NetKarma : Capture and Representation of Provenance for GENI Experiments (Spiral 2 project)

Beth Plale
School of Informatics and Computing
Indiana University Bloomington

Chris Small
Global Research Network Operations Center (GRNOC)
Indiana University

GEC-8
July 20-22, 2010
San Diego, CA

netKarma

- provenance collection tool
  - captures and stores process and data provenance.
  - based on generation of discrete provenance activities during experiment lifecycle
  - Activities aggregated to form complex data and process provenance graphs.
- Provenance adaptor : an interface that uses experiment logs and a set of rules to derive provenance events that are sent to the Karma service for storage and derivation of provenance.
- Adaptor is a generic log processing unit for GENI component log files which comprise of two sub-units:
  - Log Parser
  - Notification Generator

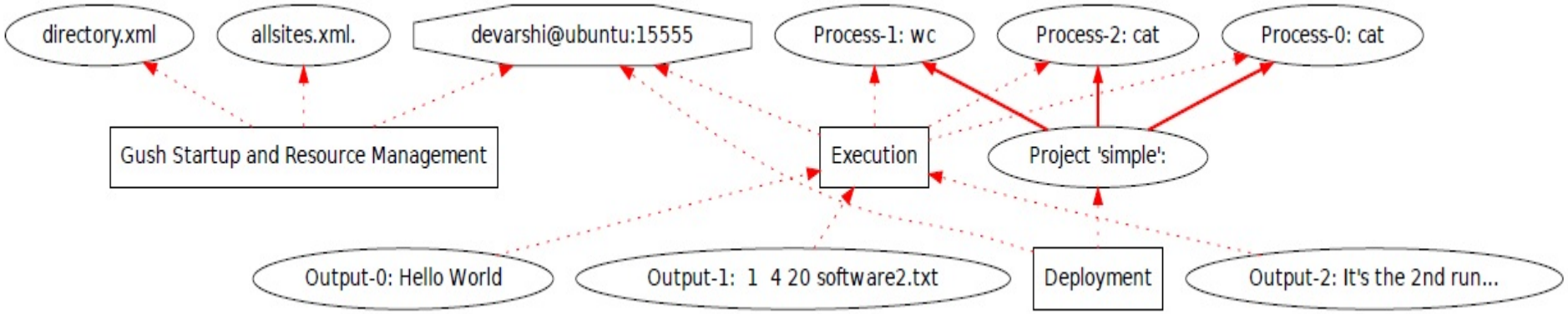
Goal: collect and represent provenance from multiple layers:

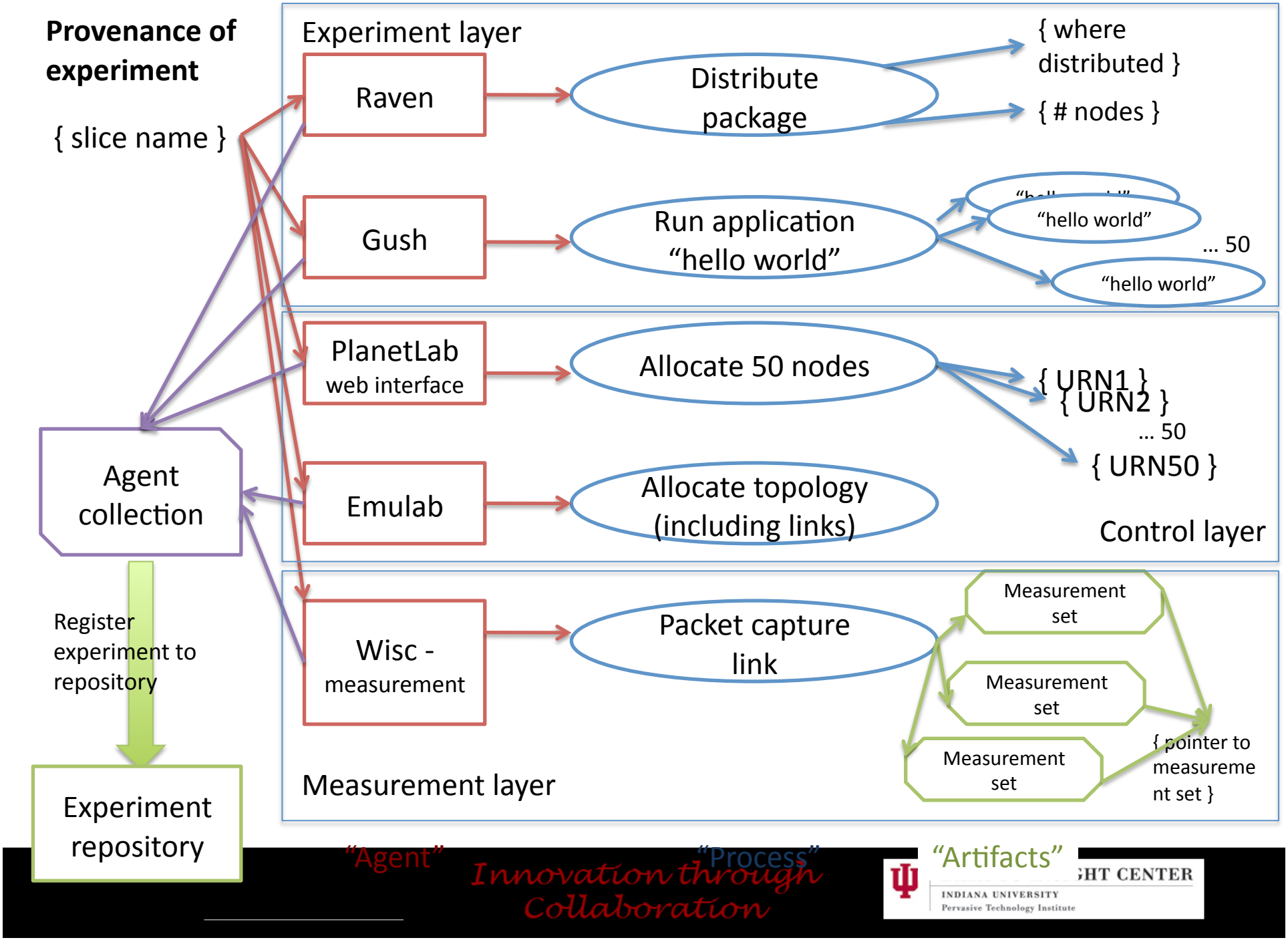
- Experimental tool commands ✓ (GUSH)
- Topology created by control frameworks
- Operational status on substrate / infrastructure
- Code and data contained in the experimental slice
- Pointers to obtained measurement data
- Annotations by experimenters

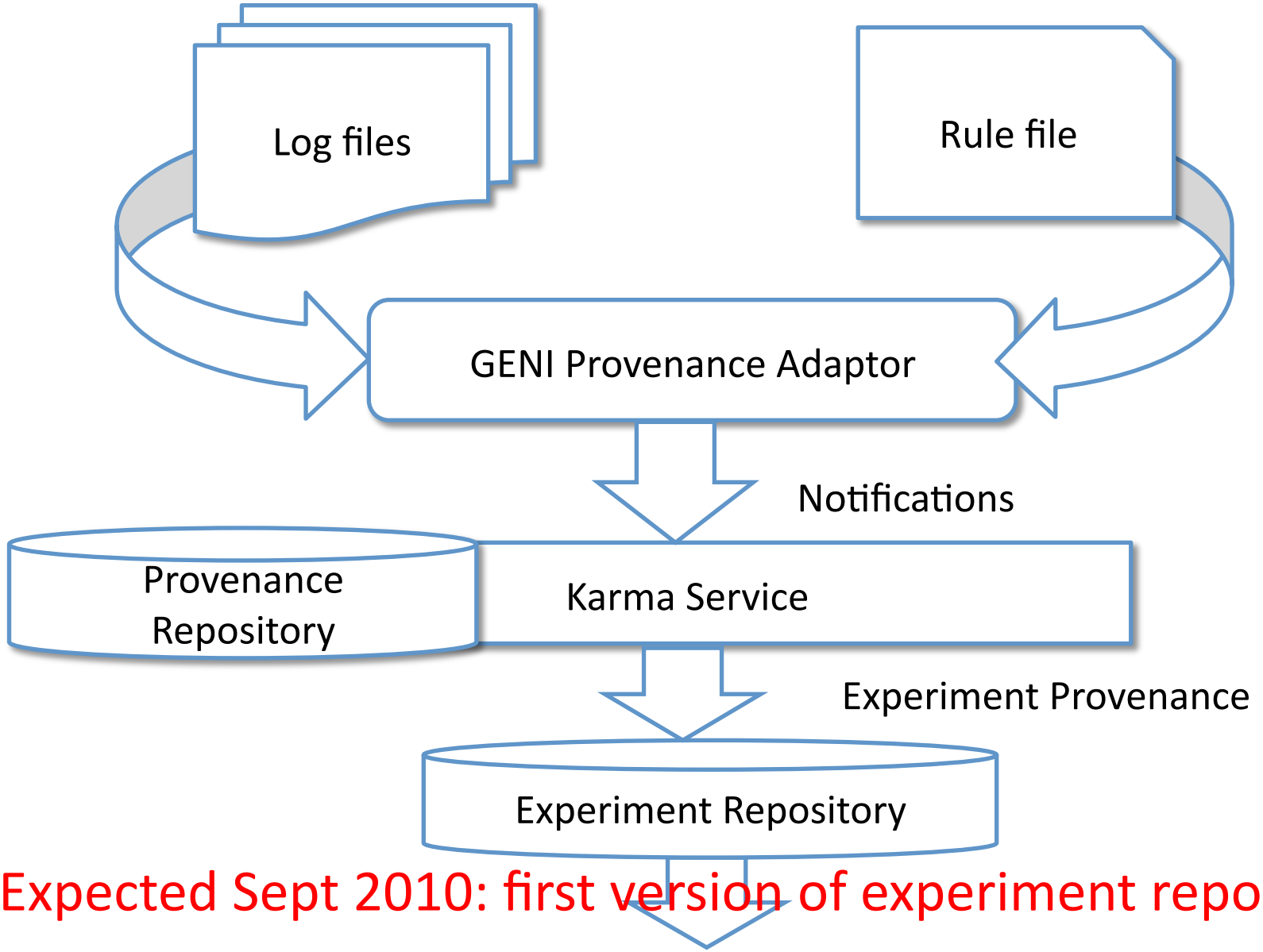
Karma Tool recent efforts

- Built first version Adaptor collection mechanism
- Working on Integrating support for RabbitMQ into architecture
  - Implements AMQP
- Expand relational database schema to handle layers of provenance (corresponding to layers of GENI infrastructure)
- Extend Existing Architecture
  - Service Core
    - Java, mySQL ← schema extension
  - Client Layer
    - Axis2/SOAP,← adding support for RabbitMQ, REST
  - Notification library
    - Provenance events ← expand

GUSH collection




Provenance of experiment
{ slice name }
Experiment layer
Raven
Distribute package
{ where distributed }
{ # nodes }
Gush
Run application "hello world"
"hello world"
... 50
"hello world"
PlanetLab web interface
Allocate 50 nodes
{ URN1 }
{ URN2 }
... 50
{ URN50 }
Agent collection
Emulab
Allocate topology (including links)
Control layer
Register experiment to repository
Wisc - measurement
Packet capture link
Measurement set
Measurement set
Measurement set
{ pointer to measurement set }
Measurement layer
Experiment repository
"Agent"
"Process"
"Artifacts"
Innovation through Collaboration
INDIANA UNIVERSITY
Pervasive Technology Institute



Expected Sept 2010: first version of experiment repository

NetKarma poster tonight presented by Devarshi Ghoshal