Workflows in Molecular Sciences

Information and Communication Technologies (ICT) are dramatically changing the way Molecular Science knowledge is produced and used. ICT offers not only computing resources but also unprecedented access to data sources and simulation tools. Molecular Science community moved towards widely using High Performance (HPC) and High Throughput (HTC) computing establishing Virtual Organizations (VO)s, Virtual Research Environments (VRE) like MOSGrid, and a Virtual Research Community (VRC) in Chemistry, Molecular Materials Sciences and Technologies using Distributed Computing Infrastructures (DCI) such as clouds and supercomputers. Molecular Science simulations run codes on DCIs exploiting their computing and data resources raising challenges for both Computational Chemistry and Distributed Computing. There are the following challenges for the Computational Chemistry community:

- building up a knowledge base of molecular simulation techniques and protocols covering the whole range of high-level *ab initio* quantum dynamics over medium scale density functional approaches to force field based methods for large simulation systems;
- bridging different molecular scales to allow quantum dynamics treatment of larger molecules and speed-up for quantum chemical tasks through force-field based preparatory steps;
- running dynamics and kinetics simulations reusing the electronic structure information generated by ab initio calculations for the database;
- standardising the formats of primary raw and processed scientific data to support sharing and using different types of computing resources and simulation tools; and,

To meet these challenges the Distributed Computing community must:

- representing simulations as workflows enabling hardware virtualization, software containerization and access to data resources to improve their reusability and reproducibility

- optimizing data processing across heterogeneous DCIs putting emphasis on dynamic scalability of computational resources and considering their security requirements;

- managing the whole data life-cycle from primary experimental data to annotated scientific data enabling data reusability and reproducibility using metadata and intelligent data storage;

- creating customized VREs as a Service (VaaS) to support domain researchers in creating and running Computational Chemistry simulations.

The presentation will outline how the Distributed Computing community can address these challenges.