Enhancing Workflow-Driven Ontologies: A Meta Model to Create Conceptual Workflows at Multiple Abstraction Levels

Leonardo Salayandía and Aída Gándara
Advisors: Ann Gates and Paulo Pinheiro da Silva

CyberShARE Center of Excellence
University of Texas at El Paso,
El Paso Texas 79968, USA

E-mail: leonardo@utep.edu, agandara1@miners.utep.edu

Abstract

This poster presents a meta-model to construct conceptual workflows from ontologies using the Workflow-Driven Ontologies (WDO) approach. The meta-model was designed based on experiences of using the proof-of-concept WDO-It! tool in different activities involving scientists from various domains. The intention of this preliminary meta-model is to serve as a starting point for an empirical study that will seek to understand the use of abstraction techniques in combination with software engineering practices towards managing and using CI resources to implement complex scientific processes.

1. Introduction

The National Science Foundation (NSF) has identified Cyber-Infrastructure (CI) as an enabler technology that is vital to enhance collaboration across interdisciplinary virtual organizations [1]. However, the impact that CI has in the sciences is hindered by the complexities of managing the vast amounts of resources accessible through CI.

Towards alleviating the complexities of managing CI resources, the Workflow-Driven Ontologies (WDO) approach [2] is being developed to facilitate the specification of scientific processes at abstraction levels that are amenable to scientists. Through the use of ontologies, the WDO approach can be used by scientists to capture concepts that represent a controlled vocabulary in their domain of expertise. Using such controlled vocabulary, scientists are guided through the specification of scientific processes. Furthermore, as scientists start collaborating with colleagues with other areas of expertise, new controlled vocabularies emerge, and these are fused into the ontology to create different perspectives and levels of abstractions about a given scientific process. Ultimately, the goal is to map scientific processes produced by scientists in the form of conceptual workflows to resources available over CI to create workflows that can yield computable scientific applications; hence facilitating the management of CI resources towards automating scientific tasks.

WDO-It! (http://trust.utep.edu/wdo/downloads) is a prototype tool that implements the WDO approach. Based on experiences of using the WDO-It! tool in a workshop and two ongoing interdisciplinary projects housed at the CyberShARE Center of Excellence at the University of Texas at El Paso, this paper presents a preliminary meta-model that addresses some of the limitations of the WDO-It! tool (and WDO approach) identified to date. Additional details about the implementation of the meta-model using the Web Ontology Language (OWL) are presented as well.

The intention of documenting the initial meta-model is to serve as a starting point for an empirical study that will seek to understand the use of abstraction techniques in combination with software engineering practices towards managing and using CI resources to implement complex scientific processes.

2. Experiences

The preliminary meta-model proposed to enhance the WDO-It! tool has been guided through experiences gained by using the tool in the following activities:

1) **The 2007 Summer Southwest Regional Cyber-infrastructure Workshop at the University of Texas at El Paso** [3]. This workshop brought together a combination of domain scientists, students, and educators with the objective of promoting awareness about the role of CI in scientific research. The participants were exposed to two main tools: The Kepler Scientific Workflow [4] showcased as a mature scientific workflow tool that features an executable framework, and the prototype WDO-It!, a tool. Additionally, the attendees participated in a structured focus group session to indentify features that worked well in the aforementioned tools, as well as to identify shortcomings.

2) **The Earth Sciences project housed at the CyberShARE Center of Excellence** [5]. This ongoing project brings together scientists and students from the areas of Seismology, Computational Mathematics, and Computer Science to analyze existing Seismology techniques to model the subsurface of the Earth. In addition to producing controlled vocabularies across multiple disciplines, the goal of the project is to enhance existing Earth modeling techniques and automate selected techniques through CI.
3) **Collaborations between the National Center for Atmospheric Research (NCAR) group at Boulder, Colorado** [6]**, and the TRUST research group at the Computer Science Department of the University of Texas at El Paso** [7]. This project has resulted in the specification of workflows about several semi-automated pipeline applications that are implemented over legacy systems. Given that little documentation exists about some of these applications, a crucial goal of this collaboration is to document these scientific processes in a way that is scalable to new technologies such as CI and the Semantic Web. It is expected that similarities between distinct pipeline applications will be leveraged with the introduction of standardized vocabularies.

The graphical representation of process knowledge captured in an ontology is one feature of the WDO-It! tool that users from the workshop and the projects described above have consistently highlighted as practical. Users expressed that this feature leads to ease of use across disciplines; however, when compared to graphical representations of workflows from a mature workflow execution environment (i.e., Kepler), users expressed that the graphical representation of workflows in WDO-It! had significant expressivity limitations. In particular, some users indicated that workflow graphs should provide decision points and iteration of processes. Although the intention of the WDO approach is to capture conceptual workflows that do not necessarily have all the formal constructs to execute them, it is clear that scientists need decision points and iteration constructs to express complex scientific processes.

Another desirable feature in the WDO-It! tool was the ability to further break down workflow steps into sub-workflows or steps described at a finer granularity. This feature leads to the support of multiple levels of abstraction in the specification of scientific processes. Correspondingly, data inputs and outputs of a workflow step should also be specifiable at finer levels of detail, hence providing the mechanisms to extend the controlled vocabularies according the different levels of abstraction.

Users have expressed the need to identify manual steps in a scientific process, i.e., steps that are performed by a human or that are semi-automated but that require human inspection. Identifying such manual steps is important because it helps guide the deployment efforts over CI and it provides human operating requirements of the scientific process.

Finally, users have also expressed the need for mechanisms to track provenance of executing scientific processes. Provenance refers to meta-data about data and processes that documents their origins. This is a critical point that has arisen in the NCAR-TRUST collaboration case study as legacy systems are adopting state-of-the-art technologies to enhance their scientific applications.

3. Proposed Meta-Model

A meta-model (or a model about a model) is a model about the constructs and rules used to construct a model in some domain of interest. Figure 1 shows the proposed meta-model to construct scientific processes from ontologies using the WDO approach. The meta-model has been implemented as an ontology using the Web Ontology Language (OWL), and work is underway to incorporate it into the WDO-It! tool.

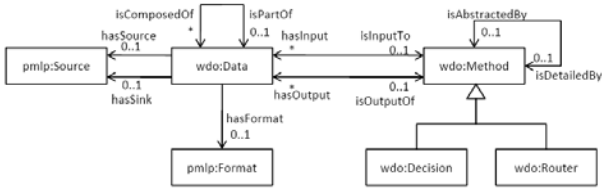

Fig. 1: The WDO meta-model

An advantage of implementing the meta-model as an OWL ontology is that the meta-model becomes more powerful when inference engines are used to interpret it. For example, given that the *wdo:hasInput* and the *wdo:isInputTo* relations are set to be inverse relations of each other, an inference engine should be able to determine that a *wdo:isInputTo* relation exists between a *wdo:Data* and a *wdo:Method* for the case where only the inverse relation is explicitly defined.

Another advantage of using ontologies (and OWL ontologies in particular) to implement the meta-model is the ability to reuse concepts (and related tools) from other existing OWL ontologies. For example, *pmlp:Format* and *pmlp:Source* are imported from the PML-P ontology. The PML-P ontology comes from the Proof Markup Language (PML) effort of the Inference Web Group at Stanford [8]. The PML-P ontology and related tools are being developed to capture distributed provenance and there are efforts under way to leverage PML in the WDO approach.

To address the feature of specifying finer levels of granularity in scientific processes (i.e., breaking workflow steps into sub-workflows), the *isDetailedBy* relation and corresponding inverse *isAbstractedBy* relation are depicted as self-associations on the *wdo:Method* concept. What is moreMoreover, given that *wdo:Decision* and *wdo:Router* are sub-concepts of *wdo:Method*, these concepts can be broken into finer abstraction levels as well. Similarly, to break data concepts into sub-data concepts, the *isComposedOf* and corresponding *isPartOf* relations are introduced.

The *wdo:Decision* concept is introduced to address the need to have decision points in workflows. For example, a decision point may be used to determine the route to follow in a workflow graph based on the application of some filter. The *wdo:Router* concept on the other hand, has the simpler task of forking or joining paths in a workflow graph. Additional details about the temporal constraints required to implement these constructs are hidden from the scientist and are intended to be defined at lower levels of abstraction.

Lastly, multiplicity constraints have been added to the proposed meta-model to facilitate its integration to the WDO-It! tool. For example, given that a *wdo:Data* concept is represented as an arrowed line, the number of *wdo:Method* concepts that can be graphically attached to a *wdo:Data* concept are two (one on each end of the line); hence, *wdo:isInputTo* and *wdo:isOutputOf* have multiplicity of 0..1. Additional work is required to determine whether the multiplicity constraints of the meta-model are adequate to model complex scientific processes, as well as to determine the effectiveness of the graphical notation currently used.

4. Conclusions

The current implementation of the WDO approach was utilized in several activities involving users from various scientific domains. Among others, a key feature that was highlighted as important from these experiences was the need to support multiple levels of abstraction in order to create conceptual workflows that can be employed to manage the complexities of CI resources. The development of the proposed meta-model intended to address the shortcomings identified thus far, while facilitating integration into the current WDO-It! tool. The meta-model was implemented as an ontology using OWL in order to leverage inference engines and other existing OWL ontologies.

Additional work is needed to integrate the meta-model into the WDO-It! tool. What is more, the WDO-It! tool is intended to be used as an instrument to conduct an empirical study that will seek to understand the use of abstraction techniques in combination with software engineering practices towards managing and using CI resources to implement complex scientific processes.

Acknowledgements

This work is funded by the National Science Foundation through grants CNS-0540592 and HRD-0734825.

References

1. J. Cummings, T. Finholt, I. Foster, C. Kesselman, and K. Lawrence, *Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations*, final report from workshops on building effective virtual organizations, National Science Foundation, May 2008.
2. L. Salayandia, P. Pinheiro da Silva, A.Q. Gates, and F. Salcedo, Workflow-Driven Ontologies: An Earth Sciences Case Study, *In Proc of 2nd Intl Conf on e-Science and Grid Computing*, Amsterdam, Netherlands, December 2006.
3. Summer Southwest Regional Cyberinfrastructure Workshop, El Paso, Texas, August 30, 2007.
4. B. Ludäscher et al., Scientific Workflow Management and the Kepler System, *Concurrency and Computation: Practice & Experience, Special Issue on Scientific Workflows*, 2005.
5. The Geosciences Subproject at the CyberShARE Center of Excellence, El Paso, Texas www.cybershare.utep.edu/research/gs
6. The National Center for Atmospheric Research, www.ncar.ucar.edu
7. The Trust Research Group at the University of Texas at El Paso, http://trust.utep.edu
8. Proof Markup Language (PML) Primer, Inference Web Group, www.inference-web.org/2007/primer