**OSDC/PIRE Research Proposal**

<u>**Smart Grid and Cloud Computing Laboratory (SGCLab) and MediaLab**</u>

The research interests of the SGCLab focus on:
1. the implementation of tools and middleware to facilitate the adoption of cyber-infrastructures by scientists from a diverse range of areas;
2. the parallelization and grid/cloud enabling of existing scientific applications;
3. the development of novel implementations of applications through the integration of emerging distributed computing technologies, and;
4. questions related to security in cyberinfrastructures.

Since 2002, the SGCLab has been operating clusters, grids and clouds, developing autonomic scientific applications and providing support to scientists. The laboratory has also participated in a number of Brazilian and European sponsored projects, and provides International Grid Trust Federation (IGTF) Certificate services to the Latin American grid research community.

## Title of the Research: Accelerating Biodiversity Science with Hybrid Computing

The United Nations have reiterated that climate change, energy, biodiversity and sustainable development are the most demanding challenges of our time and require an urgent and extraordinary global response. Coupled with this, Europe and Brazil have recognized the need to work together towards achieving the global target of significantly reducing the current rate of biodiversity loss. The project EUBrazilOpenBio - Open Data and Cloud Computing e-Infrastructure for Biodiversity plans to deploy an e-Infrastructure of open access resources supporting the needs of the biodiversity scientific community. Tackling the complexity of Biodiversity Science requires dealing with multiple multidisciplinary datasets spanning from climatology to earth and biological sciences, all of key importance in overcoming the fragmentation and focusing on uniting different European and Brazilian data sources to provide scientists with an even greater knowledge base, achieved through the integration and shared use of appropriate computing resources.

The objective of this research proposal is to assist in improving the performance of two tools used by scientists. Adequate access to taxonomic information is essential for a proper management of the global problems related to environmental change, nature conservation and the sustainable use of biological resources. One major problem in world scale taxonomy is how to integrate regional taxonomies created locally for regional floras and faunas with global taxonomies linked to global monographs and global species databases. In this regard, a cross-mapping tool is used to analyze two taxonomic checklists in order to obtain the relationships between their taxa, and identify missing species. Another tool is Ecological Niche Modeling (ENM) that is a widely used approach to predict and understand the distribution of species and to investigate, for example, the set of ecological requirements for a certain species to survive and maintain viable populations over time. Unfortunately, such scientific studies with these tools can be extremely computational expensive, where a single execution can take several weeks on a single computer. We wish to develop and evaluate new parallel implementations that take advantage of state-of-the-art computational environments available (e.g. clusters of multicore processors with hardware accelerators such as GPUs).

Candidates should be interested in programming with C/C++, MPI, OpenMP and CUDA. Knowledge of (or willingness to learn about) MapReduce and programming with databases might be useful.