

Facilitating Open Science with the DISC Visualization Collaboratory

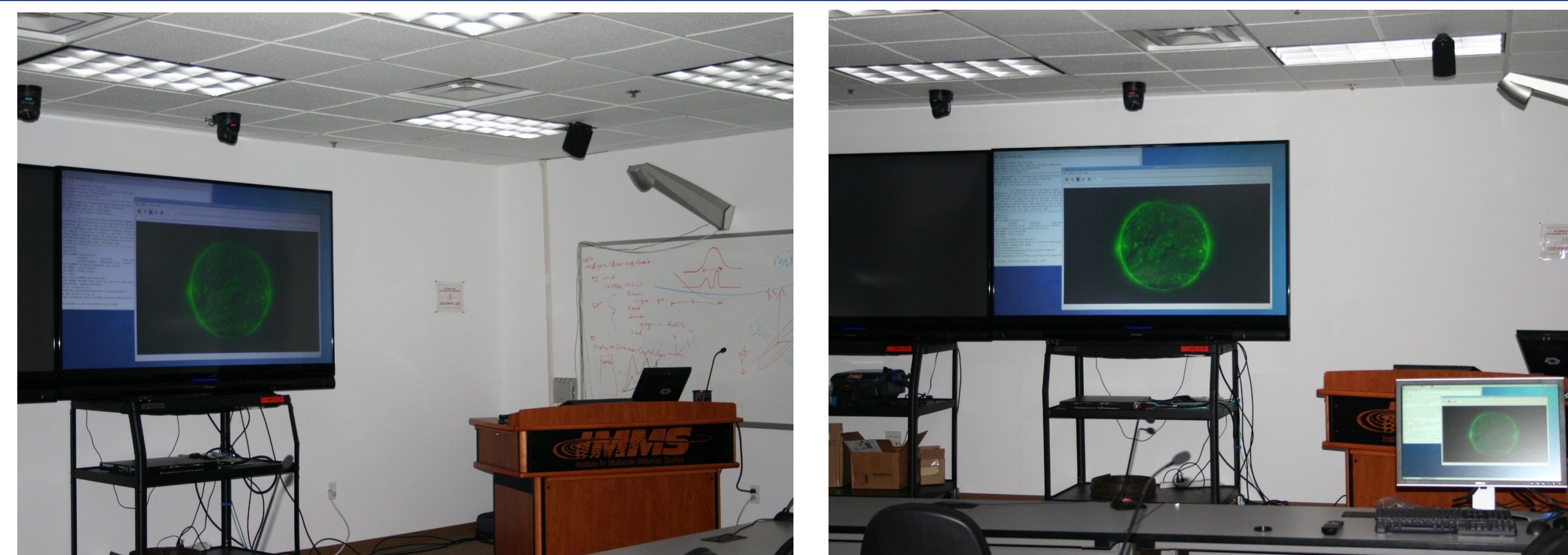
Katherine Nystrom INST-OFF Project Mentors: Gary Grider HPC-DO, Carolyn Connor HPC-5

Overview

The DISC Visualization Collaboratory is a multi-purpose facility designed to provide scientists with an open area for collaboration. The newly upgraded room has an Access Grid system, capability for 3D visualization with stereo, and conference room facilities.

As part of the upgrade network connections were added between the Data Intensive Supercomputing (DISC) cluster and the Collaboratory to provide a resource for 3D visualization and simulations. Ceiling projectors were removed and replaced by three 82" wide-screen stereo-capable TVs. The TVs are used for visualization and as displays for the Access Grid system. Future plans include upgrading to LCD screens and setting up a touch table.

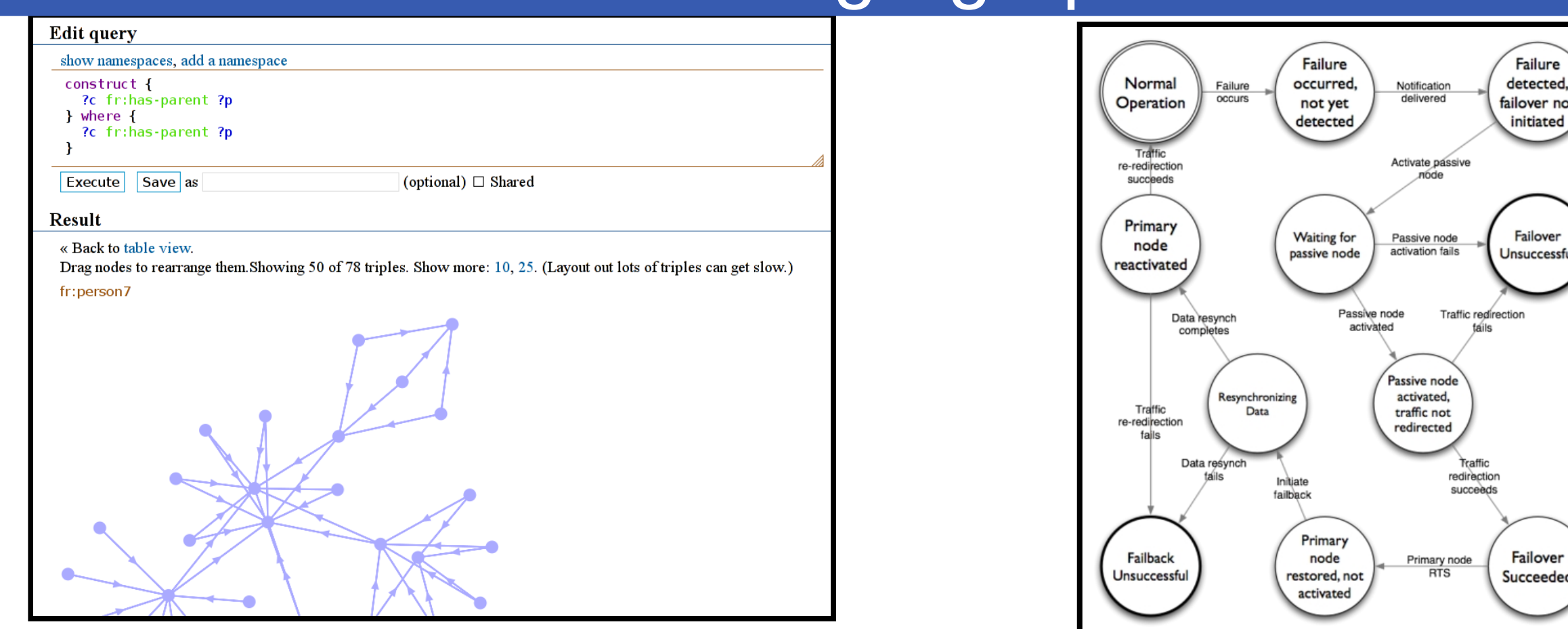
Our goal in upgrading the room is to provide the laboratory with a facility for open science 3D viz and a unique staging area to explore and test new visualization technologies.



The DISC Visualization Collaboratory is a facility designed to provide a high-quality collaboration environment in a multi-purpose space. The Collaboratory includes 3D visualization capabilities which are connected to the Data Intensive Supercomputing Cluster (DISC).

Projects using DISC Viz Collaboratory

Allegrograph

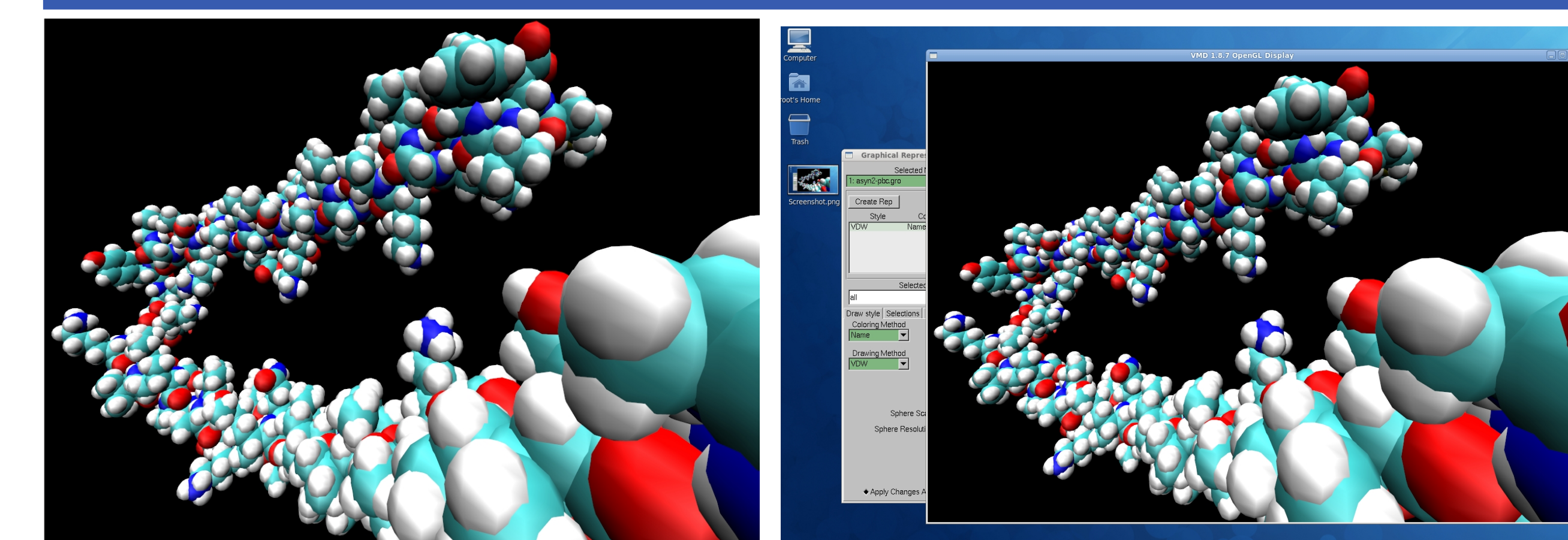


Examples of the graphical representation of data in Allegrograph.

Allegrograph is a software package for creating triple-stores which connect data in subject, predicate, object relationships. Allegrograph uses disk-based storage making it a good fit for the DISC cluster. The software has support for social network, geospatial, and semantic web applications. Since the data is stored in graph format it can be useful to look at the data graphically. The DISC Viz Collaboratory will provide this capability. We are putting a federated installation of Allegrograph on the DISC cluster and already have several interested users.

An example using Allegrograph is the "Map of Science" a project done at LANL in 2006-2008. This project created the largest usage database in the world with data on nearly 1 billion user interactions from scientific portals and websites. Allegrograph was used to store triples representing the interactions captured in the usage logs.

Biology Research on Proteins and Molecules S. Gnanakaran, T-6



Simulation of a protein believed to be a cause of Parkinson's Disease.

Many cellular proteins are intrinsically disordered and undergo folding upon binding to their physiological targets. While the absence of well-defined structures and conformational flexibilities of these intrinsically disordered proteins (IDPs) enable them to bind to multiple partners and to modulate their functional pathways, this transient nature also makes them susceptible to many human diseases. To understand the causes of these diseases and to derive effective treatments, it is crucial to understand the correlation between the underlying structures of these IDPs to their biological functions.

DISC Cluster System

- 3 segments, 255 nodes each + visualization facility
- First segment is on the yellow network with 4 TB/node for a total of 1 PB
- Second segment is targeting the turquoise network with 4 TB/node for a total of 1 PB
- Third segment divided into two parts: one with 128 nodes with 128 GB solid state drives (SSDs), other with 127 nodes with 256 GB/node for testing purposes
- Current user applications include biology, data search and filtering, network security and development of data-intensive tools
- Future applications include cosmology and image processing

CASL Hub Support

The Consortium for Advanced Simulation of Light Water Reactors (CASL) is a new collaborative effort spearheaded by Oak Ridge National Laboratory (ORNL). LANL is a partner in this effort and the DISC Viz Collaboratory will be used to facilitate interactive collaborative meetings between the two laboratories and other CASL collaborators.

Goals

- Provide an open facility for researchers to experiment with data intensive methods
- Discover new ways to use visualization tools such as fusing data from simulation and sensors
- Expand LANL's expertise in data fusion and data intensive computing for science