Full length article

AstroTaverna—Building workflows with Virtual Observatory services

J.E. Ruiz [*],[1], J. Garrido [*], J.D. Santander-Vela, S. Sánchez-Expósito, L. Verdes-Montenegro

Instituto de Astrofísica de Andalucía - CSIC, Glorieta de la Astronomía s/n, 18008 Granada, Spain

HIGHLIGHTS

- AstroTaverna is a plugin we developed for the Taverna workflow management system.
- It integrates VO web services as first-class building blocks in Taverna scientific workflows.
- Digital capture of otherwise lost procedural steps, manually performed in GUIs, is now possible.
- Integration of STIL Tool Set, Aladin VO software and SAMP-compliant astronomy tools.
- It allows enhanced documentation and comprehensive view of automated digital VO recipes.

GRAPHICAL ABSTRACT
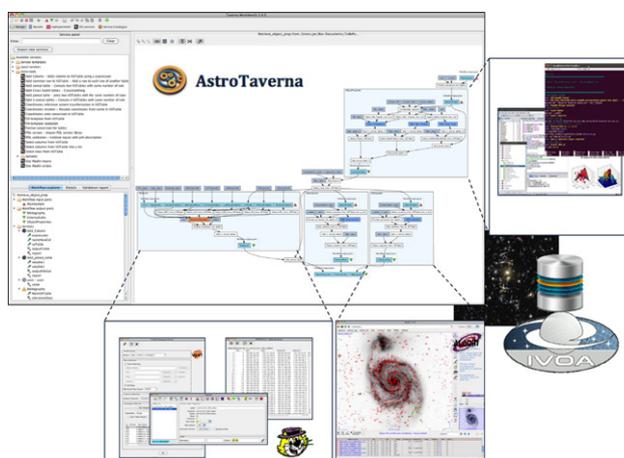
ABSTRACT

Despite the long tradition of publishing digital datasets in Astronomy, and the existence of a rich network of services providing astronomical datasets in standardized interoperable 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 Workbench 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.

© 2014 Elsevier B.V. All rights reserved.

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

* Corresponding authors.
   E-mail addresses: jer@iaa.es (J.E. Ruiz), jgarrido@iaa.es (J. Garrido), jdsant@iaa.es (J.D. Santander-Vela), sse@iaa.es (S. Sánchez-Expósito), lourdes@iaa.es (L. Verdes-Montenegro).
   [1] Tel.: +34 958 230 618.

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[2] (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) provides a way to transparently access and extract scientific knowledge from multiple archives of astronomical data, enabling standardized services discovery and access to interoperable data. It consists of a network of web-distributed data services which provides free access to astronomical catalogs, images and spectrum data, that combined with tools for local visualization and analysis (VO Apps) allow astronomers to graphically explore the different datasets, all following IVOA recommendations.

VO Apps are any kind of local software able to access the elements of the VO data network, following IVOA protocols, and sometimes permitting discovery of services available in the VO Registry (Benson and Walton, 2009) in order to provide the users with their datasets of interest. Current VO Apps are mainly of two-kinds: general purpose software for accessing a particular kind of data (catalogs, images, or spectra); and custom-built programs and scripts created by astronomers in order to perform their scientific research. Both mechanisms are typically used together for a science use case, with the user performing each step with different tools. This means that, in order to register the fully procedural protocol so it can be reproduced and possibly encapsulated in a more complex process, the user needs to manually capture the dependencies between steps, which is currently accomplished by writing scripts driving the execution of other tools or scripts.

Digital scientific workflows (Gil et al., 2007; Gil and Rey, 2008) allow capturing the actions performed in accessing distributed data, their 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[3] (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 the myExperiment[4] (Goble et al., 2010) public digital library of workflows. Moreover, its plugin architecture allows loose integration and open-source developments of plugins, decoupled from core developments of Taverna Workbench.

In this paper we present AstroTaverna, a plugin that simplifies the use of VO services from within the Taverna Workbench 2.x,

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 Section 2 we provide a very brief overview of workflow-related initiatives we are aware of in Astronomy. In Section 3 we describe the Taverna Workbench and the main VO architectural elements taken into account in the implementation of AstroTaverna. In Section 4 we provide a description of the functional drivers that led the developments of AstroTaverna as well as its main features, illustrating them all with an example use case. In Section 5 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 Section 6 the main conclusions of the exposed work are presented.

2. Related work

There have been several attempts of bringing workflow technologies to the astronomy field, and in particular based on the Taverna workflow system. The ESO Sampo[5] 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[6] subproject (Hook et al., 2009), using the Kepler engine (Altintas et al., 2004) instead. Recently, the HELIO[7] 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.

Other workflow-related software and initiatives are the Triana[8] (Taylor et al., 2007) and Pegasus[9] (Deelman et al., 2005) workflow environments; in particular the Montage[10] image mosaicing software provides specific modules for its implementation in Pegasus system. On a different angle, projects like Astro-WISE,[11] IceCore (Maisala et al., 2012) and CyberSKA[12] (Kiddle et al., 2011) have provided web interfaces for the building and execution of workflows based on proprietary services running in their server infrastructure. The ER-Flow[13] project aims to create a workflow user community across Europe in different fields of scientific research. It provides support for communities working in Astronomy and Heliophysics, with workflows developed in a variety of workflow systems, engines and languages, that are later implemented in their own cyber-infrastructure through the SHIWA Portal.[14]

In general, workflow-related approaches in Astronomy deal with very specialized proprietary data reduction pipelines, which tend to produce scientifically exploitable data, still distant from the *ready to publish* scientific insight. There has been little effort, to our knowledge, to integrate Virtual Observatory publicly accessible data and services into the workflow working methodology. In this context, the VO-France Workflow working group (Schaaff et al., 2008) aims to provide use cases to implement them as workflows with VO-enabled workflow tools, and we may also find JLOW plugin—a graphical tool that allows the design and execution of

---

2 http://www.ivoa.net/.

3 http://www.taverna.org.uk/.

4 http://www.myexperiment.org/.

---

5 http://www.eso.org/sci/software/sampo/.

6 http://www.eso.org/sci/software/sampo/reflex/.

7 http://www.helio-vo.eu/.

8 http://www.trianacode.org/.

9 http://pegasus.isi.edu/.

10 http://montage.ipac.caltech.edu/.

11 http://www.astro-wise.org/.

12 http://www.cyberska.org/.

13 http://www.erflow.eu/.

14 https://shiwa-portal2.cpc.wmin.ac.uk/liferay-portal-6.1.0/.

workflows based on Aladin (Bonnarel et al., 2000) tasks available at the Aladin plugin repository.[15]

The AstroGrid[16] project undertook 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 and Walton, 2009; Walton and González-Solares, 2009) which integrated VO Apps in the source code of the tool. Unfortunately the final result was not publicly released, though the code has been re-used in the development of a plugin for Taverna 2 workflow engine in the frame of the VAMDC[17] (Walton et al., 2011) project. This plugin allows the building and execution of workflows based on very specific services running in the project infrastructure, its use remaining again inside the scope of its own project.

3. Opening Taverna to the Virtual Observatory

3.1. 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 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 understanding is specially needed in collaborative science, where sharing executable recipes is vital 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 the workflow designer to add 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 unresolved 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.

3.2. The Virtual Observatory architecture

The Virtual Observatory infrastructure provides the community electronic access to numerous sources of information. Users consume transparently the VO-technologies by using astronomy VO-compliant local software – VO App in Fig. 1 – or using a web browser to access a VO-enabled web portal. In order to access either locally archived or remote files, they must only be aware of the different applications of their interest, and that there exist interoperable image, spectra, and catalog servers, transparently accessed from their toolset. They must also be aware that some applications can send messages and data between them, sending for instance data to be analyzed to some applications, and sending the received results to other applications for plotting.

All the elements of the architecture outside the local user environment – 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 these 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 has been designed to provide the means for easy discovery and query of VO data access services. In addition, functionality for the manipulation and visualization of VOTables has also been 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 (ConeSearch 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 the more used *simple data access protocols*[18] (ConeSearch Protocol, Simple Image Access Protocol, Simple Spectral Access Protocol). These services accept parameterized inputs encoded in one URI that is used as the entry point to query the services. Data returned are always in the VOTable format. Science data are embedded in the VOTable, or linked from data access fields.

- VO Apps: TOPCAT (Taylor, 2011b) and Aladin are 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 of these capabilities may be fully integrated in Taverna Workbench as local services performing very specific tasks based on the STIL API and STILTS Java framework (Taylor, 2006, 2011a), or executing scripts and macros in the

---

[15] http://aladin.u-strasbg.fr/java/nph-aladin.pl?frame=plugRep.
[16] http://www.astrogrid.org/.
[17] http://www.vamdc.eu/.

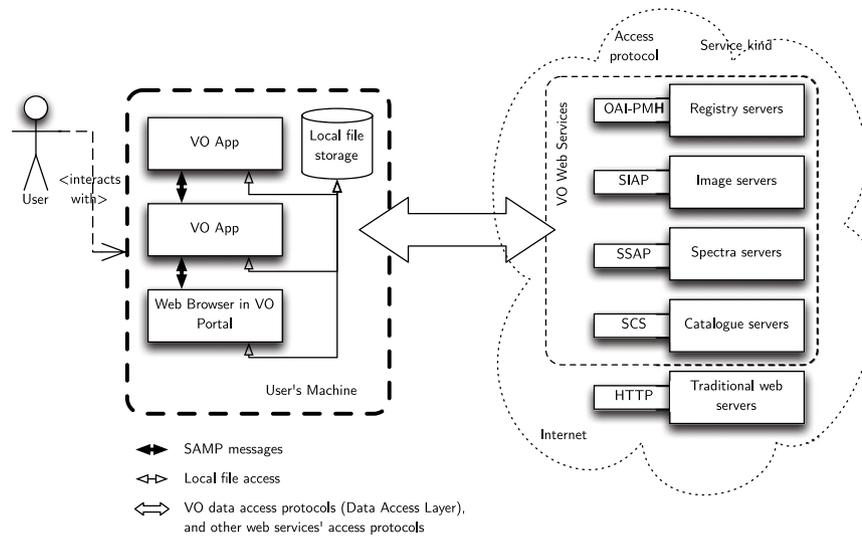[18] http://www.ivoa.net/documents/.

Fig. 1. Virtual Observatory usage and architecture.

case of Aladin. 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 like AstroTaverna plugin to send intermediate data and/or results to locally installed VO Apps for enhanced data inspection.

4. AstroTaverna design

4.1. Functional drivers

The AstroTaverna plugin targets users with no special technical knowledge of the underlying applications involved in the workflow execution process. It has been conceived to be easy to install and allow users to design, create and fully understand workflows without commissioning specialists or hiring software engineers. It provides the users with the means to design, build and execute VO-services-based workflows through the Taverna Workbench.

AstroTaverna aims to provide universal access to tabular representations of the kind of data issued by most of the VO 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 multi-band 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 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 a nice-to-have feature, but already accomplished by many scientists through different scripting languages and environments with no help of workflow tools. The main goal of AstroTaverna is improving readability and enabling reproducibility of digital VO recipes, capturing and registering provenance information of the whole experimental protocol otherwise lost in procedural steps manually performed in GUIs of VO Apps.

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

- *Data discovery, gathering and aggregation.*

  AstroTaverna aims to 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 tables 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 have considered 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 scripts (e.g. Python, bash, etc.) or local command line software.

- *VO software integration and data inspection.*

  In order to benefit from widely used and well-known functionalities of established VO Apps TOPCAT and Aladin, we decided to integrate some of their specific tasks. Moreover, in order to achieve a full understanding of final and intermediate data a proper rendering of the VOTable format was needed. Sharing these data with local VO Apps in a seamless way would allow each user to individually perform enhanced data inspection at any moment, which does not affect the reproducibility of the workflow and the final results.
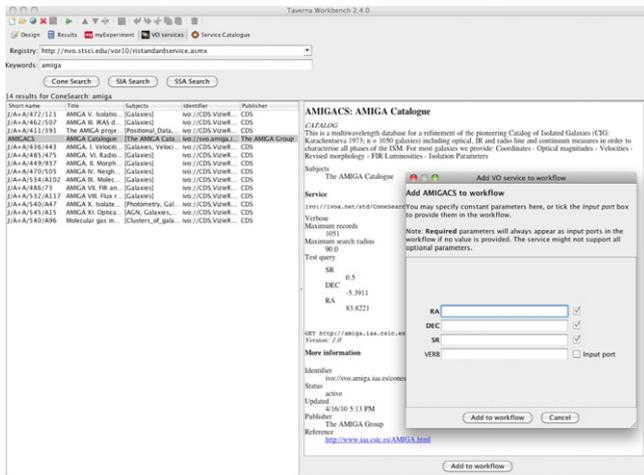
Fig. 2. AstroTaverna user interface for discovery of VO data access services. The list of services found is displayed on the left panel; the description and input parameters of a selected service are shown on the right panel and floating window.

Fig. 3. Workflow design in Taverna Workbench, showing *Astro tools* services on the upper left panel and a workflow for coordinate unit conversion on the *Design* perspective.
Source: Adapted from http://www.myexperiment.org/workflows/3515.

4.2. Data discovery, gathering and aggregation

AstroTaverna provides support for discovery of VO data access services, with the addition of a new *VO Service* tab to the 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 App.

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. Alternatively, the parameter can be supplied from a previous step in the workflow. 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 services 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.

4.3. 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 the 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 whic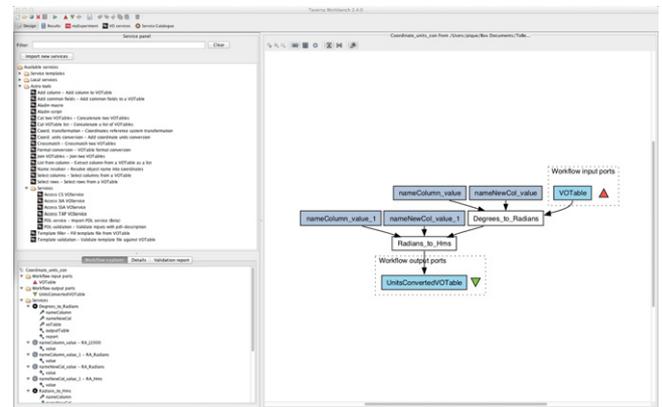h might be near a given position, with a prescribed tolerance) and table joins on common keys are also common actions that need to be covered in data manipulation basic capabilities.

AstroTaverna provides a collection of built-in data manipulation services, listed as Astro tools[19] within Taverna's services panel.

- *VOTable format conversion*
- *Concatenate two VOTables*
- *Concatenate a list of VOTables*
- *Join two VOTables*
- *Add common fields to a VOTable*
- *Crossmatch two VOTables*
- *Select columns from a VOTable*
- *Extract column from a VOTable as a list*
- *Select rows from a VOTable*

These services, shown in Fig. 3, can be dragged and added to the workflow diagram in the central design panel. They include table row filtering by Boolean algebraic expressions on field values and column subsets. The expressions for row filtering and column selection follow the STILTS syntax.[20,21]

4.4. Data transformation

On the same basis, AstroTaverna also provides a 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,[22] as well as the conversion of coordinates between different units. These are the following:

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

---

[19] http://amiga.iaa.es/p/312-astrotools.htm.
[20] http://www.star.bristol.ac.uk/~mbt/stilts/sun256/jel.html.
[21] http://www.star.bristol.ac.uk/~mbt/stilts/sun256/colid-list.html.
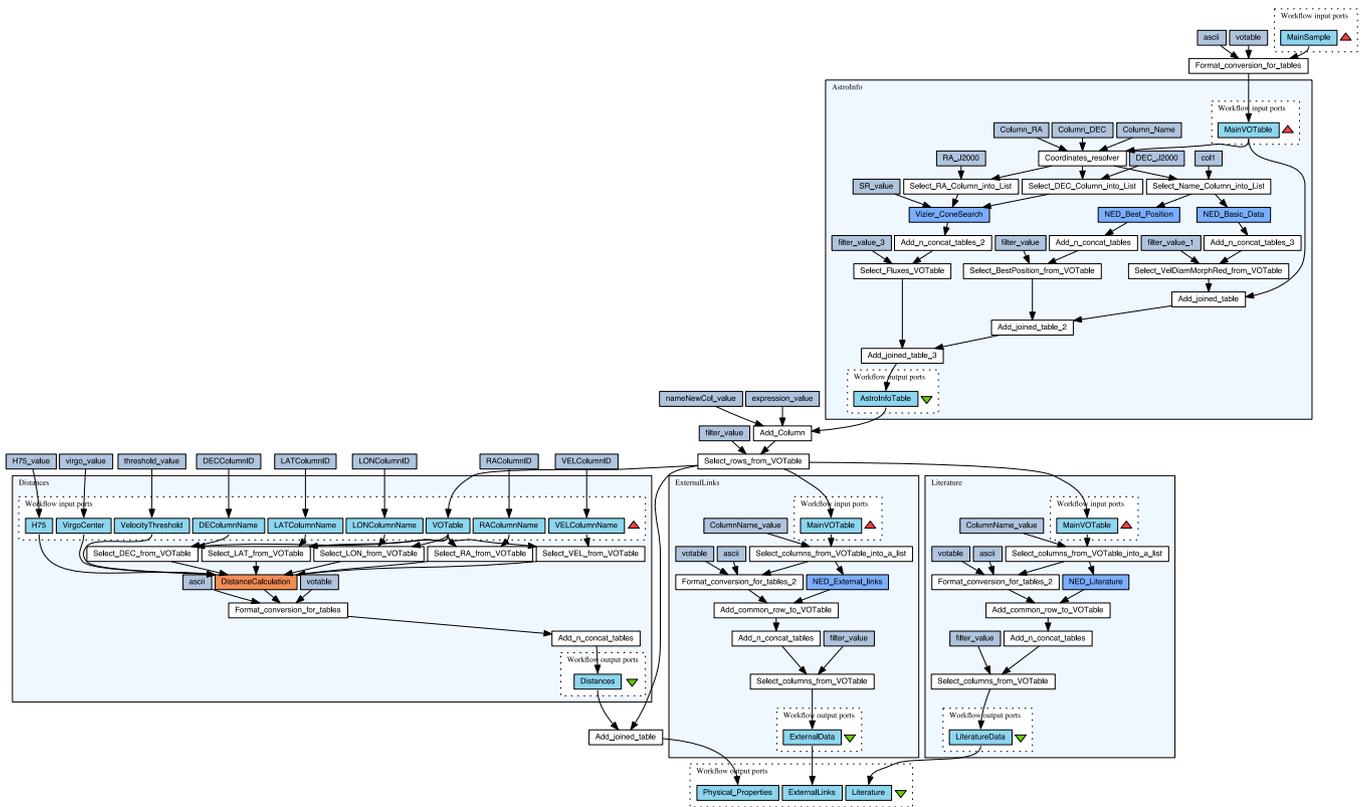[22] http://cdsweb.u-strasbg.fr/doc/sesame.htx.

Fig. 4. Diagram of AstroTaverna workflow for the example use case described in Section 4.6.

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

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

More complex data transformations may be possible using the *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 scripts (e.g. Python, bash, etc.) commonly used in astronomy research.

4.5. 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 fields metadata, in a spreadsheet-like form, as shown in Fig. 5. 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 other VO Apps, giving the user the possibility to perform post-workflow data inspection and analysis using more specific local software.

The Aladin VO App is bundled into the AstroTaverna plugin, which allows the user to execute Aladin macros and scripts using data extracted from VOTables provided by VO services. The Aladin services may be configured with the GUI option enabled, which triggers the visual automated execution of actions on Aladin VO App 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 and planes saved on the local disk during the executions of macros and/or scripts, according to the commands present on the scripts.

4.6. 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 assume 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 occasions in order to provide up-to-date information when needed to prepare a new target list, and possibly 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. 4, and it is publicly accessible[23] through the 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 the first step it queries, using the name of each galaxy as input, two different data access web services provided by
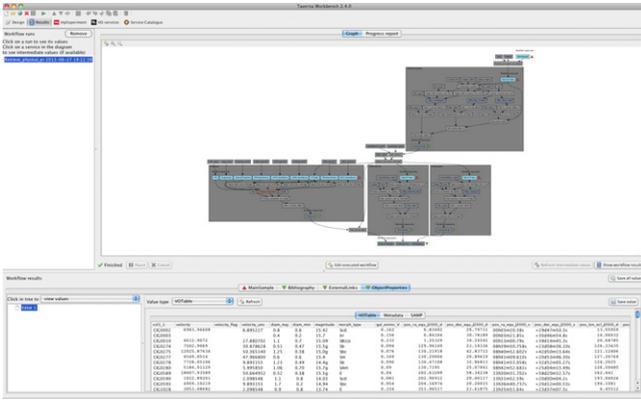

Fig. 5. Execution of AstroTaverna workflow and rendering of the tabular results for the example use case described in Section 4.6.

NASA/IPAC Extragalactic Database (NED)[24] archive (Mazzarella and 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/s.
- 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 *VOTable format conversion* AstroTaverna service for making ASCII to VOTable conversion, as well as format conversion for datavalues using STILTS expressions on the *Add column to a VOTable* 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 the Ks-band.
- Error in the flux density in the Ks-band.
- Log of the luminosity in the Ks-band.
- Error in the log of luminosity in the 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 the *Tool* service that is provided by the 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.

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.

5. 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 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.

Because of the exposed capabilities to improve readability and reproducibility of automated tasks, we think workflows could nicely complement the scarce narratives produced in the classic documentation process. The Taverna Workbench allows detailed inspection of the different workflow elements, like free text descriptions provided in the services of the workflow and input/output values for any intermediate parameters involved in one particular execution of the workflow or any other previous one.

Transitioning the experimental protocol from manually interacting with graphical user interfaces 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 Virtual Observatory. This is especially relevant with automated execution of

---

[24] http://ned.ipac.caltech.edu/.

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 example, which expedites training and avoids reinvention.

In this sense, a collection of 29 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,[25] in order to provide potential users of AstroTaverna a set of small workflows carefully designed to perform very specific actions. These kind of workflows can be considered as workflow components,[26] designed to be mixed and nested into larger workflows to perform a particular analysis.

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 others 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 the 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 the Taverna Workbench.

The AstroTaverna plugin was presented in several IVOA Interop conferences in Sao Paulo, Heidelberg, and Madrid, 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[27] and through the IVOA general mailing list, together with several video tutorials. These videos may be found in the AMIGA Group website[28] together with general information and documentation (see footnote 19) for the AstroTaverna plugin, also present in Taverna Plugins[29] website and in the specific AstroTaverna website.[30] The source code has been published in Astrophysics Source Code Library (ASCL; see Garrido et al., 2013), and can be publicly accessed at the GitHub[31] open source hosting service. Additionally a support mailing list[32] (astrotaverna-users@iaa.es) has been created for the AstroTaverna community.

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. Moreover, AstroTaverna plugin and AstroTaverna Starter Pack have been included recently at the core of specific Astronomy edition[33] of the Command Line Tool and the Taverna Workbench 2.5 release, together with project-specific plugins from VAMDC and HELIO-VO projects. 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 1.10. 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 and execution of Table Access Protocol (TAP; Dowler et al., 2010), also to handle potential multiple endpoints for a VO service in the discovery process, as well as the possibility to execute AstroTaverna workflows in the Taverna Server. 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., 2014) has been approved as an IVOA Recomendation in order to deal with web services interoperability 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.

6. Conclusions

We have presented AstroTaverna, a plugin for Taverna Workbench 2.x 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 (see footnote 30). 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 (see footnote 25) in order to encourage self-learning by example and avoid reinvention.

Future work considers providing support for discovery and execution of TAP services, a client for PDL self-described services that will allow data input validation before service invocation, and the possibility to execute AstroTaverna workflows in the Taverna Server, as well as handling potential multiple endpoints for a VO service in the discovery process.

Acknowledgments

---

25 http://www.myexperiment.org/packs/420.
26 http://www.taverna.org.uk/developers/work-in-progress/components/.
27 http://www.ivoa.net/newsletter/009/.
28 http://amiga.iaa.es/p/290-astrotaverna.htm.
29 http://www.taverna.org.uk/documentation/taverna-plugins/taverna-2-x-plugins/.
30 http://wf4ever.github.io/astrotaverna/.
31 https://github.com/wf4ever/astrotaverna.
32 http://smtp.iaa.es/mailman/listinfo/astrotaverna-users.
33 http://www.taverna.org.uk/download/workbench/2-5/astronomy/.
34 http://www.wf4ever-project.org/.

Part of the improvements is being undertaken in the framework of CANUBE Project[35] CEI2013-P-14, an Open Science project granted by the Second Call for Proposals of the Bio-TIC Campus of International Excellence[36] of the University of Granada, in Spain.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ascom.2014.09.002.

References

Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludäscher, B., Mock, S., 2004. Kepler: an extensible system for design and execution of scientific workflows. In: 16th International Conference on Scientific and Statistical Database Management, pp. 423–424.

Belhajjame, K., Corcho, O., Garijo, D., Zhao, J., Missier, P., Newman, D.R., Palma, R., Bechhofer, S., García-Cuesta, E., Gómez-Pérez, J.M., Klyne, G., Page, K., Roos, M., Ruiz, J.E., Soiland-Reyes, S., Verdes-Montenegro, L., De Roure, D., Goble, C., 2012. Workflow-centric research objects: a first class citizen in the scholarly discourse. In: Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web, SePublica2012.

Benson, K.M., Walton, N.A., 2009. AstroGrid: taverna in the virtual observatory. Mem. Soc. Astron. Ital. 80, 574–577.

Bentley, R., Brooke, J., Csillaghy, A., Fellows, D., Blanc, A.L., Messerotti, M., Pérez-Suárez, D., Pierantoni, G., Soldati, M., 2013. Helio: Discovery and analysis of data in heliophysics. Future Gener. Comput. Syst. 29 (8), 2157–2168.

Bonnarel, F., Fernique, P., Bienaymé, D., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M., Bartlett, J.G., 2000. The ALADIN interactive sky atlas A reference tool for identification of astronomical sources. Astron. Astrophys. Suppl. Ser. 143 (1), 33–40.

Deelman, E., Singh, G., Su, M.-H., Blythe, J., Gil, Y., Kesselman, C., Mehta, G., Vahi, K., Berriman, G.B., Good, J., Laity, A., Jacob, J.C., Katz, D.S., 2005. Pegasus: a framework for mapping complex scientific workflows onto distributed systems. Sci. Program. 13 (3), 219–237.

Dowler, P., Rixon, G., Tody, D., Andrews, K., Good, J., Hanisch, R., Lemson, G., Mcglynn, T., Noddle, K., Ortiz, I., Osuna, P., Plante, R., Salgado, J., Stebe, A., 2010. IVOA Recommendation: Table Access Protocol v1.0. Tech. Rep.

Garrido, J., Soiland-Reyes, S., Ruiz, J.E., Sánchez, S., 2013. AstroTaverna: tool for scientific workflows in astronomy. In: Astrophysics Source Code Library, record ASCL: 1307.007.

Gil, Y., Deelman, E., Ellisman, M., Fahringer, T., Fox, G., Goble, C., Livny, M., Moreau, L., Myers, J., 2007. Examining the challenges of scientific workflows. IEEE Comput. 40 (12), 24–32.

Gil, Y., Rey, M., 2008. From data to knowledge to discoveries: scientific workflows and artificial intelligence. Inform. Sci. 16 (4), 1–25.

Goble, C.A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., De Roure, D., 2010. myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Res. 38 (May), 677–682.

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

Hook, R., Ullgrén, M., Romaniello, M., Maisala, S., Oittinen, T., Solin, O., Savolainen, V., Järveläinen, P., 2009. ESO Reflex: a graphical workflow engine for data reduction. Mem. Soc. Astron. Ital. 80, 578–583.

Kiddle, C., Taylor, A.R., Pigat, D., Eymere, O., Rosolowsky, E., Kaspi, V., Willis, A.G., 2011. CyberSKA: an on-line collaborative portal for data-intensive radio astronomy. In: ACM Workshop on Gateway Computing Environments, pp. 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. IceCore: a web portal for workflow execution. In: Astronomical Data Analysis Software and Systems ASP XXI. In: ASP Conference Series, vol. 461. p. 99.

Mazzarella, J.M., Team, N.E.D., 2008. Web-based tools—NED VO services. In: The National Virtual Observatory Book, vol. 382. (Chapter 18).

Ochsenbein, F., Williams, R., Davenhall, C., Durand, D., Fernique, P., Giaretta, D., Hanisch, R., McGlynn, T., Szalay, A., Taylor, M., Wicenc, A., 2009. IVOA Recommendation: VOTable Format Definition v1.2. Tech. Rep. August 2004.

Ruiz, J.E., 2012. Digital science: reproducibility and visibility in astronomy. In: 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. Workflow in astronomy: the VO France workflow working group 1 experience. In: Astronomical Data Analysis Software and Systems XVII. In: ASP Conference Series, vol. 394. pp. 77–80.

Taylor, M.B., 2006. STILTS—a package for command-line processing of tabular data. In: Astronomical Data Analysis Software and Systems XV. In: ASP Conference Series, vol. 351. pp. 666–669.

Taylor, M., 2011a. Starlink tables infrastructure library tool set. In: Astrophysics Source Code Library, record ASCL: 1105.001.

Taylor, M., 2011b. TOPCAT: tool for operations on catalogues and tables. In: Astrophysics Source Code Library, record ASCL: 1101.010.

Taylor, M., Boch, T., Fitzpatrick, M., Allan, A., Paioro, L., Taylor, J., Tody, D., 2012. IVOA Recommendation: SAMP—Simple Application Messaging Protocol v1.3. Tech. Rep.

Taylor, I., Shields, M., Wang, I., Harrison, A., 2007. The Triana workflow environment: architecture and applications. In: Workflows for e-Science. pp. 320–339.

Tody, D., Dolensky, M., McDowell, J., Bonnarel, F., Budavari, T., Busko, I., Micol, A., Osuna, P., Salgado, J., Skoda, P., Thompson, R., Valdes, F., 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.A., Dubernet, M.L., Mason, N.J., Piskunov, N., Rixon, G.T., 2011. VAMDC: the virtual atomic and molecular data center. In: Astronomical Data Analysis Software and Systems XX. In: ASP Conference Series, vol. 442. pp. 89–92.

Walton, N., González-Solares, E., 2009. Astrogrid and the virtual observatory. Lect. Notes Phys. 791, 81–113. (Jets From Young Stars V).

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., Williams, A., Withers, D., Owen, S., Soiland-Reyes, S., Dunlop, I., Nenadic, A., Fisher, P., Bhagat, J., Belhajjame, K., Bacall, F., Hardisty, A., Nieva de la Hidalga, A., Balcazar Vargas, M.P., Sufi, S., Goble, C., 2013. The taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. Nucleic Acids Res. 41 (W1), W557–W561.

Zwolf, C., Harrison, P., Le Petit, F., Garrido, J., Ruiz, J.E., 2014. IVOA Recommendation: PDL Parameter Description Language. Tech. Rep.

---

[35] http://amiga.iaa.es/p/294-open-science-canube.htm.

[36] http://biotic.ugr.es/en/.