

Extended Abstract: Northeast Cyberteam - Methods, Results and Expansion via the Connect.Cyberinfrastructure Portal

John Goodhue
Massachusetts Green High
Performance Computing Center
Holyoke, MA USA
jtgoodhue@mghpcc.org

Julie Ma
Massachusetts Green High
Performance Computing Center
Holyoke, MA USA
jma@mghpcc.org

Adrian Del Maestro
University of Tennessee
Knoxville, TN USA
adel@utk.edu

Sia Najafi
Worcester Polytechnic Institute
Worcester, MA USA
snajafi@wpi.edu

Bruce Segee
University of Maine
Orono, ME USA
segee@umaine.edu

Scott Valcourt
University of New Hampshire
Durham, NH USA
s.valcourt@northeastern.edu

Ralph Zottola
University of Alabama
Birmingham, AL USA
rzottola@uab.edu

Abstract— *Computing has become an essential component of research and education for nearly every scientific discipline. Meeting the need for research computing facilitators (RCFs) - support staff who can help faculty make the best use of available computing resources - is a significant challenge for small and mid-sized institutions that do not have the critical mass to build teams of RCFs on site. The NSF-sponsored Northeast Cyberteam is addressing this challenge by building a process by which research computing facilitators can be shared across institutional boundaries while also developing self-service tools that reduce the support burden.*

The Connect.Cyberinfrastructure Portal, developed to support the Northeast Cyberteam, has enabled adoption of program methods by other communities of practice and collaboration with the broader research computing community. A pilot expansion effort was initiated in July 2020 to explore collaborative opportunities. Progress and preliminary findings are discussed.

Keywords— *workforce development, research computing facilitator, project portal, Ask.CI, MGHPC, Northeast Cyberteam*

I. THE NORTHEAST CYBERTEAM PROGRAM

Cyberinfrastructure is as important for research in the 21st century as test tubes and microscopes were in the 20th century. Familiarity with, and effective use of cyberinfrastructure at small and mid-sized institutions is essential if their faculty and students are to remain competitive. Launched in 2017 with NSF-funding, Northeast Cyberteam Program [1,3,4] aims to increase effective use of cyberinfrastructure by researchers and educators at small and mid-sized institutions in Northern New England, by making it easier to obtain support from Research Computing Facilitators (RCFs). The program combines direct assistance to computationally intensive research projects; experiential learning opportunities that pair experienced mentors with students interested in research computing facilitation; sharing of resources and knowledge across large and small institutions; and tools that enable efficient oversight and possible replication of these ideas in other regions.

RCFs combine technical knowledge and strong interpersonal skills with a service mindset and use their connections with cyberinfrastructure providers to ensure that researchers and educators have access to the best available resources. It is widely recognized that RCFs are critical to successful utilization of cyberinfrastructure, but in very short supply [2] and that this is currently one of the most significant barriers to productive use of research computing at small and mid-sized institutions.

To help meet this need while also developing a next generation workforce, the Northeast Cyberteam Program provides direct support to faculty at small and mid-sized institutions by launching projects that pair a student-facilitator with a mentor with domain expertise to address an immediate faculty research gap. Projects draw mentors and students from across the region. Mentors are usually professional research computing facilitators from research computing groups at larger institutions in the region who have subject matter expertise relevant to the project. As of October 2020, we have launched 44 projects at 21 institutions, pairing a diverse population of student RCFs with knowledgeable mentors to assist researchers and educators in the region. Concurrently, we have developed a self-service learning toolkit, described below, to provide timely access to information when it is needed.

Program direction is set by a Steering Committee composed of leaders from each of the larger institutions that serve as anchors for the Northeast Cyberteam: University of Maine, University of New Hampshire, University of Vermont and MGHPC; a program manager who coordinates day to day activity; and key personnel from a few other institutions that provide a source for student-facilitators and mentors. The Steering Committee approves all projects undertaken. For selection of projects, the Steering Committee relies less on competitive applications (though merit will naturally play a role), and more on outreach to faculty at smaller institutions who can benefit from access to cyberinfrastructure but are

either unaware of available resources or have given up after a poor experience. Care has been taken in sourcing and monitoring projects to ensure that they 1) lead to results that might not otherwise have been achieved and 2) establish a model for engagement with the Northeast Cyberteam that others at the institution can follow.

II. KNOWLEDGE SHARING AND SELF-SERVICE LEARNING RESOURCES

Providing tools to enable self-service learning is a key to our strategy of developing facilitators through experiential learning, recognizing that one of the most fundamental skills of successful facilitators is their ability to quickly learn enough about new domains and applications to then be able to draw parallels with their existing knowledge and help to solve the problem at hand. There is usually not enough time to enroll in a traditional training course or attend a seminar when a new domain or application is encountered.

The Connect.Cyberinfrastructure (Connect.CI) Portal (<https://connect.cyberinfrastructure.org>) is used to access and aggregate self-service learning resources that provide just-in-time information delivery to participants as they embark on projects in unfamiliar domains. The goal of these learning resources is to reduce the need for direct assistance; reduce duplication of effort by adapting and building awareness of available documentation, training, application software and software utilities; and supplement these resources where there are high impact opportunities. A common tagging infrastructure and voting capabilities modeled after crowd-sourced repositories such as Stack Exchange, provide a uniform underlying infrastructure. This infrastructure allows a user to click on a tag from any part of the portal and obtain a listing of all portal content, including mentor profiles, project profiles, frequently asked questions, and training resources.

The self-service learning section of the Portal is designed to accommodate three types of information commonly needed by RCFs: 1) Frequently asked questions whose answers evolve over time as technology advances, serviced by Ask.CI, a Q&A Platform for Research Computing [5]; 2) Relatively static information such as introductory training modules on linux clusters, programming languages and schedulers, serviced by a Learning Resources Wiki; and 3) Dynamic, situation-specific information needed to solve an immediate problem. Typically handled by a Help Desk at larger institutions, this need serviced by a Regional Help Desk which is staffed by Cyberteam student-facilitators.

The Northeast Cyberteam Program also relies heavily on the Connect.CI Portal for management of project workflows, recruitment of mentors and student facilitators, and recording results. Anyone interested in working on a Cyberteam project as a researcher, mentor or student facilitator can create an account and become part of the community. The management

section of the Portal captures experiences, lessons learned and impacts on a per project basis which have subsequently been used in publications and reporting.

The Portal provides mentor matching and targeted learning resource identification utilizing the underlying tagging mechanism. For example, project manager seeking a mentor with particular expertise can search by individual tags to obtain a listing of all portal users that have selected that tag as a skill. Recognizing that: 1) these matching functions benefit from broad participation because deep knowledge is widely dispersed throughout the research computing community and 2) many Portal functions have utility beyond the Northeast Cyberteam, the Portal was developed with an eye toward making it possible for other communities to adopt it while maintaining their own branding and project workflows.

III. RESULTS/LESSONS LEARNED

We have launched 44 projects over the past three and a half years. With a few exceptions, the results have been extremely promising with many successful short, 3- to 6-month projects yielding publishable results. We are also beginning to see impact beyond just the individual project level, with some smaller institutions starting to treat research computing as an ordinary part of the research and education toolkit instead of a distant luxury item. Although there is still much to do, we have enough experience to draw some preliminary conclusions, first reported in [1] at the Second Workshop on Strategies for Enhancing HPC Education and Training (SEHET19) at PEARC19 and described below.

A. *Value of RCFs to research and education at small and mid-sized institutions*

Consistent with the findings of the report that inspired the Northeast Cyberteam Program [2], the number of research projects that can benefit from RCFs is limited only by our ability to find and recruit them, which is improving over time.

B. *Ability of finite-length student projects to fill the need*

Overall, we have been impressed by the quality and responsiveness of the students who have participated in the program. Interestingly, we have had success with grade levels ranging from sophomore to post-doctoral. We have almost always been able to structure an assignment that moves the project from one reasonably well-defined state to another. Examples include a) moving from a workstation to a cluster for greater throughput; b) improving the performance or throughput of a workflow in order to generate results with faster turnaround or in greater volume; and c) adopting a new computing tool such as Jupyter notebooks.

C. *Willingness of mentors to participate*

Experience over the past three and a half years has validated our hypothesis that experienced RCFs would be willing to serve as mentors as part of their regular jobs.

D. Ability to build teams across institutional boundaries

This aspect of the program was critical to success, and we found that the two most likely obstacles have not been significant impediments. Our first concern was distance, however tools such as Zoom, Google docs and Github have made it possible to execute on projects without ever meeting in person. The second concern was administrative, as grant administrators understandably lean toward applying funds in ways that benefit students and faculty at their home institutions. While each co-PI has needed to spend some extra effort explaining the purpose and benefit of the program, this has not prevented cross-institution assignments.

E. Willingness of larger institutions to share information

In addition to providing mentoring, research computing groups at larger institutions, have contributed extensively to Ask.CI both for the initial building of the shared Q&A list, and more recently, locales.

F. Importance of active program management

The second largest expense category for the program after student stipends is support for a project lead at each Anchor Institution and the Program Manager who manages the overall program. While the value of program management is often overlooked, this investment has been critical to success. It has enabled several important outcomes, including a) efficient recruiting of projects, students and mentors; b) development of process, tools, and strategy; c) effective communication across the anchor institutions and d) attaining institutional support.

We have developed processes and tools that will reduce the need for active management and coordination. However, it seems likely that at least *some* active management, will be required for ongoing success.

IV. STATE OF THE PRACTICE

The Northeast Cyberteam (NECT) Program allows researchers at small and mid-sized institutions to take advantage of cyberinfrastructure when their work requires it. Simultaneously, it exposes a new generation of potential facilitators to this exciting and dynamic field earlier in their careers than most practicing research computing facilitators came into the practice, expanding the available pool of candidates. Our sustainability explorations have yielded opportunities to collaborate with other programs focused on workforce development for the research computing community and we continue to seek additional partners.

Collaboration is facilitated through the use of the Connect.CI Portal. In a pilot launched in July 2020, the NSF-funded CAREERS, Great Plains Network, RMACC, SWEETER, Kentucky, TRECIS Cyberteams; and Colorado

School of Mines have been exploring use of the Portal as a management tool for their related programs.

Each of the Expansion Pilot participants is exploring aspects of the Portal that meet their program needs. For example, the Kentucky Cyberteam is exploring use of the Regional Help Desk and its associated ticketing capability to track engagements with hundreds of constituents. TRECIS will use an Ask.CI Locale as the entry point into their Cyberteam resources. In perhaps the most comprehensive effort, the CAREERS Cyberteam adopted the NECT experiential learning model and engaged the NECT Manager to serve as a Co-Program Manager to ensure efficient transfer of knowledge. New features are also being added to the Portal based on input from the Expansion Pilot participants. Representatives from each program meet monthly to exchange experiences, prioritize feature requests and to curate the Tag taxonomy.

While the Portal expansion is still in its initial stages, the preliminary results are promising and support the premise that broad participation in the Portal will yield greater cross-program collaboration and the timely location of subject matter experts when needed. We welcome participation from the broader community as individual “at large” members or as additional communities of practice.

ACKNOWLEDGMENTS

We thank the many Cyberteam student facilitators for their participation in this effort and the Cyberteam mentors, researchers and educators who have created experiential learning experiences for them; the Ask.CI site and locale moderators, and the hundreds of contributors to Ask.CI; and to the leaders of the CAREERS, Great Plains Network, Kentucky, RMACC, SWEETER, TRECIS Cyberteams; and Research Computing leaders at the Colorado School of Mines Research Computing, that enabled the participation of their respective constituents in the Connect.CI Portal Expansion pilot.

REFERENCES

- [1] J. Goodhue, J. Ma, A. Del Maestro, B. Segee, S. Valcourt, S. Najafi, R. Zottola, *Northeast Cyberteam Program - a Workforce Development Strategy for Research Computing* 2019 JOCSE DOI: <https://doi.org/10.22369/issn.2153-4136/11/1/2>
- [2] G. Monacom, G. Huntoon et al. *The Role of Regional Organizations in Improving Access to the National Computational Infrastructure* NSF June, 2016
- [3] J. Goodhue, J. Ma et al, *Northeast Cyberteam: Workforce Development for Research Computing at Small and Mid-sized Institutions* 2020. PEARC20 ACM Proceedings, NY, NY.
- [4] J. Goodhue, J. Ma et al, *Northeast Cyberteam – Building an environment that encourages publishing, sharing best practices, and solutions*, 2020 IEEE HPEC Proceedings.
- [5] J. Ma, J. Goodhue et al, *Ask.Cyberinfrastructure.org: Creating a Platform for Self-Service Learning and Collaboration in the Rapidly Changing Environment of Research Computing*, 2020 JOCSE, Accepted for publication.