



Publication Year	2016
Acceptance in OA	2020-05-22T16:25:55Z
Title	Milky Way Analysis through a Science Gateway: Workflows and Resource Monitoring
Authors	SCIACCA, Eva, VITELLO, FABIO ROBERTO, BECCIANI, Ugo, COSTA, Alessandro, Hajnal, Akos, Kacsuk, Peter, MOLINARI, Sergio, DI GIORGIO, Anna Maria, SCHISANO, EUGENIO, LIU, Scige' John, ELIA, Davide Quintino, CAVUOTI, STEFANO, RICCIO, GIUSEPPE, BRESCIA, Massimo
Handle	http://hdl.handle.net/20.500.12386/25109
Serie	CEUR WORKSHOP PROCEEDINGS
Volume	1871

Milky Way Analysis through a Science Gateway: Workflows and Resource Monitoring

Eva Sciacca, Fabio Vitello, Ugo Becciani, Alessandro Costa * Akos Hajnal, Peter Kacsuk [†]
Sergio Molinari, Anna Maria di Giorgio, Eugenio Schisano, Scige John Liu, Davide Elia [‡]
Stefano Cavuoti, Giuseppe Riccio, Massimo Brescia[§]

* INAF-Osservatorio Astrofisico di Catania, Italy

[†]Laboratory of Parallel and Distributed Systems SZTAKI, Budapest, Hungary

[‡]INAF-Istituto di Astrofisica e Planetologia Spaziali, Roma, Italy

[§]INAF-Osservatorio Astronomico di Capodimonte, Napoli, Italy

Email: eva.sciacca@oact.inaf.it

Abstract—This paper presents the latest developments on the VIALACTEA Science Gateway in the context of the FP7 VIALACTEA project. This science gateway operates as a central workbench for the VIALACTEA community in order to allow astronomers to process the new-generation (from Infrared to Radio) surveys of the Galactic Plane to build and deliver a quantitative 3D model of our Milky Way Galaxy. The final model will be used as a template for external galaxies to study star formation across the cosmic time. The adopted AGILE software development process allowed to fulfill the community needs in terms of required workflows and underlying resources monitoring. The scientific requirements arose during the process highlighted the needs for easy parameter setting, fully embarrassingly parallel computations and large-scale input dataset processing. Therefore the science gateway based on the WS-PGRADE/gUSE framework has been able to fulfill the requirements mainly exploiting the parameter sweep paradigm and parallel jobs execution of the workflow management system. Moving from the development to the production environment an efficient resource monitoring system has been implemented to easily analyse and debug sources of failure due to workflows computations. The results of the resource monitoring system are exploitable not only for IT experts administrators and workflow developers but also for the final users of the gateway. The affiliation to the STARnet Gateway Federation ensures the sustainability of the presented products after the end of the project, allowing the usage of VIALACTEA Science Gateway to all the stakeholders and not only to the community members.

Keywords—Workflow Systems; Science Gateways; Collaborative Environments; Astrophysics; DCIs; Milky Way Analysis; Infrastructure Tests; Monitoring

I. INTRODUCTION

The Milky Way is a complex ecosystem where a cyclical transformation process brings diffuse baryonic matter into dense unstable condensations to form stars, that produce radiant energy for billions of years before releasing chemically enriched material back into the InterStellar Medium in their final stages of evolution. Although considerable progress has been made in the last two decades in the understanding of the evolution of isolated dense molecular clumps toward the onset of gravitational collapse and the formation of stars and planetary systems, a lot remains still hidden.

The aim of the European FP7 VIALACTEA project is to exploit the combination of all new-generation surveys of the

Galactic Plane to build and deliver a galaxy scale predictive model for star formation of the Milky Way. This model will be used as a template for external galaxies and studies of star formation across the cosmic time. Usually the essential steps necessary to unveil the inner workings of the galaxy as a star formation engine (such as the extraction of dust compact condensations or robust reconstruction of the spectral energy distribution of objects in star-forming regions) are often carried out manually by the astronomer, and necessarily over a limited number of galactic sources or very restricted regions.

Therefore scientists required new technological solutions able to deal with the growing data size and quantity coming from new-generation surveys (from Infrared to Radio wavelength). Moving to the Big Data era, allows to overcome the current challenges pushing the envelope of the current state of the art both from technological and scientific point of view. The extraction of the meaningful informations contained in the available data required an entirely new approach (the new paradigm of “data driven scientific discovery”) which resulted in a novel framework based on advanced visual analytics techniques, data mining methodologies, machine learning paradigms and Virtual Observatory (VO) based data representation and retrieval standards. All the underlying pipelines required by this framework (e.g. knowledge base catalogue creation, maps making for visual analytic) are available through the VIALACTEA Science Gateway.

The gateway (described in Section III) is based on the WS-PGRADE/gUSE [1] portal framework which provides several ready-to-use functionalities off-the-shelf. It allows development of scientific workflows composed of nodes corresponding to almost any kind of application in a convenient graphical user interface. Workflows can be executed in parallel in a wide set of Distributed Computing Infrastructures such as grids, clusters, supercomputers, and clouds. It enables sharing, importing and exporting workflows and managing credentials (and robot certificates), or gathering workflow execution statistics. Beyond these features, the portal is extensible, in fact WS-PGRADE/gUSE offers a number of interfaces to add new applications and portlets to its base capabilities.

This paper presents the latest developments on the

VIALACTEA Science Gateway including the workflows designed for the community and the resource monitoring system. The workflows (see Section III-B) are mainly focused for performing intensive computations: map making, i.e. the formation of sky images from the instruments data; data mining to obtain band-merged catalogues relating sources with associated counterparts at different wavelengths; filamentary structure detection and extraction from images.

Due to the diverse variety of software and computing capabilities required by the workflows, a novel monitoring system has been developed within the gateway to test the status of the whole system. The monitoring covers different levels of tests (see Section IV) checking the gateway interoperability with the computing infrastructures and the workflow submission and execution processes. These tests are performed periodically and the resulting reports are published on the gateway so that also final users are aware of any failure of the system avoiding waste of time in debugging their work. Furthermore e-mail alerts are sent on any failure to the infrastructures administrators to promptly fix the problem.

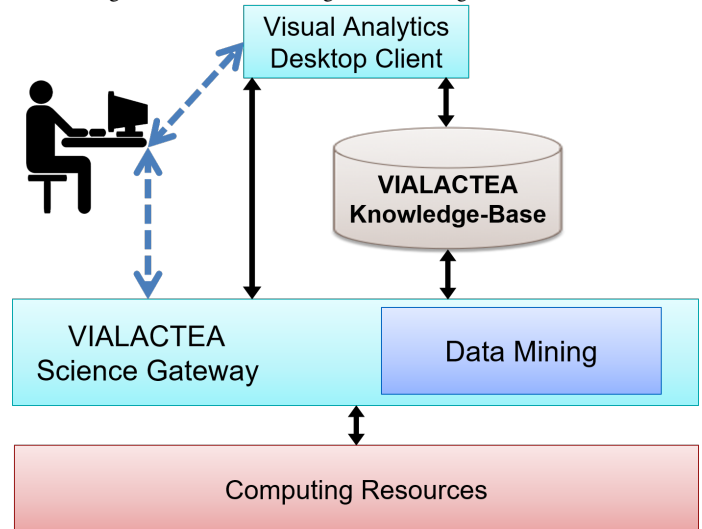
II. VIALACTEA REQUIREMENTS AND TECHNOLOGICAL ARCHITECTURE

In order to deliver a model of our galaxy with quantitative star formation laws, it is necessary to reveal and analyse throughout the galaxy the dense filamentary clouds where star-forming clumps are found. These clumps are found in very different environments and in different evolutionary stages and their properties are characterized through detailed modelling of their Spectral Energy Distribution. Their exact location is determined using the most up to date distance estimators and all these pieces need to be assembled to get a new view of our Galaxy.

The Galactic distribution of Star Formation Rate (stellar mass produced per unit time) and Efficiency (stellar mass produced per unit mass of available dense gas) can be quantitatively related to the variety of physical agents that drive star formation in the Galaxy. The timely exploitation of the huge amount of data available requires new technological solutions able to overcome the current challenges pushing the envelope of the current state of the art both from technological and scientific point of view. Therefore it has been implemented a novel system based on advanced visual analytics techniques, data mining pipelines, VO-based standards and science gateway technologies. The implemented framework can be seen as an integrated workspace where the Visual Analytics Desktop Client, the Science Gateway embedding the Data Mining pipelines and the VIALACTEA Knowledge Base can be employed both as independent actors or as interacting components (see Figure 1).

The VIALACTEA Knowledge Base (VLKB) includes a combination of storage facilities, a Relational Data Base (RDB) server and web services on top of them. It allows easier searches and cross correlations between data and currently contains: 2D surveys, catalogue sources and related band merged information; structural informations such as filament

Fig. 1. VIALACTEA integrated technological framework.



structures or bubbles; and Radio Datacubes with search and cutout services. Data-mining and machine-learning pipelines are embedded within the Science Gateway as workflows and employed to carry out building of Spectral Energy Distributions, distance estimate and Evolutionary classification of hundreds of thousands of star forming objects on the Galactic Plane. All these produced results are then ingested to the VLKB. The Visual Analytics tool allows the interaction with the VIALACTEA data and to carry out complex tasks for multi-criteria data/metadata queries on the VLKB, subsample selection and further analysis processed over the science gateway, or real-time control of data fitting to theoretical models.

Due to the cross-domain scientists involved in the community (computer scientists, technologists and astronomers) AGILE software development approach has been adopted. This approach in fact promotes adaptive planning, evolutionary development, early delivery, and continuous improvement, and it encourages rapid and flexible response to change. Cross-disciplinary face-to-face meetings have been organized to promote an iterative, incremental and evolutionary framework based on several cycles of requirements and feedback sessions.

The science gateway is exploited by the scientists to configure and run the VIALACTEA workflows implementing the pipelines developed by the community (see Section III-B). Furthermore the science gateway allows the Visual Analytic tool to submit workflows through the usage of Remote API [2]. This API provides also methods for checking the workflow's status, and for downloading the outputs. The scientists required easy parameter setting, fully embarrassingly parallel computations and large-scale input dataset processing. Therefore the science gateway, based on the WS-PGRADE/gUSE framework¹, has been able to fulfill the requirements (see

¹gUSE Web Page <http://guse.hu>

Section III-A).

III. VIALACTEA SCIENCE GATEWAY

The usage of the science gateway provides user-friendliness (intuitive user interface), efficiency (fast response time even for complex user requests), scalability (fast response time even for a large number of simultaneous user requests), robustness (keeps working under any circumstances and recovers gracefully from exceptions) and extensibility (easy to be extended with new interfaces and functionalities).

The VIALACTEA Science Gateway² is based on a customized version of WS-PGRADE/gUSE version 3.7 and is affiliated with the STARnet Gateway Federation [3]. STARnet envisages sharing a set of services for authentication, a common and distributed computing infrastructure, data archives and workflow repositories. Each STARnet gateway provides access to specialized applications via customized workflows. The affiliation to the STARnet Gateway Federation also ensures the sustainability of the whole products after the end of the project. This will allow the usage of the science gateway by all the future possible stakeholders and not only by the VIALACTEA community.

A. *gUSE Key Features*

This section outlines some characteristics of gUSE that have been identified as key feature for the VIALACTEA science gateway.

Parallelism: gUSE supports four levels of parallelism on workflow execution. The lowest level, or node-level parallelism, is where the application itself is prepared to utilize the benefits of multicore processors (multithread) or cluster systems (e.g. parallel execution using MPI). Besides this option, gUSE supports parallel execution of different jobs placed at different parallel branches of the workflow graph as the most intuitive and simple concurrent execution (branch-level parallelism). A third level of parallelism covers the situation when one algorithm should be executed on a large parameter field, generally called parameter study or parameter sweep (PS) execution. The highest level of parallelism is where the execution of the same workflow is done in parallel. In fact, such a parallel execution of workflows can also be initiated by the user submitting the same workflow with different configurations.

DCI and Storage access: gUSE can access various DCIs using the DCI Bridge [4] and different data storages via the Data Avenue [5]. It provides flexible and versatile access to all the important applied DCIs within Europe supporting a wide range of different middleware types (Clusters, Grids, Supercomputers, Desktop grids, Clouds). The file transfer among various storages and workflow nodes can be handled automatically/transparently using Data Avenue service.

Workflow Management System: the workflow creation and parameter setting can be performed from the web interface by importing the workflows from the repository or by creating

their new one using a web-based graph editor. The graph editor has been recently improved (see [6]) replacing the three-stage process of creating, configuring and submitting workflows to a single stage process allowing workflow creation, instant configuration and submission within a single portlet.

B. *The VIALACTEA Workflows*

The available VIALACTEA workflows are mainly designed for: map making, i.e. the production of high quality images from the raw instruments data; data mining to obtain band-merged catalogues, whose entries consist of sources with associated counterparts at different wavelengths; filamentary structure detection and extraction from images. Specifically the following workflows have been identified.

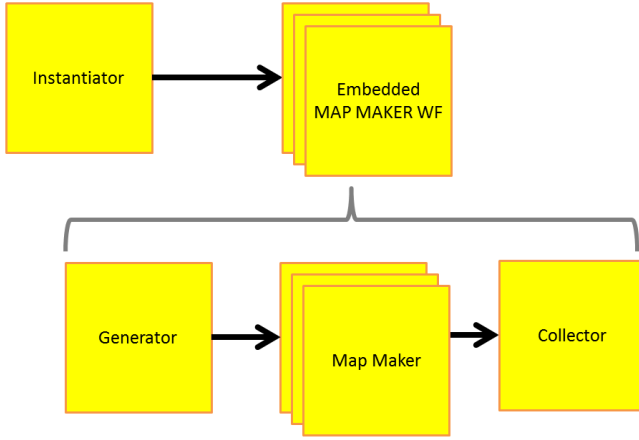
MOSAIC: The MOSAIC workflow employs Unimap [7] as map maker software to produce high quality mosaic images from the raw instruments data of the infrared imaging photometers onboard of the ESA Herschel satellite. The employed applications are coded in IDL, Matlab and Bash scripting language. The workflow has been implemented as a parameter sweep workflow [8] embedding a parameter sweep map maker workflow. This allows a full parallelization of the processes to be executed. See Figure 2 for the schema of the workflow. The inputs specify the tiles to be processed (longitude and wavelength) and the parameters of the Unimap application. The workflow automatically imports the required data from the Herschel infrared Galactic Plane Survey (Hi-GAL) [9], [10]. The Instantiator job prepares the input tiles to be processed by the map maker embedded workflow, which computes each tile separately. The Generator job prepares the sub-tiles to be processed by the Map Maker job (Unimap). Finally, the output is given by the Collector job of the map maker embedded workflow and contains the maps in FITS (Flexible Image Transport System) file format.

PPMAP: The PPMAP workflow executes a Point Process MAPPING (PPMAP) [11] which is a Bayesian procedure that uses images of dust continuum emission at multiple wavelengths to produce resolution-enhanced image cubes of differential column density as a function of dust temperature and position. The employed applications are coded in Fortran90, IDL and Bash scripting language. As for the MOSAIC workflow, this workflow has been implemented using the parameter sweep submission schema as shown in Figure 2. The inputs specify the tiles to be processed and the parameters (one for each input tile) to be sent to the PPMAP application. The workflow automatically imports the required data from the Hi-GAL Survey. The output is given by the collector job of the workflow and contains the maps as FITS file formats.

Q-FULLTREE: The Q-FULLTREE workflow performs compact source identification through band-merging. The application is based on the positional cross-match among sources at different wavelengths. It is configured as a multi threaded job splitting the single-band input catalogues into a user-chosen number of small sub-catalogues, with a user-selected percentage of overlapping entries in order to avoid the loss of merged sequences related to borderline entries. FT-Recap

²VIALACTEA Science Gateway: <http://via-lactea-sg00.iaps.inaf.it:8080>

Fig. 2. MOSAIC Workflow.



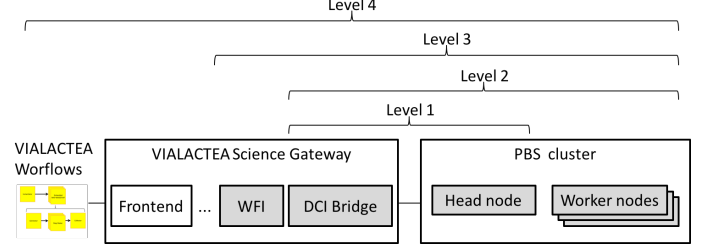
(FullTree-Recap), a post-processing application associated to the band-merging workflow, is submitted to re-organize the output of the Q-FULLTREE in order to fulfil the Spectral Energy Distribution visualization expectations of the Visual Analytic Desktop Client. The employed applications are coded in Python and make internal use of the STILTS public library [12]. The inputs of the workflow are: a TAR archive containing the sources at different wavelength in CSV format and two text files specifying the setup and the configuration for the application.

Filamentary Structure Detection: The workflow is designed to perform filament extraction. The underlying application [13] identifies filamentary-like extended structures on astronomical images and determines their morphological and physical parameters. The workflow is developed as a three steps processing, one for feature detection, one for filament extraction and a final one for filtering of artifacts and creation of the final catalogue. All these applications are implemented in IDL. The first step performs the detection of candidates through advanced image analysis techniques based on mapping of eigenvalues of the local Hessian Matrix computed from the input map. The second step analyses the region of interests with the support of morphological operators that decompose the initial binary mask into simpler units. Finally the third step analyses the candidate list and filters out low elongated structures and possible artifacts building up the final candidate filamentary catalogue.

IV. RESOURCE MONITORING

Continuous monitoring of the operational status (“health”) of the underlying distributed computing infrastructures (DCIs) connected to the science gateway is of high importance, as they serve as the actual platform performing scientific calculations for the VIALACTEA Science Gateway, which functionality was yet missing from the base WS-PGRADE/gUSE portal

Fig. 3. Monitoring levels and the covered components.



framework. DCIs are though much more powerful but somewhat less reliable than standalone desktop PCs due to their inherent complexity (remote execution, data staging, environment changes, etc); moreover, the running time of scientific applications varies widely, can take minutes but even days or weeks to complete. Any outage of the underlying DCIs might cause breaking the flow of calculations, and in spite of built-in failover mechanisms (re-submit jobs on failure), on error, it can be very difficult to localize without having information about proper behavior of the computing infrastructure. Debugging complex workflows can be a very tedious and time consuming, which require re-running the application several times, with slight modifications, added logging. Furthermore, sometimes these errors are not even repeatable; temporary blackouts or when worker nodes run out of disk space may prevent the job scheduling and submission system even to record a notice about the actual cause of the failure.

Using a DCI monitoring system such as the one designed and implemented in the VIALACTEA Science Gateway, workflow developers can make sure that all the related DCIs operate normally prior to starting long running calculations; and also on error, by revising historical data of monitoring records, be sure in that the error is not caused by the failure of the underlying infrastructure, respectively. System administrators can also benefit from using resource monitoring, as they can quickly overview all the systems under their supervision, and due to e-mail alerting option, they can react to the corrupted behavior as soon as possible. Due to historical data trustiness of the computing resources can be assessed; potential improvements, measures to prevent the same failures in the future can be initiated. At the moment, resource monitoring restricts to monitoring DCIs of type Portable Batch System (PBS) used in the context of the VIALACTEA project. To help in better identifying the location of errors different levels of monitoring activities had been designed, which run periodically for all connected DCIs.

Four levels of resource monitoring had been implemented: Level 1 (PBS cluster infrastructure head node monitoring); Level 2 (PBS cluster worker nodes environment monitoring); Level 3 (Portal PBS cluster interoperability monitoring); and Level 4 (VIALACTEA, domain-specific, workflow operational monitoring).

The lowest level, level 1 (called “PBS head nodes”), checks that the DCI is indeed accessible from the portal (head node

responds to ping, successful SSH connection can be established) and all the essential middleware commands (`qsub`, `qstat`, `pbsnodes`, etc.) operate as expected. Level 2 tests (called “PBS worker nodes”) scan through all worker nodes available in the DCI (they are all candidates of potential job execution) and checks, one-by-one, that the expected execution environment is available, such as enough disk space, necessary runtime environments and libraries (Java, IDL, Matlab, Python, etc.). Level 3 tests involve testing both the portal’s and the DCI’s functionality (“Portal-PBS interoperability”), which executes a probe workflow, composed of a single job. Level 4 tests (“Vialactea base workflows”) submit pre-created, domain-specific workflows having characteristics and requirements similar to other applications used in the customized portal, though with parameters resulting in less load compared to other full-fledged computations. Note that once tests on a certain level fail, tests at higher levels will fail too; thus the lowest level of failures help in locating the source of the problem as precisely as possible.

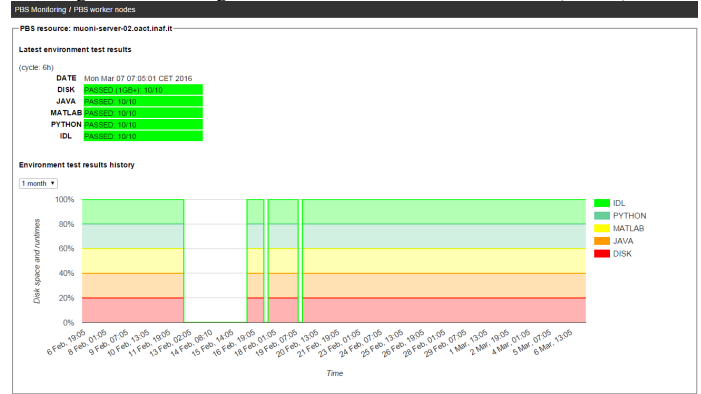
Figure 3 illustrates which main components of the system are covered by the different levels of tests (the higher the level, the more components are covered by the monitoring test).

For the different levels of monitoring activities different frequency can be specified, i.e. how often and at what time they are to be executed. For example, in the current VIALACTEA portal, level 1 tests are set to be executed every 3 hours; level 4 tests run once a day, at midnight. It makes possible to tune and schedule the load caused by monitoring system itself to avoid performance degradation might be experienced during normal use of the portal as much as possible. All results are recorded, so operational status of each resource can be traced back for the specified period of time (30 days, by default). Also, for error events e-mail alerting can be requested for any number of e-mail addresses (primarily, system administrator is notified).

Monitoring data can be viewed by any user of the portal; changes to settings are however allowed by the portal administrators only. Monitoring results are summarized and visualized in the form of tables and charts on a web interface. Figure 4 shows level 2 resource monitoring results. The table (on the top of figure 4) shows the latest results and the frequency of these tests (6 hours). PASSED 10/10 means all tests (free disk space, Java, Matlab, Python, IDL) had passed on all 10 worker nodes available in cluster “muoni-server-02.oact.inaf.it”. The chart (on the bottom of figure 4) shows test results of the last 30 days, indicating outage on dates 12–16, 17, and 19 February; the DCI worked properly at other times.

As a result of the introduced monitoring service, in practice, gateway users could now verify, prior to running their workflows, whether the infrastructure of their choice is indeed available, operational, responsive, and, if not, they still had the possibility to choose another DCI. Also, on error, they could check past records to clarify whether the failure was due to infrastructure problems thus avoiding costly debugging. Administrators were always notified about DCI outages on time, so they could fix issues as soon as possible and inform

Fig. 4. Monitoring results of “PBS worker nodes” tests (level 2).



portal users about the incidents and expected time of recovery, respectively. Unwanted side effects of software configuration changes in clusters were also detected by the monitoring tool automatically (e.g., corrupted Matlab, IDL paths in worker nodes). Finally, due to historical data, it turned out that the capacity of one PBS cluster was insufficient; response times were at an acceptable level at weekends only, at very low load. The given infrastructure was then re-installed with more processors and more worker nodes to fulfill users’ needs.

V. RELATED WORK

To deal with the data deluge that the Astrophysics community is facing, different science gateways and workflow technologies are being exploited. Apart from the WSGRADE/gUSE framework that has been extensively employed by the authors, see e.g. [14], [15], [16], different approaches have been followed to allow the end users to easily interact with the applications ported on the DCIs.

In [17], the authors present an approach based on the Taverna Workbench³ [18] and the Astrotaverna plugin⁴ [19] to perform kinematical modelling of galaxies as an example of analysis task required by the SKA project (which aims to build an instrument that will be the worlds largest radio interferometer, able to reach data rates in the exa-scale). The Apache Airavata⁵ [20] environment on XSEDE⁶ resources have been used in [21] to produce multiple synthetic sky surveys of galaxies and large-scale structure in support of Dark Energy Survey analysis. The underlying technologies described in those works are well suited to be ported into a science gateway such as the VIALACTEA one, but requires time and extra IT effort for coding web services (as wrapper) on top of each application of interest of the astronomers.

The Kepler⁷ scientific workflow system [22] has been employed in [23] to implement automatic data reduction pipelines. This approach could have been very useful within

³Taverna web site: <http://www.taverna.org.uk>

⁴AstroTaverna plugin: <http://amiga.iaa.es/p/290-astrotaverna.htm>

⁵Apache Airavata web site: <http://airavata.apache.org>

⁶XSEDE web site: <https://www.xsede.org>

⁷Kepler project web site: <https://kepler-project.org>

the VIALACTEA project but again it requires IT effort to build the required Kepler actors for each application.

Finally, to our knowledge, none of the above solutions included a resource monitoring system able to check the status of the overall gateway interacting components, including the required runtime, as required by the VIALACTEA community.

There exist several resource monitoring tools available, such as Ganglia[24], Nagios⁸, Zabbix⁹, Prometheus¹⁰ to mention a few, shipped with numerous probes out-of-the-box to monitor typical host and service metrics such as availability, CPU, network utilization, memory, disk space usage, service checks, etc. Beyond that they require individual installation, administration, and considerable expertise to manage, they seemed not to be easily adaptable in our special case, as worker nodes, behind head nodes in PBS clusters, are inaccessible from outside (they reside in private network); their monitoring was possible only through submitting dedicated PBS jobs. Also, verifying the results of workflow execution, which can only be done using the "remote API" of the portal, seemed to be difficult to realize using such tools. Our implementation, and its integration into the portal has other advantages as well: it uses the same monitoring source (host of the gateway) and mechanisms (software libraries, SSH connections, PBS commands) as the portal, so it tests resources from an identical environment. Nevertheless, we connected our tool to Zabbix to record workflow execution time metric, and we used Zabbix triggers, notifications, and chart visualization.

VI. CONCLUSIONS AND OUTLOOK

In this paper we have introduced a new framework that allow astronomers to process the new-generation surveys of the Galactic Plane to build and deliver a quantitative model of Milky Way Galaxy. The presented science gateway operates as a central workbench for the VIALACTEA community allowing to deal with the growing data size and quantity coming from new-generation surveys. The extraction of the meaningful informations contained in the available data required an entirely new approach (the new paradigm of data driven scientific discovery) which resulted in a novel framework based on advanced visual analytics techniques, data mining methodologies, machine learning paradigms and Virtual Observatory based data representation and retrieval standards.

The focus of the presented workflow applications is on map making, i.e. the formation of sky images from the instruments data; data mining to obtain band-merged catalogues relating galactic sources with associated counterparts at different wavelengths; and filamentary structure detection and extraction from sky images. Furthermore we have highlighted how the usage of WS-PGRADE/gUSE framework have been able to fulfil the project requirements thanks to its key features: user-friendliness, efficiency, scalability, robustness and extensibility.

⁸Nagios: <http://www.nagios.org>

⁹Zabbix: <http://www.zabbix.com>

¹⁰Prometheus: <https://prometheus.io>

This paper also described a novel resource surveillance component integrated into WS-PGRADE/gUSE portal capable of checking operational status of the employed computational infrastructures based on Portable Batch Systems (PBS). The monitoring covers different levels of tests checking the gateway interoperability with the computing infrastructures and the workflow submission and execution processes. These tests are performed periodically and the resulting reports are published on the gateway so that also final users are aware of any failure of the system avoiding waste of time in debugging their work.

Amongst the things deserving further studies is the evaluation of MetaBrokering service of WS-PGRADE/gUSE which is capable of distributing and balancing the load among different distributed computing infrastructures. This will be exploited for parameter sweep jobs, such as the map making computations, avoiding excessive load of one resource with respect to other having higher capacity.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 607380 (VIALACTEA).

REFERENCES

- [1] P. Kacsuk, Z. Farkas, M. Kozlovsky, G. Hermann, A. Balasko, K. Karóczkai, and I. Marton, "Ws-pgrade/guse generic dc gateway framework for a large variety of user communities," *Journal of Grid Computing*, vol. 10, no. 4, pp. 601–630, 2012.
- [2] A. Balasko, Z. Farkas, and P. Kacsuk, "Building science gateways by utilizing the generic WS-PGRADE/gUSE workflow system," *Computer Science*, vol. 14, no. 2, pp. 307–325, 2013.
- [3] U. Becciani, E. Sciacca, A. Costa, P. Massimino, F. Vitello, S. Cassisi, A. Pietrinferni, G. Castelli, C. Knapic, R. Smareglia *et al.*, "Creating gateway alliances using ws-pgrade/guse," in *Science Gateways for Distributed Computing Infrastructures*. Springer, 2014, pp. 255–270.
- [4] M. Kozlovsky, K. Karóczkai, I. Márton, P. Kacsuk, and T. Gottdank, "Dci bridge: Executing ws-pgrade workflows in distributed computing infrastructures," in *Science Gateways for Distributed Computing Infrastructures*. Springer, 2014, pp. 51–67.
- [5] Á. Hajnal, Z. Farkas, P. Kacsuk, and T. Pintér, "Remote storage resource management in ws-pgrade/guse," in *Science Gateways for Distributed Computing Infrastructures*. Springer, 2014, pp. 69–81.
- [6] G. A. McGilvary, M. Atkinson, S. Gesing, A. Aguilera, R. Grunzke, and E. Sciacca, "Enhanced usability of managing workflows in an industrial data gateway," in *e-Science (e-Science), 2015 IEEE 11th International Conference on*. IEEE, 2015, pp. 495–502.
- [7] L. Piazza, L. Calzoletti, F. Faustini, M. Pestalozzi, S. Pezzuto, D. Elia, A. di Giorgio, and S. Molinari, "unimap: a generalized least-squares map maker for herschel data," *Monthly Notices of the Royal Astronomical Society*, vol. 447, no. 2, pp. 1471–1483, 2015.
- [8] P. Kacsuk, K. Karóczkai, G. Hermann, G. Sipos, and J. Kovacs, "WS-PGRADE: Supporting parameter sweep applications in workflows," in *Workflows in Support of Large-Scale Science, 2008. WORKS 2008. Third Workshop on*. Ieee, 2008, pp. 1–10.
- [9] S. Molinari, B. Swinyard, J. Bally, M. Barlow, J.-P. Bernard, P. Martin, T. Moore, A. Noriega-Crespo, R. Plume, L. Testi *et al.*, "Hi-gal: The herschel infrared galactic plane survey," *Publications of the Astronomical Society of the Pacific*, vol. 122, no. 889, p. 314, 2010.
- [10] D. Elia, S. Molinari, Y. Fukui, E. Schisano, L. Olmi, M. Venezianni, T. Hayakawa, M. Pestalozzi, N. Schneider, M. Benedettini *et al.*, "The first hi-gal observations of the outer galaxy: A look at star formation in the third galactic quadrant in the longitude range 216. 5 225. 5," *The Astrophysical Journal*, vol. 772, no. 1, p. 45, 2013.

- [11] K. Marsh, A. Whitworth, and O. Lomax, "Temperature as a third dimension in column-density mapping of dusty astrophysical structures associated with star formation," *Monthly Notices of the Royal Astronomical Society*, vol. 454, no. 4, pp. 4282–4292, 2015.
- [12] M. Taylor, "Stilts-a package for command-line processing of tabular data," in *Astronomical Data Analysis Software and Systems XV*, vol. 351, 2006, p. 666.
- [13] E. Schisano, K. Rygl, S. Molinari, G. Busquet, D. Elia, M. Pestalozzi, D. Polychroni, N. Billot, S. Carey, R. Paladini *et al.*, "The identification of filaments on far-infrared and submillimeter images: Morphology, physical conditions and relation with star formation of filamentary structure," *The Astrophysical Journal*, vol. 791, no. 1, p. 27, 2014.
- [14] U. Becciani, E. Sciacca, A. Costa, P. Massimino, C. Pistagna, S. Riggi, F. Vitello, C. Petta, M. Bandieramonte, and M. Krokos, "Science gateway technologies for the astrophysics community," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 2, pp. 306–327, 2015.
- [15] E. Sciacca, M. Bandieramonte, U. Becciani, A. Costa, M. Krokos, P. Massimino, C. Petta, C. Pistagna, S. Riggi, and F. Vitello, "Visivo science gateway: a collaborative environment for the astrophysics community," in *5th International Workshop on Science Gateways, IWSG 2013*. CEUR Workshop Proceedings, 2013.
- [16] A. Costa, P. Massimino, M. Bandieramonte, U. Becciani, M. Krokos, C. Pistagna, S. Riggi, E. Sciacca, and F. Vitello, "An innovative science gateway for the cherenkov telescope array," *Journal of Grid Computing*, vol. 13, no. 4, pp. 547–559, 2015.
- [17] S. Sanchez Exposito, P. Martin, J. E. Ruiz, L. Verdes-Montenegro, J. Garrido, R. S. Pardell, A. Ruiz Falco, and R. Badia, "Web services as building blocks for science gateways in astrophysics," in *Science Gateways (IWSG), 2015 7th International Workshop on*. IEEE, 2015, pp. 80–84.
- [18] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher *et al.*, "The taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud," *Nucleic acids research*, p. gkt328, 2013.
- [19] J. Ruiz, J. Garrido, J. Santander-Vela, S. Sánchez-Expósito, and L. Verdes-Montenegro, "Astrotavernabuilding workflows with virtual observatory services," *Astronomy and Computing*, vol. 7, pp. 3–11, 2014.
- [20] M. E. Pierce, S. Marru, L. Gunathilake, D. K. Wijeratne, R. Singh, C. Wimalasena, S. Ratnayaka, and S. Pamidighantam, "Apache airavata: design and directions of a science gateway framework," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 16, pp. 4282–4291, 2015.
- [21] B. Erickson, R. Singh, A. E. Evrard, M. R. Becker, M. T. Busha, A. V. Kravtsov, S. Marru, M. Pierce, and R. H. Wechsler, "Enabling dark energy survey science analysis with simulations on xsede resources," in *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*. ACM, 2013, p. 16.
- [22] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao, "Scientific workflow management and the kepler system," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 10, pp. 1039–1065, 2006.
- [23] W. Freudling, M. Romaniello, D. Bramich, P. Ballester, V. Forchi, C. García-Daból, S. Moehler, and M. Neeser, "Automated data reduction workflows for astronomy-the eso reflex environment," *Astronomy & Astrophysics*, vol. 559, p. A96, 2013.
- [24] M. L. Massie, B. N. Chun, and D. E. Culler, "The ganglia distributed monitoring system: design, implementation, and experience," *Parallel Computing*, vol. 30, no. 7, pp. 817–840, 2004.