

Digital Object Registry

Corporation for National Research Initiatives

Project Status Report – Aug 16, 2011 through Nov 4, 2011

1. Major Accomplishments

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 in Year 1, and extend that functionality to additional frameworks, and to additional functions, in Years 2 and 3. We have successfully adapted the Digital Object Registry and related technologies to build a GENI Federated Clearinghouse, a Distributed Hash Table for Seattle, aka Million Node GENI, led by Justin Cappos, and 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 4, in addition to supporting the above services, includes documenting necessary extensions to the MDA prototype based on discussions with the GENI I&M members, and subsequently implementing those extensions, if the time and the resources are available. Authentication and access policy integrations are also part of the scope for extensions. Finally, demonstrations of the system to experimenters and I&M members are also part of the scope.

During this reporting period, we continued participating in various GENI activities and programs, including attending the GEC 12 held in Kansas City, MO, and also continued our collaboration with GENI members and System Engineers as part of the I&M Working Group. We:

- Continued our collaboration with OnTimeMeasure and INSTOOLS project members for integrating the prototype with their measurement tools.
- Demonstrated the Prototype at GEC 12. Prototype documentation that discusses the internal data management details, APIs, and access points is now available on the project wiki page.
- Attended the various meetings held on I&M topics during the GEC 12.
- 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, S4.a, as discussed below.

We demonstrated the MDA Prototype at the GEC 12 during the poster session and discussed it at length during several I&M meetings and sessions held during the same conference. We highlighted how user workspaces can be integrated into an archive, and discussed the various environments in which those workspaces could exist. We presented examples of value-added functions that make an archive effective for long-term use. We led discussions during the I&M meeting to define metadata schemas and also define where enforcement of schemas should take place. We also led discussions to collectively define version 2 of the prototype. This meets the goals for the S4.a milestone.

1.b. Deliverables: During this period, we made available the poster we displayed, and the slides we presented during the I&M session at the GEC. A video demonstration of the prototype was previously made available on the wiki page.

2. Description of Work Performed

2.a. Activities and Findings During This Period

Measurement Data Archive Prototype

We demonstrated the MDA service in GEC 12 during the poster session and highlighted the project during other I&M related sessions in the same conference event. We divided the MDA service into two sub-services corresponding to the User Workspace and Persistent Archive components. Each of those sub-services provides its own storage and data filtering mechanism via its search capability.

The User Workspace component is an entry point for users (e.g., experimenters, instrumentation researchers, etc.) to store and transfer both measurement data, which could be in a variety of forms (e.g., formatted datasets, raw files, etc.), and metadata describing the measurement data. Users can then curate the data and metadata held in the workspace, e.g., make changes to the files, delete the files, etc.

Data and metadata files managed in the user workspace can be archived for long-term storage in the Persistent Archive. Once data is archived, a persistent and unique identifier is created and assigned to it. Additionally, a persistent reference to the data is provided. Discovery and filtering of the archived data is enabled using a search service.

Both components are based on the Digital Object (DO) Repository software that CNRI has made publicly available. DO Repository includes an open, flexible, secure, and scalable protocol and software suite that provides a common interface for interacting directly with all types of Digital Objects. DO Repository normally has a storage module directly attached, where the data to be managed is stored in a data structure that complies with the DO model. In the specific case of the User Workspace, however, the repository connects to a workspace account, made

available to interested GENI members for storing measurement data in the form of directories and files, and maps the measurement data into digital objects. The flexibility of the repository to manage incoming data as digital objects or map existing data into digital objects is one of the many benefits of using the technology for managing information.

While the repository interface enables storage and retrieval of digital objects, additional interfaces are also available as part of this prototype. These interfaces will allow potential I&M users to integrate the workspace and the archive into their daily workflows.

Details on the prototype can be found on the project wiki.

During the I&M session in the GEC 12, several focus areas were identified and sub-groups were formed to deal with each of those focus areas. Metadata and archive services, among other areas, were identified as core for I&M. Sub-groups are formed with Giridhar Manepalli as the leader of those two. Preliminary discussions are to be held between GEC 12 and GEC 13 on these topics.

The sub-group on Metadata and Registry would focus on identifying the various types of objects that get archived, the metadata definitions for each type of object, archival policies of those objects, i.e., whether the objects are encrypted, etc., whether metadata is enforced before archiving, etc., and access control policies.

The goal of the User Workspace and Archive Services sub-group is to identify how several workspaces can be supported to fit with a single archive, how data across users is shared between workspaces, and how many kinds of data are supported for transfer into an archive.

CNRI also plans to participate in the operational-services sub-group discussions on lookup services for the various I&M components, system requirements, etc., for bringing the I&M suite of applications, including MDA, into production.

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. CNRI displayed a poster during GEC 12. CNRI also gave a presentation on the need for archiving during the I&M session and demonstrated the developed MDA Prototype, the slides of which are available on the GENI wiki page for the project.

2.d. Outreach Activities

Giridhar Manepalli attended the GENI Engineering Conference held in Kansas City, MO, and participated in a variety of discussions with GENI members and System Engineers. He participated in the poster session, discussing the GENI services that CNRI offers. He gave a presentation highlighting the project during the I&M session.

2.e. Collaborations

CNRI continued to collaborate with University of Ohio, University of Kentucky, and BBN members, to integrate MDA into their individual environments.

2.f. Other Contributions

Production Services

CNRI continued to support the GENI Federated Clearinghouse and the Distributed Hash Table services on hardware purchased during this contract year. The new hardware is in a collocation facility that features redundant power and air conditioning units, physical security, etc. A 100Mbps network pipe is dedicated to the machine.