







**Partnership for International Research and Education** A Global Living Laboratory for Cyberinfrastructure Application Enablement

# **Project Title: Mining Processes from Provenance of Scientific Workflows**

**Student:** Reng Zeng, Ph.D Candidate, Florida International University **Research Advisor:** Dr. Xudong He, Florida International University **CI-PIRE Partner Advisor:** Dr. Wil van der Aalst, Eindhoven University of Technology



# **I. Research Overview and Outcome**

### **Problem Statement**

Scientific Workflow Management Systems provide scientists tools to build scientific workflows manually and execute them automatically, however, it remains difficult for scientists to build and update a scientific workflow. This project aims at developing and evaluating methods to build and update scientific workflows automatically by using process mining techniques on provenance from either system-level monitoring or workflow-based systems.



#### **Provenance**

- □ Provenance, in scientific workflows community, refers to the sources of information, including entities and processes, involving in producing or delivering an artifact.
- Open Provenance Model (OPM), by International Provenance and Annotation Workshop
- □ Another working-in-progress Provenance model by W3C Provenance Incubator Group
- □ In existing Scientific Workflow Management Systems:



### **Motivation and Impact**

Scientific workflows have emerged as a new paradigm for representing and managing complex distributed scientific computations and data analysis, have accelerated many scientific analysis process.

□ It is an evolving process to solving scientific problem. Scientists may fine-tune the experiment by reiterating their process for many times with various data and different parameters. This evolving process can reveal the nature of scientific problems.

□ Provenance describes evolution of an object over time and it describes evolving scientific processes.

Mining a process from provenance of scientific workflow can offer insight into workflows to help rerunning the process by simulation, validating the process against desired property, improving the process and revealing the



Taverna	Pegasus	Triana
□Kepler	U VisTrails	Trident

□In system level systems:

□ Provenance Aware Service Oriented Architecture(PASOA), U. Southampton

- □ VisTrails, general to a wide range of applications, Univ. of Utah
- □ Wings/Pegasus, produce application-level provenance, USC
- □ Provenance Aware Storage Systems (PASS), Harvard

□ Provenance doesn't keep track of control flows that don't result in data flow.



□ Successfully get the causality relationship between event 1 and event 2 above. □ Event 1 and event 2 above are clustered indicating their low significance. Give a zoom-able view of scientific workflow to provide insight of event relationship in different abstraction level.

#### **II. International Experience**





**Eindhoven City Center** 

Campus at 10:30pm

A collaborator Dr. Fabrizio Maggi

## III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065, HRD-0833093. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Dorm, so called "spacebox"