



Contribution ID: 150

Type: Oral

## Workflows Description, Enactment and Monitoring through SAGA

*Thursday 15 April 2010 10:00 (20 minutes)*

Most existing workflow systems are tightly integrated with middleware, which limit wide scale adoption and may limit efficient execution of workflows. Whereas the use of SAGA enables the use of multiple heterogeneous resources, there exists an initial but somewhat limited support for workflows using the SAGA specification. Thus to reconcile the advantages of the SAGA paradigm with broad usage of workflows as a way of composing applications, this project investigates approaches that extend the SAGA-workflow package (Digedag) to support workflows such as neuroimaging analysis in neugrid.

### Detailed analysis

Digedag will be extended to support the efficient execution of broad range of applications such as neuroimaging in neugrid, in an extensible, flexible and robust manner. Currently digedag does not have support to coordinate with a EGEE workload management system to schedule a series/sequence of jobs as per the requirements of a workflow. This limits the effective management, monitoring and executions of workflows. The proposed system will implement the following in the Digedag workflow engine:

1. Users may write workflows using the SAGA workflow package which can be enacted and scheduled using appropriate middleware adapters.
2. The submitted workflows can be managed and monitored as a single entity.
3. SAGA workflow package will take care of all the dependencies and execution will be subject to user preferences and dependencies
4. The sequence of jobs and order will remain intact even if a workflow is distributed across sites for execution.

Although we will initially establish and demonstrate these benefits in the context of neugrid based neuroimaging, the advantages will extend to a range of DAG-based and other workflow applications, thus impacting a number of communities.

### Conclusions and Future Work

Using neuroimaging as an exemplar, SAGA will provide support for workflow enactment. Users will translate workflow descriptions into a SAGA API, which can then be enacted using the appropriate execution engine and scheduled, distributed and executed in multi-site environments.

### Impact

The proposed solution will reduce job submission and scheduling latencies. The user will not have to break the workflow into jobs and then submit and schedule them individually. The user will write the workflows in a workflow authoring environment which will be translated into appropriate SAGA structures. Users will also be able to monitor and manage hundreds of jobs in a workflow as a single entity. This will help the users to keep track of execution, failures and outcome of dependencies. Scheduling will also become effective as a higher level enactor, Digedag, will coordinate with a Grid wide scheduler to plan and schedule the jobs.

As a consequence of the proposed project, user workflows can be developed programmatically, and can be enacted to use any available middleware without restricting the specification to a particular execution engine. This flexibility and extensibility along with robustness is critical to enable applications to utilize infrastructure at scale. Users will be able to get features such as “write once and run anywhere”.

## **Keywords**

SAGA, Workflow Description, Monitoring and Enactment

## **URL for further information**

<http://forge.gridforum.org/sf/projects/saga-rg>

**Author:** Dr ANJUM, Ashiq (UWE Bristol, UK)

**Co-authors:** Dr MERZKY, Andre (Louisiana State University, USA); Mr HABIB, Irfan (UWE Bristol,UK); Prof. MCCLATCHEY, Richard (UWE Bristol, UK); Dr JHA, Shantenu (Louisiana State University, USA)

**Presenter:** Dr ANJUM, Ashiq (UWE Bristol, UK)

**Session Classification:** Workflow Management

**Track Classification:** Programming environments