

North East Cyberinfrastructure Consortium

For this current visit, the AAAS panel was asked to review and provide guidance on the five-state North East cyberinfrastructure collaborative (NECC) in which VGN participates. Established in 2006 through Track-2 grants from NSF EPSCoR to programs in Delaware, Maine, New Hampshire, Rhode Island, and Vermont, the NECC aims to 1) identify and promote the shared use of research facilities across the region; and 2) assess and address cyber-infrastructure needs. The University of Vermont has been designated the lead institution for the NECC.

Funding for the NECC is heavily leveraged, with federal contributions from NSF and NIH, and is used to support the purchase and lighting of physical fiber, personnel to maintain the fiber network and enable collaborative research, and two regionally distributed data centers (in Delaware and Maine). In addition, the NECC has created a funding stream to support cyber-enabled pilot research projects involving collaboration between researchers throughout New England.

To better understand the NECC, its relationship to VGN, and its impact on the region, the AAAS panel met with stakeholders from the University of Vermont as well as members of the NECC Executive Committee from each of the participating 5 states.

NECC is currently building out the high-speed fiber optic backbone which will interconnect its member institutions, as well as provide high-speed connectivity to regional and national research networks. The new fiber will create dual, diverse paths through Vermont and New Hampshire, as well as additional paths in Maine and Rhode Island. Importantly, in addition to directly connecting the member institutions, the new fiber will provide high-speed connectivity to the broader Northeast Research and Education Network (NEREN) which includes the Northern Crossroads, OSHEAN, Nysernet, the University of Maine and others.

The network is in various stages of completion within the individual states. The fiber pathways through Vermont, which will complete the loop with New Hampshire, were in the process of being laid during the AAAS visit. New fiber in Maine will complete a Maine loop via the MDI Biological Lab, as well as provide a path to the northeastern border with Canada near Presque Isle. Maine's connectivity is highly leveraged from their BTOP funding and other state networking infrastructure. Rhode Island is also highly leveraging their existing OSHEANS network infrastructure, as well as a new pathway to Kingston, and their build-out is 70% complete.

The project is ending its second year of NSF Track-2 and NIH funding. Once the installation of the remaining fiber and loop is completed, the network will be capable of providing multiple pairs of 10 gigabit/second circuits, a huge increase in available bandwidth for all of the members. The partners are providing good leverage from a variety of funding sources, including the NSF and NIH, but also others, as well as

leveraging existent WAN infrastructure. Much of the current funding is being used to fund long-term leases of fiber. Additionally, most of the institutions will be utilizing the new NECC network paths to replace their other commercial circuits and redirecting their cost-savings into the NECC. This strategy of acquiring long-term leases and redirecting cost savings bodes well for the long term sustainability of the network. All of the institutions have committed to the network operation for at least ten years.

Although the NECC is still in its early stages, the project is already having significant impact locally and regionally. Numerous examples of the collaboration and cooperation between the parties were exhibited during the visit, including several “proof-of-principle” projects. These include the Little Skate Genome and the Blue-green Algae Metagenomics projects. During the meetings with the participants, it became clear that the foundations for this project had been set through numerous prior interactions and collaboration that have been taking place throughout the region, many of these being driven by INBRE and EPSCoR programs. Cooperation within each of the institutions was equally in evidence with good vertical coordination between the researchers, the EPSCoR and INBRE directors, the Bioinformatics facility, the CIO’s office, and senior university leadership.

Regional impacts

Once the NECC network is fully operational, it should dramatically increase the member institutions ability to engage in cyber-enabled research and education by providing significantly increased network capacity and the increased availability afforded by the redundant optical network paths. In addition to creating enhanced network connectivity, the NECC is creating benefits from the increased levels of collaboration and coordination between the participants. Even in these early stages, a number of collaborative projects have been initiated and appear to be working well.

a. Two examples of collaborative projects created by the NECC are the Little Skate Genomics project and the Blue-green Algae Metagenomics project. These projects resulted in collaborative research across several states, educational opportunities for a large number of students at multiple institutions and a new genomics data base published for the broader research community. These projects would not have been possible without the NECC. To foster collaborative research, NECC has created the Regional Awards for Cyber Enabled Research (RACER) program to fund small (~\$10K) pilot projects. Funded projects must involve investigators from multiple institutions. In its first year RACER funded the start-up of the metagenomics project. In year two RACER funded a UNH/UVM project to create a metagenetics data base for eukaryotes. This year six pre-proposals have been submitted for the current solicitation, and two proposals have been funded in year 3. The first proposal is to study microRNAs expressed in response to hypoxia in, *Fundulus heteroclitus*, and the second is to develop a cyber-enabled Northeastern Monitored Lake Temperature Network. Both are aligned with the NECC mission and support environmental or bioinformatics collaborations among investigators from at least two jurisdictions within the

NECC.

The NECC has afforded the opportunity to create a number of shared resource centers across the region, most notably a pair of data centers, one in Delaware and another in Maine. These data centers are providing researchers with access to large data storage arrays. NECC has developed cybertools to house the primary data storage in Delaware and to keep a live, replicated copy in Maine. The shared data centers played a key role in the Skate Genome and Metagenomics projects. Within the framework of NECC the partners have developed a data access and sharing plan which will be beneficial to all of the members. Once the new network connections become fully operational, the demand on the shared data centers is likely to increase substantially.

Additionally, Delaware has received an MRI grant to acquire a next-gen sequencer, in substantial part due to their participation in the NECC. This will be a valuable resource for all of the partners. Delaware's sequencing capabilities are playing a key role in many of the NECC projects, such as the Skate Genome.

As a result of the NECC, the partners are beginning to institute curricular changes within their individual institutions and states. There are several initiatives to bring bioinformatics training into the educational pipeline. Rhode Island is working with the Rhode Island School of Design (RISD) on the visualization of genomics data and working around the state to develop curricula that integrate these technologies. New Hampshire and Delaware are also developing programs to introduce bioinformatics into the classroom. The NECC Watershed project is engaging undergraduates and high school students across and beyond the region. Many students from multiple institutions were engaged through the workshops associated with the Skate Genome project. Although currently unfunded, the partners have proposed the development of a cyber-based, undergraduate bioinformatics course. This course would use the shared datacenter resource to analysis large genomic data sets and use video conferencing over the NECC network for distributed, remote teaching. Projects such as this which leverage so many aspects of the NECC-provided cyberinfrastructure capabilities are laudable.

All of the NECC team members that met with AAAS panel were very positive about the collaborations, as well as the impact that the project was having on their institutions. The NECC members at several levels meet regularly using video conferencing. The organization appears to be running well and being quite productive.

Vermont impacts

The new NECC fiber will have a dramatic impact on networking at UVM. Once complete the new fiber will afford two orders of magnitude increase in available bandwidth. This dramatically increases the opportunities for UVM to engage in cyber-based research and education. Beyond the local region, the new network will provide enhanced access to other network resources including Internet2, the NEREN, the

proposed U.S. UCAN pop in Albany, and NSF and NIH resources such as XSEDE, the next generation Teragrid.

Using NIH ARRA supplements, VGN has hired a bio-IT professional to develop the shared data center. A new full time position, as well as two full-time summer internships, has been created within the Bioinformatics Core to support large scale sequencing projects for collaborative research and to build capacity in the metagenomics area.

To fully benefit from the NECC capabilities, the networking infrastructure would need to extend throughout the state. Since Vermont does not currently have a state broadband network, this remains a challenge. Bringing advanced cyberinfrastructure to other educational institutions throughout the state, including the BPIs, would be an important, but very costly, undertaking.

Recommendations:

- The RACER program to fund pilot research projects is very creative and offers a good proof of principle for the network. This program will be important to continue to spur collaboration, utility, and evolution of the network.
- Regional collaboration seems to be working well. However, the risk of interconnectivity is reliance on each state to do their part. It will be important to define how that will be monitored and enforced.
- VGN may consider asking Joy Livingston, the project evaluator, to attempt to assess the impacts of NECC.