

# Using Ontologies to Provide Different Levels of Abstraction in Scientific Workflows

Daniel de Oliveira<sup>1</sup>, Eduardo Ogasawara<sup>1</sup>, Fernanda Araujo Baião<sup>2</sup>, Marta Mattoso<sup>1</sup>

<sup>1</sup>*Federal University of Rio de Janeiro (COPPE/UFRJ), Brazil*

<sup>2</sup>*Federal University of the State of Rio de Janeiro (NP2Tec/UNIRIO), Brazil*  
{danielc, ogasawara, marta}@cos.ufrj.br, fernanda.baiao@uniriotech.br

Scientific experiments are usually modeled as scientific workflows, and may be enacted using Scientific Workflow Management Systems (SWfMS). There is a huge variety of SWfMS available, such as Taverna [1] and VisTrails [2], which enables the specification and execution of a chain of activities usually represented by programs or services.

However, while using SWfMS the workflow has to be designed, revised and tuned on top of an execution format. Such format has many data conversion components and other auxiliary tools that can pollute the workflow definition. Working on a higher level of representation lets the scientist closer to the experiment definition rather than implementation details. There is a lack of semantic help for workflow analysis, *i.e.* to query on algorithms in a group of executable scientific workflows, a scientist should know *a priori* which program is related to which algorithm of the experiment. The scientist needs to query on the executable terms rather than the abstract terms.

Representing a workflow in hierarchical abstract levels can help the process of composition and analysis in high level of abstractions, but this representation is an open issue [3]. An abstract workflow is a representation of a flow of tasks that is specified without mapping to execution resources. A concrete workflow binds workflow tasks to specific execution resources. Yet, the abstract level itself may be further decomposed into different sublevels of abstraction until each reaches the concrete level. This situation would allow for scientists to compose and tune workflows in any of those abstract levels. Nevertheless, using the current available SWfMS, scientists are limited to compose only concrete workflows.

To add semantics to executable workflows, some systems suggest associating domain ontologies [4] to the specification [1]. This allows for navigating on the ontology of one activity of a concrete workflow, but it does not help in grouping workflows that share the same algorithm, or method or software, since the domain ontology is not associated to these abstract workflow activity roles. Therefore, workflow metadata

is needed to represent the abstract workflow and to help workflow resource finding according to these roles in higher levels of abstraction. Task and domain ontologies may be complementary. Domain ontologies [4] describe the domain terminology and its relationships, while task ontologies [4] describe the vocabulary related to a generic task, thus encompassing all knowledge related to a task, independent of domain. In our context, we propose the combination of a domain-specific ontology with a workflow modeling ontology. The workflow modeling ontology, named SciFlow, may be associated to a domain ontology to relate the components of the scientific workflow with domain terminology, increasing the semantics of the modeled experiment. SciFlow maps components of the workflow with domain terminology associating roles in different levels of abstraction. These levels go down until the workflow executable level. SciFlow can explicitly represent the workflow activity chain including data dependency. Workflow composition can be guided from high abstraction levels, such as method, algorithm and software until the execution level by querying and reasoning on SciFlow.

- [1] T. Oinn, P. Li, D.B. Kell, C. Goble, A. Goderis, M. Greenwood, D. Hull, R. Stevens, D. Turi, e J. Zhao, "Taverna/myGrid: Aligning a Workflow System with the Life Sciences Community," *Workflows for e-Science*, Springer, 2007, pp. 300-319.
- [2] S.P. Callahan, J. Freire, E. Santos, C.E. Scheidegger, C.T. Silva, e H.T. Vo, "VisTrails: visualization meets data management," *Proceedings of the 2006 ACM SIGMOD*, Chicago, IL, USA: ACM, 2006, pp. 745-747.
- [3] M.C. Cavalcanti, R. Targino, F. Baião, S.C. Rössle, P.M. Bisch, P.F. Pires, M.L.M. Campos, e M. Mattoso, "Managing structural genomic workflows using web services," *Data & Knowledge Engineering*, vol. 53, 2005, pp. 45-74.
- [4] A. Gomez-Perez, O. Corcho, e M. Fernandez-Lopez, *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. First Edition, Springer, 2004.