

Workflow Systems

As grid technology has improved, its users have ported larger and more complex applications to the grid. Often the increased complexity manifests itself as a workflow, a graph of interrelated executables, that defines a complete analysis. Tool designers have recognized this need, and currently, there are a number of sophisticated workflow management systems available. The developers have tailored the systems to particular use cases and to particular scientific domains, increasing efficiency but introducing interoperability problems.

Mr. Tamas KUKLA described the traditional methods to achieve interoperability between different software systems and workflows in particular: standardization and translation. As standard languages, neither BPEL nor XPDL has been widely adopted. For workflows, using translation to achieve interoperability is difficult because of semantic differences between workflow engines. He advocates using a meta-workflow system. He presented a use case where Triana, Taverna, and Kepler workflows were combined using GEMICA and the P-GRADE portal. For users, this is a promising strategy that works with the current infrastructure.

Mr. Viet TRAN described the capabilities of the K-wf Grid workflow engine for EGEE users. As the system is based on web services technology, it required some adaptation for the gLite services. This workflow engine has been ported to the EGEE infrastructure and will be used within the earth science community. In contrast to others, it is a semantic workflow engine that can use metadata descriptions to deduce the workflow description necessary to generate the desired data. Once the metadata catalogs for a scientific domain have been created, the K-wf Grid workflow engine can relieve users from dealing with the low-level details when using workflows.

Hiding differences between grid infrastructures and federating diverse sources of data are critical for efficient production of scientific results. Dr. Maria MIRTO showed developments from the Proteomics and Genomics Grid Project to provide a grid problem-solving environment for bioinformatics. The system allows users to define complex workflows while hiding differences between gLite, Unicore, and Globus for computational resources and federating different data sources with the GRelC system.

Mr. George GOULAS presented a Java-based SchedScripter framework that allows easy implementation of applications using master-worker patterns, swarm patterns, BPEL workflows, and direct web service integration. This system uses pilot job techniques to minimize latencies and to make efficient use of resources. As a case study, an existing exam scheduling application was ported, showing the benefits of using SchedScripter and the grid infrastructure. Providing such high-level tools is another mechanism for interoperability for users.

The sophistication of workflow engines available to users of the grid infrastructure continues to improve. Unfortunately, the numerous workflow engines are largely incompatible—a situation that is likely to continue for the foreseeable future. Fortunately, solutions, like using a meta-workflow manager, allow different workflow engines to be used together and hide differences between resources, permitting users to perform their analyses efficiently while using the widest range of resources.