VisIVO Workflow-Oriented Science Gateway for Astrophysical Visualization

Eva Sciacca *, Marilena Bandieramonte *‡, Ugo Becciani *, Alessandro Costa *, Mel Krokos †, Piero Massimino *, Catia Petta ‡, Costantino Pistagna *, Simone Riggi * and Fabio Vitello *
* INAF-Osservatorio Astrofisico di Catania, Italy
† University of Portsmouth, United Kingdom
‡ Dipartimento di Fisica e Astronomia, Università di Catania, Italy

Abstract—Nowadays visualization-based knowledge discovery can play an important role in astrophysics. Collaborative visualization can enable multiple users to share visualization experiences, e.g. by interacting simultaneously with astrophysical datasets giving feedback on what other participants are doing/seeing. Further, workflow-driven applications allow reproduction of specific visualization results, a challenging task as selecting suitable visualization parameters may not be a straightforward process. This paper presents VisIVO Science Gateway, a web-based workflow-enabled framework integrating large-scale, multidimensional datasets and applications for visualization and data filtering on Distributed Computing Infrastructures (DCIs). Advanced users are able to create, change, invoke, and monitor workflows while standard users are provided with easy-to-use customised web interfaces hiding all technical aspects of the visualization algorithms and DCI configurations.

Keywords—Collaborative Tools; Workflow Systems; Grid Computing; Scientific Gateway; Scientific Visualization

I. INTRODUCTION

Several TBs are often generated by modern cosmological simulations while the highest-resolution codes executed on supercomputing infrastructures can easily produce several PBs. Further, datasets coming from large-scale astrophysical observations are already stored in archives (e.g. LSST [1], LOFAR [2] and SDSS [3]). Such large data volumes pose significant challenges in terms of data analysis, storage and access; a critical step forward in understanding, interpreting and verifying their intrinsic characteristics can be achieved through visualization. A comprehensive review of tools and algorithms for visualization in astronomy can be found in [4].

Gaining a good insight into large-scale multi-dimensional datasets typically requires very sophisticated statistical and data analysis algorithms. Often, a number of data exploration and visualization tools are employed for visual discoveries in order to identify regions of interest (e.g. see [5]). As these involve large-scale distributed datasets and computing infrastructures they require collaboration among scientists. Recently scientific workflows have gained popularity in modeling and organizing such processes [6] explicitly by specifying dependencies between underlying tasks and then orchestrating the required distributed resources appropriately.

The VisIVO Science Gateway [7] is wrapped around WS-PGRADE [8], a highly-flexible interface for the grid User Support Environment (gUSE) [9]. The gateway provides access to VisIVO Server tools [10], enabling execution of a comprehensive collection of modules for processing and visualization of astrophysical datasets on DCIs. A number of customized workflows have been configured by default, e.g. to allow local or remote upload of datasets and creation of scientific movies. These workflows are provided with specific user interface portlets to enable easy parameter setting for standard users while hiding the complexity of the underlying system and infrastructures. Advanced users are able to adapt default workflows in the context of their specific problems. This modularity, achieved by subdividing the workflow into a number of elementary tasks, ensures re-usability of subtasks and provides high flexibility to end users.

Our gateway is connected to data repositories and workflows via a mobile application for astrophysical visualization called VisIVO Mobile. The mobile application employs user accounts from the gateway and offers a platform for astrophysical communities to share results and experiences of analysis and exploration of their datasets. The remainder of the paper is organized as follows: sections II and III are brief introductions to WS-PGRADE/gUSE and VisIVO respectively. Section IV describes the undertaken VisIVO Science Gateway development and illustrates the main portlets. Section V describes details of the mobile application for smartphones and tablet devices. Section VI reports on a use case; sections VII and VIII discuss related works and conclusions.

II. WS-PGRADE/gUSE

WS-PGRADE/gUSE is a collaborative and community oriented application development environment that allows developers and end users to share layered and parameter sweep enabled workflows [11], workflow graphs, workflow templates, and ready-to-run workflow applications via a repository. There is full integration in the portal framework Liferay [12] which is highly customizable thanks to the
adoption of portlet technology defined in the Java Specification Request 168 and 286 [13], and compatible to modern web applications.

gUSE is a virtualization environment providing a large number of high-level Distributed Computing Infrastructure (DCI) services achieving interoperation among classical services and desktop grids, clouds, clusters, web services and user communities in a scalable way. gUSE is realised as a set of web services to dynamically provide services in DCIs.

WS-PGRADE uses client APIs of gUSE services to translate user requests into sequences of gUSE specific web service calls. WS-PGRADE hides conveniently all relevant communication protocols and sequences behind portlets and its user interface can be accessed through standard web browsers.

To convert the generic WS-PGRADE portal instance into a research domain specific science gateway a particular portal extension called Application Specific Module (ASM) provides an easy-to-use solution [14]. ASM consists of two components: a) a script-layer used for installing different parts of the module (e.g. data tables, services, portlets) and b) the Java-layer used as the Application Programming Interface (API) during the development of the web-interface providing programmatically most of the functionalities of gUSE.

III. VisIVO Tools

VisIVO is an integrated suite of tools and services for effective visual discovery within large-scale astrophysical datasets. VisIVO is realised as: a) Visivo Desktop, a stand alone application for interactive visualization running on standard PCs and b) Visivo Server, a grid-enabled high performance visualization platform. Users of either realization can obtain meaningful visualizations rapidly while preserving full and intuitive control of relevant visualization parameters.

This section focuses on Visivo Server which can be installed on any web server with a database repository and consists of Visivo Importer, Visivo Filter and Visivo Viewer. To create customized views of 3D renderings from astrophysical data tables a Visivo Importer is first utilized to convert user datasets into Visivo Binary Tables (VBTs). Then, a Visivo Viewer is invoked to display these renderings.

A VBT is a highly-efficient data representation realized through a header file containing all necessary metadata and a raw data file storing actual data values. Visivo Importer converts user-supplied datasets into VBTs without imposing any limits on sizes or dimensionality. It supports conversion from several popular formats such as: ASCII and CSV, VOTables or FITS Tables. Visivo Filter is a collection of data processing modules to modify a VBT or to create a new VBT from existing VBTs. The filters support a range of operations such as scalar distribution, mathematical operations or selections of regions. For example decimation or randomization are typically employed for constructing reduced VBTs to perform a sub-sampling in order to fit them in the available RAM memory. Visivo Viewer is based on the Visualization ToolKit [15] library for multidimensional visualization. It creates 3D images of datasets rendering points, volumes and isosurfaces within a bounding box used for representing the coordinate system employed. Moreover there is support for customized look up tables and visualizations using a variety of glyphs, such as cubes, spheres or cones.

The Visivo Library was developed to port Visivo Server on gLite middleware [16]. It allows a job running on a grid node to produce a set of images or movies directly using Visivo with its internal data arrays without the need to produce intermediate files. This is particularly important when running on the grid, where the user may want to have a quick look on the results during data production phases. The images in this way can be produced directly in the grid catalogue, while the user code is running in a system that cannot be directly accessed by the user (e.g. in worker nodes).

IV. Visivo Science Gateway

The existing Visivo Web [17] has been integrated within the WS-PGRADE generic gateway to offer new, easily accessible opportunities not only to specialized users, e.g. astrophysical researchers, but also to the wider public, e.g. high-school education or innovative citizen science activities.

The Visivo Science Gateway is designed as a workflow enabled grid portal that is wrapped around WS-PGRADE providing visualization and data management services to the scientific community by means of an easy-to-use graphical environment for accessing the full functionality of Visivo Server. Complex workflows can be created and executed on a variety of infrastructures (e.g. clouds, desktop and service grids, supercomputers) to obtain comprehensive exploration and analysis of large-scale astrophysical datasets.

The gateway offers role-based authorization modules and supports login with user name and password. Currently a number of main roles are implemented for access as follows: guests, standard and advanced users, and finally administrators.

Guests are not provided with an account for the science gateway. However, they can obtain information about Visivo and the features offered by the gateway. A new account can be created by a standard user or by exploiting credentials from already created Facebook or OpenId accounts.

Standard users are allowed to select default workflows, to become familiar with the tools and specific workflows offered, by using simplified user interfaces. They are also
allowed to change input and relevant parameters, to invoke and monitor workflows.

Advanced users can access additional features to create and change workflows and to set configurations of grid infrastructures.

Finally, administrators are additionally enabled to manage all credentials, individual users, organizations, and user communities.

Standard users can upload and manage their datasets through portlets without any knowledge about the (conveniently hidden) underlying grid-infrastructure and middleware. By using interactive widgets users can construct customized renderings, or store data analysis and visualization results for future reference. Their datasets are managed internally through a relational database preserving their metadata and maintaining data consistency. Figure 1 shows the main portlets of the Gateway connecting to VisIVO Importer, Filters and Viewer services.

![Figure 1. Main VisIVO Gateway portlets.](image)

Services are independent of each other and the navigation between them is possible by a menu located on the upper side of the portal web page and from the interaction with the Data Management portlet (see Section IV-B).

Both remote and local datasets can be uploaded - i.e. residing on a remote URL or locally on a user’s PC. For remote files the user must specify URL and optionally a username and password for authentication. Depending upon the size of the datasets under consideration, remote uploads could last a long period. To resolve this situation VisIVO Gateway allows an off-line mode by means of a workflow submission so that users can issue upload commands and then simply close their current session - a follow up e-mail typically gives notification once the uploading operation is completed. The workflow employed for remote importing is shown in Figure 2.

![Figure 2. Remote VisIVO Importer Workflow.](image)

The nodes of the graph, represented by yellow boxes, are called jobs and denote the activities related to individual computations (e.g. VisIVO Importer or GnuPlot). Jobs communicate with other jobs within the workflow through job-owned input and output ports. An output port (small grey boxes) of a job connected with an input port (small green boxes) of a different job is called channel; these are directed edges of the graph. A single port must be either an input, or an output port of a given job. The workflow allows to produce significant information for metadata exploration, e.g. statistics on data values, histogram calculation and plotting or a sample extraction of uploaded datasets. Such metadata is available through the Properties portlet and some can be modified by the user (e.g. renaming a VBT or the related fields).

Once the data file is uploaded a sequence of simple actions is required to rapidly obtain meaningful visualizations. Typically various VisIVO Filter operations are performed, and VisIVO Gateway automatically displays all applicable VisIVO Filter operations allowing input of the relevant parameters. Finally the VisIVO Viewer is employed for image display. A right click on any processed dataset in the Data Management portlet is used in conjunction with the View button to create user-prescribed VisIVO Viewer views.

A. Scientific Movies

VisIVO Gateway allows users to generate scientific movies. Scientific movies are useful not only to scientists to present and communicate their research results, but also to museums and science centres to introduce complex scientific concepts to a general public audience. Users can create a panoramic movie by moving a camera along a motion path of 360° in azimuth and +/- 90° in eleva-
tion within the domain of a dataset. Customized movies can be produced by intermediate snapshots specified as camera positions/orientations and the gateway generates a movie with a camera path containing these specified positions/orientations. Dynamic Movies can be created by interpolating several steps of a time evolution of a cosmological dataset. The user can browse a cosmological time evolution and choose two or more coherent datasets. The designed workflow will then produce the necessary number of intermediate VBTs by calculating particle positions and applying boundary conditions as necessary. This approach can be very useful, e.g. in revealing galaxy formation or large-scale structures such as galaxy clusters.

The creation of a movie represents a significant challenge for the underlying computational resources as often hundreds or thousands high quality images must be produced. For this reason Parameter Sweep (PS) workflows are employed. PS workflows are executed in distributed parallel computations. The gUSE/WS-PGRADE infrastructure supports special ports (generator and collector ports) which enable the PS elaboration of a workflow: a single set of input files containing more than one element associated by a port - or several input ports having this feature - may trigger the proper number of submissions of the associated job. As an example the panoramic movie is generated with the workflow shown in Figure 3, it generates four movies with different camera position paths on the generator port: from $0^\circ$ to $360^\circ$ azimuth rotation, from $0^\circ$ to $90^\circ$ elevation rotation, from $90^\circ$ to $-90^\circ$ elevation rotation and from $-90^\circ$ to $0^\circ$ elevation rotation. The generation of these four movies is executed in parallel and finally merged through a collector port.

![Figure 3. Panoramic Movie Workflow.](image)

The creation of a movie with 110 frames from 70 million particles of a cosmological N-body simulation takes 2,000 seconds using a node with 4 cores (AMD Opteron 2218 rev. F dual-core processors with a clock rate of 2.6 GHz); this time remains more or less the same for generation of up to four movies simultaneously. These results agree with the time to produce a single frame which is about 20 seconds. Overall computational times scale linearly with the data dimension.

A preliminary performance test on a Desktop Grid with 1500 Windows PCs [18] has been performed on a dataset containing a sub-box of an N-body simulation of the universe with 500 million particles. 2020 frames splitting the workload in 20 working units take a minimum time of 600 seconds mean time of 5000 seconds for each working unit.

### B. Data Management

A Data Management portlet (see Figure 4) allows users the recycling of their private staging area within the system for managing their datasets as well as images and movies produced from such datasets. The underlying data-model is implemented through a MySql RDBMS.

![Figure 4. Data Management portlet.](image)

The metadata associated with uploaded datasets as well as the produced images and movies are encapsulated by customized GenericFile objects. These objects contain attributes such as owner or creation time. The objects VisIVOTable and VisIVOVolume are defined by inheriting from GenericFile to represent VBTs. VisIVOTable objects introduce attributes such as endianity, number of fields, type and number of elements. VisIVOVolume objects introduce attributes such as number and size of cells for each dimension. Both VisIVOTable and VisIVOVolume contain one or more VisIVOFields including details and statistics on each field of the VBT. A detailed schema of the employed data base is shown in Figure 5.

### C. Implementation Details and Computing Infrastructures

The implemented portlets are developed with the Java Vaadin web Framework [19]. This open source framework has been employed to implement server side Java Servlet based web applications using the full power and flexibility of Java without taking care of the client side since it compiles the Java source code to JavaScript which can then be run on browsers.

The gUSE Application Specific Module API has been included to the portlets to be able to reuse the implemented workflows stored in the local repository of gUSE.
The VisIVO Science Gateway can access the Cometa Consortium grid (gLite v 3.1). This infrastructure is distributed in seven sites of Sicily. All sites have the same hardware and software configuration allowing high interoperability and realizing an homogeneous environment. The computing infrastructure is based on IBM Blade Centre each containing up to 14 IBM LS21 blades interconnected with the low latency Infiniband-4X network, to provide High Performance Computing (HPC) functionalities on the grid. There are currently about 2000 CPU cores and more than 200 TBs of disk storage space available on this HPC Sicilian e-Infrastructure.

D. Testing

The VisIVO Science Gateway is being currently tested under the ETICS system [20]. ETICS is a distributed software configuration, build and test system designed to fulfill the needs of improving the quality, reliability and interoperability of distributed software in general and grid software in particular. The computing service consists of a build and test job execution system based on the Metronome software [21] and an integrated set of web services and software engineering tools to design, maintain and test scenarios. The ETICS system allows taking into account complex dependencies among applications and middleware components and provides an environment to perform static and dynamic analysis of the software and execute deployment, system and interoperability tests.

Web testing is being performed on the gateway portlets. The purpose of these tests is to validate the local and the remote VisIVOImporter portlets on testable input files and input parameters and to validate the VisIVOFilter and VisIVOViewer portlets on testable imported datasets. The tests are considered as passed if the VisIVO Binary Table and the produced filtered VBTs and images are available in the Internal Data Management portlet and the user can download them and get information on the relative meta data. The employed testing tools include: Robot framework [22] (generic test automation environment) with Selenium [23] (web test automation tool) and Sikuli [24] (image based GUI test tool).

V. VisIVO MOBILE

The VisIVO Mobile application (see Figure 6) allows smartphone devices to exploit VisIVO Gateway functionalities to access large-scale astrophysical datasets residing on a server repository for analysis and visual discovery. Through interactive widgets, customized visualizations (images or movies) can be generated and stored on the remote server. The application notifies users when requested visualizations are available for retrieving on their smartphones and allows sharing of data, images and movies via e-mail or common social networks.

The current version of VisIVO Mobile is implemented in Objective-C optimized for the Apple iPhone, iPod and iPad, and, in the near future, it will be ported to other popular smartphone devices. End users can login with the same credentials as on the gateway and the application provides the password coding in SHA cryptography (employed by the Liferay environment) and queries the remote database to verify access credentials.

The configuration and submission of workflows residing on the VisIVO Gateway is performed by means of the gUSE Remote API. These API enables the interface to the core gUSE services without the WS-PGRADE user interface component. Thus, they allow running and managing scientific workflows by command line solutions consisting of curl based access wrapped in shell scripts. The API exposes usage of gUSE components through a simple web service interface, resulting in wide adaptability by a diverse set of tools and programming languages.
VI. Case Study: Particle Visualization

As a representative case study we show in this section the steps needed to visually analyse simulation data resulting from a scattering of cosmic radiation. The deflection of muonic particles present in the secondary cosmic radiation results from crossing high atomic number materials (such as uranium or other fissile materials). This technique can provide a significant improvement compared to the detection methods used so far based on X-ray scanners [25], in terms of capacity for identification and location of illicit material, even in the presence of screens designed to mask its existence. In this case visualization plays a crucial role in obtaining tomographic images of a cargo container.

We here consider a simulation of particles in a large muon tracker consisting of 4 planes, 6 metres long and 3 metres wide, for the inspection of a container carrying high atomic number material. The illicit material is shaped with the string "CT".

The data file containing the coordinates on the muon tracker planes is first uploaded to the gateway and filtered using the POCA (Point of Closest Approach) algorithm [26] to obtain the VBT containing the scattering deflection of cosmic radiations. The resulting VBT can be visualized using a point viewer as shown in the top image of Figure 7.

This dataset can be further processed through the VisIVO Filter portlet to perform a rows filtering based on a given threshold at lower bound. The resulting VBT is then converted into a volume using the deflection angle field distribution, employing the 3D Cloud-in-Cell (CIC) [27] smoothing algorithm, on an input defined regular mesh. The intermediate produced images are shown in steps 2 and 3 of Figure 7. Finally, a tomography can be performed on the produced volume VBT.

An advanced user can decide to build a customized workflow for this particular problem connecting the predefined building blocks as shown in Figure 8 and share it with the relevant scientific community.

VII. Related Works

For displaying 2D or simple 3D plots, astrophysicists use programs such as gnuplot, SuperMongo, or scripting languages such as Python, Matlab or IDL. VisIt [28] or ParaView [29] offer a combination of 2D and 3D plotting capabilities, realtime and offline analysis, scripting and graphical control.

VisIt has been provided with grid services for scientific collaborative visualization in UNICORE Grids [30]. ParaView has been extended to offer grid services [31] and a plugin has been developed to provide interactive remote visualization for collaborative environments based on video streams [32].

Nevertheless scientific visualization is a complex process involving several steps: filtering data, choosing a representation, choosing a desired level of interactivity, and customizing the manner in which the data is displayed. None of the above mentioned tools are provided with a science gateway to interface them with workflow services. Within
VisIVO Science gateway ready-to-use workflows can be downloaded, parametrized and executed under a controlled environment. The applications run on DCIs and end users can focus on their applications instead of devoting efforts in learning and managing the infrastructure they require to run their applications.

VIII. Conclusions

Traditionally the common practice among astronomers for data exploration tools was to employ small, individually created, autonomous applications. However this scenario is not applicable to modern large-scale datasets. The development of easy to use web interfaces for data analysis and visual discovery is instrumental in reaching out to the wider astrophysical community and providing effective usage of e-infrastructures.

A workflow-oriented gateway allows scientists to share their analysis workflows and identify best practices for investigating their datasets. More importantly, they can automate workflows for repeated analysis with changed parameters, which was a manual, slow and very error prone process in the past. This way scientists can focus on their core scientific discoveries as opposed to spending time in data analysis on inadequate resources.

VisIVO Gateway provides a web based portal for easily setting up, running, and evaluating visualizations in the astrophysics field for large-scale datasets exploiting DCIs resources. The gateway includes a data repository containing the results of produced images and movies on imported datasets, as well as repositories of fundamental workflows, which can be set up, used, improved and distributed by the users.

We presented several portlets running in a Liferay portal environment which enables standard and advanced users to set up simple and sophisticated VisIVO scenarios and submit them to associated grids. The results are processed and related meta data information are available into the portal as well as being accessible through mobile devices using a mobile application.

The end users do not necessarily need to know the parameter file details containing all set-up options or even be aware of the computing infrastructure operating behind the scenes.

The usage of WS-PGRADE enables end users to develop, share and reuse designed workflows. These might realize data processing jobs, data mining or advanced visualization scenarios for large-scale datasets from numerical simulations or real-world observations.

Through the shareable workflows and reusable portlets of the gateway, not just astrophysicists but also other scientific fields and citizen scientists can exploit the developed technologies.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Commission’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no 283481 (SCI-BUS). SCI-BUS (SCIfic Gateway Based User Support) aims to create a generic-purpose gateway technology as a toolset to provide seamless access to major computing, data and networking infrastructures and services in Europe including clusters, supercomputers, grids, desktop grids, academic and commercial clouds. Further information is available at http://www.sci-bus.eu/.

REFERENCES


