

USING SENSORML TO DESCRIBE SCIENTIFIC WORKFLOWS IN DISTRIBUTED WEB SERVICE ENVIRONMENTS

Terence van Zyl, Anwar Vahed, Graeme Mcferren, Petrus Shabangu, and Bheki Cwele

ICT4EO, The Meraka Institute,
Johannesburg, South Africa
tvzyl@, avahed@, gmcferren@, pshabangu@, bcewle@meraka.org.za

ABSTRACT

Scientific Workflows provides a technology that facilitates researchers by allowing them to capture in a machine processable manner the method relating to some research. This increases both provenance and repeatability of the research and allows for increased collaboration through workflow sharing. The Sensor Web is an open complex adaptive system that pervades the internet and provides access to sensor resources. One mechanism for describing sensor resources is through the use of SensorML. It is shown that SensorML can be used as a mechanism for describing Scientific Workflows and thus facilitates completely distributed workflow descriptions on the web. However in order to fully capture the requirements as relating to open and distributed environments some extensions to SensorML will be required.

Index Terms— Sensor Web, Scientific Workflows, Web Service, SensorML

1. INTRODUCTION

Scientific workflow is another type of workflow that usually underlies many complex e-science applications such as climate modeling and disaster recovery simulation [1]. Scientific workflows are used as means for modeling and enacting scientific experiments, they share many features of business workflows, but also go beyond them. Many known workflow results and techniques can be leveraged in scientific settings, and many additional features of scientific applications can be usefully deployed in business settings [2]. Scientific workflows promise to become an important area of research within workflow and process automation, and will lead to the development of the next generation of problem-solving and decision-support environments. Compared with business workflows, scientific workflow has special features such as computation, data or transaction intensity, less human interaction, and a large number of activities [3]. Some emerging computing infrastructures such as grid computing with powerful computing and resource sharing capabilities present the potential for accommodating those special features. The workflows aim to provide a simple concise notation that allows easy parallelization and supports the composition of large numbers of parallel computations [4]. As parallel

computing expands, systems such as PVM and standards such as MPI encourage scientists to construct complex distributed solutions that span the networks.

“The Sensor Web is an open complex adaptive system, organised as a network of open sensor resources, which pervades the internet and provides external access to sensor resources. By open sensor resources, we include any open system, including sensor networks, that is a source of sensor data or sensor meta-data [5].” A more functional derived consensus view of the Sensor Web is as a “coordinated observation infrastructure composed of a distributed collection of resources that can collectively behave as a single, autonomous, task-able, dynamically adaptive and reconfigurable observing system that provides raw and processed data, along with associated meta-data, via a set of standards-based interfaces [6].” The predominant standardised mechanism for describing sensor resources relating to the Sensor Web is through SensorML, an OGC standard. SensorML or Sensor Modelling language provides the XML schemas for encoding data and meta data relating to sensor resources. However SensorML is more than just a mark-up language but also allows the user to describe models, provenance, lineage and other details relating to not only physical sensors but also various virtual and composite sensor resources [7].

One core problem relating to the Sensor Web is how to make it an effective tool in the hands of scientists it is conceivable that the technologies relating to Scientific Workflows may be invaluable in this case. If the Sensor Web is to be effectively used as an underlying data source in Scientific Workflows then the predominate mechanism for describing sensor resources and processes relating to them in the Sensor Web, SensorML, will need to be investigated and probed as to how it is able to serve that purpose. Further there is no specific workflow description language that deals with the geospatial aspects of data and sensors as it relates to the Sensor Web. For this purpose SensorML is the current dominant standard.

2. METHODOLOGY

In order to assess suitability of SensorML to Scientific Workflows, a number of requirements for a workflow description language for Scientific Workflows on the Sensor web are established. The requirements outlined extend previous work by the authors [8]. SensorML is then

assessed as to how well it is able to meet these requirements. In order to do the assessment a workflow system is constructed that uses SensorML as a workflow description language. Various types of Scientific Workflows are tested and evaluated.

3. CONCLUSIONS

SensorML is a good candidate as a workflow description language for Scientific Workflows on the Sensor Web. However there are still some extensions to SensorML that will be required when relating to the dynamic nature of Scientific Workflows in open environments. Also SensorML is well suited to the Sensor Resource portion of the Scientific Workflows but does not fully support the standard requirement of workflows in general. This includes alerts, branching, and alternate flows for errors.

4. REFERENCES

- [1] Chen J. Media Talk On Scientific Workflow (with Prof. Dr. W.M.P. van der Aalst from Eindhoven University of Technology, The Netherlands) in the newsletter of IEEE Technical Committee on Scalable Computing, 2007.
- [2] Singh M, Vouk M. *“Scientific Workflows: Scientific Computing Meets Transactional Workflow”*, NSF Workshop on Workflow and Process Automation in Information Systems, 1996
- [3] Davidson S,B. Boulakia S,C. Eyal A. Ludäscher B. McPhillips T,M. Bowers S. Anand M,K. Freire, J. *“Provenance in Scientific Workflow Systems”*. IEEE Data Eng. Bull, 2007
- [4] Zhao Y. Raicu I. Foster I,T. *“Scientific Workflow Systems for 21st Century e-Science, New Bottle or New Wine?”* CoRR abs,2008
- [5] T. L. van Zyl, I. Simonis, and G. Mcferren, "The sensor web: systems of sensor systems," *International Journal of Digital Earth*, vol. 99999, no. 1, pp. 1-14, 2008.
- [6] Various, "ESTO/AIST sensor web pi meeting," The NASA Earth Science Technology Office, 2007.
- [7] M. Botts, OpenGIS sensor model language (SensorML) implementation specification. Technical report, Open Geospatial Consortium Inc. 2006.
- [8] G. McFerren, T. van Zyl, M van der Merwe and M. du Preez "User Requirements For Sensor Web Based Scientific Workflows In The Cholera Research Domain" *IGARSS 2008 proceedings*, 2008.