



Jun 26th, 10:40 AM - 12:00 PM

Principle-based, Semi-automatic Ontology Generation to Support Cross-Domain Interoperability of Data Sets and Models

Scott Dale Peckham

University of Colorado, Boulder, scott.peckham@colorado.edu

Maria Stoica

University of Colorado, Boulder, Maria.Stoica@colorado.edu

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Peckham, Scott Dale and Stoica, Maria, "Principle-based, Semi-automatic Ontology Generation to Support Cross-Domain Interoperability of Data Sets and Models" (2018). *International Congress on Environmental Modelling and Software*. 12.

<https://scholarsarchive.byu.edu/iemssconference/2018/Stream-A/12>

This Oral Presentation (in session) is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Principle-based, Semi-automatic Ontology Generation to Support Cross-Domain Interoperability of Data Sets and Models

Scott D. Peckham¹ and Maria Stoica¹

¹University of Colorado, Boulder

(Scott.Peckham@colorado.edu, Maria.Stoica@colorado.edu)

Abstract: Every data set and computer model has its own internal vocabulary (i.e. names or labels) for referring to its input and/or output variables. It is therefore difficult to know, even for an expert, whether a variable stored/computed in a given digital resource is equivalent to one needed by another resource. Experts can typically figure this out through a process that may involve examining the equations that are used, being familiar with domain jargon, reading documentation (e.g. source code, manuals and papers) or talking to the developer of the resource. However, this is time-consuming, frustrating and inefficient. The only way to automate this *semantic mediation* task is with an accurate, one-time mapping of these internal names to variable names in a standardized vocabulary that can be utilized by machine (i.e. accessed via function calls in a program). This task of mapping internal variable names to standardized names is known as "semantic annotation". Once completed, it is possible to automatically perform "semantic alignment" every time that resource is selected for use in a workflow, allowing variables to be correctly passed between coupled resources. We will describe efforts to semi-automatically generate standardized variable names for different domains by building on the foundational and rule-based principles of the Geoscience Standard Names ontology (geoscienceontology.org). Our initial focus will be on measurement concepts in the realms of agriculture, social science, economics, transportation networks and demographics. This work is funded by a project called MINT (Model INtegration) that is part of the World Modelers program.

Keywords: *Semantic mediation; ontology generation; model integration; variable names; semantic annotation*