

DRACO and the Italian Participation in Virtual Observatory Activities

Fabio Pasian

INAF - Osservatorio Astronomico di Trieste, Trieste, Italy

Leopoldo Benacchio

INAF - Osservatorio Astronomico di Padova, Padova, Italy

Abstract. The Italian Data Grid for Astrophysical Research (DRACO: Datagrid for Italian Research in Astrophysics and Coordination with the virtual Observatory) is a concept having the aim of providing the scientific community with a distributed multi-functional environment allowing the use of specialized (computing, storage, observational) Grid nodes. DRACO provides the framework through which the Italian astrophysical community participates in the international Virtual Observatory (VO) effort.

1. Introduction - The Grid, the Data Grid and the VO

A new paradigm for accessing and exploiting network-distributed facilities, the Grid, has recently gained importance and momentum; it is common belief that the concept of Grid will be the natural extension of the web, since it allows, besides the passive access to the resources (mainly information) available on the net, an active usage of the resources themselves, e.g. allowing the user to access distributed computing resources. The Data Grid extends this concept, being based on two fundamental services, namely: storage and processing systems, and metadata and communications management. The astrophysical community is implementing this structure by defining, at an international level, the concept of “Virtual Observatory”, which is tightly coupled to the Data Grid paradigm.

2. DRACO and the Grid

An Italian national Grid for research is being developed as a coordinated action involving research institutions (CNR, INFN, INAF, Universities, ...). In 2002, a project called “Enabling platforms for high-performance computational Grids oriented towards scalable virtual organizations” (short name: Grid.it) has been approved and funded by the Italian Fund for Basic Research (FIRB), and provides a technical and organizational framework so as to allow the various projects to operate as virtual organizations, re-organizing as needed the Grid structure (the network and the computing facilities) into logical sub-Grids.

The project, funded by FIRB, has generated a national cooperation named IG-BIGEST (Italian Grid for Business, Industry, Government, E-Science and Technology); the participation of the astrophysical community in this structure is granted by the National Institute for Astrophysics (INAF). From this national framework and as a subset of the IG-BIGEST effort, the concept of an Italian Data Grid for Astrophysical Research (DRACO - Datagrid for Italian Research in Astrophysics and Coordination with the virtual Observatory) has emerged. A plausible scenario in DRACO could foresee a scientist monitoring an observation on a remote telescope, processing in a distributed fashion the data gathered, while comparing them with archived data (i.e. extracted from the Virtual Observatory). It is to be noted that the purpose of the astrophysical section of the Grid.it project (work-package 10) is the implementation of three demonstrators which are to prove the feasibility of porting astrophysical applications within the framework of a national Grid structure. The aim of DRACO is therefore to demonstrate the feasibility of providing the scientific community with a distributed multi-functional environment allowing the use of specialized (observational, computing, storage) Grid nodes.

The three first DRACO nodes are the INAF Observatories of Padova, Trieste and Naples, having the following tasks:

- Padova and Trieste are to provide access to the prototype TNG Long-Term Archive (LTA) developed in the framework of a dedicated pilot project, and to the GSC-II Catalog Consultation System (a web-based application for the access to large astronomical catalogs, mainly the GSC 2.2); interoperability with other data providers and repositories, by co-ordination with international efforts (the OPTICON and AVO projects, funded by the EU) is a strict requirement.
- Naples is to provide astronomers with effective tools for remote access to the VST/OmegaCAM data reduction pipeline.
- Trieste is also involved in providing a Grid-enabled system allowing a transparent access to observing facilities for the monitoring of observations and managing of targets of opportunity.

The foreseen extensions to the original DRACO framework include the extension in the number of nodes to the Observatories of Catania and Rome, Universities of Naples and Salerno, and the integration of other data processing and scientific applications. In particular, the inclusion of visualization (Cosmo.Lab, ASTROMD) tools and "machine learning" applications, which encompass the use of neural networks, genetic algorithms, fuzzy-C sets (the AstroMining tool) is foreseen. A proposal for funding has been successfully submitted to the Italian Ministry for Education and Research.

3. DRACO and the VO

From the data centers included in the DRACO kernel based on the Grid.it project (the TNG Long-Term Archive prototype and the GSC-II Catalog), an extension to other data-providing nodes is foreseen:

- access to the data from high-energy missions and their reduction; in particular, access to the ASI Science Data Center (ASDC, containing among

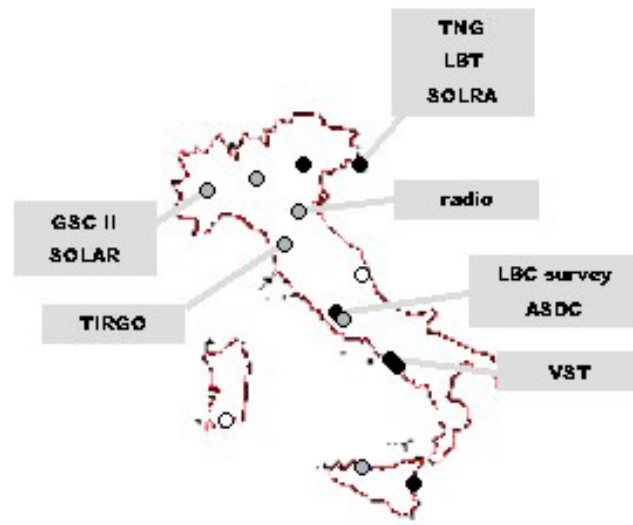


Figure 1. Italian sites in astrophysics: back dots represent DRACO sites active in grid development, grey dots sites hosting data; white dots can be considered as user sites.

others data of the Beppo-SAX satellite) and to the DIANA (Distributed Italian Astronomy Network Archive) system;

- access to a set of distributed radio and infrared data (EVN catalogue, Tirgo archive, ...);
- the archive of the observations made with the Large Binocular Camera (LBC) during the Italian time at the LBT;
- studying access to small archives and databases.

It is to be noted that the TNG LTA pilot will soon become a full-fledged archive providing service to the community, and work on the LBC archive will be started. All centers plan to use the standards defined within the IVOA collaboration to ensure interoperability.

A funding request to ASI has been recently made to coordinate a “distributed center” for scientific data acquired by space-borne instrumentation. Dedicated INAF funding has been set aside for a data center providing access to TNG and LBC data, and is expected to be made available within 2003.

4. DRACO Approach to Interoperability within the VO Framework

An example of the way DRACO intends to proceed in its coordination with international VO efforts can be the prototype TNG Long-Term Archive development activity (held between June 2001 and September 2002). One of the aims of this project was to provide tools and expertise to create and achieve a high degree of interoperability with other archives at the national and international level. The following steps have been followed:

- use of internationally-defined standards in the implementation of the prototype: FITS (the Flexible Image Transport System), ASU (the Astronomical Server URL, a standard to generate queries for retrieving tabular data and catalogues in astronomy), Astrores (a tool describing Astronomical Catalogues and Query Results with XML), a preliminary version of VOTable (an extension of Astrores allowing to take binary data into account);
- installation of a name resolver using the SIMBAD facility;
- definition of a higher level of interoperability based on the availability at the CDS of the TNG catalog of observations.
- link with international projects actively pursued by participation of the TNG LTA group to the Science WG of the AVO project and to the various WGs of the International Virtual Observatory Alliance (IVOA).

5. National and International status

IG-BIGEST participates in an EU FP6 project for the enabling of a pan-European grid for research (EGEE); the astrophysical community (INAF) participates as an unfunded partner. INAF furthermore participates in European VO activities, coordinating with similar activities throughout Europe. DRACO participates in the International VO Alliance (IVOA) and is represented in the IVOA Executive Committee. Coordination of DRACO Interoperability efforts with all other VO data providers and repositories is achieved by co-ordination within international working groups.

Acknowledgments. The original core of DRACO was funded within the Grid.it project by the Italian Fund for Basic Research (FIRB); its extension on funds of the Italian Ministry for Education and Research. The prototype TNG Long-Term Archive was funded by the National Consortium for Astronomy and Astrophysics (CNAA).

References

- Lama, N., Vuerli, C., Smareglia, R., Gasparo, F., Pasian, F. 2004, this volume, 400
- Longo, G., Donalek, C., Raiconi, G., Staiano, A., Tagliaferri, R., Sessa, S., Pasian, F., Smareglia, R., Volpicelli, A. 2002, in: Data Analysis II, SPIE proceedings 4847, 265
- Pasian, F., Smareglia, R., Benacchio, L. 2003, in: Toward An International Virtual Observatory, P.Quinn ed. ESO Astrophysics Symposia, Springer Verlag, in press
- Pasian, F., Smareglia, R., Vuerli, C., Zacchei, A., Lama, N., Benacchio, L. 2003, Mem. S.A.It. Supplements, in press
- Smareglia, R., Pasian, F., Zacchei, A., Caproni, A., Vuerli, C., Longo, G., Bacciani, U., Gheller, C. 2002, in: Virtual Observatories, SPIE proceedings 4846, p.158-169