DAME: Astrophysical Data Mining & Exploration on GRID infrastructure

M. Brescia[1,2], S.G. Djorgovski[3], G. Longo[2], O. Laurino[2]

1. INAF – Astronomical Observatory of Capodimonte, Napoli, Italy

2. Department of Physics Sciences, University Federico II, Napoli, Italy

3. California Institute of Technology, Pasadena, California USA

Abstract

Modern scientific data mainly consist of huge datasets that are gathered by a very large number of techniques and stored in very diversified and often incompatible data repositories. More in general, in the e-science environment, it is considered as a critical and urgent requirement to integrate services across distributed, heterogeneous, dynamic "virtual organizations" formed by different resources within a single enterprise and/or external resource sharing and service provider relationships. The Astronomy and Astrophysics environment has become an immensely data rich field due to the evolution of detectors (plates to digital to mosaics), telescopes and space instruments. Modern approach consists into the federation under common standards of all astronomical archives available worldwide, as well as data analysis, data mining and data exploration applications. The concept is that having this infrastructure completed, its exploitation will allow a new type of multi-wavelength, multi-epoch science which can only be barely imagined, but also poses unprecedented computing problems. The astrophysical DAME project inserted into the domain of SCoPE GRID project, extends this fundamental target, by integrating it in a service oriented infrastructure, including the implementation of advanced tools for Massive Data Sets (MDS) exploration, soft computing, data mining (DM) and Knowledge Discovery in Databases (KDD). The DAME project (http://voneural.na.infn.it), run jointly by the Department of Physics of the University Federico II, INAF (National Institute of Astrophysics) Astronomical Observatory of Napoli, and the California Institute of Technology, makes use of distributed computing environments (e.g. S.Co.P.E. infrastructure) and matches the international IVOA standards and requirements. The result is a service-oriented architecture, obtained by using appropriate standards and incorporating GRID paradigms and restful Web services frameworks, that will have as main target the integration of interdisciplinary distributed systems within and across organizational domains. Currently, the DAME project, already completed in terms of the engineering design aspects, is under implementation of the first official version of the complete data mining & processing tool package, deployed on SCoPE GRID Infrastructure. But a prototype is already available (please visit project website) and successfully tested on significant science cases, as well demonstrated by the already published scientific papers.