

CC*DNI DIBBs: Multi-Institutional Open Storage Research InfraStructure (MI-OSiRIS)

*PI: Shawn McKee, NSF CC*DNI DIBBs grant #1541335*

Overview

The OSiRIS project is exploring the creation of a multi-institutional storage infrastructure based upon Ceph and inter-connected by Software Defined Networking (SDN) capable components. The project goal is to support multi-institutional collaboration on large, diverse or dynamic data, allowing scientists at any participating institution to be able to directly access the same data as their collaborators from well-connected computing locations on campus.

Challenges Met

During the first year of the project we made extensive progress in understanding the components we are using for OSiRIS and the overall architecture we are implementing.

- **Equipment:** Defined a set of hardware that will meet OSiRIS needs; fully deployed and tested.
- **Ceph:** its configuration, operating modes, failure modes, performance and tuning. Demo at SC16 showed we could extend OSiRIS between Michigan and Salt Lake City. More to learn.
- **Network Management Abstraction Layer (NMAL):** we have designed how we will discover, monitor and model the network for use by OSiRIS. Prototypes deployed.
- **Authentication and Authorization(AA):** The initial model of AA has been developed and builds upon InCommon and Shibboleth. Consultation with CTSC ongoing (<http://trustedci.org/>)
- **Documentation:** We develop our code in GitHub, maintain project web pages and Wiki pages on the details of our work (<http://www.osris.org/>), and manage the project using OpenProject.

Challenges Remaining

After getting the basic OSiRIS components defined, prototyped and deployed we are focusing on bringing in our first two science domains: high-energy physics and physical ocean modeling. In general how best to incorporate and optimize OSiRIS for our science users is our primary challenge. However we have many details that need resolution:

- **Equipment:** Are the hardware choices made still optimal, new features
- **Ceph:** Upgrades, use of erasure-encoding vs replicas, use-case specific tuning, monitoring
- **NMAL:** Must work toward production quality; must incorporate network control/management.
- **AA:** Continued development, testing and evolution to match science domain requirements
- **Documentation:** Track project developments, how-tos and add science domain use-cases

Future Directions

Improving the process to integrate new science domains and optimize their interface to OSiRIS provided storage services. Implementation of data life-cycle management metadata automation. Expanding the toolkit of system configurations available to meet science domains use-cases. Expanding OSiRIS to new institutions. Creation of a set of whitepapers on how to deploy an OSiRIS site and enable a science domain. Implementation of SDN components to orchestrate the networks in use by OSiRIS.