BioVeL: A BiodiVersity Laboratory exploiting EGI computing infrastructure.

Tuesday, 20 May 2014 11:30 (15 minutes)

BioVeL is a virtual e-laboratory that supports research on biodiversity issues using large amounts of data from cross-disciplinary sources. BioVeL offers the possibility to use computerised "workflows" to process data, be that from one's own research and/or from existing sources.

A researcher can build his own workflow by selecting and applying successive "services" (data processing techniques), or he can re-use existing workflows available from BioVeL's library. This virtual e-laboratory cuts down research time and overhead expenses. BioVeL also provides access to a world-wide network of expert.

Within the project several activities have been carried out to exploit successfully the computing resources that EGI infrastructure is able to provide. We will show the development done in order to exploit standard grid computational resources within a Taverna workflow, together with services deployed in the EGI Federated Cloud Task Force.

Wider impact and conclusions

Thanks to the activities carried on in BioVeL project the researchers involved in the biodiversity studies are able to exploit several different computing infrastructure, like: EGI grid, Desktop Grid, local cluster, cloud resources provided by EGI Cloud Task Force, dedicated servers or public available services. All those facilities are easily and seamless available to the end users and could be used to build complex data processing workflows allowing also non expert users to process big dataset of data without the need to learn the complexity of each computing infrastructure. The choice to use standard interface and technologies like a web services and cloud based resources, give the possibility to re-use this kind on infrastructure also in other context with different users communities and different tools. Moreover the work done at the level of user interface has greatly improved the users experience reducing the time needed to achieve the analysis results.

Description of work

BioVeL community is interested in building complex workflow by means of Taverna workflow manager and re-use those workflow in a simple web interface like the one provided by Taverna Server.

In this work we will describe the activities carried on within the BioVeL project in order to build robust web services that could handle the users requests both in terms of application execution and data management.

The choice to use standard REST and SOAP web services interfaces guarantee the users to be able to exploit in the same workflow public available services (GBIF, google refine, BioSTIF, etc), or ad-hoc developed services that could be deployed on: dedicated server such as GPU enabled server, local batch farm, EGI grid, EGI Federated cloud task force, boinc enable desktop, etc.

In this way we can easily handle different kind of application from a fast and simple file parsing to a huge campaign of comparing two big dataset of biological data, or a long-running MPI multi CPU application.

We will show also the work done in order to let BioVeL to exploit successfully several services deployed within the EGI Federated Cloud Task Force. The deployment of those services in the EGI Task Force is of great importance in order to provide dynamically the required computational power to the analysis workflows.

The Taverna portal is also deployed on a OpenStack based IaaS, we will describe here, also, the relevant activities done in this field.

We will show also the development done in order to ensure a simple and user friendly interface to the final users providing the possibility to easily exploit already developed data processing workflows with small intervention useful only to customize the single execution.

The framework build within BioVeL project, is able to deal with automatic job submission, managing and resubmission to hide the complexity of computing infrastructures.

Primary authors: GOBLE, Carole (MANCHESTER); Dr SIPOS, Gergely (EGI.EU); Dr DONVITO, Giacinto (INFN); FERREIRA, Nuno (EGI.EU); PINTO, Salvatore (EGI.EU)

Presenter: Dr DONVITO, Giacinto (INFN)

Session Classification: Environmental science on grids and clouds

Track Classification: Success stories in using e-Infrastructures for research (Track Leaders: E. Katragkou, P. Castejon)