Grid infrastructures for e-Science: a use case from Latin America and Europe

Roberto Barbera[1,2,*], Francisco Brasileiro[3], Diego Carvalho[4], Marcio Faerman[5], Philippe Gavillet[6], Herbert Hoeger[7], Maria Jose Lopez Pourailly[8], Bernard Marechal[6,9], Rafael Mayo Garcia[6], Leandro Neumann Ciuffo[2], Alexandre Nobrega Duarte[3], Raul Ramos Pollan[6], Diego Scardaci[2], and Michael Stanton[5]

[1]Dipartimento di Fisica e Astronomia dell'Università di Catania, Catania, Italy
[2]Istituto Nazionale di Fisica Nucleare, Sezione di Catania, Catania, Italy
[3]Universidade Federal de Campina Grande, Campina Grande, Brazil
[4]Centro Federal de Educação Tecnológica, Rio de Janeiro, Brazil
[5]Rede Nacional de Ensino e Pesquisa, Brazil
[6] Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, Spain
[7]Universidad de Los Andes, Merida, Venezuela
[8]Cooperación Latino-Americana de Redes Avanzadas
[9]Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

The EELA (E-infrastructure shared between Europe and Latin America) Project and its successor EELA-2 (E-science grid facility for Europe and Latin America), co-funded by the European Commission, aim at building a bridge between consolidated e-Infrastructure initiatives in Europe and emerging ones in Latin America. This paper presents the experience gained by creating and operating a grid infrastructure for e-Science and discusses how such an environment can cost-effectively address regional problems and enhance research in developing countries.

***Index Terms*— Grid, e-Infrastructure, e-Science, National Research and Education Networks (NREN), Joint Research Unit (JRU), National Grid Initiative (NGI).**

I. Introduction

WORKING AS a researcher in the so-called "emerging" or "developing" countries is not always easy. Often, severe budget constraints prevent academic and research institutions to compete on terms of equality with institutions from developed countries. Such a situation is well emphasised by the comparison of the yearly numbers of patents and scientific publications from these countries. In the field of computing, a striking example is given by the ranking of the top 500 supercomputers in the world, issued on June 2008 [1], where only 1 out of 500 comes from South America (although it is 138th in the ranking). One of the well-known side effects of this situation is the brain drain phenomenon.

However, emerging grid technologies are becoming to play an important role in providing substantial distributed computing resources to scientific collaboration members, regardless of their geographical location. Thus, institutions without enough funding to afford supercomputers or to set up computer farms can benefit from available grid infrastructures to perform their research activities in suitable conditions.

As African and Latin American countries have many similar objectives in various research fields, exchange of experience on developing grid infrastructures to address regional problems, e.g. in areas of common interest, such as drug discovery, biomedicine and weather forecast, can be very valuable for both regions. Indeed, the development of these e-Infrastructures in Latin America and Africa can open to the scientific communities not only the access to High Performance Computing (HPC) platforms but also the possibility of deploy applications of special interest in their regions having a high social impact.

With this viewpoint, this paper is presenting the EELA and EELA-2 Projects experience of creating, operating and supporting a powerful and functional grid infrastructure for enhanced e-Science collaboration between Europe and Latin America.

II. Objectives

Since the late 80's, the federation of European National Research and Education Networks (NREN) under the auspices of the European Commission (EC) resulted in a Pan-European high-speed network with full administrative and operational support. The existence of such a network catering for the

* Speaker and corresponding author (email: roberto.barbera@ct.infn.it).

European Research community has paved the way for the development of an overlying grid infrastructure providing distributed computing power and storage sharing across geographical domains. This integrated network and processing/storage environment – e-Infrastructure –, provides a platform for new methods of global collaborative research – e-Science. The leading initiatives in the e-Infrastructure field are the GÉANT [2] and EGEE [3] projects.

The objective of the EELA Project [4] has been to build a Latin American e-Infrastructure at the level of European standard, by means of an EC Specific Support Actions (SSA) program. With the networking infrastructure reaching stability through the ALICE [5] and RedCLARA [6] projects, the focus of EELA was on setting up a grid infrastructure and deploying on it some exemplar EU-LA scientific applications. Over its 2-year duration EELA proved that the deployment of an EU–LA e-Infrastructure was not only viable but it was also responding to a real need of several scientific communities.

EELA-2 [7] is a so-called "Combination of Collaborative Projects & Coordination and Support Actions" EC project. Its objective is to move a step further i.e. building, on the EELA e-Infrastructure, a high capacity, production-quality, scalable grid facility providing round-the-clock, worldwide access to distributed computing, storage and network resources for a wide spectrum of applications from European and Latin American scientific communities.

The project especially focuses on two goals:

- Provide an empowered grid facility with versatile services fulfilling application requirements;
- Ensure the long-term sustainability of the e-Infrastructure beyond the 2-year term of the project.

These ambitious goals translate into specific objectives defining the EELA-2 mission:

Building the grid facility by:

- Expanding the former EELA e-Infrastructure to consist of about 30 Resource Centres mobilising about 3000 computing nodes, 700 Terabytes of storage space;
- Providing, in collaboration with related projects (e.g. EGEE), the full set of grid services needed by all types of applications in their scientific environment;
- Supporting various applications:

o Selected against well defined criteria (including grid added value, suitability for grid deployment, scientific outreach, potential social and industry impacts);

o Belonging to various classes, e.g. from classical offline data processing up to control and data acquisition of scientific instruments, in several fields such as Life Sciences, High Energy Physics, Earth Sciences, etc.

Ensuring the sustainability of the grid facility, especially in Latin America:

- Through contacts with policy / decision makers, collaborating with RedCLARA and NRENs and supporting the ongoing creation of National Grid Initiatives (NGI);
- Building the support of the e-Infrastructure to provide a complete set of Global Services from a Central Operation Centre, thus preparing the creation of Regional Operation Centres in Latin America;
- Attracting new applications with the success stories of the early adopters;
- Making available knowledge of the EELA-2 grid facility to all potential users, developers, and decision makers through an extensive Training and Dissemination programme;
- Creating knowledge repositories federated with the EGEE ones.

III. The EELA-2 Consortium

The EELA-2 Consortium encompasses 16 partners (6 from Europe, 9 from Latin America and 1 International Institution CLARA), from 14 countries (5 from Europe, 9 from Latin America), as shown in Figure 1, corresponding to 53 Institutions, most of them clustered in 9 national JRUs, assembling all grid efforts in Latin America and in several European member states. It is coordinated by the same institution (CIEMAT - Spain) as EELA, thus preserving a experienced team on the Latin American – European stage. The fact that the Consortium has grown with respect to the core EELA Institutions, integrating new scientific collaborations, is easing the crucial phase of starting up the various activities. EELA people and institutions with proven competences generally keep their responsibilities in the EELA-2 activities. The majority of the newcomers are from new communities with as main objective the running of their application(s) on the grid. The changes suggested by past experience and also by the migration of project type from "Support Specific Action" to "Combination of Collaborative Projects & Coordination and Support Actions" have been worked out. Therefore, the necessary steps have been well prepared to guarantee an efficient "mise en oeuvre" and running of the various activities.

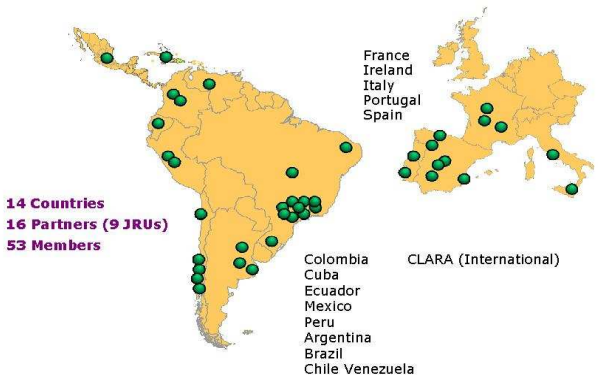

Figure 1: Geographical distribution of EELA-2 Partner / Member Institutions.

IV. Applications

The applications selected to run on the EELA-2 e-Infrastructure have been chosen against criteria maximising the

anticipated added value, the potential impact in the scientific community and in the social context, the possible Industry interest and the expected visibility towards decision makers. This choice was a result of combining the following selection criteria:

• Suitability to be deployed on the grid due to very stringent computing requirements;

• Maturity, i.e. easiness to be ported to the grid;

• Outreach relevance to the user community, especially in Latin America.

About 50 applications have so far been selected from Life Sciences (45%), High Energy Physics (HEP) (14%), Earth Sciences (14%), engineering, e-Learning, civil protection, environmental sciences and other fields, as pie-charted in Figure 2.

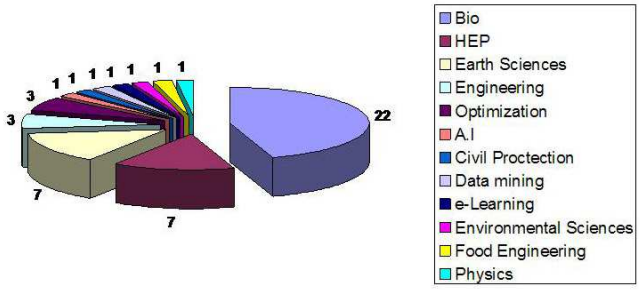

Figure 2: Breakdown of EELA-2Applications per scientific area.

The expected impact of running applications on the EELA-2 e-Infrastructure can be estimated from the expectations expressed by the application's users themselves, as summarized as follows:

• Scientific outcome: There is a large consensus that the grid model is (will be) very useful and should open new perspectives for Applications;

• Social impact: About half of the applications will have positive social outcomes, especially in the Biomedicine (82%) and Earth Sciences (70%) domains, as the grid will contribute to enhance their socially useful results by providing more accurate predictions, more precise diagnosis, faster sampling rates, more complete measurements, etc.;

• Industry spin off: A rather large fraction of applications (80%) could trigger Industry spin- offs. Arguments are:

o Biomedicine: Although there is no clear link yet with industry, the belief is that grid computing should (will) help developing better medical methods and new tools for observation, diagnosis, therapy, etc.;

o HEP: grid is already boosting simulation studies, leading to more efficient detector R&D often made in collaboration with Industry;

o Earth Sciences: grid simulations and processing of real data is definitively enhancing forecasts (seismic, weather, water reserves, etc.) for agriculture and industry activities.

In reality, Industry is usually outsourcing studies (through University-Industry contracts) to University teams already used to grids. The intention of all grid projects is now to bring Industry (especially SMEs) to directly use their grid infrastructures.

V. Outcomes

At the final EELA Project review, in February 2008, the European Commission acknowledged that the major EELA outcomes have been to:

• Successfully build a reliable and almost "production quality" e-Infrastructure in Latin America;

• Provide European and Latin American researchers with early access to a new, well-supported e-Infrastructure and enable them to speed up the processing of their scientific data;

• Favour the entry of, or consolidate Latin American communities in worldwide collaborations;

• Launch an aggressive Dissemination & Training program to develop the grid expertise necessary to autonomously support the Latin American grid on the long term;

• Amplify the e-Infrastructures relevance, blazing the trail towards Latin American e-Science initiatives and/or NGIs.

As a result, EELA has gotten the highest European Community ranking. The EELA-2 mission is to extend and consolidate the e-Infrastructure by:

• Expanding the current EELA infrastructure resources;

• Providing the full set of grid services needed by all types of applications;

• Collaborating with NRENs and creating autonomous Regional Operation Centres;

• Supporting actively the creation of National Grid Initiatives (NGIs) to take the responsibility of managing grid infrastructures;

• Assessing the financial & management schemes to operate and support the e-Infrastructure on the long range beyond the project term.

VI. Conclusions

With its very dynamic work plan, EELA-2 is drastically changing the perspective about e-Science in Latin America, "building" a human network by creating/fostering scientific collaborations across the Atlantic.

Thanks to the EELA project, a mature grid improved the scientific collaborations already established between European and Latin American scientists and helped creating new ones, further developing the research in both continents. Moreover, the e-Infrastructure built paved the way for a more ambitious project, EELA-2, which is in charge of building an enhanced e-infrastructure that should be used by strongly interacting scientific communities distributed worldwide.

Starting from scratch, the EELA Consortium learned how to deal with the European Commission rules, how to submit a proposal, how to negotiate it and how to run a multi-purpose project that involved the set up of a grid infrastructure, the use of large bandwidth networks, the "gridification" of

applications, the training of new communities and the dissemination of the "grid culture" amongst academia, industry and decision makers.

What has been done in Latin America can perfectly be "exported" to Africa and the EELA-2 Consortium is willing to share with African institutions/communities/countries resources and knowledge. It is worth mentioning that 3 EELA-2 partners, RedCLARA [6], RNP [8] in Brazil and REUNA [9] in Chile, are ready to help African initiatives in networking matters and that EELA-2 partners, CNRS [10] in France and INFN [11] in Italy, are already collaborating with African Institutions in the field of grid computing.

In fact EELA-2 is open to new member Institution through the signature of a Memorandum of Understanding (MoU) defining the terms of the collaboration, such as the conditions of use of the EELA-2 Infrastructure, the possibility to get trained to deploy and run application(s) on it and the means to receive technical support in order to integrate local computing resources into the EELA-2 grid.

Two sides of the triangle "Africa - Europe - Latin America" already exist and EELA-2 can contribute to build the "Africa - Latin America" one.

A complete technical description of the project can be found in its work plan [12].

REFERENCES

[1] www.top500.org
[2] www.geant.net
[3] www.eu-egee.org
[4] www.eu-eela.org/first-phase.php
[5] www.alice.dante.net
[6] www.redclara.net
[7] www.eu-eela.eu
[8] www.rnp.br
[9] www.reuna.cl
[10] www.cnrs.fr
[11] www.infn.it
[12] http://documents.eu-eela.org/getfile.py?docid=1066&name=EELA-2-DoW-v1.4_Public&format=pdf&version=1