

Northeast Cyberteam: Workforce Development for Research Computing at Small and Mid-sized Institutions

John Goodhue

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

Julie Ma

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

Adrian Del Maestro

University of Vermont, Burlington,
Vermont, USA
Adrian.delmaestro@uvm.edu

Sia Najafi

Worcester Polytechnic Institute,
Worcester, Massachusetts, USA
snajafi@wpi.edu

Bruce Segee

University of Maine, Orono, Maine,
USA
segee@maine.edu

Scott Valcourt

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

Ralph Zottola

University of Alabama, Birmingham,
Alabama, USA
rzottola@uab.edu

ABSTRACT

Computing has become an essential component of research and education for nearly every scientific discipline. Meeting the need for support staff who can help faculty make the best use of available computing resources is a significant challenge for small and mid-sized institutions. The NSF-sponsored Northeast Cyberteam is addressing this challenge by building a pool of research computing facilitators that can be shared across institutional boundaries while also developing self-service tools that reduce the support burden.

CCS CONCEPTS

• **Social and professional topics** → Professional topics; Computing profession; Computing occupations; Computing education; • **Information systems** → Information systems applications; Collaborative and social computing systems and tools; World Wide Web; Web applications; Crowdsourcing.

KEYWORDS

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

ACM Reference Format:

John Goodhue, Julie Ma, Adrian Del Maestro, Sia Najafi, Bruce Segee, Scott Valcourt, and Ralph Zottola. 2020. Northeast Cyberteam: Workforce Development for Research Computing at Small and Mid-sized Institutions. In *Practice and Experience in Advanced Research Computing (PEARC '20)*,

July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 5 pages.
<https://doi.org/10.1145/3311790.3396662>

1 INTRODUCTION

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 (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. In our experience, this is currently the most significant barrier to productive use of research computing at small and mid-sized institutions.

To help meet the need, and to leverage professional RCFs 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 with a professional mentor to address an immediate faculty need. Projects draw mentors and students from across the region. We

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PEARC '20, July 26–30, 2020, Portland, OR, USA

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6689-2/20/07...\$15.00

<https://doi.org/10.1145/3311790.3396662>

have launched 35 projects at 17 institutions to date, pairing a diverse population of student RCFs with knowledgeable mentors to assist researchers and educators in the region.

The core of our strategy is to build a pool of RCFs in the region and a process to share them across institutional boundaries. Concurrently, we are providing experiential learning opportunities for students interested in becoming RCFs and developing a self-service learning toolkit to provide timely access to information when it is needed. For efficiency and to provide a single point of entry, we have developed the *Cyberteam Portal* for management of project workflows. The portal also facilitates replicating these ideas in other regions.

We have maintained a regional focus for two principal reasons. First, the region has several unique characteristics: it combines three EPSCoR states with one of the most research-intensive cities in the country; it contains academic institutions that range from some of the highest ranked universities in the world to under-served public and private colleges in both rural and urban settings; and it spans four states occupying an area that is smaller than many single states in other parts of the country. This diversity of institutions in close proximity created a microcosm we have leveraged as a proving ground for the program. Second, distances are short enough to allow for occasional face to face meetings for strategic planning and cross institution resource sharing, which were critical, particularly during program start up. Initially we thought that face to face communications would also be necessary for optimal mentoring and team building, and have been pleasantly surprised to find that now ubiquitous high-quality desktop videoconferencing has been a very effective substitute.

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 MGHPCC; a program manager who coordinates day to day activity; and key personnel from a few other institutions that provide a source for 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 BUILDING A REGIONAL POOL OF RESEARCH COMPUTING FACILITATORS

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 with robust advanced computing resources and experienced facilitators can anchor regional efforts to increase the use of cyberinfrastructure and advance science throughout the area. 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, often recruited from the anchor institutions, provide subject matter expertise, and guide the project in a direction that will yield results over a 3- to 6-month period. This gives the student the 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 RCFs). 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.

3 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 Cyberteam Portal is used to access 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. Using a common tagging infrastructure and voting capabilities modeled after crowd-sourced repositories such as Stack Exchange, 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 RCFs:

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, crowdsourced Q&A site specifically curated for the research computing community. While this started as a resource for Northeast Cyberteam members, it quickly became clear that making the platform available to the broader community would be mutually beneficial, and we launched the site to the

public at PEARC18. As researchers, facilitators, staff, students, and other users ask and answer questions on Ask.CI, they are creating a shared knowledge base and alleviating the burden of a central resource to provide support to a user base. For smaller institutions, this provides a wealth of knowledge that was not previously available in an easily searchable, archived Q&A format. For larger institutions, this self-service model frees up time for facilitators and cyberinfrastructure engineers to focus on more advanced subject matter, thereby elevating the practice. We address the evolution of answers by including a voting mechanism that allows users to indicate the “best” answer to a question, which might change as new information emerges over time.

Since launch, Ask.CI has attracted over 250,000 page views (currently averaging nearly 5000 per week), over 400 contributors, hundreds of topics, and a broad audience that spans the US and parts of Europe and Asia. While Ask.CI has shown steady growth in both contributions and audience, it is still in its early stages. Finding ways to continue to grow audience participation through innovation and outreach is an ongoing focus of attention.

2) Relatively static information such as introductory training modules on linux clusters, programming languages and schedulers

In our second year, we introduced a learning resource repository designed to help facilitators and researchers come up to speed on particular topics when needed, by providing pointers to publicly available, relevant, and vetted training resources. The modules are self-paced and clearly described, requiring varying levels of expertise. Student facilitators post modules which were beneficial in coming up to speed for their projects, along with the context in which they were used, as a form of documentation and vetting. As the resource repository grows, the common tagging infrastructure makes it easy to find modules relevant modules when a new project in a similar domain or with common characteristics is initiated.

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, accessible from the Cyberteam Portal to anyone in the community. Any individual with an account on the portal can submit a ticket that is then handled by Northeast Cyberteam participants and the ticketing system keeps track of responses, timeliness and participation.

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

4 RESULTS/LESSONS LEARNED

We have launched 35 projects over the past two 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. Our goal for participation in that workshop was to find opportunities to replicate part or all of the Northeast Cyberteam Program in other regions as part of our strategy for sustaining the program beyond the end of our NSF funding. The outcome of this effort was quite fruitful, as described below.

4.1 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. 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.

4.2 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 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.

4.3 Willingness of mentors to participate

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

4.4 Ability to apply students and mentors across institutional boundaries

This aspect of the program was critical to success, and we are pleased that the two most likely obstacles have not been significant impediments. Our first concern was distance - while occasional face-to-face meetings are possible, most work must be done remotely, even if the student is separated from a project by just a few

miles. Tools for collaboration such as high-quality desktop video-conferencing, shared document repositories, and flexible source control systems have made it possible to maintain communication and trust. 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 delayed or prevented cross-institution assignments.

4.5 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 “locales” that expose internal Q&A lists outside their home institutions. In a similar vein, the regional help desk and the training resources repository have benefitted from contributions by research computing groups at larger institutions.

4.6 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 a) efficient recruiting of projects, students and mentors; b) development of process, tools, and strategy; c) effective communication across the anchor institutions and d) 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, both at the anchor institutions and for overall program management will be required for ongoing success.

5 REPRODUCIBILITY

The Northeast Cyberteam Program has been underway for two and a half 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 Cyberteam Portal, the Ask.CI Q&A site, the Regional Help Desk, and the Training Resources Repository have been designed with an eye towards reproducibility/expansion. All have been built using open source platforms and we encourage use of our Github repository, both for platform work and for tools/processes developed for individual projects. Even the logo was designed to be easily adapted to other geographic regions or domains. While there is still some work to be done, we plan to make these tools available in reusable form

via Github in the future. We have leveraged these elements in key elements of our sustainability strategy, described below.

6 SUSTAINABILITY

The Steering Committee has spent considerable effort exploring opportunities to sustain the Northeast Cyberteam after grant funding has been depleted, as the value of this program has been clearly demonstrated for our individual organizations as well as surrounding small and mid-sized institutions. Our exploration has been in four focus areas, described below.

6.1 Expansion beyond the Northeast

As noted above, the tools and methods developed for the Northeast Cyberteam Program were designed with reproducibility in mind. With a larger community, it should be possible to spread the cost of tools and supporting material over a wider base. Toward that end we have engaged on several fronts:

- 1) We are currently embarking on a pilot effort with several other Cyberteam grant recipients interested in leveraging the portal to support their project goals while enriching common elements such as Ask.CI and the Training Resources Repository. These include the Great Plains Cyberteam, Rocky Mountain Advanced Computing Consortium Cyberteam, and the Kentucky Research Computing Cyberteam.
- 2) We are in discussions with the Texas-based SWEETER team on possible collaboration.
- 3) Steering Committee member Ralph Zottola moved from the University of Massachusetts system to the University of Alabama at Birmingham during the project and is building a coalition of schools to implement some of the methods described here.
- 4) We have partnered on one grant application that if funded, would both replicate and add some new dimensions to our methodology in another region, and we are actively seeking opportunities to partner on other grants.

6.2 Memberships

We have developed and started to socialize a model where small and medium-sized institutions pay a modest annual membership fee which is applied toward the cost of administering the program. Membership provides three benefits:

- 1) Researchers at the participating institution can submit cyberteam projects which are then staffed by students and mentors recruited by the Cyberteam Steering Committee.
- 2) Students at member institutions may participate as student facilitators and the member institution receives marketing support to promote cyberteam opportunities to their students.
- 3) Researchers and educators at the member institution receive preferred access to HPC resources available at participating anchor institutions.

Each of these benefits is attractive to participating institutions as a way to offer additional resources/opportunities to their constituents – students, educators and researchers. Several institutions

that participated in projects funded by the Northeast Cyberteam have expressed interest in participating in a pilot.

6.3 Internship and education funding for student participants

We are exploring avenues for supporting student research computing facilitator assignments via internship programs, research experience programs and other sources of funding for experiential learning.

6.4 Industry partnerships

The shortage in Research Computing Facilitation resources extends beyond academia to industry, particularly at small and medium sized enterprises which, like their academic counterparts, may not be well-positioned to support critical mass of research computing facilitators needed to support the growing need to use high performance computing. We have started a trial partnership with a Massachusetts company, leveraging a state-sponsored internship program that would support a student over the summer, with the goal of expanding to other states after the trial has been successfully completed.

7 NEXT STEPS

The Northeast Cyberteam 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, significantly expanding the available pool of candidates. Our sustainability explorations are beginning to yield opportunities to collaborate with other groups focused on workforce development for the research computing community and we continue to seek additional partners. Collaboration can take many forms, ranging from posting a topic on Ask.CI to launching a cyberteam in other areas of the country. We welcome any and all inquiries to explore opportunities to leverage the process and tools we have built to make research computing more widely available to researchers and educators while building the workforce for the next generation.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. OAC-1659377, “CC* Cyber Team: Improving Access to Regional and National Cyberinfrastructure for Small and Mid-Sized Institutions.

We thank the many Cyberteam researchers, mentors, and student facilitators for their participation in this effort; the Ask.CI site and locale moderators, whose tireless efforts to build and maintain the site are beginning to yield significant results; and the hundreds of contributors who have generously shared knowledge and experience on Ask.CI.

REFERENCES

- [1] J. Goodhue, J. Ma, A. Del Maestro, B. Segee, S. Valcourt, S. Najafi, R. Zottola. 2020. Northeast Cyberteam Program - a Workforce Development Strategy for Research Computing. *Journal of Computational Science Education* (July 2020) DOI: <https://doi.org/10.22369/issn.2153-4136/11/1/2>