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 to develop 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.

Scientific Workflow
- JobD
- CSVRowPath
- CSVRowPath
- CSVRowPath

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:
  - Taverna
  - Kepler
  - VisTrails
  - Trident
- Provenance Aware Service Oriented Architecture (PASOA), U. Southampton
- Provenance Aware Storage Systems (PASS), Harvard
- Provenance doesn’t keep track of control flows that don’t result in data flow.

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 provide view of scientific workflow to provide insight of event relationship in different abstraction level.

Results and Future Work
- 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.
- Future work can employ existing data dependency to help on workflow mining.

II. International Experience

Brief Description for Photos

1. Watch World cup in the Netherlands. Dutch people won the semi-final and had a great celebration party.
2. Amsterdam is a city where you can enjoy its beauty simply by bicycle, canals are wonderful.
3. Paris is a fascinating city you can not miss in Europe, where you will love Louvre Palace and Notre Dame Cathedral.
4. Switzerland’s trains make an unforgettable trip, landscape along the railway is always beautiful and trains are always on time.

PIRE offered me a very unique opportunity to gain new experience in collaborative research and to feel European culture.

Collaborative Research
1. Working with experts in process mining research area is a great augment to my research topic.
2. Valuable feedback speed up my research much more than simply reading related papers.
3. Exposure to new research ideas which is helpful just like attending a lot weekly seminars.

PIRE Partner: Dr. Wil van der Aalst, Eindhoven University of Technology

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.