

Digital Object Registry

Corporation for National Research Initiatives

Project Status Report – Jan 1, 2011 through Mar 31, 2011

1. Major Accomplishments

The scope of work on this project is to adapt the Handle System as well as 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 extending 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 in building a GENI Federated Clearinghouse and a Distributed Hash Table for Seattle, aka Million Node GENI, led by Justin Cappos. We have offered these services by deploying them in production servers with high availability and network bandwidth.

The Year 3 scope of work, in addition to supporting the above services, is to design a prototype of the Measurement Data Archive service and coordinate with GENI members to integrate their instrumentation and measurement projects into the prototype.

During this quarter, we continued participating in various GENI activities and programs, including attending the GEC 10 held in San Juan, PR, and also continued our collaboration with GENI members and System Engineers as part of the I&M Working Group. We:

- Developed and Deployed the MDA Prototype
- Demonstrated the Prototype at GEC 10.
- Extended the MDA Prototype design document to include the internal details, APIs, and access points.
- Coordinated with the I&M Working Group in the design and specification of a Measurement Data Archive service during the tenth GENI Engineering Conference held in Washington, DC.
- 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: We demonstrated the MDA Prototype at the GEC 10 during the poster session and also during the I&M session. We received initial

feedback from the I&M members during the demonstration. We were also successful in signing up a few early adopters for the prototype service. We will engage and discuss with the GPO any feedback we get from these adopters. By demonstrating and successfully signing up early adopters, we believe we have met the requirements for milestone S3.c.

We have extended the prototype design document, which was released as a deliverable as one of the earlier milestones, to include the implementation choices made, the APIs, and access details. That document is attached to this report. This meets the requirements for milestone S3.d. This also meets CNRI's deliverable milestone considered for the GENI Integration Release (GIR) Phase 3.1

1.b. Deliverables: During this quarter, we made available an updated version of the Measurement Data Archive prototype details (attached and also made available on the GENI wiki page for our project). We also 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 is also on the wiki page.

2. Description of Work Performed

2.a. Activities and Findings During This Quarter

Measurement Data Archive Prototype

Based on the discussions we had with Harry Mussman and other GENI members in this quarter, CNRI designed and implemented the Measurement Data Archive (MDA) Prototype. We divided the MDA service into two sub-services corresponding to the User Workspace and Persistent Archive components.

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 the components are implemented based on the Digital Object (DO) Repository software that CNRI has made available to public. 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 allows for storing and retrieving digital objects, additional interfaces are also made 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.

Early adopters from the GENI community will be given user accounts on the system that in turn enable a file system share that can be mounted on a variety of machines (e.g., personal laptops, server machines, etc.) using the 'SMB' protocol. Once the file system share is mounted, users can copy data into their account mount points. Optionally, users may transfer files securely using SFTP. Users can also access a secure shell (SSH) to modify permissions and group settings, and perform other operations on the transferred data.

Data stored in the workspace, which are mapped to objects, can be searched, browsed, and retrieved using a web front-end, which is available at <http://mda.doregistry.org/>.

Further details on the prototype can be found in the attached document.

2.b. Project Participants

CNRI discussed its project activities with a number of other GENI participants, but all work done this quarter was done 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 the GEC 10. CNRI also presented on the need for archiving during the I&M session and demonstrated the developed MDA Prototype, the slides and video of which are available on the GENI wiki page for the project.

2.d. Outreach Activities

Larry Lannom and Jim French attended the GENI Engineering Conference held in San Juan, PR, and participated in a variety of discussions with GENI members and System Engineers. Larry participated in the poster session discussing the GENI services that CNRI offers. He also gave a presentation on archiving to the I&M members. Giridhar Manepalli participated in the I&M session remotely and demonstrated the workings of the MDA Prototype.

Giridhar Manepalli is supervising an undergraduate student from University of Virginia, who will be responsible for some of this year's GENI work.

2.e. Collaborations

Larry Lannom is also involved in discussions with the DataCite (<http://datacite.org/>) group, an international consortium to establish easier access to scientific research data, in the evolution and specification of a standard for describing and interacting with datasets in a standard fashion. The Measurement Data Archive could leverage this connection for attaining global interoperability across related technologies.

2.f. Other Contributions

Production Services

We continued to support the GENI Federated Clearinghouse and the Distributed Hash Table services. During this reporting period we purchased dedicated server hardware for supporting all GENI activities at a more robust level, including the new MDA prototype. We virtualized the new server into several Linux machines to support the GENI Federated Clearinghouse, the Distributed Hash Table, and the MDA prototype services.

The new server machine is in a collocation facility that is physically in a protected environment, and has redundant power supplies, air conditioning units, etc. A 100Mbps network pipe is dedicated to this machine.