Project Overview

Our work as proposed for the NSF DIBBS award #1443040 will provide extended functionality for the Tripal toolkit (http://tripal.info), which is an open-source, freely available software package that provides a framework to assist research groups publish genomic, genetic and related biological data in an online searchable format within Drupal, a popular content-management toolkit. This DIBBS award will allow us to provide greater cyberinfrastructure to these communities through development of three modules: 1) Data exchange module for a global federated network, 2) scientific workflows module for large-scale data analysis using the Galaxy workflow tool and 3) a module for improved data transfer methods, including integration with national research networks (i.e. Internet2).

Accomplishments

Journal Articles

Conference Papers & Presentations

Software Releases
- Tripal v3 alpha (Jan 2016)
- blend4php (Jun 2016)
- BDSS v1.0.1b1 (August 2016)
- BDSS v1.0.1b2 (Nov 2016)

Other Resources
- Galaxy Instance for analytics: https://galaxy.bioinfo.wsu.edu/
- Tripal v3 demo site: http://demo.tripal.info/3.x/

GitHub Repositories
- Tripal v3 development: https://github.com/tripal/tripal/tree/7.x.x
- Tripal Galaxy module: https://github.com/tripal/tripal_galaxy
- blend4php: https://github.com/galaxyproject/blend4php
- BDSS: https://github.com/feltus/BDSS
- Galaxy Workflows: https://github.com/MingChen0919/docker-galaxy-dibss/tree/master/my_workflows
- Galaxy Workflows in Docker: https://github.com/MingChen0919/docker-galaxy-dibss/tree/master/my_workflows

Example Community Databases
The following sites are just a few example community databases that use Tripal. The technologies developed by our DIBBS work can have immediate impact for these sites and more...

Module #1 Data Exchange

Tripal v3.0 Gene page via web services

Module #2 Scientific Workflow Execution: Integration with Galaxy.

Our goal is to ensure best transfer speeds for data exchange (module #1) and movement of data for scientific workflow execution (module #2). Leveraging an NSF ACI-REF award (#1341935) with collaborators at Clemson we hope to develop protocols for Software Defined Networking (SDN) to support Tripal-site data exchange. We have developed Big Data Smart Socket (BDSS) software to use knowledge of existing network paths, remote repositories transfer protocols, and client software available on the local machine to make smart decisions about data transfer. The graph to the left demonstrates use of BDSS on a CloudLab-based DTN with a parallel file system transferring data between Clemson and University of Utah over Internet2 AL2S.

Module #3 Data Transfer

Our goal is to ensure best transfer speeds for data exchange (module #1) and movement of data for scientific workflow execution (module #2). Leveraging an NSF ACI-REF award (#1341935) with collaborators at Clemson we hope to develop protocols for Software Defined Networking (SDN) to support Tripal-site data exchange. We have developed Big Data Smart Socket (BDSS) software to use knowledge of existing network paths, remote repositories transfer protocols, and client software available on the local machine to make smart decisions about data transfer. The graph to the left demonstrates use of BDSS on a CloudLab-based DTN with a parallel file system transferring data between Clemson and University of Utah over Internet2 AL2S.