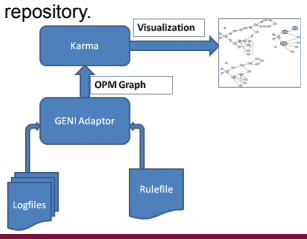
NetKarma: GENI Provenance Registry

NetKarma Model

GENI Adaptor provides an interface that uses the GENI experiment logs and a set of rules to derive provenance information and maps them into the Karma repository.



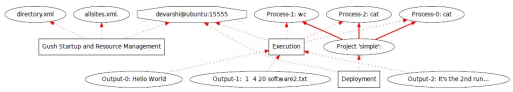
Open Provenance Model

The Open Provenance Model (OPM) is a community effort to standardize on a representation of provenance graphs.

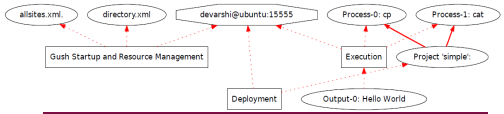
- Entities
  - Artifact : an immutable piece of state.
  - Process: action resulting in new artifact.
  - Agents: entity acting as a catalyst of process.

- Gush OPM Graphs

Graph-1

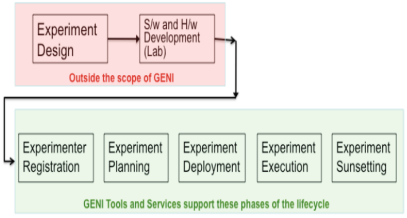


Graph-2

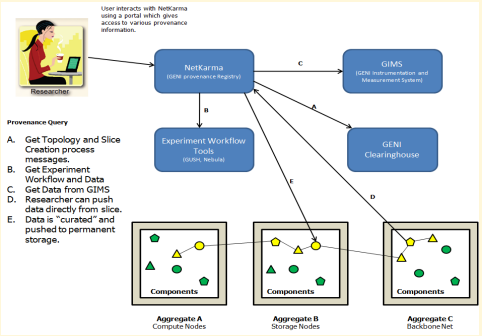


GENi Experiment Life Cycle

Diagram below shows the GENI Experiment life cycle. [3]



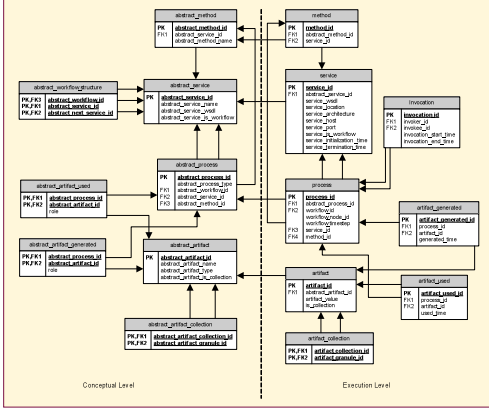
Architecture



- Cyberinfrastructure frameworks for science are an accepted way of accessing analysis tools, computational and data resources on the Internet. Provenance plays an increasingly valuable role in scientific discovery by capturing a record of activities performed that led to the creation of data

- Karma is a workflow provenance collection framework developed by Indiana University. It captures and stores both process provenance and data provenance. The framework is based on generating discrete provenance activities during the lifecycle of a workflow execution that can be aggregated to form complex data and process provenance graphs that can span across workflows.

Karma Provenance Database Schema



GENI Relevance

- NetKarma attempts to leverage the structure of GENI and use the messages and code that create a "slice" to record the workflow of an experiment. Since GENI is a virtualized infrastructure the creation and running of a slice provides a rich set of data to be used to generate an experiment's provenance.
- NetKarma provides researchers a way to understand their own experiments. It also allows others to understand (and possibly rerun experiments) after "sunset" and resources are released.
- NetKarma provides a way to integrate data collected as part of different areas of GENI such as experimental tools, operations, control planes and measurement and tie them together in a single bundle that is easily referenced with a handle such as a DOI reference. All data collected in the experiment and the conditions of the experiment are recorded.
- NetKarma will explore concepts that may ultimately be used in the creation of an "Experiment Definition Language" to describe experiments

Data Sources

- NetKarma will use a variety of different data sources to obtain information describing GENI experiments and the conditions of the experiment. These may include:
  - Experimental tool commands
  - Topology created using the control frameworks
  - Operational status on substrate/ infrastructure
  - Code and data contained in the experimental slice
  - Measurement data obtained from the GENI
  - Annotations by experimenters

References

1. Simmhan, Y.L., Plale, B., and Gannon, D., Karma2: Provenance Management for Data-Driven Workflows, *International Journal of Web Services Research*, 5(2): 1-22, 2008.
2. The Open Provenance Model (v1.01), http://eprints.ecs.soton.ac.uk/16148/1/opm-v1.01.pdf.
5. Lifecycle of a GENI Experiment . GENI-SE-SY-TS-UC-LC-01.2 http://groups.geni.net/geni/wiki/ExperimentLifecycleDocument