Digital Object Registry

Corporation for National Research Initiatives

Project Status Report – Oct 26, 2012 through Apr 2, 2013

1. Major Accomplishments

During the first three years, the scope of work on this project was to adapt the Handle System, along with components of the CNRI Digital Object Registry, to create a clearinghouse registry for principals, slices, and/or components in at least one GENI Spiral 1 control framework, capable of supporting limited operations. We have successfully adapted the Digital Object Registry and related technologies to build a GENI Federated Clearinghouse and a Distributed Hash Table for Seattle, aka Million Node GENI, led by Justin Cappos. The scope of work in Year 4 was to discuss, design, and develop a prototype of the Measurement Data Archive (MDA) service. We made these services available by deploying them in production servers with high availability and network bandwidth.

The scope of work in Year 5 is to continue the discussions with the Instrumentation and Measurement (I&M) working group members to standardize the metadata schema used by the MDA service, evaluate the integration of the Handle System into the MDA service developed by the UNC team, and review and finalize a GENI experimental data plan.

During this reporting period, we continued participating in various GENI activities and programs, including attending GEC 16 held in Utah, and also continued our collaboration with GENI members and System Engineers as part of the I&M Working Group. We:

- Attended the I&M sessions held at GEC 16. Discussed with UNC and BBN team members how to further the integration of the Handle System into the iRODS framework that is being used for implementing the MDA service. Discussed with BBN team members how to simplify the metadata schema that will be used within the MDA service.

- Designed and finalized the metadata schema. Schema and examples are on the wiki page for the project.

- Continued to make available the prototype of the MDA service.

- Continued to make available the Distributed Hash Table service for the Million Node GENI project, led by Justin Cappos.

- Continued to make available the GENI Federated Clearinghouse service that federates the information from the ProtoGENI clearinghouse and makes that information available via the Digital Object Registry interfaces.

1.a. Milestones achieved: During this period we completed one milestone, S5.b, as discussed below.

- Attended GEC 16 and participated in various I&M sessions and discussions on GENI experiments.
- Designed and finalized version 2 of the GENI object descriptor. Metadata schema and examples are on the wiki page for the project.
- Together with the GENI community, we reviewed the GENI experimental data plan for identifying, archiving, searching and sharing measurement data objects, consistent with current best practices in the research community. Any updates and recommendations will be made during the next few months.
- We discussed with UNC and other I&M members how to integrate the Handle System into the MDA service. A preliminary version has been implemented by UNC, which has not yet been made available to CNRI. We expect UNC to release the implementation for review in the next few months.
- To demonstrate our interest and support for GENI, we have agreed to support University of Kentucky in their proposal to GENI on building experimental tools that use the Handle System and CNRI-designed metadata schema.

1.b. Deliverables: During this period, we made available the metadata schema and related examples on our wiki page for the project.

2. Description of Work Performed

2.a. Activities and Findings During This Period

Measurement Data Object Descriptor

BBN and CNRI participated in several telephone and webex sessions prior to GEC 16 to discuss and simplify the metadata schema identified during Year 4. The goal of this discussion was to identify what metadata elements were absolutely necessary and what metadata elements can be ignored for this version of the metadata schema release. While complex metadata schema might capture more detailed and hence better quality metadata, it is usually a deterrent for users who contribute metadata. Simplifying metadata would lower the barrier to entry for contributors. Acknowledging the power of keeping the schema simple, BBN and CNRI identified the metadata elements within the schema that are important.

The Metadata descriptor for the GENI I&M objects, as well as for other kinds of GENI objects including those used specifically for experimentation, is designed to accommodate the following objectives:

1. An experiment, instrumentation, or measurement object should support a collection of one or more artifacts (log files, configuration files, data files, source code, documents, etc.). In our parlance, we call each of those individually identifiable files artifacts.
2. An object should support the inclusion of artifacts from multiple runs of the same experiment. For instance, it may be important to know and archive how the experiment evolved over a period of time versus just the final 'golden' run of the experiment.
3. An object should support the inclusion of artifacts from multiple experiments. For instance, multiple experiments could be conducted that are related to each other, especially if each of those experiments meet specific requirements or tasks identified for a project. As such, collecting artifacts from inter-related experiments would give a holistic view of the experiment, and therefore archiving all experiments together will be quite useful.
4. An object should support the inclusion of project-specific details in addition to details about experiments.

   The above objectives imply that an object could have the following hierarchy: Project->Experiments->Steps->Artifacts. However, as stated above, an object could be just an artifact or artifacts, or step or steps with artifacts, or experiment or experiments with steps, or project encompassing the rest of the hierarchy.

5. The metadata design should allow describing each of the hierarchical elements, project, experiment, step, and artifact, separately. In other words, one should be able to describe, via metadata schema, a project separately from an artifact, or vice-versa. This way, descriptions about each of the elements (Project, Experiment, Step, and Artifact) could change independent of the other.
6. The metadata design, congruently, should allow inheriting element-specific descriptions by the upper level in the order of the hierarchy. For example, a 'step' description should be able to inherit its artifacts' descriptions. Likewise, an 'experiment' description should be able to inherit its steps' descriptions. This way, depending on the level at which a GENI object is formed, a holistic view of the metadata should be possible for associating with the object.
7. Once archived, an object should support inclusion of some generic metadata fields to allow a global understanding of the purpose of the archived object.

The designed schema and examples that showcase each of the aforementioned objectives are on the wiki page for the project.

2.b. Project Participants

CNRI discussed its project activities with a number of other GENI participants, but all work accomplished this quarter was performed by CNRI alone, or with the

cooperation of the I&M members and Harry Mussman. Names and email addresses of CNRI participants are available on the GENI wiki page for the project.

2.c. Publications

No publications were produced this quarter.

2.d. Outreach Activities

Giridhar Manepalli attended the GENI Engineering Conference held in Utah and participated in a variety of discussions with GENI members and System Engineers.

2.e. Collaborations

CNRI continued to collaborate with the I&M working group members to standardize the metadata schema and integrate the Handle System into the MDA service.

2.f. Other Contributions

Production Services

CNRI continued to support the GENI Measurement Data Archive prototype and the GENI Federated Clearinghouse and the Distributed Hash Table services on hardware deployed in a collocation facility that features redundant power and air conditioning units, physical security, etc. A 100Mbps network pipe is dedicated to the machine.