Mem. S.A.It. Vol. 80, 520
© SAIt 2009
Memorie della

IGI (the Italian Grid initiative) and its impact on the Astrophysics community

F. Pasian, C. Vuerli, and G. Taffoni

INAF - Information Systems Unit and INAF - O.A.Trieste, Via G.B.Tiepolo 11, I-34131 Trieste, Italy, e-mail: pasian@oats.inaf.it

Abstract. IGI – the Association for the Italian Grid Infrastructure – has been established as a consortium of 14 different national institutions to provide long term sustainability to the Italian Grid. Its formal predecessor, the *Grid.it* project, has come to a close in 2006; to extend the benefits of this project, IGI has taken over and acts as the national coordinator for the different sectors of the Italian e-Infrastructure present in EGEE.
IGI plans to support activities in a vast range of scientificdisciplines – *e.g.* Physics, Astrophysics, Biology, Health, Chemistry, Geophysics, Economy, Finance – and any possible extensions to other sectors such as Civil Protection, e-Learning, dissemination in Universities and secondary schools.
Among these, the Astrophysics community is active as a user, by porting applications of various kinds, but also as a resource provider in terms of computing power and storage, and as middleware developer.

Key words. grid – grid computing – data grid – middleware – porting of applications

1. Introduction

IGI is the association for the Italian Grid Infrastructure. It has been established as a consortium of 10 different national institutions and 4 consortia to provide long term sustainability for the Italian Grid. The partners are: the National Institute for Nuclear Physics (INFN), the National Research Council (CNR), the ELETTRA Consortium, the National Institute for Alternative Energies (ENEA), National Institute for Astrophysics (INAF), the National Institute for Geophysics and Vulcanology (INGV), the University of Calabria (UNICAL), the University of Naples 'Federico II' (UNINA), the University of Perugia (UNIPG), the University of Piemonte (UNIPNM), the Sicilian COMETA Consortium, the Sardinian COSMOLAB Consortium, the Consortium for the Italian Research Network (GARR) and the SPACI Consortium.

IGI focuses on setting up and operating a common e-Infrastructure for the Italian Sciences, and includes the main public Resource Providers and Computing Centers, in addition to various Regional Initiatives and other related projects. IGI will provide a consistent and coordinated Italian strategy as a step towards the European Grid Initiative (EGI) and as an interface to the EU Grid infrastructure projects e-IRG (e-Infrastructure Reflection Group) and ESFRI (European

Send offprint requests to: F. Pasian

Strategy Forum on Research Infrastructures), and to other international activities as the need arises.

2. Previous experience

2.1. INFNGrid

The INFNGrid project, approved in late 1999, has been one of the first Grid initiatives in Europe. It has developed and deployed the first Italian Grid Infrastructure, based on GARR, the Italian research network. It is composed of more than 30 sites, both INFN and some among the most important Italian universities.

Although primarily focused on physics, it has been, since the beginning, open to other fields of research (bio-medicine, earth observation, etc.) and to industry.

Among its activities, training and support are of particular importance. GILDA (Grid Infn Laboratory for Dissemination Activities) is a virtual laboratory to demonstrate/disseminate the strong capabilities of grid computing. The Grid2Win project aims at introducing the Grid for Windows Users, by creating a User Interface (UI) and a Compute Element (CE) within the gLite middleware and running on Microsoft Windows. These initiatives have been broadened from the national INFNGrid environment to the whole EGEE user community.

2.2. Grid.it

With a grant received from the MIUR (the Italian Ministry for Education and Research) on the FIRB budget line dedicated to investment in fundamental research, the Grid.it project was implemented in the period 2002-06. This allowed a number of National Research Institutions to pave the way for the development of a common production grid infrastructure supporting the Italian Research Area.

This project had a strong interdisciplinary character, and was aimed at defining, implementing and applying innovative solutions for network computing enabling platforms, oriented towards scalable Virtual Organisations based on the Grid Computing paradigm.

The project was a combination of development and deployment of a Production Grid. Development was particularly focused towards the fields of security, grid-oriented optical switching paradigms, high-performance photonic test-beds, knowledge services for intensive data analysis, data-intensive core services. The work on deployment included the implementation of portals, grid-enabled scientific libraries, programming environment and the porting of applications in the fields of astrophysics, Earth observation, biology, geophysics and molecular virtual reality.

3. IGI

3.1. The transition from Grid.it to IGI

Currently, the Grid.it infrastructure is composed of 50 sites ("resource centers"), 4500 CPUs for a total of 4.5 MSPECINT, over 1 PByte storage, all connected through a network having a 2.5 Gbps bandwidth.

Grid.it has come to a close in 2006 as a development project; to extend its benefits, IGI has taken over and acts as the national coordinator for the different pieces of the Italian e-Infrastructure present in EGEE.

While Grid.it was active, a combination of National and European funding was invested under the framework of the MIUR PON programme to enhance the computational infrastructures in southern Italy. Five grids at the regional level were created: initially the SPACI Consortium (Southern Partnership for Advanced Computational Infrastructures) was established in 2004; two years later it was the turn of the COMETA Consortium (running the TriGrid and PI2S2 projects - Becciani (2008)), of the Cosmolab Consortium (CyberSar project - Porceddu (2008)), of the University of Naples (SCoPE project - Longo et al. (2008)), and of ENEA (CRESCO project). These projects, associated as GRISÙ (GRId del SUd), share the same middleware (the INFNGrid distribution): as a consequence, a complete interoperability among the different projects/consortia and

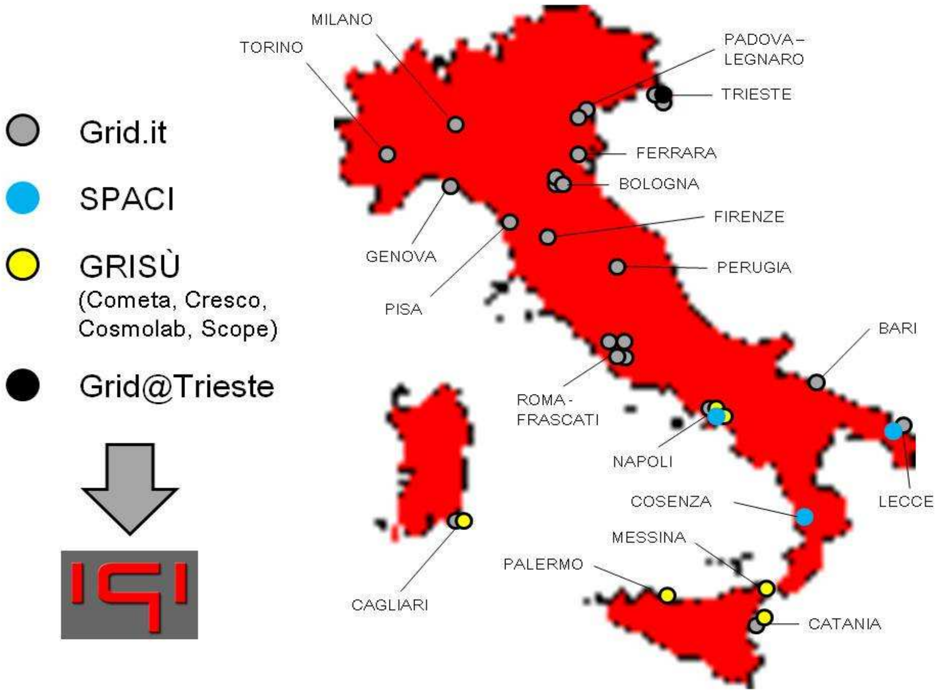

Fig. 1. The Grids in Italy – the expansion from Grid.it and regional/metropolitan Grids to the national infrastructure (IGI)

with the Grid.it/IGI/EGEE infrastructure has been achieved.

In the near future, resources from MIUR will be invested to expand the PON experience to the rest of Italy. This will allow other regional or metropolitan Grids, either existent or planned (e.g. the Grid@Trieste project), to join and share resources within the IGI framework.

IGI builds upon previous work. It plans to leverage upon the large integration and standardization effort of international and national research projects such as EDG, LCG, Grid.it and EGEE, in order to establish a coherent platform of interoperable Grid Services customized for Italian user applications, and make it available in support of the Italian scientific end user communities and, possibly, early commercial adopters.

3.2. The IGI middleware distribution

IGI plans to have its own middleware distribution. The idea is to expand the EGEE standard gLite distribution to better tackle the specific needs of the Italian Grid community.

This initiative builds on the INFNGrid experience: it is totally compatible with respect to the gLite distribution implemented by CERN, but has additional components and specific configurations. The distribution, among other things, includes the following specificities:

- support for 20 Virtual Organisations,
- the StoRM system for the access to storage,
- a Plug and Play (P’n’P) User Interface,
- configured support for the MPI protocol,
- the GridICE system for the monitoring of Worker Nodes.

Before a component is allowed to be inserted in the release, a Board (the Executive Board of IGI) receives the requests and evaluates a number of items: the functionality and added value of component, the reference User/VOrg community, the usage experience and the tests performed, the evolution/development/sustainability of the component. At the end, the Board decides which are the additional components which can be inserted in the release, defining the related priorities.

A Release Team performs a technical evaluation and defines the activity plan and schedule. A tight collaboration with the developers is mandatory for the smoothness of the integration and release process.

Among the distribution planned for spring 2008 there is an INAF-developed product, G-DSE, which is described in the next sections.

3.3. IGI in the future

In spring 2008, MIUR has asked the research institutions to provide their roadmap for Italian research infrastructures. All institutions associated within IGI have jointly submitted a proposal for the Italian Grid Infrastructure. The proposal has been built so as to be coordinated with the section of the ESFRI roadmap dealing with e-infrastructures, and to be complementary with another proposal for the MIUR roadmap, ISI, dealing with the need of an appropriate national High-Performance Computing (HPC) facility.

At this stage, it is expected that appropriate funding will be provided by MIUR to support the IGI infrastructure.

IGI plans to support activities in a vast range of scientific disciplines – e.g. Physics, Astrophysics, Biology, Health, Chemistry, Geophysics, Economy, Finance – and any possible extensions to other sectors such as Civil Protection, e-Learning, dissemination in Universities and secondary schools.

4. Astrophysics as an IGI community

4.1. Activities

Among the various scientific disciplines involved in IGI, the Astrophysics community is active as a user, by porting applications of various kinds, but also as a resource provider in terms of computing power and storage, and as middleware developer.

Taking a step backwards, the experience of the Italian Astrophysics community in the field of Grids dates back to 2002, when a dedicated work-package (WP10) was included in Grid.it. In the 2003-05 period, the DRACO (Data grid for Research in Astrophysics and Coordination with the virtual Observatory) project was active; the goals were porting of scientific applications to the Grid and starting connections with the VObs, seen as a Data Grid. The project was very successful and allowed Italy to join the International Virtual Observatory Alliance (IVOA). Since 2006, two separate initiatives have been supported and funded by INAF for the Virtual Observatory and the Grid for astrophysics: VObs.it and DRACO2 (relying on the Italian grid infrastructure), respectively.

To give an idea of the community's experience in the porting of scientific applications, it is to be noted that within the Grid.it and DRACO projects, quite a number of Montecarlo processes, high-resolution N-body simulations and data processing tasks (in particular, related to OmegaCam) were ported to GILDA first, and then to the Grid.it infrastructures, as was documented in a dedicated national workshop held at INAF Headquarters in November 2005 (Benacchio and Pasian eds. (2007)). Later, within the framework of the EGEE programme to support applications (EGAAP), complete simulations of the Planck mission were run, comprising the generation of ideal microwave skies, the observation by means of a realistic model of the instruments on-board, and the generation of 'observed' full-sky maps. More recently, on the Grid infrastructure of the COMETA Consortium, among other applications, some HPC tasks (e.g. the FLY and FLASH codes) were run. Finally, the porting on IGI (and consequently on EGEE) of the FRANEC stellar evolution code is allowing to populate the BaSTI archive (Cassisi (2008)).

The good experience of the Italian community in the porting of scientific applications on the Grid led, as a consequence, to the leadership of a dedicated Astronomy and Astrophysics cluster within the EGEE-III project, funded by the EU under the Seventh Framework Programme.

But the Italian astronomical community has also gathered some experience in the development of Grid middleware: things started in WP10 of Grid.it, where the community's requirements were defined. Afterwards, an important step to allow astronomers to smoothly use the Grid was achieved by interfacing the well-known CFITSIO library with the gridftp and EGEE file exchanging mechanisms, through the development of the appropriate drivers (Taffoni et al. (2007)).

The key accomplishment in middleware production was the development of the G-DSE system Taffoni et al. (2006), originally in collaboration with the INFN-CNAF group in Bologna, and then brought ahead by the INAF-SI group in Trieste. G-DSE allows 'native' Grid access to relational databases and has been applied (and optimised) for astrophysics. Quite thorough independent benchmarking and comparison with other systems managing DBs on the Grid (AMGA, OGSA-DAI, GRelC) has evidenced the competitiveness of G-DSE, which is by far the most efficient system for large queries (i.e. > 1000 tuples), as evidenced in Vuerli et al. (2008).

It is to be noted that this experience has led, as a consequence, to the leadership of the EGEE Database Group, assigned to Giuliano Taffoni of INAF.

4.2. Looking forward

Up to now, the Grid has mainly delivered computing power, the main issue that its implementers, mostly involved in large High-Energy Physics experiments (e.g. at CERN), needed to solve. Furthermore, last-mile network bottlenecks have prevented many astronomy users from performing massive data transfers, hence hampering the efficiency the Grid could provide.

A common way of looking at e-Infrastructures is to consider the network, the computing facilities, the data infrastructure and the applications as a set of layers standing on the top of each other.

It is clear that scientists are interested in obtaining scientific results: to do so, they interact with their application and, although using them massively, do not really want to interact with the underlying data, computing and network infrastructures. All of these complexities should be hidden by an appropriate layer.

¿From the implementer's point of view, efficiency is the word. The best way to achieve this goal, at the same time satisfying the user requirements, is to integrate all facilities within a single e-infrastructure. Therefore, when defining the roadmap for the development of an e-infrastructure for Astronomy, the network, computing and data components, and to a certain extent also some applications, need to be thought as integrated, or at least fully interoperable (Pasian (2008)).

Assuming these concepts as the basis, the following goals of the Italian Astronomy community for the next years can be defined:

- improving the GARR Network Infrastructure by increasing the bandwidth and eliminating the bottlenecks;
- improving the accessibility of the IGI Infrastructure, while increasing the number of contributed Astro nodes;
- consolidating a Data Infrastructure on top of the Network and Computing Infrastructures;
- implementing domain-specific applications on top of the Data Infrastructure;
- achieving the full interoperation of the relevant applications (data processing, analysis, mining, bibliography, semantic web) to start building a Knowledge Infrastructure for Astronomy.

In some of these areas (Data Infrastructure, interface to Computing, applications, Knowledge Infrastructure) the Virtual Observatory expects to bring in its own standards, tools and facilities. Some of them are likely to be layered on top of Grid middleware, both to facilitate the required integration and to avoid re-inventing the wheel. The Italian Astronomical community will certainly push in this direction.

Acknowledgements. The IGI initiative is led by INFN and the Head of the IGI Coordination

Committee is Mirco Mazzucato. The authors wish to thank Antonia Ghiselli for her friendly continuous support.

References

Becciani U. 2008, The COMETA Consortium and the PI2S2 Project, in: 'Italian HPC-Grid Workshop', Mem. SAIt Supplements, in press

Benacchio L., Pasian F. eds. 2007, Grid-enabled Astronomy, Polimetrica

Cassisi S. 2008, these proceedings

Longo G. et al. 2008, The PON-SCoPE GRID Infrastructure, in: 'Italian HPC-Grid Workshop', Mem. SAIt Supplements, in press

Pasian F. 2008, (Super)computing within integrated e-Infrastructures, in: 'Italian HPC-Grid Workshop', Mem. SAIt Supplements, in press

Porceddu I. 2008, The CyberSAR Project, in: 'Italian HPC-Grid Workshop', Mem. SAIt Supplements, in press

Taffoni G., et al. 2006, Astronomical Data Analysis Software and Systems 15, C.Gabriel, C.Arviset, D.Ponz, & E.Solano eds., Publications of the Astronomical Society of the Pacific, Vol. 351

Taffoni G., et al. 2007, Astronomical Data Analysis Software and Systems 15, C.Gabriel, R.A.Shaw, F.Hill & D.J.Bell eds., Publications of the Astronomical Society of the Pacific, Vol. 376

Vuerli C. Pasian F., Sponza M., Taffoni G. 2008, The Grid in INAF, in: 'Italian HPC-Grid Workshop', Mem. SAIt Supplements, in press