

Model Task Force  
First Provenance WG F2F

Satya Sahoo, Paolo Missier, Khalid  
Belhajjame, Jun Zhao

Boston, MA

# Background

- Objective: Define the provenance model to be used in PIL
- Tasks
  - Identify and provide description of provenance concepts using
    - Provenance XG final report
    - Journalism example scenario
  - Use Semantic Web technologies to define the provenance concepts
  - Provide formal semantics of the provenance model

# Agenda for first F2F

- A set of prioritized provenance concepts identified (ref: journalism example)
  - Thing
  - Invariant View or Perspective on Thing
  - Process Execution
  - Derivation
  - Use
  - Generation
  - Agent
  - Temporal ordering
- Reach agreement on the definitions of prioritized provenance concepts

# Agenda for first F2F

- Work completed prior to first F2F
  - Review, curation, and consolidation of prioritized concept definitions
  - Enumerate outstanding issues related to prioritized concepts
- Issues for discussion:
  - Finalize definitions of prioritized concepts (reconcile outstanding issues)
  - Evolve consolidated list of prioritized concepts towards a report/descriptive document
  - Review status of concepts not prioritized for F2F
  - Coordinate with provenance access and query task force

# Agreement on Provenance Concepts

- Agreement reached during weekly meetings among WG members
  - The subject of provenance may be anything, whether physical, digital, conceptual or otherwise
  - To define provenance-related concepts independently of the web architecture in a first instance
  - Definition of **thing** and the **invariant view or perspective of thing**
  - Definition of **process execution** (differentiate between execution and specification, has duration, and start of execution is always in past)
  - Definition of **derivation**
  - PIL is an assertion language
  - Use constraints introduced in definitions of concepts as building blocks for understanding PIL

# Outstanding Issues

- Resolve outstanding issues for concepts with agreement over their definitions
  - IVPT
  - Process execution
  - Derivation
- Agree on definitions of following prioritized concepts
  - Agent
  - Use
  - Generation
  - Ordering of process execution
  - Time

# Next steps for the Model TF

- Formalization of the prioritized provenance concepts
- Describe the journalism example using the prioritized concepts
  - Represent the description using Semantic Web technologies (RDF)
  - Define semantics of the encoding
- Rephrase provenance queries in journalism example using prioritized concepts
- Primer on provenance model