Mem. S.A.It. Suppl. Vol. 4, 86

Memorie della
Supplementi

Virtual Observatory and Grid-related projects: national and international status

F. Pasian

INAF/Osservatorio Astronomico di Trieste, Via G.B.Tiepolo 11, I-34131, Trieste, Italy

Abstract. The Virtual Observatory as a uniform facility for accessing archived data, and the Grid paradigm for exploiting network-distributed facilities have rapidly gained importance, and are the two environments the Italian astrophysical community will have to deal with in the future. This paper summarizes the status of the related national and international projects, and points out to the logical connections between these two fields.

Key words. data processing – catalogs – astronomical data bases: miscellaneous – data grid – virtual observatory

1. Introduction – Data-intensive computing in astrophysics

An important part of the computing challenges in astronomy are related to the handling, processing and modelling of large quantities of data. In particular, processing of huge quantities of data (large detectors, mosaics, images with high time resolution) is typical of the optical and solar communities. The amount of computations needed to process the data is impressive, but often "embarassingly parallel" since based on local operators, with a coarse-grained level of parallelism. In such cases, the "memory footprint" of the applications allows to subdivide data in chunks, so as to fit the RAM available on the individual CPUs and to have each CPU to perform a single processing unit.

Send offprint requests to: F. Pasian
Correspondence to: via G.B.Tiepolo 11, I-34131 - Trieste - Italy

Which kind of resources are necessary to tackle the processing and analysis of large quantities of data in astrophysics?

In most cases, where we can sub-divide the data to be processed as evidenced in the above paragraph, "distributed supercomputers", i.e. a local cluster of PCs such as a Beowulf machine, or a set of PCs distributed over the network, can be an effective solution to the problem. In this case, the **Grid** can be considered to be an important step forward in the provision of the computing power needed to tackle the new challenges.

As for data, the concept of "distributed archives" is already familiar to the average astrophysicist. The leap forward in this case is to be able to organize the data repositories to allow efficient, transparent and uniform access: these are the basic goals of the Virtual Observatory (**VO**). In more than a sense, the VO is an extension of the classical Computational Grid; it fits perfectly the Data Grid concept, being based on storage and processing systems, and metadata and communications management services.

2. The Grid

The GRID is a collaborative set of computing, data storage and network resources belonging to different administrative domains that has knowledge about the status of its components through an active, distributed information service.

A number of project at international and national level have lately become active to allow the Grid infrastructure to become a reality at the users' fingertips.

2.1. Grid.it and astrophysics

Grid.it is a project involving the whole Italian scientific community, funded in 2002 by the Ministry of Education and Research through the Italian Fund for Basic Research (FIRB). The full name of the project is "Enabling platforms for high-performance computational Grids oriented towards scalable virtual organizations" and provides a technical and organizational framework so as to allow various projects to operate as virtual organizations. The project has been financed with a total of 8 M over three years.

Grid.it has bootstrapped a collaboration called Italian Grid for Business, Industry, Government, E-Science and Technology (IGBIGEST), whose task is to coordinate the Italian Grid infrastructure and the participation of the Italian Grid community in international projects. The participation in this structure of the astrophysical community is guaranteed by INAF.

The Italian astrophysical community participates in Grid.it through the development of three applications, to be used as demonstrators (Work-Package 10):

- access and browsing on astronomical archives and databases (Trieste, Padova)
- access to a pipeline facility for the calibration of wide-field astrophysical images (Naples)
- distributed monitoring of astronomical observations (Trieste)

The key point of this participation is the strict technical coordination, obtained by using the hardware and software standards defined jointly within the Grid.it project (through its Work-Package 4).

The main purpose of the activities carried out within WP10 is porting the software already available for the three above-mentioned applications onto a grid structure, to verify the requirements the applications themselves impose on the grid middleware. As a matter of fact, since the applications are developed within an international framework, there are standards, data representations, software, etc. to be used and that can be considered as constraints for the integration in a grid structure.

Up to now, a number of activities common to the three applications have been carried out within WP10:

- analysis of documentation and of the software platform to be used;
- setup of a prototype grid site using available hardware;
- specification of the hardware and software environment for the standard astrophysical node of Grid.it;
- preliminary gathering of requirements coming from the three applications, with specific reference to needed features in the grid middleware;
- analysis of the security requirements for astrophysics and setup of the necessary structure (Registration Authority, certificates, ...);
- installation of a user interface and support to the community.

2.2. DRACO, and its extension

The astrophysics community has leveraged on the Grid.it project to "create" an astrophysical subset, which was named DRACO (Datagrid for italian Research in Astrophysics and Coordination with the virtual Observatory), to be expanded from its original core through other sources of funding.

A proposal to extend the three Grid.it WP10 sites to include three additional sites (Rome, Naples/Salerno, Catania) to build the core of an Italian Astrophysical Grid was submitted for funding in early 2003. The proposal

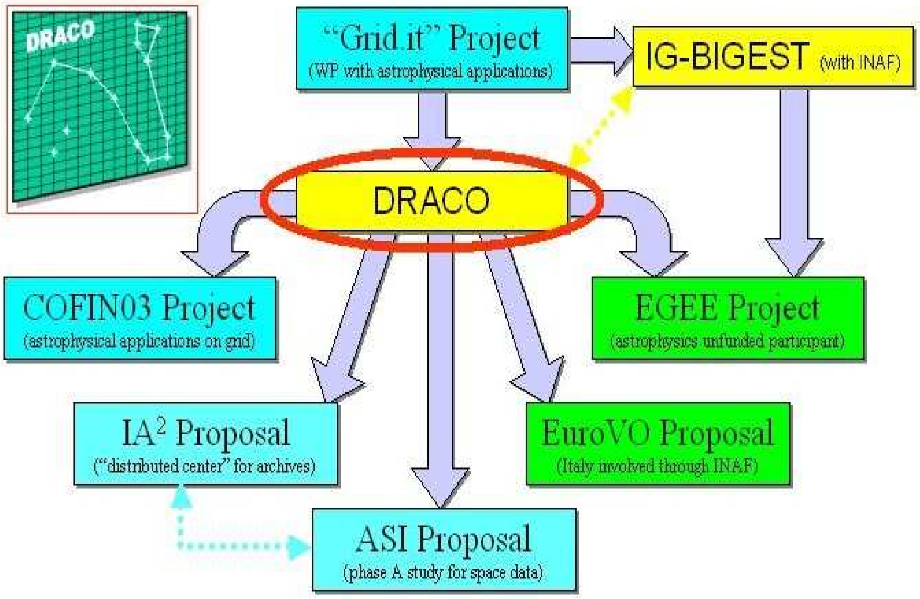

Fig. 1. The various Italian projects on Grid and Virtual Observatories and the logical flow for their generation. Blue boxes represent Italian projects, green boxes European ones. Yellow boxes represent organizations.

was successful and the project was funded by the Ministry of Education and Research through its Cofin03 budget line for a total of 370 K over 2 years.

The focus of the project is on allowing existing applications (visualization, data mining, data access, etc.) to become grid-enabled. All partners in the project will provide both Storage Elements and Computing Elements and therefore provide data and computing power. Use of hw/sw standards defined by Grid.it

2.3. EGEE

IG-BIGEST, together with partners from almost all European countries, have submitted to the EU 6th Framework Program (EU/FP6) a proposal called Enabling Grids for E-science in Europe (EGEE). The purpose of this project is to develop a seamless pan-European grid infrastructure for the support of research by integrating existing national, regional and thematic grid efforts. The project, coordinated by CERN, has been funded with a total of 32 M over 3 years and the kick-off date is expected to be 1st April 2004.

Astrophysics did not participate, being already partners in the Euro-VO EU/FP6 proposal, and also being strictly an application on top of the grid infrastructure to be deployed by EGEE. Scientific domains included in EGEE are particle physics and biomedicine, however generic "other applications" can be supported in their datagrid needs. In this framework, INAF has accepted to participate in EGEE as an unfunded partner and is defining, with European partners, specific applications to be ported on the Grid as demonstrators.

3. The Virtual Observatory

The Virtual Observatory (VO) is a paradigm to utilize multiple archives of astronomical data in an interoperating, integrated and logically centralized way, so to be able to "observe a virtual sky" by position, wavelength and time. Not only data actually observed are included in this concept: theoretical and simulated information can be included as well.

3.1. The International Virtual Observatory Alliance

The International Virtual Observatory Alliance (IVOA) was created in June 2002 to facilitate the international coordination and collaboration necessary for the development and deployment of the necessary tools, systems and organizational structures.

The key results IVOA needs to achieve are not only technical, but have to do mainly with organizational and policy aspects to be harmonized among all partners. Each of the existing national and international projects will have its own particular set of science objectives and technology drivers and constraints. Obviously, also policy and funding issues will be different at the individual national levels. This diversity, while being respected, needs to be harmonized.

On the other hand, there are also elements of the international effort that must be common and agreed upon. Mostly, they deal with standards for data and interfaces, software packages, source code libraries, and development tools.

All funded VO projects may join IVOA, provided they commit to:

– support full exchange of technical developments;
– participate in the definition and evaluation of standards;
– implement VO-compliant interfaces and documentation.

The services IVOA provides to the participating members are already relevant, even in this preliminary phase. Among those, seamless and transparent interrogation of data archives, new powerful analysis and visualization tools, a standard framework for data publishing and delivering, and in general services using published data. Extensions have naturally been planned for the future.

In any case, the key to the IVOA is standardization: of data and their description (metadata), of data exchange methods, of registry services.

3.2. European efforts towards a Virtual Observatory

The Astrophysical Virtual Observatory Project (AVO) is a project financed by the Fifth Framework Programme of the European Commission (EU/FP5) to conduct a research and demonstration (Phase-A) programme on the scientific requirements and technologies necessary to build a VO for European astronomy.

The plans are for the AVO project to be extended from a Phase-A study into a full-fledged system (Euro-VO) to serve the whole astrophysical community. For these purpose, three proposals to EU/FP6 were envisaged: two were submitted in spring 2003, with the purpose of defining a Data Centre Alliance federating data providers and a Facility Centre allowing to act as a reference point for data providers and consumers. Partners in the proposal are: the European Southern Observatory, the European Space Agency, the Centre Données Strasbourg, the British AstroGrid project (represented by the University of Edinburgh), the German Astronomical VO (represented by MPE), INAF, the Dutch NOVA organization, and the Spanish LAEFF institute.

The submitted proposals were not successful: in the 2003 round, the EU Commission policy was to fund projects aiming at providing infrastructure facilities, such as EGEE and DEISA); applications were given lower priority. The third proposal (the Euro-VO Technical Centre) was not submitted. A comprehensive proposal is planned to be re-submitted in late 2004; the bridging phase between the end of the AVO project and the production Euro-VO is being planned by extending AVO facilities c/o ESO, ESA and the national partners.

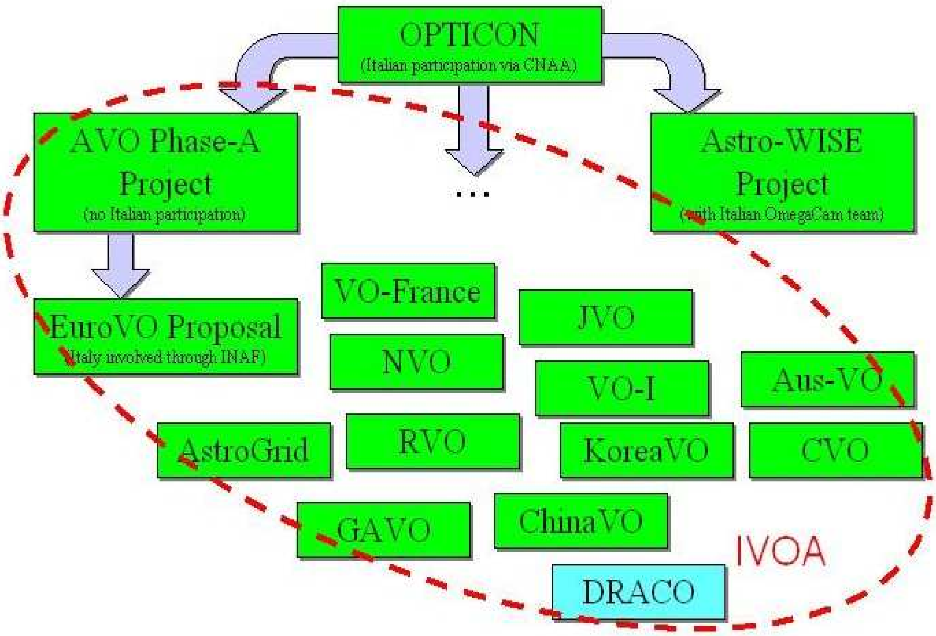

Fig. 2. The astrophysical EU/FP5 OPTICON project and its successors, Astro-WISE and AVO. Other international VO projects forming IVOA (including DRACO) are also shown.

Italy is, and will be, represented in the new Euro-VO EU/FP6 proposals by INAF. The INAF role in the proposal is *"Network member, national node for all Italian data centres"* and theplanned contributed competence is *"Coordination for the Italian data centres, provision of applications"*.

The contributions to the Euro-VO project are envisaged in the following topics:

– access to the prototype TNG Long-Term Archive (LTA (http://wwwas.oat.ts.astro.it/lotar/ Documentation.htm)), located in Trieste, and soon to the full-fledged service;
– access to the ASDC (http://www.asdc.asi.it/) archive of high-energy data, located at Frascati, and in general to the DIANA (http://www.mi.astro.it/docM/reports/ ann2001/ren01/node64.html) system;
– access to a set of distributed IR and radio data (EVN catalogue, Tirgo archive, ...);
– visualization, using the utilities provided by the Cosmo.Lab (http://cosmolab.cineca.it/) project, financed by EU/FP5, and in particular ASTROMD (http://arxiv.org/abs/astro-ph/0006402), a software package for visualization and analysis allowing the possibility of 3D graphical representation;
– "machine learning", which encompasses the use of neural networks, genetic algorithms, fuzzy-C sets; the "AstroMining" (http://wwwas.oat.ts.astro.it/draco/ DRACO-home_file/AstoMining.htm) data mining tool will be used.

The participation of the Italian community assumes an already-established national VO; this is accomplished only partially since, as mentioned before, INAF exploited the Grid.it project, expanded through Cofin03, to "create" DRACO.

3.3. Towards an Italian Virtual Observatory

As a matter of fact, there are indeed plans to build the core of an Italian VO, within the so called Italian facility for Astronomical Archives (IA^2).

In Italy, acute is the issue of handling and storing data coming from the two main Italian optical observing facilities: TNG, located in La Palma, Canary Islands and the Large Binocular Camera to be installed in 2004 at the LBT (Mt.Graham, Arizona). A proposal has been submitted to INAF to build an Italian facility for Astronomical Archives (IA^2), which has as its main purpose the archiving of all TNG and Italian LBC data.

However, there are also some extension plans: IA^2 is proposed to act as a "distributed center", coordinating user access to Italian archives through the international standards defined within the IVOA collaboration.

The steps to establish this first version of an Italian VO are listed below. The purpose is to store and provide uniform access to data of the following facilities:

1. TNG technical and scientific data;
2. provision for LBC / LBT technical and scientific data (in this case an agreement with the LBT Corporation is needed);
3. existing TIRGO observations, radio catalogues and data;
4. harmonization with high-energy space data (through ASDC);
5. mirrors archives (GSC-II data, DPOSS) and solar data;
6. LBC and VST surveys;
7. Other (e.g. local).

The items above are listed in rough order of priority. The priority is defined taking into consideration a number of factors, ranging from data availability (current, future, planned) to technical difficulties in the integration of the archive, to challenges in merging scientifically different classes of data.

An important part of the IA^2 project is to integrate the data access facilities with data processing applications. Reusing developments from other projects (IVOA, DRACO, EGEE) becomes natural. In particular, the integration of the applications is foreseen: visualization (AstroMD), "Machine learning" (neural networks, fuzzy C-sets, genetic algorithms, etc.). The development is planned to follow the Grid paradigm and the IVOA standards.

4. Conclusions – a plan for action

As a conclusion, a basic plan of action can be drawn for the development of the Grid and the VO in Italy, strictly connected to the international endeavuors in these fields.

- Follow closely IVOA technical developments;
- Develop IA^2 to incorporate a number of archives, using IVOA standards and including applications;
- Harmonize IA^2 and high-energy archives through close collaboration between INAF and ASI through the ASDC;
- Harmonize IA^2 and solar archives;
- Use grid technologies as applicable;
- Develop DRACO by integrating applications;
- Participate in EGEE as proper by proposing and integrating relevant applications in an international context;
- Contribute to IVOA and to the Euro-VO FP6 proposal.

In any case, it is fundamental to understand that, to avoid duplication of work, to foster collaboration and, especially, to obtain a more efficient facility for the scientists to use, the Italian Astrophysical Grid and VO activities should be considered to be a single logical project.

Acknowledgements. It is a pleasure to thank Leopoldo Benacchio, who leads the INAF effort in the Grid.it project, and is one of the firmest believers in the need to support research in astrophysics through a grid structure.
The INAF participation in all EU activities is coordinated through its Third Department (Infrastructure and National Facilities); its Director, Giampaolo Vettolani, and the Commissario Straordinario of INAF, Piero Benvenuti, are warmly acknowledged for their suggestions and guidance. Many thanks

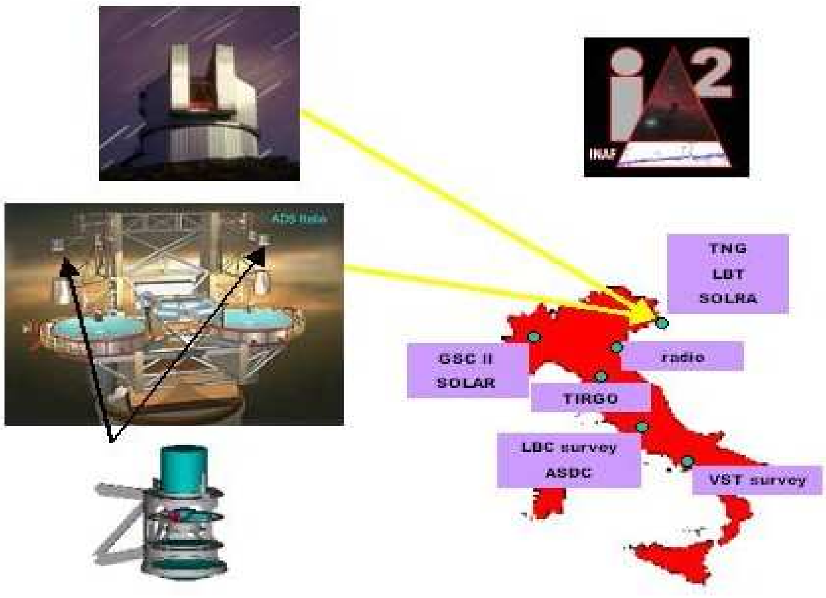

Fig. 3. The Italian facility for Astronomical Archives (IA^2), and its logical evolution to support a number of available data archives covering a large range of the spectrum.

also to Peter Quinn, Andy Lawrence, Françoise Genova and Mirco Mazzuccato for their openness in discussing the needs of the Italian astrophysics community in the framework of European projects, Enzo Valente for important suggestions on networking issues, and Paolo Giommi for useful discussions on the harmonization of INAF archives with the ASDC. Italian Grid and VO activities are funded by the Italian Ministry of Education and Research (MIUR).

References

Benvenuti P. (2002) in: Organizations and strategies in astronomy III, Andr Heck (ed.), Astrophysics and Space Science Library, Vol. 280, p. 107

Butler D. (2002) Nature, Volume 417, Issue 6891, p. 777

Pasian F. , Smareglia R., Benacchio L. (2003) in: Toward An International Virtual Observatory, P.Quinn ed., ESO Astrophysics Symposia, in press

Pasian F. et al. in: Mem. SAIt vol.3 368 2003

Pasian F.,Benacchio L. in: Astronomical Data Analysis Software and Systems XIII, in press

Smareglia R. et al. (2002) in: Virtual Observatories, SPIE proceedings 4846, p.158-169