

GENI Project Progress Report

Project Title: Intelligent Data Movement Service on GENI

Principal Investigators:

Ezra Kissel <ezkissel@indiana.edu>

Martin Swany <swany@iu.edu>

1. Major accomplishments: 8/2014–3/2015

These milestones reflect the revised Year 2 statement of work for two quarters covering GEC21 and GEC22 (Oct 2014 and March 2015).

Milestones achieved:

- **GEC21.d.1** - Expand data plane connectivity to include ExoGENI storage VMs.
 - Made use of new AL2S (and existing ION) stitching to connect EG storage nodes with IG shared vlans.
 - Replaces previous method that leveraged the OF mesoscale.
- **GEC21.d.2** - Use external Phoebus Gateways to bridge public R&E connectivity, stitched into IDMS data plane.
 - Available Internet2 gateway nodes were used to bridge between R&E networks and the IDMS private data plane using ION circuits. This allows non-GENI users to make use of IDMS without connecting via a shared vlan.
- **GEC21.d.3** - Make use of public IPs at GENI racks where available.
 - Available R&E addressable IPs at Nysernet and RENCi aggregates were used as an alternative to the external I2 gateway approach.
 - The routable data plane IP configuration at these two racks can hopefully be a model for more widespread availability.
- **GEC21.d.4** - Demo EODN usage at the GEC21 demo session involving external clients.
 - Integrated with USGS EarthExplorer bulk-data download client to demonstrate IDMS storing and downloading Landsat imagery.
- **GEC22.e.1** - Improve client support and add AA mechanisms.
 - New Java download/upload clients were created to support non-CLI mechanisms for interacting with IDMS.
 - Web-GUI developed to track storage resources and infrastructure measurements.
 - Web-GUI includes download visualization to highlight file distribution across storage nodes.
 - Initial AA additions to our schemas and backend have been introduced but not yet fully implemented or exposed via user interfaces.
- **GEC22.e.2** - Incorporate Ceph-based object storage into IDMS experiment alongside existing IBP depots.
 - Prototype Ceph object store image developed and tested.
 - Integration with IDMS has been delayed due to time constraints.
- **GEC22.e.3** - Demonstrate automatic load balancing of user data within IDMS network.
 - Early load balancing functionality is present in DMS dispatcher: evenly distributing new uploads across storage resources based on available capacity. Dispatcher allows for policy to stage specified workflow data at specific storage sites.
- **GEC22.e.4** - Upload documentation detailing experiment setup and usage.
 - IDMS tutorial wiki has been updated to include recent topology and available resources.

Deliverables made:

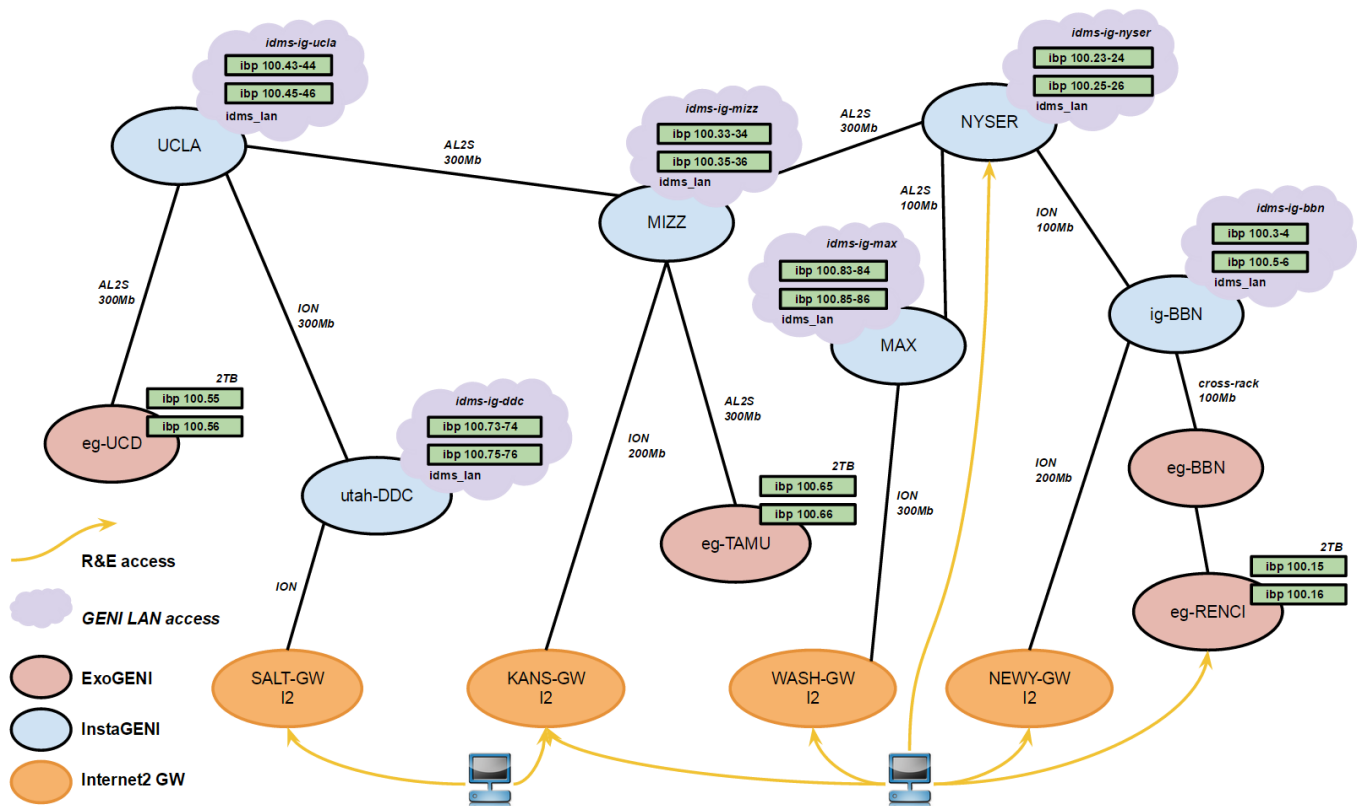
- Demonstrations at GEC21 and GEC22
- Updated appliance images for storage service.
- Updated documentation describing custom image creation and conversion for InstaGENI and ExoGENI.
- Long-lived IDMS slice(s) created to support external experimenter use. Tutorial page published and updated on project Wiki.

2. Description of work performed

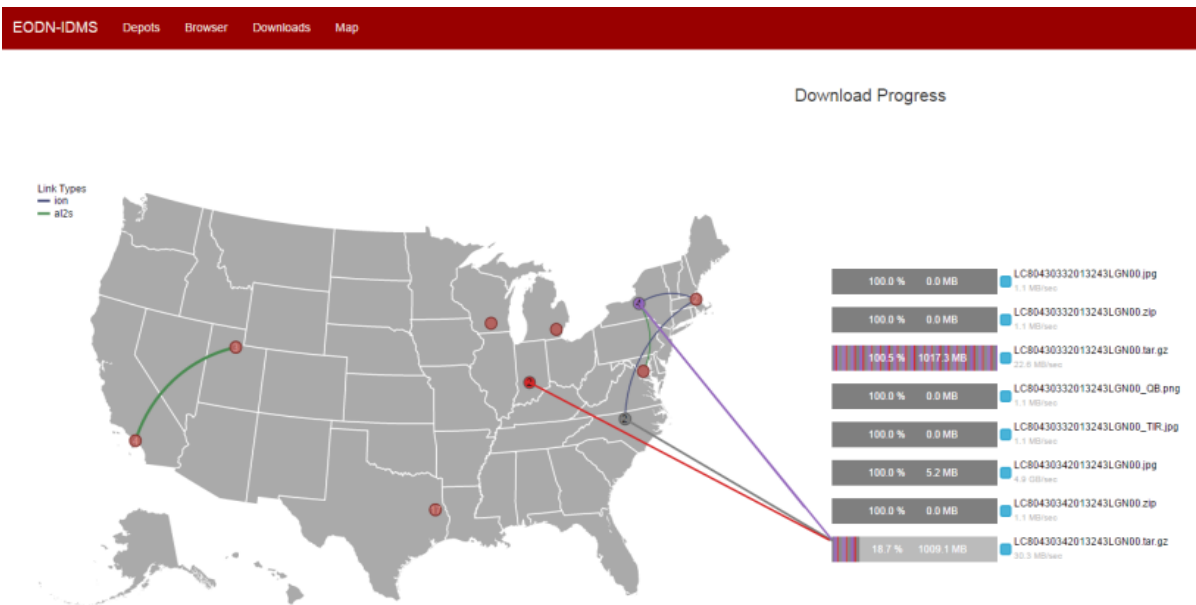
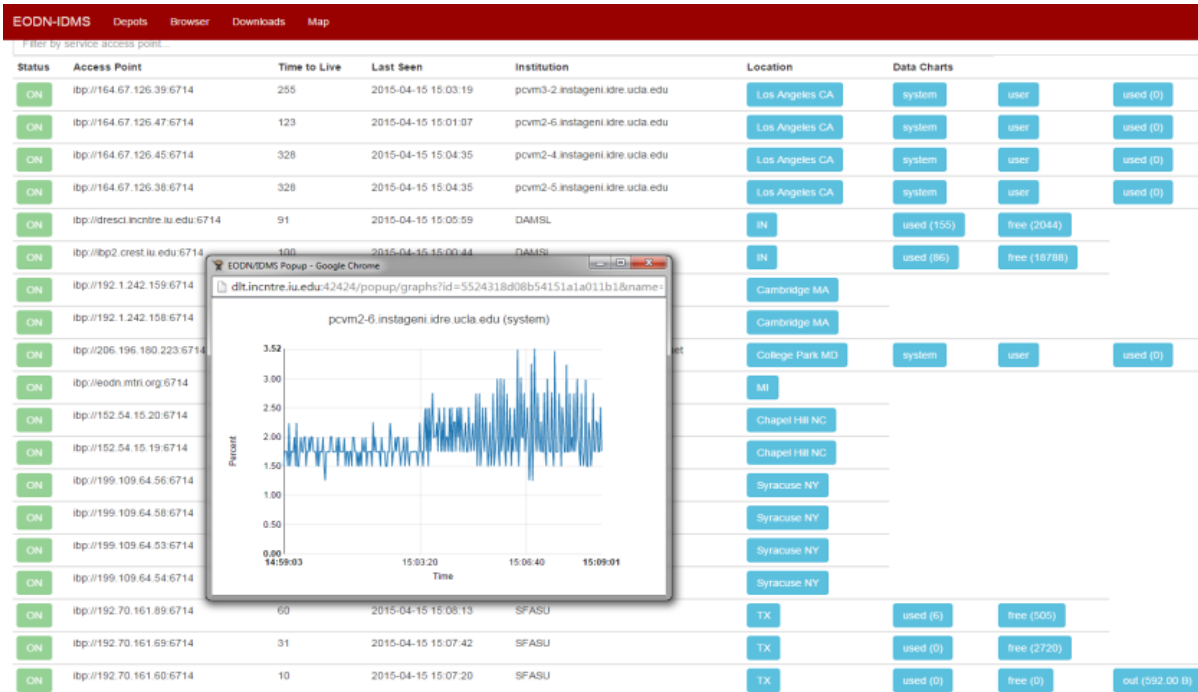
Activities and findings:

We focused on the following development areas since the previous period (GEC20):

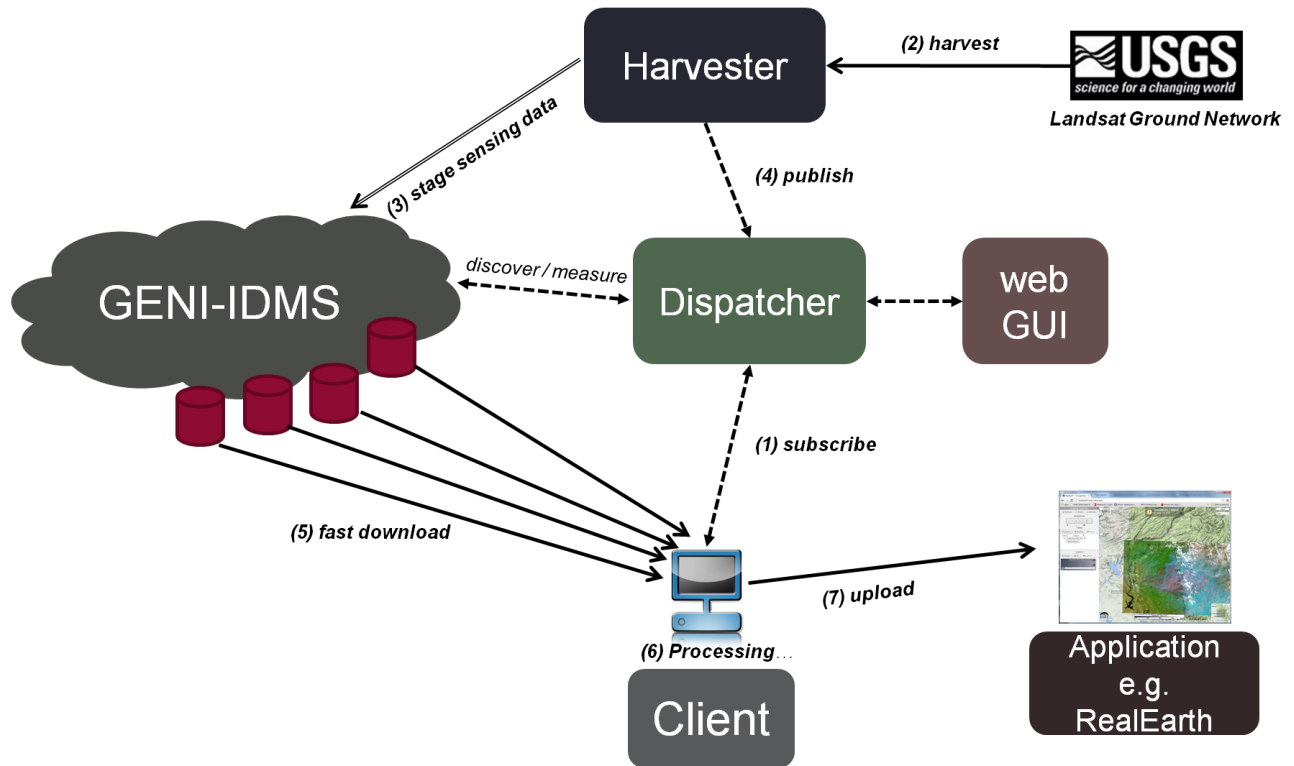
1. A significant effort was made to allow external access to the IDMS service in addition to expanding the number of shared vlan sites for GENI-attached users. ION circuits from Internet2 “gateway” machines were used to forward external, R&E routable traffic into the IDMS data plane. The data plane itself now uses a combination of ION and AL2S stitched links to connect each rack exposing a shared vlan. As ION support is phased out, we plan to migrate to all AL2S stitching where possible.



2. An IDMS web-GUI was created to support our project integration efforts. Key highlights include:
 - a. "Depot status page" – Tracks all registered storage services in IDMS in addition to externally running services that inter-operate with IDMS. Exposes measurement collection with the ability to chart storage, network, and system/cpu utilization.
 - b. "Exnode Browser" – A web-based file browser for querying and selecting files staged in IDMS for download.
 - c. "Download progress" – A view for tracking active uploads/downloads from IDMS, displaying storage locations active for a given transfer.
 - d. Java clients for upload and download of files that are launched from web-GUI.



- Our collaboration with Dr. Sam Batzli (UW-Wisconsin) and his RealEarth application allowed IDMS to drive a specific scientific workflow involving Landsat satellite imagery. Command line interface tools were developed to publish and stage a specific set of imagery from USGS into IDMS, which was then processed and uploaded to a persistent RealEarth instance at Wisconsin for visualization.



Project participants:

Ezra Kissel ezkissel@indiana.edu (PI)
 Martin Swany swany@iu.edu (co-PI)
 Akshay Dorwat adorwat@indiana.edu (RA)
 Jeremy Musser jemusser@indiana.edu (RA)
 Prakash Rajagopal prakraja@indiana.edu (RA)
 Rohit Khapare rkhapare@uemail.iu.edu (RA)
 Joseph Cottam jcottam@uemail.iu.edu (developer)

Publications:

N/A

Internal project documents:

- Available on request from IDMS git repository.

External publications:

- Published on project Wiki: IDMS experimenter tutorial, custom image creation/conversion HOWTO, and updated image table with example RSpec.

Outreach activities:

- Presented GENI-IDMS service to AmericaView community as an integrated, extended storage service for the prototype Earth Observation Depot Network (EODN).

Collaborations:

- Continued knowledge exchange with Paul Ruth (Science Shakedown) and ExoGENI team concerning our shared goals across respective sol4 experiments, e.g., image creation, stitched topologies, and storage.
- Science Shakedown has successfully used IDMS in their workflow as a GENI-attached user.
- Continued outreach with geni-help, orca-users, and protogeni-users communities to ensure IDMS operation across leading aggregate types and backbone nodes.
- Success with RealEarth integration has spurred new work in integrating with their workflow, and specifically with the challenge of specifying policies to stage imagery at locations in IDMS that minimize latency and time to acquisition.

Other Contributions:

- Development of IDMS software has led to improved Python support in Omni as well as driving a subset of features in the “stitcher” code base.