Leveraging Semantics to Improve Reproducibility in Scientific Workflows

Idafen Santana-Perez[1], Rafael Ferreira da Silva[2], Mats Rynge[2]
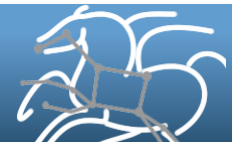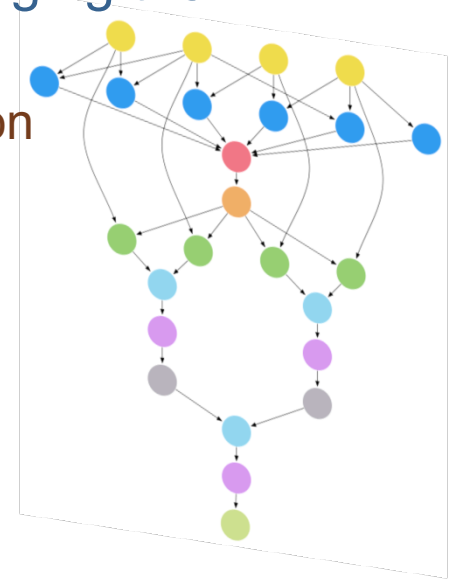Ewa Deelman[2], María S. Pérez-Hernández[1], Oscar Corcho[1]

The reproducibility@XSEDE workshop
July 14th 2014, Atlanta, GA

[1]Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
[2]University of Southern California, Information Sciences Institute, Marina Del Rey, CA, USA

USC Viterbi
School of Engineering
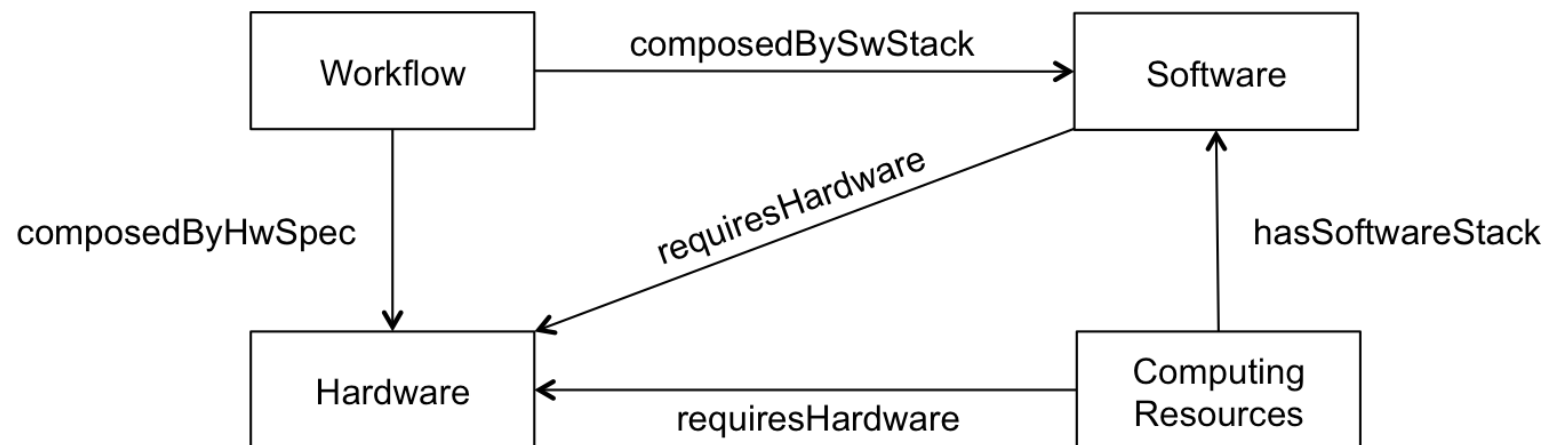
Information Sciences Institute

Introduction

- Scientific workflows are a useful representation for managing the execution of large-scale computations
  - Facilitates the creation and management of the computation
  - Provenance is captured at runtime

- Are *data*, *code*, and the *workflow description* enough to achieve reproducibility?
  - What about the underlying infrastructure?

- Two approaches for conserving the environment of an experiment
  - Physical conservation: the real object is conserved
  - Logical conservation: the object description is conserved (**Semantics**)
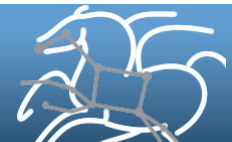
Semantic Modeling

- We propose a logical-oriented approach to conserve computational environments
  - The capabilities of the resources are <u>semantically</u> described

- We propose <u>semantic models</u> for describing the main domains of a computational infrastructure and how it is related to the workflow:
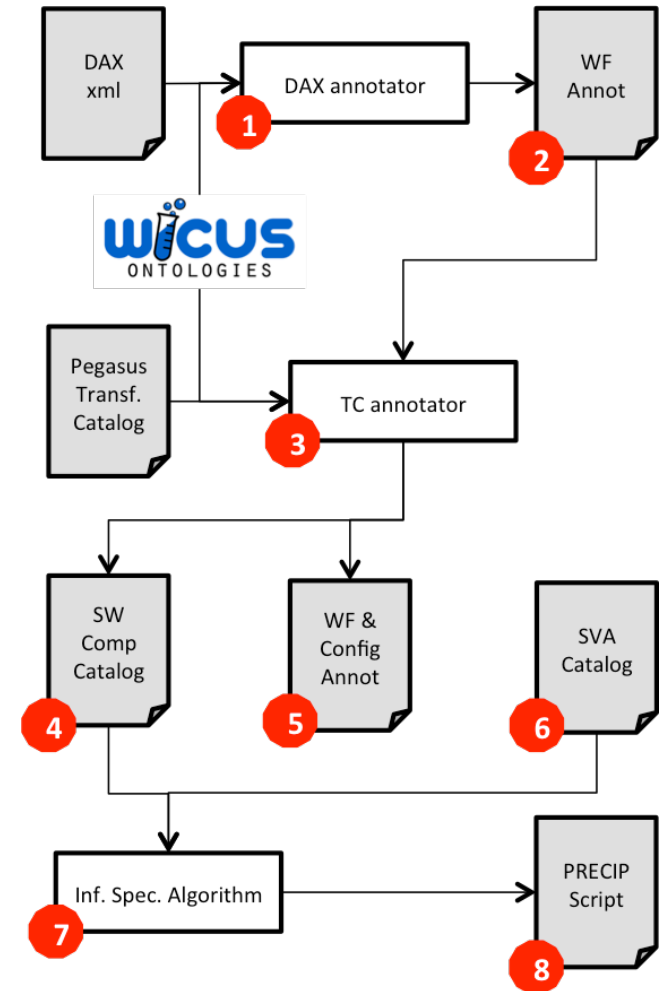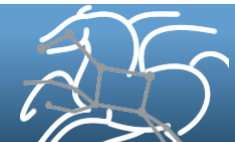

Overview of the Ontology Network

Reproducibility Process

- Experiment Setup
    - **Montage** Workflow (Astronomy)
    - **Pegasus** Workflow Management System
    - Workflow execution on **FutureGrid**

- Semantic Analysis
    - Annotations generated using **WICUS**
    - Infrastructure specification translated as a **PRECIP** script

Overview of the reproducibility process for scientific workflows

Leveraging Semantics to Improve Reproducibility in Scientific Workflows

Thank you

rafsilva@isi.edu

"Semantic annotations combined with scripting functionality is a strong approach for attaining reproducibility of computational environments."

Research Object http://pegasus.isi.edu/publications/xsede-reproducibility

PRECIP http://pegasus.isi.edu/precip

WICUS http://purl.org/net/wicus

Pegasus WMS http://pegasus.isi.edu

USC Viterbi
School of Engineering

Information Sciences Institute