ESCOGITARE Project: Towards a Grid-based e-Science Infrastructure for the Italian Agricultural Research Council (CRA)

P. Faccioli[2], D. Laforenza[1], R. Lombardo[1], M. Maggini[3], M. Scarpellini[1], M. Serranò[1]

[1] Information Science and Technologies Institute (ISTI) National Research Council, Pisa, ITALY

[2] Agricultural Research Council (CRA)-Istituto Sperimentale per la Cerealicoltura, Sezione di Fiorenzuola d' Arda, ITALY

[3] Agricultural Research Council (CRA), ITALY

Abstract

Advances in agricultural research have resulted in massive increases in the volume and variety of available data. In particular, the complexity of biological information necessitates powerful bioinformatics computational resources both at the level of storage and analysis of data. Moreover, the heterogeneous bioinformatics-related knowledge of scientists involved in agricultural research makes it fundamental to provide an in-situ training program dedicated to the most utilized computational tools. All these challenges are being tackled by the "$ESCOGITARE$" project which should provide an e-Science based, virtual lab structure for plant and animal genomics to help in the understanding of important traits such as yield, stress resistance and quality. The recently founded Italian Agricultural Research Council can be considered as a good example of application of such a collaborative research approach due to the number of the belonging institutes and to the wide range of competences characterizing each of them.

Aim of the "$ESCOGITARE$" Project is thus to provide a Grid-based infrastructure to CRA Institutes in order enable them to find and share their know-how, data and computational power in a collaborative work context. This target will be reached in this project by using both Grid and videoconference technologies. In particular, we are implementing Grid infrastructure prototype to allow discovery, management and secure access to a set of high-level services compliant to the Web Services Resource Framework (WSRF) specification. In order to avoid starting from scratch $GlobusToolkit4$ Core middleware was chosen. Web Services can be used to virtualize computational and data resources in order to allow workflow composition by using an orchestration approach. End users can create a specific workflow by using the $Active\ BPEL\ Designer$ tool and a collection of dynamically discovered high level services description. The output of this operation is an abstract sharable workflow where each node corresponds to a web service invocation. At execution time every action in the

workflow will be scheduled to a particular instance of a related general Web Service description and the workflow result can be made available on the portal. We are also implementing a set of administration and configuration tools in order to easily add and manage a node to the Grid. Future works foreseen wrapping of mostly used bio-informatics tools like MPI-Blast as an high-level service. In order to make easier the workflow composition to scientists, we intend to develop a web-based workflow designer tool embedded in the "$ESCOGITARE$" portal.