BIOMEDICAL APPLICATIONS

Biomedicine is a major application area for the EGEE project. With more than 20 applications deployed and being ported, it is subdivided into three areas: medical image processing, biomedicine, and drug discovery, all of which already have many individual applications deployed on the EGEE infrastructure.

These applications stress the middleware with specific requirements, especially related to security (sensitivity of data), data management (complex data structures and distribution), and execution of large numbers of small jobs with data intensive needs. The biomedicine applications are now established as regular users of the infrastructure: as a consequence, the virtual organization gathering users from the life sciences and medical research communities is the biggest consumer of the infrastructure resources after the four LHC experiments.
Below is an overview of the biomedical applications currently deployed on the EGEE infrastructure.

The **medical imaging** sector targets the computerised analysis of digital medical images. It includes medical data federation, compute-intensive medical procedures, processing large data sets and statistical studies over large populations.

- **GATE** is a Monte Carlo-based simulator for planning radiotherapy treatments based on patient images. It uses the EGEE Grid infrastructure to reduce the time to needed complete Monte Carlo simulations to a value reasonable for clinical use.
- The Clinical Decision Support System (**CDSS**) uses image classification based on expert knowledge to aid clinical decisions. The Grid is exploited both for collecting large data sets and for efficiently training classification software over these large data sets.
- The **Pharmacokinetics** application studies the diffusion of a contrast agent in the liver from sequences of magnetic resonance images. Artefacts due to the movement of the patient make images directly incomparable. However the parallelised image co-registration computations running on the Grid allow analysis of the sequence in a reasonable time.
- **SiMRI3D** is a Magnetic Resonance Imaging simulation for the production of artificial but realistic 3D Magnetic Resonance (MR) images to analyse images from perfectly known sources, study artefacts, and further develop and optimise MR sequences.
- The **gPTM3D** application allows interactive reconstruction of 3D medical images, e.g. for the volume reconstruction of large or complex organs. The quality-of-service required for interactivity means that some sites on the Grid have to define a high-priority for this class of jobs.
- The **Bronze Standard** is an application to evaluate medical image registration algorithms. The amount of data to manipulate and the cost of computations are out of reach for standard computers, yet the application can easily be distributed over a Grid.
- The **SPM** software package is used by the neurological research community for the early diagnosis of Alzheimer's disease. It is based on the comparison of the candidate case to a large set of normal cases. Grid technologies allow easy access to distributed data as well as to distributed computational resources.
- **SEE++** is a software for the biomechanical 3D simulation of the human eye and its muscles. It simulates the common eye muscle surgery techniques in a graphic interactive way that is familiar to an experienced surgeon to deal with the support of diagnosis and treatment of strabismus.
- **ThIS** is a Therapeutic Irradiation Simulator based on the GEANT4 toolkit. It simulates the irradiations of living tissues with photons, protons or light ions beams for cancer therapy. Monte Carlo simulation is parallelized over grid resources for efficiency.
- The Medical Data Manager (**MDM**) is a high level middleware service, tightly coupled to the gLite middleware, for secured medical data management. It proposes a DICOM-to-grid data management system interface, medical metadata management and high security.

Last Update: 20/09/2007

EGEE-II is a project funded by the European Commission - contract number INFSO-RI-031688

The **bioinformatics** activity targets gene and protein analysis, including genomics, proteomics and phylogeny. The 21 bioinformatics applications are now established as regular users of the infrastructure. Our main goal is to build a bioinformatics community of scientists using the Grid and to provide them with common biological databases and tools on the EGEE platform. We are collaborating with related projects BIOINFOGRID (EU-FP6), EELA (EU-FP6), SwissBioGrid (NGA) to port a broad spectrum of their applications to the EGEE production grid, and also with the European Network of Excellence EMBRACE (EU-FP6). Some common works are, for example, porting and running the PyBioS application about System Biology on EGEE Grid (Coll. MPI-MG & CNRS IBCP), or refining the whole protein structures of the PDB database (Coll. CNRS LCP and IBCP, CMBI, SIB, Uppsala Univ.).

- **GPS@, Grid Protein Sequence Analysis**, is a Bioinformatics Web portal providing biologists with several Bioinformatics Resources on the Grid through the proven NPS@'s convenient Web interface. The Grid complexity is completely hidden for several famous databases (SwissProt, TrEMBL, PROSITE) and tools (BLAST, FASTA, SSearch, ClustalW, …).
- **Systems Biology on the Grid**. PyBioS is a simulation and modelling platform, automating construction of large biological network models through a database interface to common biological pathway databases.
- **BioDCV** is a molecular oncology application for the analysis of microarray and proteomic data with Support Vector Machine (SVM) classifiers. Its outcomes are predictive classification models, ranked lists of biomarkers and estimates of accuracy on novel data.
- **SPLATCHE** (SPatiaL And Temporal Coalescences in Heterogeneous Environment) is a cellular tool for genome evolution modeling, allowing the reconstruction of the global spread of past humans in a genographically realistic landscape.
- **BiG** is a Grid-enabled service to access large-scale BLAST operations, trough Web portal of Web Service interfaces . It uses mpiBLAST as processing engine and permits launching BLAST operations on different databases simultaneously.
- **Superlink-online**. The group is developing a tool for genetic linkage analysis, Superlink online. Genetic Linkage Analysis is about hunting for disease-provoking genes.
- **Mlalign2D and MLrefine3D** is the port to the Grid of these two key image processing applications of electron microscopy, which permit structural characterization of macromolecular assemblies in distinct functional states.

The **drug discovery** sector aims to help speed up the process of finding new drugs through *in silico* simulations of protein structures and dynamics.

- The **WISDOM** initiative runs large-scale computations for the *in silico* drug discovery against emerging and neglected diseases. These molecular docking calculations determine how well certain drugs attach to specific sites on the target virus – those which dock are more likely to be active against the virus. It has been successfully deployed against malaria and avian influenza, with exciting results confirmed in vitro..
- **GridGRAMM** is a simple interface to do molecular docking on the Web. Results include a quality score and various access methods to the 3D structure of the complex. Molecular docking can be used for the study of molecular interactions, to analyze enzyme-substrate interactions, for drug design and to understand morbid mutant behaviour.
- The goal of **GROCK** (Grid Dock) is to provide an easy way to conduct mass screenings of molecular interactions using the Web by allowing users to screen one molecule against a whole database of known structures.

EGEE is keen to consider other applications. For further information on how to participate, as well as more information about the applications running on EGEE, visit the User and Application Portal at http://egeena4.lal.in2p3.fr/.

Last Update: 20/09/2007