ELSEVIER

CrossMark

Full length article

An integrated visualization environment for the virtual observatory: Current status and future directions

E. Sciacca [a,*], U. Becciani [a], A. Costa [a], F. Vitello [a], P. Massimino [a], M. Bandieramonte [a,b], M. Krokos [c], S. Riggi [a], C. Pistagna [a], G. Taffoni [d]

[a] *INAF - Astrophysical Observatory of Catania, Via Santa Sofia 78, 95123, Catania, Italy*
[b] *Dipartimento di Fisica e Astronomia, Università degli Studi di Catania, Italy*
[c] *School of Creative Technologies, University of Portsmouth, United Kingdom*
[d] *INAF - Astronomical Observatory of Trieste, Via G.B. Tiepolo 11, 34143, Trieste, Italy*

ARTICLE INFO

ABSTRACT

Visual exploration and discovery applications are invaluable tools to provide prompt and intuitive insights into the intrinsic data characteristics of modern astronomy and astrophysics datasets. Due to the massively large and highly complex datasets, various technical challenges are involved to reach, e.g. interactivity, integration, navigation and collaboration. This paper describes a number of approaches to address these challenges, and focuses on the current status of VisIVO (Visualization Interface for the Virtual Observatory) concentrating on the provided tools ranging from a desktop application to a science gateway and a mobile application. We emphasize the latest developments made in the context of past and current international European funded projects and highlight planned future developments towards further integration within the framework of the Virtual Observatory.

1. Introduction

Modern astronomy and astrophysics produce massively large data volumes (in the order of petabytes) coming from observations or simulation codes executed on high performance supercomputers. Such data volumes pose significant challenges for storage, access and data analysis, leading to the development of a fourth data intensive science paradigm (Hey et al., 2009) (also known as eScience). A critical aspect in understanding, interpreting, and verifying the outcome of automated analysis and data mining processes is the visualization of the scientific results. Data visualization is a fundamental, enabling technology for knowledge discovery, and an important research field that covers a number of different topics such as: optical and radio imaging (Nita et al., 2011), simulation results (Labadens et al., 2012), multi-dimensional exploration of catalogues (Beaumont et al., 2013) and public outreach visuals (Kent, 2013).

Visual exploration of big datasets poses some critical challenges that must drive the development of a new generation of graphical software tools, specifically:

Interactivity The majority of existing astronomical analysis and processing solutions lack the ability to deal with datasets exceeding the local machine's memory capacity while visual exploration and discovery in complex, multi-dimensional datasets is more effective through real-time interaction although sizes may not fit the available memory (Jin et al., 2010);

Integration Most of the data analysis systems are implemented as a set of separate independent tasks that can interact and exchange information via stored files only. This will be a significant factor which delays or even prohibits day-to-day data analysis tasks over big data sizes. Visualization tools should be ideally fully integrated within the scientists' toolkit for seamless usage, abstracting from technical details freeing scientists to concentrate in doing science (e.g. see Goodman, 2009);

Navigation Some of the current data processing techniques depend on parameters tuning, which may not be easy to achieve with large data sizes due to processing power limitations. Adopted solutions should allow intuitive and sophisticated navigation among datasets by exploiting ubiquitous environments (e.g. see Greensky et al., 2008 and Keefe, 2010), such as tablets or motion controllers,

---

* Corresponding author.
E-mail address: eva.sciacca@oact.inaf.it (E. Sciacca).

offering new Human–Computer Interaction paradigms to better tune the processing parameters;

Collaboration It will no longer be an easy job to develop a simple script or program to deal with such data. These tasks usually require a deep knowledge of technicalities and programming experience which are typical of computer scientists rather than of astronomers. Tools should be built into the processing pipelines in order to facilitate visualization, processing and analysis of big data in a collaborative manner (Mark et al., 2002; Balakrishnan et al., 2008; Isenberg et al., 2011).

Therefore new visualization tools and software should be designed to overcome the limits and the barriers of traditional software by exploiting the latest technological opportunities. Based on the aforementioned challenges a number of proposed objectives to design new generation graphical software tools by exploiting the latest technological advances (e.g. science gateways) are the following:

Interactivity — For complex visualizations the relevant computations should be performed close to the data to avoid time consuming streaming of large data volumes. This can be achieved via flexible distributed architectures striking a balance between local interactive exploration tools and remote services hiding data complexity.

Integration — Tools should be coupled with advanced high performance computing (HPC) resources to deal with requests to archives through scientifically meaningful lightweight versions of the datasets obtained by analysis/processing operators since full data sizes may not fit the available memory to allow real-time interaction.

Navigation — Local exploration tools should enable interactive visualization optimized for ubiquitous computing environments, intuitively controlling the resulting visualization.

Collaboration — Tools should be combinable within e.g. science gateway technologies (Kacsuk, 2014) to allow collaborative activity between users and provide customization and scalability of data analysis/processing workflows, hiding underlying technicalities.

A number of different analysis and visualization tools and software have been developed for astronomers and astrophysicists during the last decades (refer to Hassan and Fluke, 2011 for a comprehensive review on existing tools). Many of them are designed to run on desktop computers, which have a finite memory size, and they are not suitable for exploring large scientific datasets. A solution to this problem is the use of distributed visualization, where a networked computing cluster shares the visualization tasks. For example, ParaView[1] or VisIt[2] can be used either as stand-alone local tool or they can run on a cluster in a client–server configuration. Remote visualization can imply losing real-time interactivity due to data transfer over the network.

Another important aspect is the collaborative visualization that enables multiple users to share a visualization experience. To successfully achieve this aim, it is necessary to provide high-speed network connections and effective communication protocols, e.g. in AstroSim (Nakasone et al., 2009).

When dealing with a large dataset, additional benefits may be achieved using workflow driven applications, e.g. in VisTrails (Freire et al., 2014). A workflow manages data and metadata of visualization products and it simplifies the complex problem of creating visualizations that becomes subsequent tasks of a workflow engine.

To reduce the output data size some authors propose an approach based on in situ visualization. For example Kageyama and Yamada (2014) developed a method for the interactive analysis of in situ visualization images produced by a batch simulation job, however this approach lacks integrated data analysis functionalities for effective data exploration.

A high-performance, GPU-based framework is presented in Hassan et al. (2013) for the analysis and visualization of terabytes sized 3D images. The framework was one of the first distributed environments but it lacks any collaborative facility and, most importantly, it relies only on powerful HPC infrastructures i.e. GPU clusters.

Finally, the astronomy and astrophysics community often relies on customized codes and scripts that are created to overcome limitations in other software to handle specific astronomy data formats or other domain specific issues related e.g. to the kind of visualization they want to obtain, see for example Labadens et al. (2012) and Szalay et al. (2008).

VisIVO software is an innovative tool that accounts for all the criteria highlighted above. VisIVO offers a unique integrated ecosystem for visualization including: services for collaborative portals, mobile applications for visualization and data exploration. It includes a number of key components such as workflow applications, analysis and mining functionalities, which can be re-thought under the big-data paradigm (see Section 5), focusing on: (1) an effective merger of the filtering/visualization pipeline so that the user can see in real time the effect of operations or interactions with the visualization results and (2) optimizing data movement and memory usage.

This paper presents the latest developments of VisIVO. Section 2 describes the basic technical details and features of VisIVO software tools and Section 3 discusses the latest developments made in the context of past and current international European funded projects. Section 4 is dedicated to present some scientific cases in which the support of data analysis and visualization is of primary importance. Section 5 presents the planned developments of VisIVO in the context of the Virtual Observatory framework to provide an integrated visual analytics tool for big data analysis and visualization.

2. The VisIVO ecosystem

VisIVO is an open source collection of graphics applications, which blend high performance multidimensional visualization, data exploration and visual analytics techniques. VisIVO tools have been developed to allow multidimensional data exploration of very large-scale datasets in order to visualize and discover a priori unknown data characteristics e.g. searching for outliers, characteristic regions or special properties. The underlying design based on commonly used toolkits (e.g. the Visualization ToolKit[3]) makes VisIVO a visualization framework useful also in contexts outside astronomy and astrophysics.

The main VisIVO tools are the following:

- VisIVO Desktop (Comparato et al., 2007), a stand-alone application for interactive visualizations running on PCs;
- VisIVO Server[4] (Becciani et al., 2009), a grid-enabled high performance data exploration and visualization command line toolkit;
- VisIVO Library (Becciani et al., 2012b), an application programming interface that has been developed to port VisIVO Server on gLite middleware[5] but can be installed on common Distributed Computing Infrastructures (DCIs);

---

[1] www.paraview.org.
[2] https://wci.llnl.gov/codes/visit.
[3] http://www.vtk.org.
[4] http://sourceforge.net/projects/visivoserver.
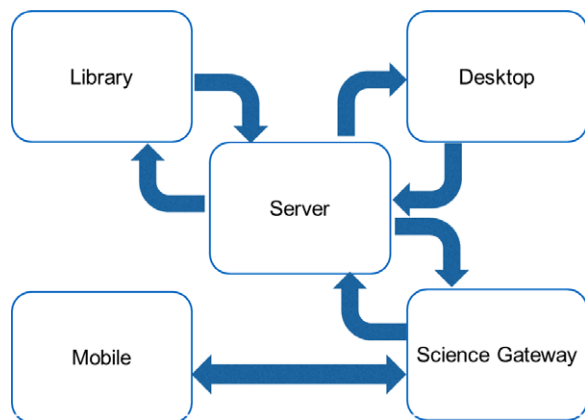[5] http://glite.cern.ch.

Fig. 1. VisIVO integrated set of tools: the Server tool is employed by the Library, the Desktop and the Science Gateway; the Mobile application connects through the Science Gateway to perform the VisIVO Server processing.

- VisIVO Science Gateway[6] (Becciani et al., 2014), a workflow enabled web portal wrapped around WS-PGRADE/gUSE framework (Kacsuk et al., 2012) providing visualization and data management services to the scientific community by means of an easy-to-use graphical environment for accessing the full functionalities of VisIVO Server.
- VisIVO Mobile[7] (Vitello et al., 2014), a mobile application connected to a workflow enabled framework to perform visualization and data filtering of large-scale, multidimensional datasets on DCIs.

All these tools represent an integrated environment as shown in Fig. 1. Some of the VisIVO tools are independent stand-alone software (e.g. VisIVO Desktop or VisIVO Server), others are used in co-operation for example to process some data on the server and make them accessible via the mobile application or the science gateway.

Users rapidly obtain meaningful visualizations (Johannessen, 2008) while preserving full and intuitive control of relevant visualization parameters. Each VisIVO tool is based on the following core modules: the VisIVO Importer, the VisIVO Filters and the VisIVO Viewer. The VisIVO Importer converts user-supplied datasets (e.g. generic formats such as ASCII and CSV or specialized formats such as Gadget[8] and HDF5[9]) into an internal representation called VisIVO Binary Table (or VBT). The VisIVO Filters are a collection of data processing modules to assist the extraction of hidden properties of the dataset. Finally, the VisIVO Viewer is the visualization core component based on the Visualization ToolKit(VTK) and Splotch (Dolag et al., 2008), a customized volume ray casting algorithm recently redesigned in view of the CUDA paradigm[10] to fully exploit modern HPC infrastructures (Rivi et al., 2014). VisIVO Viewer can also be used to produce images from a given set of camera positions that can be assembled externally to generate movies.

VisIVO Desktop was one of the first visualization tools that operated within the Virtual Observatory framework (Costa et al., 2008) and its development followed the International Virtual Observatory Alliance[11] (IVOA) recommendations and issues. It was able to interact with other VO compliant tools implementing PLASTIC[12] (PLatform for AStronomy Tool InterConnection) (Boch et al., 2008)

interoperability standards. The VisIVO Desktop interface with the PLASTIC hub allows the user to have different and complementary points-of-view of the data, and not only the ones provided by VisIVO: the user can get images making VisIVO interoperable with Aladin,[13] or directly from loaded table data using TOPCAT.[14] VisIVO Desktop is also able to retrieve data directly from available VO services (e.g. VizieR[15]) implementing a web-service based approach to interact with remote resources.

The VisIVO Desktop is now being redesigned as a tool for visual analytics in the context of the VIALACTEA project (see Sections 3 and 5). Within this project the role of the VO is of fundamental importance. In fact the usage and test of VO standards and tools can increase the scientific productivity and encourage the development of automatic pipelines to explore existing VO-compatible databases and archives. Seamless interaction with already existing DB/archive will be achieved without changing their internal structure, but adding a Data Access Layer to make them VO-compatible. The access to those data is performed using cURL[16] calls to the TAP service.[17] Although it is being developed to access the project databases, it is possible to perform queries, both synchronous and asynchronous, to any TAP services. The next step will be to implement upon SAMP[18] (Simple Application Messaging Protocol). SAMP is a messaging protocol, that enables astronomy software tools to interoperate and communicate, built on the prior messaging protocol, PLASTIC. Another important aspect regards the possibility to manage data saved on a remote or distributed storage. In the VO framework the interface over the distributed storage is the VOSpace (Graham et al., 2014). The current version is not supporting a VOSpace client and it is not able to read and store data on a VOSpace remote storage, therefore, although it is able to manage data on a local filesystem, a remote http and ftp resource, a VOSpace access will be developed to overcome this limitation.

In all VisIVO Ecosystem there is support to VOTable[19] that is an XML standard for the interchange of data represented as a set of tables. In this context, a table is an unordered set of rows, each of a uniform structure, as specified in the table description (the table metadata). Each row in a table is a sequence of table cells, and each of these contains either a primitive data type, or an array of such primitives.

3. Current status

In the last few years, the VisIVO ecosystem has been in continuous development thanks to a number of national and European funded projects as described below.

VO-Tech[20] (2005–2008) was an FP6 European funded project[21] aimed at completing all technical preparatory work necessary for the construction of the European Virtual Observatory (Euro-VO). Within this project the core modules of VisIVO Desktop, based on the Multimod Application Framework—MAF (Viceconti et al., 2007) and VTK, have been originally developed as a collaboration between the Italian National Institute for Astrophysics (INAF)—Astrophysical Observatory of Catania and CINECA (the largest Italian academic high performance computing centre).

---

6   http://visivo.oact.inaf.it:8080.
7   https://itunes.apple.com/it/app/visivo-mobile/id914448996?l=en&mt=8.
8   http://www.mpa-garching.mpg.de/gadget.
9   http://www.hdfgroup.org/HDF5.
10 http://developer.nvidia.com/cuda/.
11 http://www.ivoa.net.
12 http://www.ivoa.net/documents/latest/PlasticDesktopInterop.html.

13 http://aladin.u-strasbg.fr/.
14 http://www.star.bris.ac.uk/~mbt/topcat.
15 http://vizier.u-strasbg.fr.
16 http://curl.haxx.se.
17 http://www.ivoa.net/documents/TAP/20100327/REC-TAP-1.0.html.
18 http://www.ivoa.net/documents/SAMP.
19 http://www.ivoa.net/documents/VOTable.
20 http://ec.europa.eu/research/infrastructures/pdf/votech_en.pdf.
21 http://ec.europa.eu/research/fp6/index_en.cfm.

The EGI-Inspire project[22] aimed to join together the new Distributed Computing Infrastructures (DCIs) such as clouds, super-computing networks and desktop grids, for the benefit of user communities within the European Research Area. During this project, VisIVO was ported on grid infrastructures and the VisIVO Library was developed to port VisIVO Server on gLite middleware.[23] The VisIVO library allows a job running on a grid node to produce a set of images or movies directly using VisIVO from its internal data arrays without the need to produce intermediate temporary files. This is particularly important when running on the grid, where the user may want to have a quick look at the results during data production phases. In this way, the images can be produced directly in the grid catalogue, while the user code is running on a computing system that cannot be accessed directly (e.g. in worker nodes).

The Science and Technology Facilities Council in UK in 2012 supported the development of SpaceMission (Massimino et al., 2013) under their scheme "Science in Society Small and Science Centre Award", for a project entitled "Mobile Visual Discovery for Public Engagement in Modern Astrophysics". SpaceMission is a mobile application (iOS) game, especially designed for young people, offering hands-on experience of astrophysical concepts using scientific simulations (see details in Section 4.3).

The VisIVO Science Gateway and VisIVO Mobile were designed and developed in the framework of the SCI-BUS[24] project. The SCI-BUS project was a large European infrastructure FP7 project.[25] The project provided operation and maintenance of a generic-purpose science gateway framework which accesses to major computing, data and networking infrastructures and services including clusters, supercomputers, grids, desktop grids, academic and commercial clouds.

The Muon Portal project[26] is assembling a prototype of a large area detector for cargo containers inspection in harbours in order to counteract the illicit carrying of fissile material using of the muon tomography technique. In this context VisIVO packages play an essential role for cargo inspection, automated object recognition and in discriminating different types of material.

The aim of the VIALACTEA project[27] is to exploit the combination of all the new-generation surveys of the Galactic Plane, from space missions and ground-based facilities, to build and deliver a quantitative model of the VIALACTEA Galaxy. To reach this goal it is employed a novel data and science analysis paradigm based on 3D visual analytics and data mining framework, to build and deliver a quantitative 3D model of our galaxy as a star formation engine that will be used as a template for external galaxies and study star formation across the cosmic time. Within this project the VisIVO Desktop application is being further extended with 3D-aided visual analytic functionalities.

Finally, within the Gaia European Space Agency (ESA) space mission,[28] the Gaia Data Processing and Analysis Consortium (DPAC) is proposing to use VisIVO for the data exploration and interactive visualization of the Gaia mission outcomes following the Archive Access Co-Ordination Unit (CU9) model O'Mullane (2009).

4. Case studies

The VisIVO tools help researchers in the analysis of different scientific problems. Here, we select a key sample of VisIVO applications that show the benefit of using VisIVO ecosystem for scientific research.

4.1. Visual comparison of large scale cosmological simulations

In the standard cosmological paradigm, General Relativity (GR) is assumed to be the correct theory of gravity on all scales. According to GR, the expansion of the Universe should have slowed down. However, the accelerating expansion of our universe was discovered in 1998 by supernovae measurements (Riess et al., 1998; Perlmutter et al., 1999), which was awarded the Nobel prize in 2011. Understanding the physics of the cosmic acceleration is one of the most challenging problems in theoretical physics today. In the framework of GR, the acceleration may result from Dark Energy (DE), an unknown energy component (see Weinberg et al., 2013 and references therein for the DE review). Alternatively, DE may not exist but GR itself might break down on cosmological scales. Actually, assuming GR on all scales is a significant extrapolation of our limited knowledge of gravity, and GR remains untested on cosmological scales using conventional methods. Therefore, the discovered cosmic acceleration may suggest that GR, as well as the standard cosmology model based on GR, needs to be modified on cosmological scales.

Viable modified gravity (MG) models, for example, scalar–tensor theories, can mimic the LCDM model (GR plus a cosmological constant as dark energy) at the background level (for a recent review of MG models, please refer to Clifton et al., 2012). However, since the equations governing the evolution of cosmological perturbations are modified as well, such models generally predict different patterns for the growth of cosmic structure, which leaves characteristic imprints on cosmological and astrophysical observables. On highly-nonlinear scales where the local density is sufficiently high, the effect of MG can be shielded by the chameleon (Khoury and Weltman, 2004) or (Vainshtein, 1972) screening mechanisms. This is why such MG models can evade the solar system tests. One working example of such models is the f(R) gravity model, where the Ricci scalar R is generalized to a function of R in the Einstein–Hilbert action (Hu and Sawicki, 2007).

Observing the large scale structure of the universe could in principle provide new test of GR on cosmic scales. This kind of test cannot be done without the help of simulations as the structure formation process is highly non-linear. Large-scale simulations are thus performed for modified gravity models (Zhao et al., 2011), e.g. from the Large Simulation for Modified Gravity (LaSMoG) consortium.[29]

The workflow shown in Fig. 2 implements a customized VisIVO processing and visualization for aiding analysis of modified GR simulations, more specifically inspecting datasets to discover anomalies by comparing appropriately with datasets coming from standard GR models.

The main computational steps are summarized as follows:

- Two datasets corresponding to snapshots of standard gravity (D_S) and modified gravity (D_M) model simulations in RAMSES format (Teyssier, 2002) are imported by means of VisIVOImporter tool each having 16 777 216 particles.
- Sub-samples of the point distributions with a reduced number of points in the two datasets are generated with a VisIVOFilter processing using a *decimator* operation. Then, for each of these sub-samples a panoramic movie is created (as shown in Fig. 3) using a frame rate of 10 fps.
- A *point distribute* operation is performed on D_S and D_M to create new volume datasets (V_S and V_M respectively) using a field distribution NGP algorithm (Hockney and Eastwood, 2010) on a regular mesh of size $32 \times 32 \times 32$.

22 https://www.egi.eu/about/egi-inspire.
23 http://glite.cern.ch.
24 http://www.sci-bus.eu.
25 http://ec.europa.eu/research/fp7/index_en.cfm.
26 http://muoni.oact.inaf.it:8080.
27 http://cordis.europa.eu/project/rcn/188856_en.html.
28 http://www.esa.int/Our_Activities/Space_Science/Gaia.

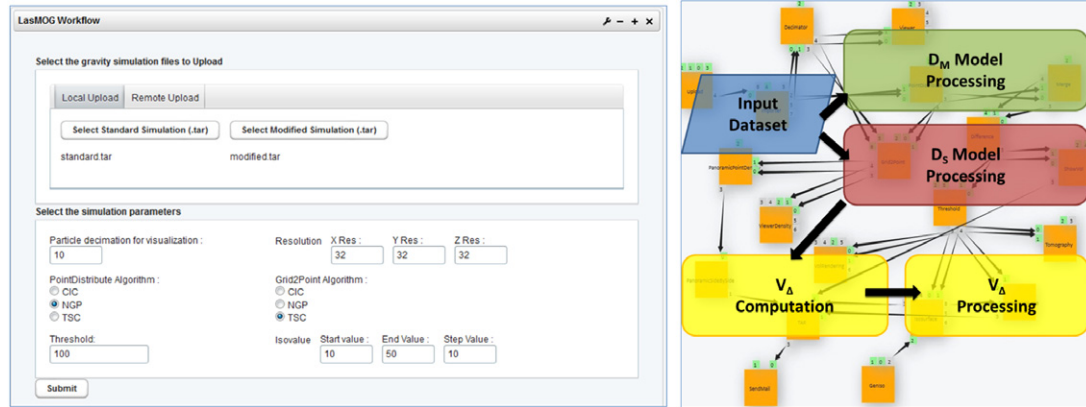29 http://icosmology.info/Nbody_Simulation.html.

Fig. 2. Large Scale Cosmological visualization produced by VisIVO Science Gateway: portlet interface (figure on the left) and workflow (figure on the right).
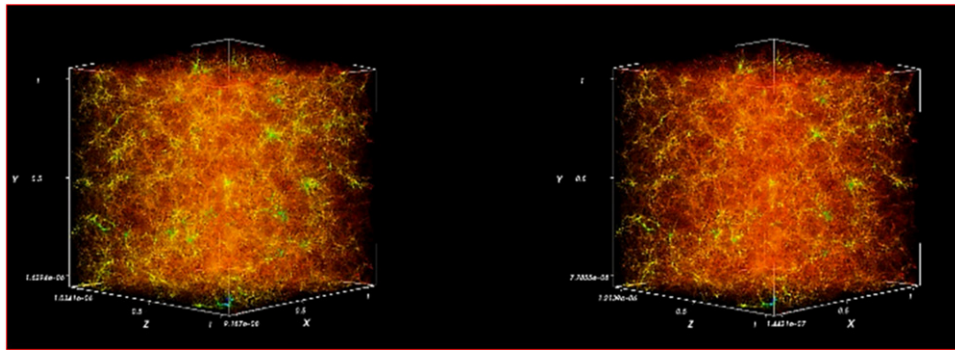

Fig. 3. Selected results of the Large Scale Cosmological visualization produced by VisIVO Science Gateway: panoramic movie of the standard gravity and modified gravity model simulations.

- A volume property on the same computational domain is distributed on a regular mesh producing a density field.
- A new volume V_Δ is computed where each of its voxels shows a difference of values in the density between V_S and V_M. It is then filtered with a lower bound threshold and all the voxels satisfying the filters are saved in a text file for further analysis purposes (as shown in resulting right image of Fig. 4).
- Several renderings of V_Δ are performed:
  - Volume rendering;
  - Isosurface rendering of the density field to produce panoramic movies using different iso-values (as shown in the resulting left image of Fig. 4);
  - Ortho-slice rendering i.e. orthogonal slice planes through the volume dataset.

4.2. On-line visualization of large scale cosmological simulations

One of the most challenging tasks in performing large-scale cosmological simulations is monitoring the outcomes, so that if anomalies are observed the simulation run is corrected appropriately. Typically, such monitoring requires visualization of several snapshots at different redshift, each being tens of gigabytes.

VisIVO Server tools have been employed to perform a large-scale structure of the universe simulation (100 Mpc h^{-1}) containing 800^3 particles (Becciani et al., 2012a). The run was performed on the COMETA Consortium Grid[30] using FLY simulation code (Becciani et al., 2007). The simulation was carried out over three months using 250 CPU cores; each data snapshot was 12 GB. Finally 8 million particles have been extracted to identify a subregion (15 Mpc h^{-1}) and interpolated in the time domain, producing
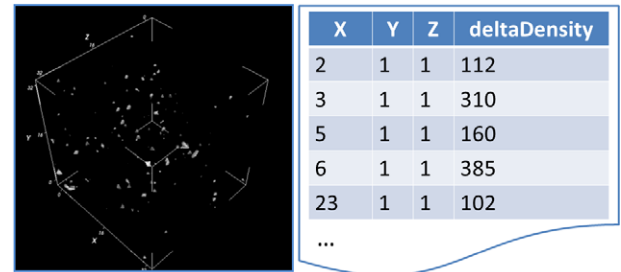

Fig. 4. Selected results of the Large Scale Cosmological visualization produced by VisIVO Science Gateway: isosurface rendering (figure on the left) and textual output (figure on the right) of V_Δ.

a large number of intermediate frames that have been used in creating a movie showing the fluid time evolution. Overall, 1500 intermediate datasets were created requiring a CPU time of two months. The final movie employed a frame rate of 10 fps. The whole data archive handled for the movie creation was in excess of 18 TB.

4.3. Public outreach

A VisIVO based mobile application game named SpaceMission has been developed to offer hands-on experience of astrophysical concepts using scientific simulations (Massimino et al., 2013). The goal of this game is to find four regions of interest inside a cosmological simulation: an elliptical galaxy, an elliptical galaxy at the centre of a galaxies cluster, the formation of elliptical galaxy, a spiral galaxy. During the investigation, players are provided with tools to create movies of their explorations. Such movies are rendered by exploiting (in a seamless way) grid infrastructures.

Thanks to the usage of WS-PGRADE/gUSE framework, a new mobile application named VisIVO Mobile (Vitello et al., 2014) has

---

30 http://www.pi2s2.it/.

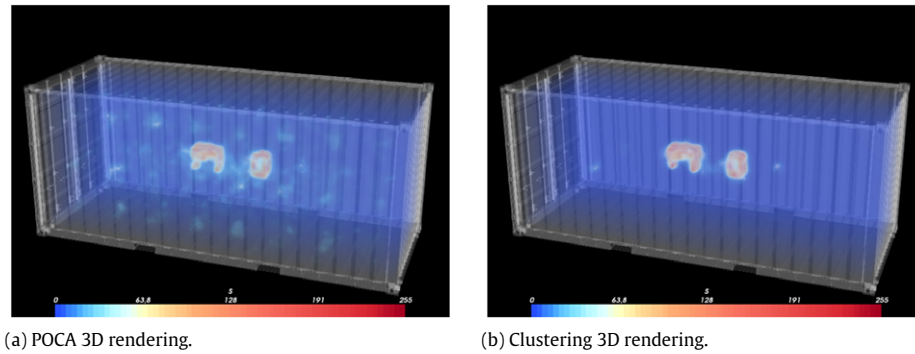(a) POCA 3D rendering.

(b) Clustering 3D rendering.

Fig. 5. Tomographic imaging of the simulated scenario obtained with the POCA and Clustering methods with 3D volume rendering of the entire container.

been designed in order to load, analyse and visualize any user provided dataset and then processing the data not only on grid infrastructures but also on any other DCIs handled by gUSE such as Linux clusters or clouds.

In order to disseminate the developed tools to new communities and common citizen and attract more end-users, a science visualization contest has been recently set up. The competition is a call to the worldwide scientific community to use technologies provided by the VisIVO Science Gateway and VisIVO Mobile application to produce images and movies from multi-dimensional datasets coming either from observations or numerical simulations. Although VisIVO tools are specialized for astrophysical datasets the competition was open to all scientists and common citizens alike aiming to investigate datasets related to other fields, e.g. life sciences or physics.

4.4. Particle tracking visualization

The flexibility of VisIVO allowed to successfully adopt it also in applications not strictly related to astronomical case studies. In fact, it has been used also in nuclear physics 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 (Z) 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 (Katz et al., 2007), 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 suitable images of a cargo container and common tools used for this kind of simulations (e.g. ROOT[31]) does not provide the needed visualizations such as the volume rendering, isosurfaces or tomographic images.

As an example we show a simulated scenario that contains a "MUON" shape built with voxels of size 10 cm × 10 cm × 10 cm inserted at the centre of a container filled with layers of washing machine-like elements. Each letter is made of different materials: M = Uranium, U = Iron, O = Lead, and N = Aluminium. The "washing machines" are made by an aluminium casing with an iron engine inside with relative support bars and a concrete block. The container load relative to the scene is ~3500 kg. Using a realistic energy spectrum with range 0.1–100 GeV, we simulated 5×10^5 muon events for this scenario, corresponding to ~10 min scanning time.

The data file containing the coordinates on the muon tracker planes is first imported and filtered using the POCA (Point of

Closest Approach) algorithm (Sunday, 2004) to obtain the VisIVO Binary Table containing the scattering deflection of cosmic radiations. The resulting VBT can be visualized using a volume rendering algorithm as shown in the left image of Fig. 5.

This dataset can be further processed through the VisIVO Filter to perform a clustering algorithm. Materials with high atomic number are elaborated with the POCA method as regions of higher densities of unspecified shape with respect to the background. It is therefore a natural choice to employ density-based clustering methods in the tomography reconstruction. Clustering algorithms, working directly on points, are useful to detect the presence of suspicious items inside the container, acting as a filter for a preliminary analysis of the data.

As can be seen in 5(a), the scenarios are successfully identified. Due to the intrinsic resolution of the POCA method a persistent halo is present and consequently the imaged objects are slightly increased in size with respect to the real dimensions, particularly along the vertical z-axis. A considerable noise, related to the engine elements, is present in this dense environment scenario. In Fig. 5(b) we report the results obtained with the clustering method. As it can be seen, a good accuracy is achieved in the identification of the material with high atomic number and the noise due to the washing machines engine is significantly reduced. The clustering algorithm in fact acts as a filter, removing all spurious hits even if they have high scattering angles, because they are spatially isolated from the other POCA events.

This application was embedded into a workflows oriented Science Gateway: the VisIVO Science Gateway. Any Gateway user can decide to build a customized workflow connecting the predefined VisIVO Workflows as building blocks (see Fig. 6). The workflow performs the importing and filtering of the dataset comparing the POCA and Clustering methods and produces several visualizations as well as panoramic movies inspecting the cargo container automatically generated by a camera motion path of 360° in azimuth and ±90° in elevation within the dataset's domain.

5. Future directions

Future developments within the VisIVO ecosystem are foreseen to move from a data exploration paradigm to a visual analytics one. These developments are mainly driven from the needs of the VIALACTEA consortium. In fact, the VIALACTEA scientists expect to employ a 3D-aided visual analytic environment that allows them to interactively analyse the multi-dimensional Galactic Plane.

A visual analytic toolkit will equally integrate data analysis and visualization. 3D-aided visual analytics enable the user to, e.g., steer the analysis algorithms while receiving feedback about areas of interest in the data, or investigate the data in all its detail by providing a tight coupling between the user and machine algorithms.
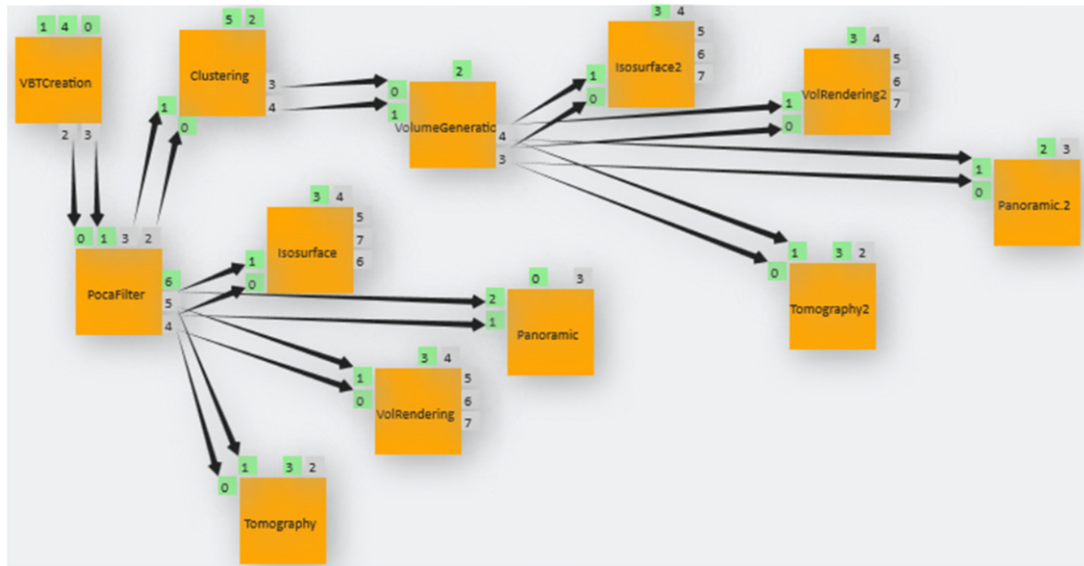
---

[31] http://root.cern.ch/drupal/.

Fig. 6. Muon processing workflow: dataset importing, POCA filtering, Clustering and production of visualizations and panoramic movies.

To achieve this goal we expect to move towards a new generation conceptual design, underpinning the overall ecosystem, consisting of three layers and to make use of specific technologies and infrastructures, such as GPUs and Cloud computing.

5.1. Conceptual architecture

In the near future we envisage the VisIVO tools as an integrated framework with a multi-layer architectural design (Sciacca et al., 2014) in order to provide a visual analytics application for big data analysis and visualization where datasets, models, analysis, data-mining/machine-learning tools and information will be enclosed in a standardized, homogeneous and interoperable framework handled via Science Gateway technologies.

This architectural design will be implemented as Service Oriented Architecture (SOA), offering data as a service and computing as a service environment. An SOA environment helps to hide the infrastructure complexity and provides a highly accessible solution distributed user community. The remote services can be connected to different distributed computing infrastructures (DCIs) ranging from HPC clusters, to grids and clouds. This allows maximum flexibility for the data processing tasks to be performed. On the other hand it can be connected to the data archive using available standard protocols such as the ones of the International Virtual Observatory Alliance.

The adopted multi-layer design provides a generic and scalable solution for data analysis and visualization and can be summarized into three conceptual layers of the VisIVO Desktop application as shown in Fig. 7:

The application layer includes the visual analytics client: a user interface along with controllers for user interaction events. It also contains representations of the current data and handles user actions, such as zooming and panning delegating them to a singleton shared module that communicates with the remote server using an asynchronously paradigm. They are used to view filtered data using the received underlying dataset from the data layer or, alternatively, using the whole raw dataset from the stream data layer, thanks to a data manager component. This way, the top-level application can be independent and can choose at runtime either an optimized data source from a data layer or the entire raw dataset.

The data layer reads and parses the different supported file formats and supplies the application layer with requested data. A reader component is in charge of reading the data also from the catalogue. When it receives a request from the top-level application layer it produces the required dataset if it already has the data necessary for producing them in a local mirror, otherwise it relies on the stream layer to obtain the new data necessary for the computation.

The stream layer is responsible for supporting random access to sections of dataset (i.e. subset and specific regions) accessed by VO compliant protocols. The SAMP protocol allows VisIVO Desktop to interoperate and communicate with other VO enabled tools such as Topcat or Aladin. The access to the knowledge based database is enabled through SQL and TAP (Nandrekar-Heinis et al., 2014) queries which also allows the application to access VO catalogues.

A Science Gateway interface module allows the desktop application to submit workflows through the usage of Remote API (Balasko et al., 2013). This provides methods for checking the workflow's status, and for downloading the outputs. Remote API is implemented as a simple servlet that can be installed as one of the gUSE services. Servlets can be called from anywhere such as web interfaces (Javascript), standalone applications, different web services or anything that supports HTTP or HTTPS communication protocols. As a result Remote API is a powerful tool to exploit gUSE capabilities independently from the WS-PGRADE portal interface.

5.2. Technological details

Data mining techniques based on machine learning algorithms will be investigated: unsupervised, supervised or mixing the two approaches (Ball and Brunner, 2010) and implemented as workflows. For example, supervised techniques such as artificial neural networks or decision trees could give good insight into classification of photometric objects. Some other examples are computation of indexes for quick data access, advanced clustering methods to enable faster transfer and strategies for data selection or simplification (such as the adaptive mesh refinement techniques). Those operators add quantitative analysis to the qualitative visualization
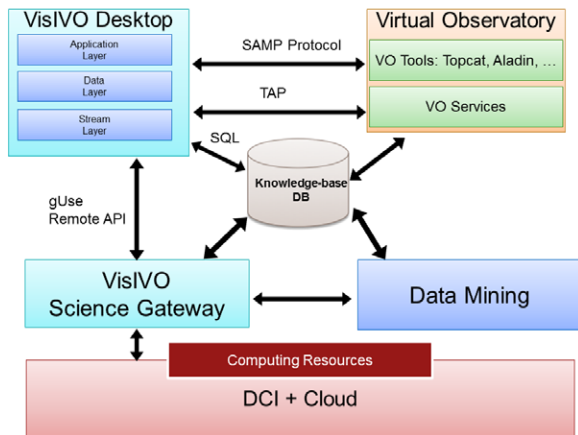
Fig. 7. Visual analytics VisIVO Desktop architecture for big data analysis and visualization.

and optimize the data movement between the framework components minimizing data transfer costs. In order to cope with the volume, velocity and variety of data produced, a streaming model will be investigated where analysis of data has to use low memory and process items only once, therefore only the summary and statistical information are kept reducing enormously the storage requirements in distributed computing infrastructures. Stream Processing Engines (SPE), e.g. the Apace! S4 (Neumeyer et al., 2010), have shown to be a more flexible and powerful tool for dealing with big data in distributed stream computation (Amini et al., 2006), thus a natural evolution from the batch MapReduce (Dean and Ghemawat, 2008) tools. The advantage of SPEs is that of processing an unbound, possibly infinite flow of data and returning results in any time.

An aspect of possible improvements will be to exploit HPC architectures which are nowadays increasingly populated with GPUs. Our framework is based on a backend solution that filters and deploys the required data and a client solution that translates user visualization parameters that are fed into the computational server. Therefore, the client needs an additional computational time that is required to render efficiently the received data on the screen. An optimization would be to split and pipeline the rendering into multiple channels. The Splotch rendering algorithm for exploration and visual discovery included into the VisIVO packages has been already redesigned for exploiting HPC architectures by using the CUDA programming paradigm (Rivi et al., 2014). This approach is currently fully supported by VTK, so it should be easy to incorporate it in the proposed framework.

Finally, cloud computing will be exploited in order to offer highly flexible, scalable, on-demand computational storage and networking resources for the large-scale distributed computing tasks. For this purpose we have identified the example of CANFAR[32] federated cloud operated in Canada which offers an infrastructure oriented to the astronomy and astrophysics community joining together IaaS and the standards and services developed by the IVOA e.g. for user authentication and authorization, data sharing, access to data and archives, and finally data processing.

6. Conclusion

Nowadays astronomy and astrophysics produce massively large data volumes coming from observations or simulation codes executed on high performance supercomputers. Data visualization

is a fundamental, enabling technology for knowledge discovery, involving a number of technical challenges, e.g. allowing real-time interaction although data sizes may not fit the available memory; providing tools that are fully integrated within the scientists' toolkit, sophisticated navigation in ubiquitous environments and finally collaborative mechanisms.

We have described a number of approaches towards addressing these challenges, focusing on the latest technological advances within VisIVO which is unique integrated ecosystem for visualization providing comprehensive services for collaborative portals and mobile applications for visualization and data exploration.

We gave an overview on the developments of the VisIVO ecosystem within a number of national and international European funded projects and presented scientific cases on particle tracking visualization to identify high atomic number materials, comparing on line visualizations of large scale cosmological simulations, and disseminating results for public outreach.

We presented the planned developments within the context of the Virtual Observatory framework to provide an integrated visual analytics ecosystem for big data analysis and visualization. Finally, we outlined future developments in terms of a new generation layered conceptual design, underpinning the overall ecosystem and also in making use of emerging technologies and infrastructures, such as exploitation of GPUs/coprocessors and cloud computing.

References

Amini, L., Andrade, H., Bhagwan, R., Eskesen, F., King, R., Selo, P., Park, Y., Venkatramani, C., 2006. SPC: a distributed, scalable platform for data mining. In: Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms. ACM, pp. 27–37.

Balakrishnan, A.D., Fussell, S.R., Kiesler, S., 2008. Do visualizations improve synchronous remote collaboration? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 1227–1236.

Balasko, A., Farkas, Z., Kacsuk, P., 2013. Building science gateways by utilizing the generic WS-PGRADE/gUSE workflow system. Comput. Sci. 14, 307–325.

Ball, N.M., Brunner, R.J., 2010. Data mining and machine learning in astronomy. Internat. J. Modern Phys. D 19, 1049–1106.

Beaumont, C., Robitaille, T., Goodman, A., Borkin, M., 2013. Multidimensional data exploration with glue. In: Proceedings of the 12th Python in Science Conference, SCIPY 2013.

Becciani, U., Antonuccio-Delogu, V., Comparato, M., 2007. FLY: MPI-2 high resolution code for lss cosmological simulations. Comput. Phys. Comm. 176, 211–217.

Becciani, U., Antonuccio-Delogu, V., Costa, A., Petta, C., 2012a. Cosmological simulations and data exploration: a testcase on the usage of grid infrastructure. J. Grid Comput. 10, 265–277.

Becciani, U., Costa, A., Comparato, M., Caniglia, G., Gheller, C., Krokos, M., Grillo, A., Munzone, R., Vitello, F., 2009. VisIVO: new integrated services. In: Astronomical Data Analysis Software and Systems XVIII. p. 555.

Becciani, U., Costa, A., Ersotelos, N., Krokos, M., Massimino, P., Petta, C., Vitello, F., 2012b. VisIVO: a library and integrated tools for large astrophysical dataset exploration. In: Astronomical Data Analysis Software and Systems XXI. p. 505.

Becciani, U., Sciacca, E., Costa, A., Massimino, P., Pistagna, C., Riggi, S., Vitello, F., Petta, C., Bandieramonte, M., Krokos, M., 2014. Science gateway technologies for the astrophysics community. Concurr. Comput.: Pract. Exper. doi:10.1002/cpe.3255.

Boch, T., Comparato, M., Taylor, J., Taylor, M., Winstanley, N., 2008. Plastic—a protocol for desktop application interoperability version 1.00. IVOA Applications WG Working Draft.

Clifton, T., Ferreira, P.G., Padilla, A., Skordis, C., 2012. Modified gravity and cosmology. Phys. Rep. 513, 1–189.

Comparato, M., Becciani, U., Costa, A., Larsson, B., Garilli, B., Gheller, C., Taylor, J., 2007. Visualization, exploration, and data analysis of complex astrophysical data. Publ. Astron. Soc. Pac. 119, 898–913.

Costa, A., Manzato, P., Becciani, U., Comparato, M., Costa, V., Gasparo, F., Gheller, C., Grillo, A., Molinaro, M., Pasian, F., et al., 2008. The tvo archive for cosmological simulations: web services and architecture. Publ. Astron. Soc. Pac. 120, 933–944.

Dean, J., Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. Commun. ACM 51, 107–113.

Dolag, K., Reinecke, M., Gheller, C., Imboden, S., 2008. Splotch: visualizing cosmological simulations. New J. Phys. 10, 125006.

Freire, J., Koop, D., Chirigati, F., Silva, C., 2014. Reproducibility using vistrails. In: Implementing Reproducible Computational Research. Chapman & Hall/CRC, Boca Raton, FL.

Goodman, A., 2009. Seeing science. ArXiv Preprint arXiv:0911.3349.

Graham, M., Morris, D., Rixon, G., Dowler, P., Schaaff, A., Tody, D., Major, B., 2014. VOSpace service specification. URL: http://www.ivoa.net/documents/VOSpace/.

Greensky, J.B., Czech, W.W., Yuen, D.A., Knox, M.R., Damon, M.R., Chen, S.S., Kameyama, M.C., 2008. Ubiquitous interactive visualization of 3D mantle convection using a web-portal with Java and Ajax framework. Vis. Geosci. 13, 105–115.

Hassan, A., Fluke, C.J., 2011. Scientific visualization in astronomy: towards the petascale astronomy era. Publ. Astron. Soc. Aust. 28, 150–170.

Hassan, A., Fluke, C.J., Barnes, D.G., Kilborn, V., 2013. Tera-scale astronomical data analysis and visualization. Mon. Not. R. Astron. Soc. 429, 2442–2455.

Hey, A.J., Tansley, S., Tolle, K.M., et al., 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. Springer, Berlin, Heidelberg.

Hockney, R.W., Eastwood, J.W., 2010. Computer Simulation Using Particles. CRC Press.

Hu, W., Sawicki, I., 2007. Models of f (r) cosmic acceleration that evade solar system tests. Phys. Rev. D 76, 064004.

Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.L., Hagen, H., 2011. Collaborative visualization: definition, challenges, and research agenda. Inf. Vis. 10, 310–326.

Jin, Z., Krokos, M., Rivi, M., Gheller, C., Dolag, K., Reinecke, M., 2010. High-performance astrophysical visualization using Splotch. Procedia Comput. Sci. 1, 1775–1784.

Johannessen, R.A., 2008. Interactive visual analysis and exploration of astrophysical data. In: Seminar in Visualization (INF358).

Kacsuk, P. (Ed.), 2014. Science Gateways for Distributed Computing Infrastructures. Springer International Publishing.

Kacsuk, P., Farkas, Z., Kozlovszky, M., Hermann, G., Balasko, A., Karoczkai, K., Marton, I., 2012. WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities. J. Grid Comput. 10, 601–630.

Kageyama, A., Yamada, T., 2014. An approach to exascale visualization: interactive viewing of in-situ visualization. Comput. Phys. Comm. 185, 79–85.

Katz, J., Blanpied, G., Borozdin, K., Morris, C., 2007. X-radiography of cargo containers. Sci. Glob. Secur. 15, 49–56.

Keefe, D.F., 2010. Integrating visualization and interaction research to improve scientific workflows. IEEE Comput. Graph. Appl. 30, 8–13.

Kent, B.R., 2013. Visualizing astronomical data with blender. Publ. Astron. Soc. Pac. 125, 731–748.

Khoury, J., Weltman, A., 2004. Chameleon cosmology. Phys. Rev. D 69, 044026.

Labadens, M., Chapon, D., Pomarède, D., Teyssier, R., 2012. Visualization of Octree adaptive mesh refinement (AMR) in astrophysical simulations. In: Astronomical Data Analysis Software and Systems XXI. p. 837.

Mark, G., Kobsa, A., Gonzalez, V., 2002. Do four eyes see better than two? Collaborative versus individual discovery in data visualization systems. In: Sixth International Conference on Information Visualisation, 2002. Proceedings. IEEE, pp. 249–255.

Massimino, P., Costa, A., Becciani, U., Krokos, M., Bandieramonte, M., Petta, C., Pistagna, C., Riggi, S., Sciacca, E., Vitello, F., 2013. Learning astrophysics through mobile gaming. In: Astronomical Society of the Pacific Conference Series. p. 113.

Nakasone, A., Prendinger, H., Holland, S., Miura, K., Hut, P., Makino, J., 2009. AstroSim: collaborative visualization of an astrophysics simulation in second life. IEEE Comput. Graph. Appl. 29, 69–81.

Nandrekar-Heinis, D., Michel, L., Louys, M., Bonnarel, F., 2014. The table access protocol: providing standard access to astronomical data. Astron. Comput..

Neumeyer, L., Robbins, B., Nair, A., Kesari, A., 2010. S4: distributed stream computing platform. In: 2010 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, pp. 170–177.

Nita, G.M., Fleishman, G., Gary, D., Kuznetsov, A., Kontar, E., 2011. GX_Simulator: an interactive IDL widget tool for visualization and simulation of imaging spectroscopy models and data. Bull. Am. Astron. Soc. 1811.

O'Mullane, W., 2009. Gaia catalogue and archive, plans and status. In: SF2A-2009: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics, p. 65.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R., Nugent, P., Castro, P., Deustua, S., Fabbro, S., Goobar, A., Groom, D., et al., 1999. Measurements of ω and λ from 42 high-redshift supernovae. Astrophys. J. 517, 565.

Riess, A.G., Filippenko, A.V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P.M., Gilliland, R.L., Hogan, C.J., Jha, S., Kirshner, R.P., et al., 1998. Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. 116, 1009.

Rivi, M., Gheller, C., Dykes, T., Krokos, M., Dolag, K., 2014. GPU accelerated particle visualization with Splotch. Astron. Comput. 5, 9–18.

Sciacca, E., Pistagna, C., Becciani, U., Costa, A., Massimino, P., Riggi, S., Vitello, F., Bandieramonte, M., Krokos, M., 2014. Towards a big data exploration framework for astronomical archives. In: Proceedings of the 2014 International Conference on High Performance Computing & Simulation, HPCS 2014.

Sunday, D., 2004. Distance between lines and segments with their closest point of approach. URL: http://softsurfer.com/Archive/algorithm_0106/algorithm_0106.htm.

Szalay, T., Springel, V., Lemson, G., 2008. GPU-based interactive visualization of billion point cosmological simulations. ArXiv Preprint arXiv:0811.2055.

Teyssier, R., 2002. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. Astron. Astrophys. 385, 337–364.

Vainshtein, A.I., 1972. To the problem of nonvanishing gravitation mass. Phys. Lett. B 39, 393–394.

Viceconti, M., Zannoni, C., Testi, D., Petrone, M., Perticoni, S., Quadrani, P., Taddei, F., Imboden, S., Clapworthy, G., 2007. The multimod application framework: a rapid application development tool for computer aided medicine. Comput. Methods Programs Biomed. 85, 138–151.

Vitello, F., Sciacca, E., Becciani, U., Costa, A., Massimino, P., Takacs, E., 2014. Developing a mobile application connected to a science gateway. In: 6th International Workshop on Science Gateways, IWSG 2014, IEEE Xplore.

Weinberg, D.H., Mortonson, M.J., Eisenstein, D.J., Hirata, C., Riess, A.G., Rozo, E., 2013. Observational probes of cosmic acceleration. Phys. Rep. 530, 87–255.

Zhao, G.B., Li, B., Koyama, K., 2011. N-body simulations for f (r) gravity using a self-adaptive particle-mesh code. Phys. Rev. D 83, 044007.