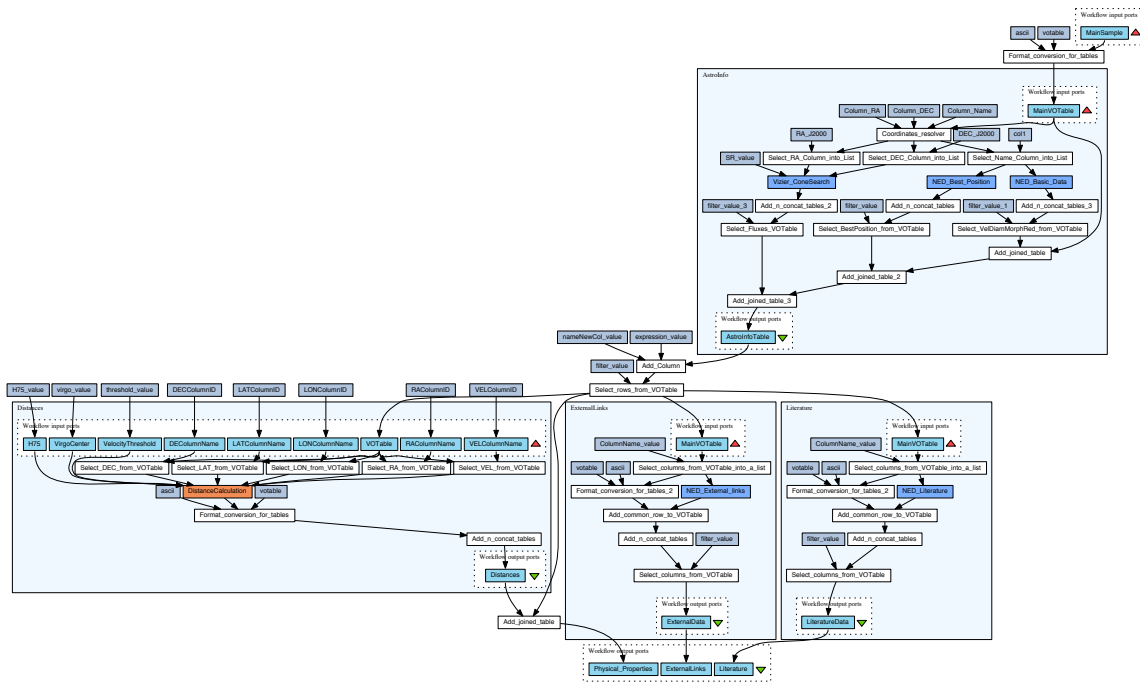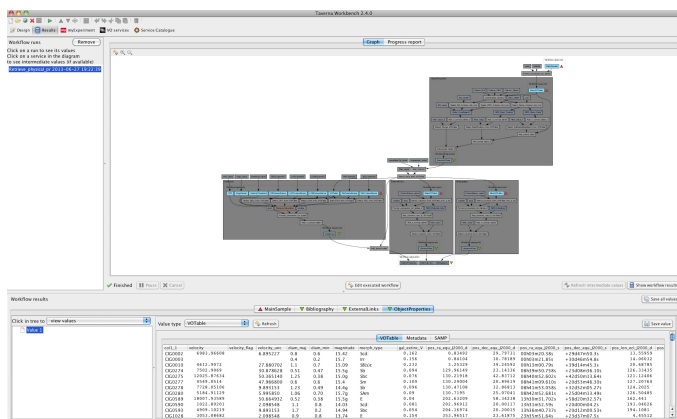**Fig. 3.** Diagram of AstroTaverna workflow for the example use case described in Sect. 2.4.

Fig. 4. Execution of AstroTaverna workflow and rendering of results for the example use case described in Sect. 2.4.THIS IS A MOVIE WHEN PAPER READ IN ELSEVIER WEBSITE

Service Short Name: NED_external
Service Id: `ivo://NED/Basic_External_Links_by_object_name`
Service Output VOTable Fields:

- Object Name
- NED's link to related on-line astronomical services that are specific to the survey or catalog associated with the NED name.
- URL address of related on-line astronomical service.

3. Results and discussion

The AstroTaverna plugin opens a window to astronomy archives by enabling easy design and execution of VO services-based workflows. These strongly rely on the very basic foundation of

the Virtual Observatory: public data access, data interoperability and minimized network transfers with lightweight metadata. AstroTaverna workflows also benefit from seamless connectivity with all the ecosystem of local SAMP-enabled VO Apps. All these concepts are key in the future development of digital science in Astronomy, where we are already facing the advent of a plethora of astronomical public archives, and transfer network latency is more than an issue.

The focus of AstroTaverna lies in enhancing the documentation of the scientific process, where the transparency of the method is exposed in an interactive monitor showing the progress of the execution of the workflow. The automation of the process is just a nice-to-have feature, long time accomplished by scientists through different scripting languages and environments. The main goal of AstroTaverna workflows is improving readability and enabling reproducibility of digital VO recipes with a small additional effort, complementing the scarce narratives produced in the classic documentation process. In addition, they capture and register provenance information otherwise lost in procedural steps manually performed in GUIs of VO Apps.

The capture of the experimental protocol related with GUI software into a more automated digital flow has made us notice the potential impact of workflows as living tutorials, explicitly exposing how to take advantage of existing rich infrastructure of data and services provided by the VO. This is especially relevant with the automated execution of Aladin macros and scripts in GUI mode. Scientists may visualize the actions performed by the workflows as they progress in their executions, allowing them to practice self-learning by the example, which expedites training and avoids reinvention.

In this sense, a collection of 26 elemental workflows showcasing the potential of AstroTaverna (workflow snippets) has been developed and published in the MyExperiment portal with the name of AstroTaverna Starter Pack[18], in order to provide po-

---

[18] http://www.myexperiment.org/packs/420.html

tential users of AstroTaverna a set of small workflows carefully designed to perform very specific actions. Digital libraries of workflows could increase the visibility of the scientific outcome, hence its discovery, re-use and a more efficient exploitation of present astronomical archives, computational infrastructures and observational facilities (Ruiz 2012). Astronomy is a collaborative science, and it has also become highly specialized, as many other disciplines. Sharing, preservation, discovery and a much simplified access to resources in the composition of scientific workflows will enable astronomers to greatly benefit from each other's highly specialized knowhow, pushing them to share and publish not only results and data, but also processes and methodologies.

Going into more technical considerations, by developing AstroTaverna as a user downloadable plugin, loosely coupled with Taverna Workbench, we can more quickly iterate and add additional capabilities as separated independent developments. In particular, given the dependency of AstroTaverna on the STILTS framework, we can use updated versions of it, bringing bug fixes and new features. By creating the AstroTaverna plugin, we enable the VO as a core building block for Taverna Workbench.

The AstroTaverna plugin has been presented in the last IVOA Interop Conferences in Sao Paolo and Heidelberg, as well as in the 8th VO France Workflow Working Group Meeting held in Paris-Meudon Observatory. AstroTaverna has been advertised in Issue 009 of IVOA Newsletter and through the IVOA general mailing list, together with a Video Tutorial[19]. Information on the AstroTaverna plugin may be found in AMIGA Group[20] and Taverna Plugins[21] websites, as well as in the specific AstroTaverna Website[22]. The source code has been published in Astrophysics Source Code Library (ASCL) Garrido et al. 2013, it may be publicly accessed in GitHub[23] open source hosting service and an AstroTaverna support mailing list (*astrotaverna-users@iaa.es*) has been created. All this work has triggered the interest of VAMDC and ER-Flow EU FP7 funded projects in order to broaden the users community of AstroTaverna and study potential collaborations for developments in the field of workflows. In addition, a strategy for dissemination and community engagement is being set up with the Spanish Virtual Observatory, where specific use cases are being migrated to AstroTaverna workflows.

At the moment of writing these lines, the latest version of the AstroTaverna plugin is 1.08. New versions of the plugin are periodically released, not just with bug-fixes, but implementing additional capabilities (e.g. additions of new local services, improvements on VOTable rendering and on the VO Services Discovery GUI). We plan to add support for discovery, design and execution of Table Access Protocol (TAP) in future versions of the module, as the VO is moving towards the more generic and capable TAP services for general tabular data access. We have explored the possibilities of adding a richer Python support through Jython, but the lack of a pure-python Numpy has precluded that possibility. The orchestration of external services in workflows would greatly benefit from the adoption of web services interoperability standards. In this context, the PDL (Parameter Description Language; Zwolf et al. 2013) has been proposed in IVOA forums in order to deal with web services interoperabil-

ity in the VO. Next versions of the AstroTaverna plugin will provide a client for PDL self-described services, allowing data input validation before service invocation, providing service metadata description, and services interoperability in the design process of workflows. Other improvements consider straightforward access to the Starter Pack from the Taverna Workbench.

4. Conclusions

We have presented AstroTaverna, a plugin for Taverna Workbench 2.4 that provides the means to build astronomy web services-based workflows upon VO services discovery and VOTable efficient manipulation. It integrates SAMP-enabled VO Apps, as well as the possibility to execute Aladin scripts and macros. The plugin is easy to install for any user and operating system, installation instructions may be found in the AstroTaverna Website[24]. The plugin has been illustrated with a specific use case that considers a workflow making use of local services for VO data gathering, manipulation, transformation and inspection.

AstroTaverna enables astronomers to capture, in a digital reproducible workflow, the experimental protocol and provenance information otherwise lost in procedural steps manually performed in GUIs of VO Apps. A collection of workflow snippets has been produced with the name of AstroTaverna Starter Pack in order to alleviate self-learning by the example and avoid reinvention.

Future work considers support for discovery, design and execution of TAP services, provide a client for PDL self-described services that will allow data input validation before service invocation, and straightforward access to the Starter Pack from the Taverna Workbench.

References

Altintas, I., Berkley, C., Jaeger, E., et al. 2004, in 16th International Conference on Scientific and Statistical Database Management, 423–424

Belhajjame, K., Corcho, O., Garijo, D., et al. 2012, in Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web, SePublica2012

---

[19] https://vimeo.com/51095004
[20] http://amiga.iaa.es/p/290-astrotaverna.htm
[21] http://www.taverna.org.uk/documentation/taverna-plugins/taverna-2-x-plugins/
[22] http://wf4ever.github.io/astrotaverna
[23] https://github.com/wf4ever/astrotaverna

[24] http://wf4ever.github.io/astrotaverna
[25] http://www.wf4ever-project.org
[26] http://amiga.iaa.es/p/294-open-science-canube.htm
[27] http://biotic.ugr.es/en/

Benson, K. M. & Walton, N. A. 2009, Memorie della Società Astronomica Italiana, 80, 574

Bentley, R., Brooke, J., Csillaghy, A., et al. 2013, Future Generation Computer Systems

Bonnarel, F., Fernique, P., Bienaymé, D., et al. 2000, Astronomy & Astrophysics Supplemet Series, 143, 33

Deelman, E., Singh, G., Su, M.-H., et al. 2005, Sci. Program., 13, 219

Dowler, P., Rixon, G., Tody, D., et al. 2010, IVOA Recommendation: Table Access Protocol v1.0, Tech. rep.

Garrido, J., Soiland-Reyes, S., Ruiz, J. E., & Sánchez, S. 2013, in Astrophysics Source Code Library, record ascl:1307.007

Gil, Y., Deelman, E., Ellisman, M., et al. 2007, IEEE Computer, 40, 24

Gil, Y. & Rey, M. 2008, Information Sciences, 16, 1

Goble, C. a., Bhagat, J., Aleksejevs, S., et al. 2010, Nucleic Acids Research, 38, 677

Hanisch, R. 2007, IVOA Recommendation: Resource Metadata for the Virtual Observatory v1.12, Tech. rep.

Hook, R., Ullgrén, M., Romaniello, M., et al. 2009, Memorie della Società Astronomica Italiana, 80, 578

Kiddle, C., Taylor, A. R., Pigat, D., et al. 2011, in ACM workshop on Gateway computing environments, 65–72

Lagoze, C., Van de Sompel, H., Nelson, M., & Warner, S. 2002, The Open Archives Initiative Protocol for Metadata Harvesting, Tech. Rep. v2.0, Open Archives Initiative

Maisala, S., Oittinen, T., Takala, T., Solin, O., & Huovelin, J. 2012, in Astronomical Data Analysis Software and Systems ASP XXI - ASP Conference Series, Vol. 461, 99

Mazzarella, J. M. & Team, N. E. D. 2008, in The National Virtual Observatory Book, Vol. 382

Ochsenbein, F., Williams, R., Davenhall, C., et al. 2009, IVOA Recommendation: VOTable Format Definition v1.2, Tech. Rep. August 2004

Ruiz, J. E. 2012, Highlights of Spanish Astrophysics VII, Proceedings of the X Scientific Meeting of the Spanish Astronomical Society (SEA)

Schaaff, A., Petit, F. L., Prugniel, P., Slezak, E., & Surace, C. 2008, in Astronomical Data Analysis Software and Systems XVII - ASP Conference Series, Vol. 394, 77–80

Slominski, A. 2010, in Services (SERVICES-1), 2010 6th World Congress on, 326–327

Taylor, I., Shields, M., Wang, I., & Harrison, A. 2007, in Workflows for e-Science, 320–339

Taylor, M. 2011, in Astrophysics Source Code Library, record ascl:1101.010

Taylor, M., Boch, T., Fitzpatrick, M., et al. 2012, IVOA Recommendation: SAMP - Simple Application Messaging Protocol v1.3, Tech. rep.

Taylor, M. B. 2006, in Astronomical Data Analysis Software and Systems XV - ASP Conference Series, Vol. 351, 666–669

Tody, D., Dolensky, M., McDowell, J., et al. 2012, IVOA Recommendation: Simple Spectral Access Protocol v1.1, Tech. rep.

Tody, D., Plante, R., & Harrison, P. 2009, IVOA Recommendation: Simple Image Access Specification v1.0, Tech. Rep. October 2008

Walton, N. & González-Solares, E. 2009, Lecture Notes in Physics, 791, 81

Walton, N. A., Dubernet, M. L., Mason, N. J., Piskunov, N., & Rixon, G. T. 2011, in Astronomical Data Analysis Software and Systems XX - ASP Conference Series, Vol. 442, 89–92

Williams, R., Hanisch, R., Szalay, A., & Plante, R. 2008, IVOA Recommendation: Simple Cone Search v1.03, Tech. Rep. February

Wolstencroft, K., Haines, R., Fellows, D., et al. 2013, Nucleic Acids Research, 41, W557

Zwolf, C., Harrison, P., Le Petit, F., Garrido, J., & Ruiz, J. E. 2013, IVOA Proposed Recommendation: PDL Parameter Description Language, Tech. rep.