Northeast Cyberteam Program – A Workforce Development Strategy for Research Computing

John Goodhue

MGHPCC
Holyoke, MA
jtgoodhue@mghpcc.org

Julie Ma

MGHPCC
Holyoke, MA
jma@mghpcc.org

Adrian Del Maestro

University of Vermont
Burlington, VT
adrian.delmaestro@uvm.edu

Sia Najafi

Worcester Polytechnic Institute
Worcester, MA
snajafi@wpi.edu

Bruce Segee

University of Maine
Orono, ME
segee@maine.edu

Scott Valcourt

University of New Hampshire
Durham, NH
scott.valcourt@unh.edu

Ralph Zottola

University of Alabama
Tuscaloosa, AL
rzottola@uab.edu

ABSTRACT

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.

The Northeast Cyberteam Program is a 3-year NSF-funded regional initiative 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.

Research Computing Facilitators 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 Research Computing Facilitators are critical to successful utilization of cyberinfrastructure, but in very short supply. The Northeast Cyberteam aims to build a pool of Research Computing Facilitators in the region and a process to share them across institutional boundaries. Concurrently, we are providing

experiential learning opportunities for students interested in becoming Research Computing Facilitators, and developing a self-service learning toolkit to provide timely access to information when it is needed.

Keywords

workforce development, research computing facilitator, project portal, Ask.CI, MGHPCC, Northeast Cyberteam

1. INTRODUCTION

The Northeast Cyberteam Program is a National Science Foundation (NSF)-funded initiative to increase effective use of cyberinfrastructure by researchers and educators at small and mid-sized institutions in Northern New England (Maine, Massachusetts, New Hampshire, Vermont). 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.

2. STRATEGY AND METHODS

The core of our strategy is to build a regional pool of research computing facilitators (RCFs) and a process to share them across institutional boundaries, augmented by knowledge sharing and self-service learning tools that increase the effectiveness of Research Computing Facilitators. To encourage the face-to-face communication necessary for effective mentoring and cross

institution resource sharing, we have maintained a regional focus, with oversight from anchor institutions in each participating state. For efficiency, and to open the possibility of replicating these ideas in other regions, we have developed a portal for management of project workflows.

2.1 Building a Regional Pool of Research Computing Facilitators

Research Computing Facilitators 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 Research Computing Facilitators (RCF) are critical to successful utilization of cyberinfrastructure, but in very short supply[1].

Since most small and mid-sized institutions cannot individually support a research computing department, the Northeast Cyberteam aims to develop a sustainable pool of facilitators who can work across institutions in the region.

The project gains further leverage by partnering with the large research universities in the Massachusetts Green High Performance Computing Center (MGHPCC) consortium, and with national programs such as the Campus Champions.

To deliver direct assistance to research and education projects while giving students experiential learning opportunities, we developed a model where researchers are paired with student facilitators, typically individuals with an affinity for computationally intensive research, but often with little or no domain expertise relevant to the project. Mentors provide subject matter expertise, and guide the project in a direction that will yield results over a 3-6 month period. This gives the student an opportunity to practice facilitation skills, gain some hands on experience with advanced computing resources, and learn a new domain.

This method of exposing a student to a new scientific domain, with a mentor who provides a safety net of subject matter expertise while modeling how facilitation should be provided, expands the student's domain knowledge and ability to apply computing skills in new situations (a common *modus operandi* for Research Computing Facilitators).

By matching students, mentors, and projects across institutional boundaries, the program expands the skill sets available to all participants in the pool, and provides 'bench depth' that makes it easier to manage turnover, handle bursts of activity, and foster communication among peers to accelerate professional growth.

2.2 Knowledge Sharing and Self-Service Learning Tools

Providing peer-validated tools to enable self-service learning is a key to our strategy of developing facilitators through experiential learning. We recognize 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. This is especially true of researchers who may face a particular computational roadblock in their pursuit of a result.

The Cyberteam Portal is used to access the self-service learning resources developed to 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, and reduce duplication of effort, by adapting and building awareness of available documentation, training, application software, and software utilities, and by supplementing these resources where there are high impact opportunities.

Using a common tagging infrastructure and voting capabilities modeled after crowd-sourced repositories such as StackExchange, we are building a uniform underlying structure. This allows a user to click on a tag from any part of the portal and obtain a listing of all 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 research computing facilitators:

1. Frequently-asked questions whose answers evolve over time as technology advances. We partnered with the Campus Champions and research computing groups at large and small institutions to develop Ask.CI (https://ask.ci), a collaborative, crowd-sourced Q&A site specifically curated for the research computing community. Principal goals for the site are to: 1) reduce RCF workload at institutions of all sizes by pooling questions and answers on an open, searchable, archived site, and 2) make Q&A content available to smaller institutions that do not have the resources to maintain their own internal repositories. We address the evolution of answers over time by including a voting mechanism that allows users to indicate the "best" answer to a question, which might change as new information emerges.

2. Relatively static information such as introductory training modules on Linux clusters, programming languages and schedulers. We are developing a resource repository designed to help facilitators come up to speed on particular topics when needed by providing pointers to publicly available, relevant, and vetted training resources. The modules that we are collecting are self-paced, and clearly defined, requiring varying levels of expertise.

3. Dynamic, situation specific information needed to solve an immediate problem, typically handled by a Help Desk at larger institutions. We are piloting a Regional Help Desk that is accessible via the portal. Any user in the region can submit a ticket that is then handled by Northeast Cyberteam participants.

2.3 Regional Focus

National scale initiatives are an important starting point, but cannot efficiently reach thousands of smaller institutions. On the other hand, expecting every small and mid-sized institution to develop advanced computing capacity on its own invites unsustainable cost and duplication of effort. The Northeast Cyberteam strategy is based on the premise that larger institutions

[1] [1] Gregory E. Monaco, Gwendolyn Huntoon, David Swanson, Donald F. McMullen, Henry Neeman, Jennifer Leasure, Joni Blake, Kate Adams. The Role of Regional Organizations in Improving Access to the National Computational Infrastructure. National Science Foundation, June 2016.

with robust advanced computing resources and experienced facilitators can anchor *regional* efforts to increase the use of cyberinfrastructure and advance science throughout the area.

2.4 Oversight

Program direction is set by a Steering Committee that includes leaders from each of the larger institutions that serve as "anchors" for the Northeast Cyberteam Program, in this case, University of Maine, University of New Hampshire, University of Vermont, and MGHPCC. The steering committee also includes a program manager who coordinates day to day activity, and key personnel from other institutions that have provided students and mentors. The Steering Committee as a whole 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 lead to results that might not otherwise have been achieved, and blaze trails that others can follow.

2.5 Program Management Portal

The program relies heavily on the Northeast Cyberteam Portal for management of project workflows, recruitment of mentors and student facilitators, and recording results. The management section of the portal also encapsulates the experience that we are gaining, with the goal of making it possible to replicate the methodology in other regions.

The process for managing a project through its life cycle follows a standard set of steps, all of which are managed via the portal.

1) A Steering Committee member introduces the project, usually planned to be 3-6 months in duration, for approval.

2) If approved, the project is posted on the portal and Steering Committee members collaborate to recruit a mentor and a student RCF. The student and mentor both register on the Portal and become members of the Northeast Cyberteam. Individuals can also register on the portal in advance of a project assignment and become part of the Cyberteam pool that are considered first when new projects are recruiting.

3) The student RCF executes the project with support from the mentor, reporting on progress at monthly Cyberteam videoconference meetings.

4) At the end of the project, the Cyberteam Program Leader conducts exit interviews and the Steering Committee reviews lessons learned.

3. RESULTS/LESSONS LEARNED

We have launched 28 projects over the past two years, most of them lasting 3-6 months, and many of them supporting generation of publishable results. We are also beginning to see impact beyond 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.

1. Value of Research Computing Facilitators to research and education at small and mid-sized institutions: Consistent with the findings of the report that inspired the Northeast Cyberteam Program[1], the number of research projects that can benefit from Research Computing Facilitators is limited only by our ability to find and recruit them, which is improving over time.

Based on feedback from exit interviews, we are starting to think more systematically about how to assess project readiness. We have seen a spectrum of readiness levels - at one end there are faculty who have a clear idea about what they need to get to a new level of sophistication, while at the other end there are faculty who need help but are unable to engage productively. Over time, we expect to develop an explicit set of readiness criteria, and will gain more experience on how to respond when a project is not yet ready.

2. 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 (1) moving from a workstation to a cluster for greater throughput; (2) improving the performance or throughput of a workflow in order to generate results with faster turnaround or in greater volume; and (3) adopting a new computing tool such as Jupyter notebooks.

3. Willingness of mentors to participate: Experience over the past two years has validated our hypothesis that experienced Research Computing Facilitators would be willing to serve as mentors as part of their regular jobs. The opportunity to evaluate potential new hires is a practical motivator, but it also helps that people who become RCFs generally enjoy teaching others, and that teaching is central to the culture of academic institutions.

4. Ability to apply students and mentors across institutional boundaries: This aspect of the program has been critical to success. We are pleased that two initial concerns have not been significant impediments. Our first concern was distance – while occasional face-to-face meetings are possible (and necessary), most work must be done remotely, even if the student is separated from a project by just a few miles. We have found that tools for collaboration, such as high quality desktop videoconferencing, shared document repositories, and flexible source control systems, are sufficient to maintain communication and trust when combined with face-to-face contact. 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 every co-PI has needed to spend some extra effort explaining the purpose and benefits of the program, this has not delayed or prevented cross-institution assignments.

5. Willingness of larger institutions to share information: The Ask.CI project has received considerable support from Research Computing groups at larger institutions, both for the initial idea of building a shared Q&A list, and the more recent idea of "sandboxes" that expose internal Q&A lists outside their home institutions. In a similar vein, the regional help desk and the

training information repository have benefitted from contributions by research computing groups at larger institutions.

6. Importance of active program management: The second largest expense category for the project (after student support) 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: (1) efficient recruiting of projects, students and mentors; (2) development of process, tools, and strategy; (3) effective communication across the anchor institutions; and (4) the ability to explain the purpose and benefits of the program to grant administrators who have expressed initial skepticism about supporting this kind of collaboration across institutions. We have gained some recruiting momentum, and 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.

4. REPRODUCIBILITY

The Northeast Cyberteam Program has been underway for just over two years. It took some time for our steering committee to get into a regular rhythm of meeting times, project submissions and approvals, but we now have a reasonably well-established system that is delivering on the goals of moving science forward while giving potential student facilitators real world experiential training in the field of research computing.

All of the tools that we have developed, including the Portal, Ask.CI Q&A site, Regional Help Desk, and Training Resources Wiki, have been designed with an eye towards reproducibility/expansion. Even the logo was designed to be easily adapted to other geographic regions or domains.

5. Northeast Cyberteam and SEHET

Our goal in participating in the SEHET19 workshop is to find opportunities to collaborate with other groups focused on workforce development for the Research Computing community. Collaboration can take many forms, beginning with small steps such as posting a topic on Ask.CI or adding links to our Training Resources Wiki. A more ambitious collaboration would involve launching cyberteams in other areas of the country, anchored by a large institution (or group of institutions) where advanced research computing is a priority, and outreach to the surrounding institutions is encouraged. Leveraging the Northeast Cyberteam model and tools will allow researchers at surrounding smaller institutions to take advantage of cyberinfrastructure when their work requires it. Simultaneously, it will expose a new generation of potential facilitators to this exciting and dynamic field earlier in their careers, significantly expanding the available pool of candidates.

6. ACKNOWLEDGEMENTS