Impact of Blue Waters Education and Training

William Kramer
University of Illinois
Urbana, IL
wtkramer@illinois.edu

Scott Lathrop
University of Illinois
Urbana, IL
lathrop@illinois.edu

Steven I. Gordon
The Ohio State University
Columbus, Ohio
sgordon20@gmail.com

Robert M. Panoff
Shodor Education Foundation, Inc.
Durham, NC
rpanoff@shodor.org

Aaron Weeden
Shodor Education Foundation, Inc.
Durham, NC
aweeden@shodor.org

Lizanne DeStefano
Georgia Institute of Technology
Atlanta, GA
ldestefano6@gatech.edu

ABSTRACT
The Blue Waters proposal to NSF, entitled "Leadership-Class Scientific and Engineering Computing: Breaking Through the Limits," identified education and training as essential components for the computational and data analysis research and education communities. The Blue Waters project began in 2007, the petascale computing system began operations on March 28, 2013, and the system served the community longer than originally planned as it was decommissioned in January 2022. This paper contributes to the Blue Waters project's commitment to document the lessons learned and longitudinal impact of its activities.

The Blue Waters project pursued a broad range of workforce development activities to recruit, engage, and support a diverse mix of students, educators, researchers, and developers across the U.S. The focus was on preparing the current and future workforce to contribute to advancing scholarship and discovery using computational and data analytics resources and services. Formative and summative evaluations were conducted to improve the activities and track the impact.

Many of the lessons learned have been implemented by the National Center for Supercomputing Applications (NCSA) and the New Frontiers Initiative (NFI) at the University of Illinois, and by other organizations. We are committed to sharing our experiences with other organizations that are working to reproduce, scale up, and/or sustain activities to prepare the computational and data analysis workforce.

Keywords
HPC, Computational science and engineering, Data analytics, Education, Training, Higher education, STEM workforce

1. INTRODUCTION
The Blue Waters proposal to NSF, entitled "Leadership-Class Scientific and Engineering Computing: Breaking Through the Limits," [1] identified education, training, and community engagement as essential components for the computational research and data analysis communities. Blue Waters emphasized working with organizations and individuals where they live and work and proactively recruited, engaged, and retained a diverse community of partners (i.e., similar to what others often refer to as "users") in utilizing the full range of Blue Waters resources and services as well as the expertise of the University of Illinois and our collaborators.

We present information on how training, education, student research experiences, and community engagement activities were organized and how these activities evolved based on external evaluations, community feedback, and attention to the needs and requirements of the target audiences. We also describe the partnerships, cyberinfrastructure, and evaluations that enhanced, supported, and improved the activities. The information presented here complements and extends previous reports we have given [9, 7].

Throughout the University of Illinois's 35+ years of providing national-scale HPC resources and services, there has been a persistent turnover of people using the resources and services (over 50% of whom are graduate students), increased utilization of HPC systems among disciplines that have not traditionally used them (e.g., humanities, arts, and social sciences), and the continued rapid evolution of hardware and software available to accelerate discovery. The critical need for HPC expertise in academia, industry, and government is growing faster than the workforce is able to sustain.

There is a persistent need to develop and deliver new training content while continuing to keep existing materials current. Higher education institutions need to identify strategies to maintain state-of-the-art courses to improve the preparation of students to contribute to and advance computational and data analytics research and scholarship. The education, training, student research experiences, and community engagement activities described herein provide foundational elements that have good potential to be scaled up and sustained.

2. EDUCATION
The education programs were developed to prepare the next generation of computational researchers, educators, and

DOI: https://doi.org/10.22369/issn.2153-4136/13/2/5

[1] "Leadership-Class Scientific and Engineering Computing: Breaking Through the Limits" supported by the National Science Foundation awards OCI-0725070 and ACI-1238993, the State of Illinois, and as of December, 2019, the National Geospatial-Intelligence Agency.

practitioners attending higher education institutions. Blue Waters provided allocations on the Blue Waters system for programs across the U.S. needing access to petascale class resources.

The majority of the activities focused on the creation and/or delivery of formal credit courses, curricular materials, and exercises. The materials that were developed continue to be used in hands-on workshops for faculty and students to demonstrate how computational thinking and computational science can be incorporated into the undergraduate curriculum and enhance STEM education. The materials are disseminated through multiple repositories such as HPC University [19] and the Computational Science Education Reference Desk [18]. The ability for instructors to adopt and/or adapt well tested and vetted materials provides a good foundation upon which to learn and build. The community provides feedback to improve the materials, and they in turn contribute materials tailored to their own teaching and learning needs for dissemination through the repositories.

2.1 Curriculum Development

Blue Waters launched its education initiative even before the Blue Waters computing system was in operation to help faculty begin to think about how to prepare their students for working in a highly parallel high-performance computing environment. The Shodor Education Foundation, Inc., led the effort to engage undergraduate faculty in the development of 30 undergraduate course modules for use in undergraduate classrooms. The materials were designed to prepare students for utilizing a petascale class system to run interactive models and simulations to learn STEM principles. Once the Blue Waters system was deployed, these modules were updated by the Shodor team and collaborating faculty to incorporate the petascale capabilities of the system.

The Shodor team utilized these materials during curriculum development workshops that were conducted from 2009–2011 with over 700 undergraduate faculty across the U.S. By using computational models that scale to hundreds (if not thousands) of processors, the team was able to demonstrate how faculty can teach the complexities of developing parallel codes that cannot be realized using systems with four to eight processors. The workshops supported faculty in adopting and/or adapting these materials into their own courses. In return, the faculty provided feedback for improving the materials, and many contributed their materials to the repositories. Shodor also conducted workshops that were tailored to high school and middle school teachers incorporating computational thinking into their classrooms using many of these same materials.

During 2020, the Shodor team led a group of 22 faculty to develop additional shorter modules in applied HPC modeling and simulation by updating and adapting the previous 30 modules as well as materials previously developed for the Blue Waters Intern Petascale Institute described below in the Student Research Experiences section. Approximately 60 small modules were created. The curricular materials are available [17].

2.2 Virtual School of Computational Science and Engineering

The Blue Waters team prototyped the delivery of shared graduate-level credit courses among a diverse mix of institutions. The project focused on delivering HPC-related topics that were not able to be provided at collaborating academic institutions. Six graduate credit courses for the Virtual School of Computational Science and Engineering were conducted from 2013 to 2016. These courses were modeled upon the distributed courses offered by the Big Ten Academic Alliance Online Course Sharing Program. Computational science experts taught HPC topics to graduate students at 31 collaborating academic institutions. Three of the institutions were Minority Serving Institutions, and 12 were in EPSCoR [14] jurisdictions. Faculty at the collaborating institutions were the official class instructors, allowing a total of 633 students to register and earn graduate credit at their own institutions. This approach made it possible for the high-end computational science course topics to be offered at colleges and universities that would otherwise not have been able to offer them.

Flipped classrooms were employed for a combination of synchronous and asynchronous learning. Flipped sessions provided students with materials to study on their own followed by live (in-person and virtual) discussion groups with their local instructors. A number of the faculty indicated that they planned to continue to use the flipped model even when not conducting shared virtual classes.

The sharing of courses among institutions brought changes to the participating institutions. The faculty became more informed about the subject matter and gained confidence subsequently to teach the subject matter on their own. The students earned credit from their own institutions even though the content was developed and delivered by faculty at other institutions.

These efforts demonstrated the potential for scaling up access to advanced topics and the potential to prepare more faculty to teach these advanced topics on their own. To be sustainable, the benefits need to be recognized and supported by multiple faculty, and the courses must become institutionalized by the department chairs with support from their deans. In addition, inter-institutional agreements are needed for cost-sharing among participating institutions. Other organizations continue to explore the scalability and sustainability of these types of initiatives. A report of lessons learned delivering online graduate credit courses is available [5].

2.3 Education Allocations

The Blue Waters education allocations provided access to the computing system, enabling U.S.-based organizations to engage people in a variety of computational science and data enabled learning opportunities. The community proposed innovative teaching and learning activities that would benefit from access to petascale computing resources. The types of sessions that were supported included workshops, institutes, hackathons, undergraduate and graduate courses, research experiences for undergraduates, internships, and graduate fellowships. Formative feedback led to streamlining the process for submitting requests and receiving timely decisions within two weeks. Applications could request up to 25,000 node-hours (a measure of the number of computing elements used times the amount of time used), although a few exceptions were made for large credit courses with over 100 enrolled students.

Education allocations supported a broad range of learning activities, and we encouraged innovative approaches to support learning, with 188 education allocations supporting over 5,000 participants at 160 institutions, including 41 institutions in EPSCoR jurisdictions and 14 Minority Serving Institutions. The education allocations caused minimal interference to the research allocations running alongside them on the computing system. The graduate fellowships were the only categories that requested and received increased allocations based on documented progress reports and ongoing computational research plans.

3. TRAINING

The audience for training activities included undergraduate and graduate students, faculty, researchers, and staff from among academic institutions, industry, business, government agencies, decision makers, and other sectors of society. Graduate students represented the largest fraction (over 50%) of participants. Anecdotally, we learned from high school teachers that some of the training materials were used in their classes. Training activities focused on developing and delivering materials to enable participants to learn about cyberinfrastructure tools, methods, and resources via informal learning events to empower participants to conduct research and scholarship at scale.

In-person training sessions were offered at conferences, institutes, hackathons, workshops, the two-week Blue Waters Intern Petascale Institute, the International HPC Summer School events, the annual Blue Waters Symposium, on multiple campuses, and at other events that brought people together in one location.

While in-person training events are considered to be very effective, the majority of the Blue Waters partners/users were spread across the country, and most of them were unable to travel to in-person training events. The most time- and cost-effective method for serving the community was by providing the majority of the training virtually. Face-to-face virtual sessions can be conducted effectively using high-definition video conferencing infrastructure. The ability for the presenters and participants to see each other and interact in real-time, even though remote from each other, enriches the teaching and learning environment. High-definition video conferencing, interactive communications tools, and expert online advisors support learning at scale at reasonable costs.

While many people asked that sessions be recorded, our experience was that a very small fraction (less than 5%) of participants accessed these recordings. Our experience with the uptake of self-paced training tutorials and webinars demonstrates that it is valuable to the community when the recordings are split into smaller modules, which does incur post-production labor costs.

We evaluated the training events to capture feedback and recommendations from participants for improving the delivery and identifying additional topics to offer. However, there is an ongoing challenge to identify quantitative and qualitative methods for assessing the long-term benefits and a return on investment (ROI) for providing training.

3.1 Virtual School of Computational Science and Engineering

Virtual School of Computational Science and Engineering (VSCSE) summer schools were conducted from 2008 through 2016 in cooperation with the University of Michigan. The infrastructure team enhanced the video conferencing quality to utilize high-definition video conferencing facilities to broadcast the sessions among multiple sites simultaneously. The program delivered training content to an initial set of four locations in the first year and expanded over time with experience and improved infrastructure to deliver content to 24 locations simultaneously. Speakers presented from various sites, with participants able to conduct face-to-face Q&A sessions from their home institutions. The video conferencing training rooms were updated over time with additional cameras and screens to enhance the environment for instructors and participants. We gained confidence in the ability to scale up support to accommodate the management and coordination of more than 20 sites simultaneously.

This approach facilitated access by participants at 146 different academic institutions, including 10 Minority Serving Institutions and 63 institutions in EPSCoR jurisdictions, reaching a total of 5,865 people. Three papers about the VSCSE including lessons learned are available [4, 3, 29].

3.2 Petascale Computing Institutes

Building on the VSCSE experiences, two week-long Petascale Computing Institutes were conducted in 2017 and 2019 and engaged over 600 registered participants at 21 sites, including two international sites, all simultaneously. Individuals were able to participate from their home/office/lab to view the broadcasts and pose questions, although these individuals did not receive accounts on the institute HPC systems since we could not validate their identities. The institutes were possible due to collaborations with content providers from multiple HPC centers and the cooperation of hosts at multiple campuses. These events demonstrated that it is practical and cost-effective to engage a large and diverse community of learners.

Improvements were made to facilitate additional online Q&A sessions, allow individuals to participate from their own offices, spend more time testing exercises in advance, and ensure that participants could use the computing systems before hands-on activities began. More time was also spent to test A/V connections and to allow time for presenters to become familiar with the physical setup in advance of their presentations. The advance testing helped to reduce problems during the presentations and hands-on sessions.

3.3 International HPC Summer School

Blue Waters contributed support to the in-person International HPC Summer School for multiple years by providing access to an education allocation on the Blue Waters system as well as providing technical staff to teach sessions and serve as mentors to the graduate students participating in the program. Each summer school supported 80 graduate students from the U.S., Canada, Europe, and Japan.

Over time, the Summer School provided more depth of technical training in response to the participant feedback and evaluations, brought back students from previous summer schools as peer mentors, provided more in-depth constructive advice on being good mentors, and increased the time for mentoring. Backup technology plans were added to address challenges experienced in previous years.

3.4 Hackathons

Blue Waters hosted hackathons to focus on helping selected research teams of two-to-six members with improving the performance of their codes on the Blue Waters system. An emphasis was placed on helping the participants take advantage of the GPU capabilities of the system. Mentoring assistance was provided by NVIDIA staff and academic professionals and researchers.

It is important to identify mentors with both technical and scientific knowledge appropriate for each team. It is also important that the mentors are able to spend a considerable amount of time with their assigned team. A number of teams benefitted from spending time optimizing their CPU codes and realizing considerable speedup before considering incorporating GPU capabilities.

3.5 Webinars

The Blue Waters team organized a series of nearly 200 webinars, most of which lasted about an hour, although a few sessions were two or three hours in length. All sessions were recorded for access by people who were not able to attend at the scheduled time. As the number of webinars grew, they were categorized to assist people in finding recordings of the sessions of interest. The webinar content spanned technology, research, and workforce development topics. YouTube was used to maximize ease of access to live and recorded sessions and for providing automatic transcriptions.

There were more than 45,000 views of the webinars, including a large number of views of the recordings. This made the time and effort to record the sessions worthwhile. We were advised to allow at least three weeks of advance notice to allow time for people and organizations to promote the webinars. Based on the positive feedback, the webinar series will continue past the Blue Waters funded period with support and leadership from the New Frontiers Initiative.

3.6 Self-Paced Training

For many people, just-in-time training is essential to their learning needs. We have recommended Cyberinfrastructure Tutor [25] and the Cornell Virtual Workshop [1] as two popular sources for self-paced HPC learning resources. Reports from these providers indicate that these tutorials are very heavily used. Evaluations indicate that some viewers are looking for answers to specific questions rather than going through the full tutorial, which makes short, indexed modules quite useful to the community.

3.7 Training Event Planning

The following are lessons we have learned to help make training sessions more effective.

3.7.1 Audience

There is a persistent challenge that no matter how well one describes the prerequisites, content, and learning goals, one will still likely have an audience with diverse background and knowledge. For a session with a large audience, one needs to decide strategically how slow/fast one plans to proceed and how to address an audience with a broad base of novice and advanced skills. Regardless of one’s choices, some members may become disengaged, so it is important to make content and target audience clear from the outset. For a session with a small or targeted audience, it can be useful to conduct a short survey of the registered attendees to gain a better sense of their knowledge so that you can better tailor the content based on the audience’s knowledge and topics they want to learn.

Everyone should be made to feel welcome to encourage asking questions they may feel are too “simple.” Attendees are more comfortable asking questions if they know their background and experience are comparable to those of other participants. Initial ice breakers will make them more comfortable asking questions and talking with fellow participants.

The ability for participants to post questions and receive quick responses throughout each session enhances the learning experience. While this can be accomplished through direct interaction with the instructor, Q&A interactions with content experts through collaborative tools (e.g., Slack) can be equally useful and less disruptive to the instructional flow. These collaborative tools also help to reduce shyness and encourage people to ask questions they may be uncomfortable verbalizing.

3.7.2 Content

All sessions should be clear about the learning goals, content to be covered, and prerequisites. Sessions that encourage participants to bring their own codes/applications to use during laboratory sessions will enhance their ability to practice what they learn with their own code and data during and after the sessions.

No matter how much content you may want participants to learn, it is not possible to cover everything you feel that they “may” need to know. Do not pack sessions tight; leave time for people to absorb the content, practice with exercises, talk with the instructors and each other, and have ample time for breaks. Providing sufficient food helps people relax. For multi-day events, plan to end early on Friday so people can go home and relax; the audience will likely thin out early regardless of the schedule.

3.7.3 Instruction

Good teaching practices can be more effective than having the most technically qualified presenter. The level of content (i.e., introductory, intermediate, or advanced) should be made clear to the instructors in addition to the participants. We strongly encourage instructors to practice what they teach. When teaching students problem solving approaches using interactive tools, the instructors should practice with the same tools and approaches.

3.7.4 Exercises

A process for assisting participants with running exercises should be established in advance. This should include ensuring there is an adequate level of technical support relative to the number of participants. For remote participants, plan for effective mechanisms for “looking over the shoulder” of participants to help diagnose any problems they encounter. Exercises allow the participants to ascertain they are learning the concepts. It is essential that exercises are well tested in advance to work on the platforms that are being utilized by the participants.

Some participants appreciate the ability to earn a “recognized” badge of knowledge gained that they may add to their resume or biography. Some participants appreciate the ability to combine badges and/or earn a certificate of knowledge gained that is verified through some type of exam or quiz (as opposed to a certificate of participation that some people need for their jobs).

3.8 Impact

We can conservatively report that Blue Waters engaged over 45,000 participants from 224 academic institutions in 49 states, Puerto Rico, and 22 countries. There were participants from 20 Minority Serving Institutions and 68 institutions in EPSCoR jurisdictions. There were over 250 education and training events, 188 education allocations, and over 90 course modules. Presentations were made at numerous national and international conferences, institutes, summer schools, and workshops each year for which no records of the number of participants were reported.

Many of the materials, slides, and video recordings from the training sessions described above are available [26] along with the slides and video recordings from the Blue Waters webinar series [27] and a more detailed report of the Blue Waters training experiences [11].

4. STUDENT RESEARCH EXPERIENCES

Student engagement programs were conducted to directly engage undergraduate and graduate students in computational science and data analytics research projects. The goal was to enhance their motivation to pursue advanced studies and careers to advance research and scholarship. The programs included year-long student internships for 139 undergraduate students as well as graduate fellowships for 55 PhD students from across the U.S.

4.1 Graduate Fellowships

The Blue Waters Graduate Fellowship program engaged 55 PhD students across the U.S. from 2013 through 2022. The fellowship program was modeled upon the NSF Graduate Research Fellowship Program but with an emphasis on supporting PhD candidates needing access to petascale class high performance computing resources to support their computational and data-enabled science and engineering research. The fellows received up to $50,000 in financial support, which included a stipend, tuition support, and travel to Blue Waters events. They also received an initial allocation of 50,000 node-hours on the Blue Waters petascale computing system.

4.1.1 Recruitment and Selection

The recruitment process began with a national awareness campaign starting in the fall of each year among people with allocations on Blue Waters and the Extreme Science and Engineering Discovery Environment (XSEDE) [24]; through social media and HPC media companies; through consortia and professional societies; and with assistance from individuals with mailing lists of faculty, researchers, professionals, and historically excluded groups. Each year, approximately 90–100 applications were each reviewed by three faculty and researchers with comparable scientific background. A technical review was also conducted to ensure that the proposed research was appropriate to be conducted on the Blue Waters system. A final review was used to ensure gender diversity, institutional diversity, and geographical diversity among candidates who had been evenly ranked.

4.1.2 Start-up

The fellows attended an inaugural meeting at NCSA in the early fall timeframe to provide an overview of their research and computational plans. We found that a two-day meeting allowed time for the fellows to consult with their points of contact (POCs) and other staff and researchers on campus. To make effective use of the petascale computational resources, the fellows benefitted from some start-up training. The fellows completed a short survey of their HPC experience that helped us to fine-tune the training to be most useful for them.

Each fellow was assigned a primary point of contact; this had already been proven to be an effective mode of support for research teams with allocations on the Blue Waters system. The fellows repeatedly reported that this support model made a significant positive impact on their ability to accelerate their computational research plans.

We encouraged the fellows to invite their advisors to attend the start-up meeting; however, few advisors attended. Recognizing that some of the fellows helped to impact the computational efforts of their advisors and fellow students, we would encourage making participation by advisors a stronger component of the program.

4.1.3 Research Experience

The fellows were able to spend the year focused on their research goals. The POCs engaged in regular (at least monthly) contact with each fellow to minimize barriers to progress that normally occur over time. Quarterly calls were arranged with all the fellows and the support staff. These were useful to share information, identify problems, and propose solutions, many of which tended to benefit multiple fellows. Each fellow submitted a quarterly report and a final report to document publications and presentations and pose questions to the staff, and the fellows were interviewed by the external evaluators.

The fellows were invited to two Blue Waters Symposia to present posters about their proposed research in the first year and to give presentations on their research findings near the end of their fellowships. Through these events, the fellows learned about common challenges from people across multiple disciplines. They indicated that their participation helped them improve their own research practices, presentations, and publication skills and improve their confidence.

While every fellow received an initial allocation of 50,000 node-hours, it became apparent that many of the fellows needed supplemental allocations of time. It was important to provide an easy process for the fellows to request supplemental allocations by providing an update on research progress to date, a justification for the quantity of resources needed, and plans for future research activities.

4.1.4 Post-fellowship

Most of the fellows continued with their computational research beyond the fellowship period while working to complete their PhD. We supported extended access to the Blue Waters system for these fellows and for fellows entering postdoc positions at accredited academic institutions. We did not extend this offer to fellows that went to national laboratories or industry positions, as we felt that those organizations could support any ongoing computational needs.

4.1.5 Impact

A total of 55 graduate students were selected through competitive national application processes. The fellows represented 38 academic institutions across 27 states. Eleven of the institutions are in EPSCoR jurisdictions and one institution is a Minority Serving Institution. Twenty-two of the fellows (40%) identified as female and/or a historically excluded race or ethnicity.

We maintained communication with 96% of the fellows through ongoing communications and surveys to learn of their career progress and to capture their advice on how the fellowship program could be improved. We conducted a final survey of all fellows in March 2022. We were pleased to learn that eighteen of the fellows reported being in a postdoc position, twelve in a professional position, and eight with a faculty appointment. Additional information about the program is available [10].

4.2 Student Internships

The Blue Waters Student Internship Program was launched in 2009 and continued through 2019 to provide undergraduate students across the U.S. with training, a year of financial support, access to leading-edge petascale HPC resources, and the opportunity to work with a mentor on a computational research project. The program was managed by the Shodor Education Foundation, Inc. We previously reported details of the program in 2014 [8].

4.2.1 Recruitment and Selection

We promoted the Student Internship Program and conducted reviews of the applications in a similar manner as the Graduate Fellowship Program. The program proactively recruited individuals from groups historically excluded from HPC. The formative evaluations helped the Shodor team to modify the selection process to introduce anonymous reviews.

Initially, traditional methods of selecting from among the applicants did not yield success engaging adequate diversity among the participants. We adjusted our methods by asking application reviewers to avoid only selecting the applicants who were the "best of the best" and already had significant experience and instead to consider which of the applicants had the greatest potential for learning and would benefit the most from participating but who had not yet had opportunities to gain experience. Our working motto in the selection process was to "promote excellence rather than just rewarding excellence." This change had the profound impact of increasing the percentage of selected students from groups historically excluded from HPC from 13% in the first two years of the program to 43% over the course of the whole program. Additional information about our early efforts is available [6].

4.2.2 Start-up

During the first year of the internship program, participants were selected for summer internships. We found the interns frequently had insufficient background or experience to contribute significantly to their research projects to which they were assigned. To address this, we extended the internships to a full year and brought all of the interns together at NCSA for a two-week Petascale Institute at the start of each internship. The institute focused on providing the students with a background in parallel programming, quantitative analysis, and computational science tools and methods.

Subsequent feedback from the mentors indicated the students were much better prepared to quickly engage with and contribute to their research projects. Extending the length of the internships to a year resulted in the students being able to accomplish much more progress than students involved in a traditional summer internship.

The materials developed for the Petascale Institutes are available [23].

4.2.3 Research Experiences

Undergraduate student engagement benefited from committed mentors. A limiting factor in being able to support enough students was having enough projects for the students to work on with mentors who cared and could devote time to supporting undergraduate students. Thus, it was important to focus recruiting efforts on reaching out to potential mentors. We encouraged mentors to recruit students from their own institutions in whom they saw potential for being able to contribute to their research. We also encouraged mentors to be open to working with students at other institutions, since many students had interest in working on computational projects for which there was no suitable faculty member at their own institution.

Every intern was strongly encouraged to submit manuscripts sharing their research findings and documenting their learning outcomes to the Journal of Computational Science Education (JOCSE) [21] or another journal in their specific discipline.

A few interns were selected each year to present posters on their research at the annual Blue Waters Symposium. Participation in the Symposia enhanced the interns' research experiences, allowed them to network with graduate students and researchers, and became a capstone event for their year-long effort. A few interns were also supported to attend the SC conference to make connections with mentors and collaborators. These events helped form bonds among the interns and build community.

4.2.4 Post-internship

A number of the interns were motivated by their experiences to return as peer mentors and instructors for subsequent Petascale Institutes. Having peer mentors was an advantage to building community and enhancing the experience for the participants.

4.2.5 Impact

While there were added costs for extending from a 10-week summer program to a full year program, the benefits of student preparedness and deeper research engagement and progress outweighed the costs. The students were also more likely to be supported by their research mentors past the internship period.

A total of 139 undergraduate students were selected for research internships through national application processes. The students represented 73 academic institutions across 32 states and Puerto Rico. Thirty-seven of the students were from institutions in EPSCoR jurisdictions and 26 of the students were from Minority Serving Institutions. Sixty-one of the interns (43%) identified as female and/or historically excluded races/ethnicities. Thirty-six of the interns published a peer-reviewed paper in the Journal of Computational Science Education (JOCSE), at least 16 others published in other journals, and at least 17 others presented posters at various meetings and conferences. A number of undergraduate and graduate students funded by XSEDE student engagement programs also participated in the two-week Petascale Institute.

Over the years, the formative evaluations allowed the Shodor team to improve the content and student experiences. The topics and materials used in the Institutes were updated each year based on formative and summative evaluation feedback from the external evaluators, comments from the students, and observations from the instructors [31]. Additional information about the internship program is available [13].

The lessons learned from the Blue Waters Student Internship Program were applied to the XSEDE EMPOWER program [30] funded by the XSEDE project and managed by Shodor. The program recognized that providing different levels of internships — from learners, to apprentices to interns — is an effective strategy to provide an on-ramp to computational science for students based on their background and experience.

4.2.6 Human-interest

One intern, Aaron Weeden (co-author of this paper), was in the first group of interns in 2009, assisted as a peer mentor and assistant instructor in the first Petascale Institute in 2010, was later a staff mentor for a project involving four interns in 2014, led the curriculum development and instruction of the last five Petascale Institutes, and coordinated the internship program in its final three years. This experience allowed Aaron to grow as a mentor and leader and have a fulfilling experience enabling many other students to have impactful internship experiences.

5. COMMUNITY ENGAGEMENT

Blue Waters was funded to serve the U.S. research and education community with access to petascale resources and services. A significant challenge was disseminating information about the resources and services to the people who would most benefit. Traditional users of previous HPC resources were generally well informed about the resources funded by NSF. Our challenge was to continue to recruit and support traditional audiences while working to raise awareness and engage organizations and individuals who were not well informed.

We placed an emphasis on engaging a diverse community. We refer to the term diversity to include: 1) historically excluded groups including women, historically excluded races and ethnicities, and people with disabilities; 2) individuals from all academic institutions including Minority Serving Institutions, institutions in EPSCoR jurisdictions, PhD granting institutions, primarily undergraduate institutions, two- and four-year institutions; and 3) researchers from among all disciplines.

To recruit, engage, and support a large and diverse community of participants, partners, and collaborators required an extensive and ongoing effort to build a diverse national list of contacts. The contacts included PIs and researchers who had allocations on the Blue Waters and XSEDE systems. Participants in Blue Waters events were added to our mailing list. Blue Waters staff visited many campuses to raise awareness among faculty, staff, administrators, and students.

Students told us that they most often learned about opportunities from their faculty and advisors; thus, we pursued multiple avenues to contact faculty, deans, and department chairs.

Considering the challenge of raising awareness nationally, we included intermediaries who in turn shared our message with individuals and groups they knew may be interested. Among our intermediaries were the Campus Champions, the Coalition for Academic Scientific Computation (CASC), the Campus Research Computing Consortium (CaRCC), media organizations (e.g., HPCwire), and professional societies (e.g., the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE)). Contacts were made with student-oriented organizations (e.g., Women in Engineering, ACM Student Chapters). To engage Minority Serving Institutions, we contacted technology leaders at the Hispanic Association of Colleges and Universities (HACU) that represents Hispanic Serving Institutions, National Association for Equal Opportunity in Higher Education (NAFEO) that represents Historically Black Colleges and Universities, and American Indian Higher Education Consortium (AIHEC) that represents Tribal Colleges and Universities. We used social media to promote the resources and services.

5.1 Building Community

The external evaluators gave special recognition and praise to the interdisciplinary activities of the Blue Waters project. The annual Symposia placed a strong emphasis on bringing together PIs, researchers, professionals, and students. Discussions with participants demonstrated that the participants realized the value of sharing common algorithms and techniques for addressing computational methods that were common across multiple disciplines. The ability of fellows and interns to display their own posters and make presentations on their research gave them greater confidence in their research activities.

The fellows consistently mentioned the Blue Waters effort to build community as very important. The fellows appreciated getting to know their fellow colleagues, the Blue Waters staff, and other researchers attending the Blue Waters Symposium and other conferences. They valued the ability to learn from one another and mentioned that some of these friendships lasted far beyond the timeframe of their fellowship. A common response from students was that more community building should be pursued. This includes greater use of social media and other communication platforms (e.g., Slack).

Activities like the Petascale Institute for the student interns included social time for the students to get to know one another and the support staff while also providing time to relax and unwind. Sharing meals helped to foster conversations and connections.

On a lighter note, many HPC-related songs have been written and sung by Bob Panoff [22]. A parody of a folk song, “The Water is Wide,” performed by Shodor intern Krista Katzenmeyer, is posted on YouTube [16].

5.2 Impact

We were successful in engaging over 45,000 participants from 225 academic institutions, 20 Minority Serving Institutions, 68 institutions in EPSCoR jurisdictions, 41 laboratories and centers, and 22 industries in 49 states, Puerto Rico, and 41 international organizations in 22 countries. Additional presentations were made at numerous national and international conferences, institutes, summer schools, and workshops every year, during which data on participation were not collected.

6. SUPPORT INFRASTRUCTURE

The intellectual efforts described above were enhanced by strategic partnerships, cyberinfrastructure, and evaluations.

6.1 Strategic Partnerships

Developing partnerships and collaborations with national and international organizations was a strategic goal for scaling up and sustaining Blue Waters programs and activities.

The broad range of learning needs and requirements of the HPC community, from introductory materials to advanced skills, exceeds the capabilities of any one institution, organization, or consortium to address. There continue to be new researchers every year — whether undergraduates, graduates, or researchers — who see the potential for HPC resources to accelerate their computational research. In addition, new hardware, tools, and applications continue to emerge for which training is needed to inform and prepare the community for making effective use of the resources.

6.1.1 Project Management Partners

Partnerships were developed with the Shodor Education Foundation, Inc., to 1) develop curricular materials for classroom instruction, 2) work with faculty to incorporate HPC resources, tools, and methods into the curriculum, and 3) engage undergraduate students with internships. The Ohio Supercomputer Center supported the Blue Waters Graduate Fellowship program, enhanced the HPC University portal, and coordinated the Virtual School of Computational Science and Engineering shared graduate credit courses. The Center for Education Integrating Science, Mathematics & Computing (CEISMC) at the Georgia Institute of Technology conducted formative and summative evaluations of the education and training activities.

Blue Waters coordinated with the XSEDE project to conduct complementary education and training activities, avoiding duplication of effort and helping to facilitate quick sharing of lessons learned. Over time, Blue Waters collaborated on multiple training activities with multiple HPC centers funded by NSF, DOE, and DOD as well as with international HPC organizations including Compute Canada, PRACE, and RIKEN.

6.1.2 Collaborative Partners

We have seen the benefits derived from collaborations with educational institutions to co-teach credit courses that could not otherwise be offered at the member institutions. Collaborations with HPC centers and higher education institutions allowed the delivery of HPC training to hundreds of participants via multi-day sessions at minimal cost to the cooperating organizations and institutions. Professional societies (e.g., ACM SIGHPC Education Chapter) are bringing together trainers, educators, and staff to foster ongoing sharing of best practices and lessons learned, collaborate on the organization of education and training workshops at major international HPC conferences, and form working groups with common goals. Professional societies are extending community cooperation beyond the sharing that occurs at annual conferences and allowing member organizations to achieve greater impact.

6.1.3 Human-interest

Of special note, the SIGHPC Education Chapter announced in August 2022 that Dr. Robert Panoff, Shodor Education Foundation, was named as the first recipient of the award for Outstanding Contributions to Computational Science Education, to be honored during the awards ceremony at the SC22 Conference.

6.2 Cyberinfrastructure

The physical infrastructure included the computing resource, the high-definition video conferencing facilities, and the learning repositories. The evaluations also played a key role in helping to identify ways to improve the activities and document the longitudinal impact of the activities.

6.2.1 Reliable and Accessible Cyberinfrastructure

The infrastructure being used to deliver the content needs to be well tested in advance. This includes confirming that all presenters and facilitators have tested and used the A/V, computing infrastructure, and communications channels (e.g., Slack) in advance of the session. A plan for testing and confirming that all participants can access, log in, and submit a simple task to the computing platform in advance of the first use of the system will reduce frustrations and delays during the instructional periods.

To accommodate large numbers of participants running short exercises, arrange special queues for small jobs to get through the HPC system quickly. Advise the system operators of the impending load well in advance to preclude any downtime, system upgrades, or large-scale jobs running at the same time.

We highly recommend that a backup plan should be established to accommodate possible technology failures or interruptions that could potentially arise that would disrupt the flow of an event. Anticipating issues and being ready to mitigate them avoids wasting time trying to fix them on the fly while people “twiddle their thumbs.”

6.2.2 Video Conferencing Infrastructure

As a member of NCSA, the Blue Waters project was able to utilize the existing human talent and physical audio/video infrastructure that existed within NCSA. As Blue Waters expanded the use of high-definition video conferencing capabilities, NCSA expanded the physical infrastructure and capabilities to match the usage and needs for full two-way communications.

NCSA’s video conferencing rooms utilize multiple cameras to ensure that the instructor and the audience members can be seen along with a view of the instructor’s desktop. Microphones are placed in front of every participant with a mute/unmute button within their control. The room monitor can mute all participants, if needed, to minimize background noise. Large screens are used to ensure that all participants can see the views presented to all locations. Each site also has the capacity to see “picture in picture” views of all locations joined into the video conference.

Specifications and setup for the NCSA video conferencing setup are available [28].

6.2.3 Learning Repositories

Many educators and instructors of formal and informal learning sessions do not know how to start to develop training content and incorporate computational methods into the curriculum. This challenge can be addressed by sharing well-tested materials and exercises via public repositories (i.e., portals, collections, or libraries). We have demonstrated significant uptake of modules from such repositories by faculty and staff who have incorporated computational methods into the curriculum.

Many people use search engines (e.g., Google) to find training and education materials they need. The curation of education and training materials enhances the ability for people to find materials that are most relevant to their needs. Curation helps to ensure that materials are high-quality and up-to-date. Curated collections may include verification, validation, and accreditation information [12]. They may also include roadmaps (sequences for learning), validated exercises, translations to multiple languages, and community reviews.

Blue Waters collaborated with Shodor and XSEDE to augment the list of training and education resources accessible through the repositories HPC University [19], the Computational Science Education Reference Desk [18], and Interactivate [20]. Over six million page views of these repositories were recorded in April 2022.

There are hundreds (if not thousands) of repositories in the world, and this exceeds the capacity for any one person or group to compile and categorize an index of materials that are available. The next most useful stage of evolution of repositories for the community is to develop common metadata and implement automatic searching across multiple repositories. The ACM SIGHPC Education Chapter is pursuing this strategy to expand access to education and training materials on an international scale.

6.3 Evaluations

From the outset of the Blue Waters project, it was determined that evaluating community engagement programs was critical to learning how to improve them. Formative evaluations are most important for identifying issues that can be addressed to improve the content, delivery, and overall experiences for participants, instructors, and facilitators.

External evaluators conducted focus groups, surveys, and one-on-one discussions. The evaluators provided anonymized reports to the Blue Waters project office and discussed strategies for addressing the findings with the project managers on a regular basis.

The formative evaluations led to many improvements in the EOT activities. Their impact is reflected in the lessons learned. The summative evaluations helped to document the overall impact through longitudinal analysis of the impact of the programs. The evaluators previously reported on their methods and findings [2].

7. SUMMARY

HPC centers place a high importance on making effective use of expensive shared computing systems since demand for access far outweighs the available resources. Training is a key component in helping to ensure that people know how to make effective use of these systems and how to maximize the potential for accelerating their discoveries. The need for training is a direct result of the lack of formal preparation of today’s undergraduate and graduate students for understanding how HPC resources can be applied to advance research and development in all disciplines.

Faculty, staff and students conducting computational science and data analysis come from many backgrounds and discover the need for computational resources at very different times in their educational and professional careers. As such, a successful program needs to incorporate many different approaches.

We have demonstrated methods by which HPC centers can work with educational institutions to provide students with better preparation by working together to co-teach courses, update the curriculum, and ensure a continuum of improvement in courses over time for all institutions to benefit students in all disciplines. We have demonstrated that a comprehensive education, training, and student engagement program can effectively engage and serve a large national audience.

The impact of the lessons learned persists. Many organizations have adopted the use of high-definition video conferencing for events, the Texas Advanced Computing Center (TACC) supports an HPC graduate fellowship program, XSEDE adapted the Blue Waters Student Internship Program for the EMPOWER program [30], XSEDE scaled up their webcast training to serve people at remote institutions, groups like PICUP [15] are offering computational science and engineering workshops for faculty using Shodor's materials, more researchers are incorporating GPUs and other emerging tools, the ACM SIGHPC Education Chapter is building community and making learning repositories more accessible, the Ohio Supercomputing Center is using NSF funding to prototype shared credit courses among Minority Serving Institutions, and there are over five million page views per month of materials in the repositories supported by Shodor.

We are pleased to share our experiences with the community and to see our best practices and lessons learned being applied by other organizations. There are more detailed internal reports on these activities [11, 10, 13]. We are committed to continuing to offer our expertise, insights, resources, and services. We welcome you to contact us to discuss these topics in greater depth.

8. ACKNOWLEDGMENTS

This work is part of the Blue Waters sustained-petascale computing project, "Leadership-Class Scientific and Engineering Computing: Breaking Through the Limits", which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) the State of Illinois, and as of December, 2019, the National Geospatial-Intelligence Agency. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

9. REFERENCES

[1] Cornell University. 2022. Cornell Virtual Workshop. Retrieved November 2022 from https://cvw.cac.cornell.edu/default

[2] Lizanne DeStefano and Jung Sun Sung. 2019. Initial impact of Evaluation in Blue Waters Community Engagement Program. *Journal of Computational Science Education* 10, 1 (January 2019), 32–39. DOI:https://doi.org/10.22369/issn.2153-4136/10/1/6

[3] Sharon C. Glotzer and Thom H. Dunning. 2008. Great Lakes Consortium for Petascale Computation Virtual School of Computational Science and Engineering: Educating a New Generation of Computational Scientists and Engineers for Discovery and Innovation in the 21st Century. (May 2008). Retrieved November 2022 from https://uofi.box.com/s/frdzwfjotla0k7bln2xsggd2rbskvrru

[4] Sharon C. Glotzer, Bob Panoff, and Scott Lathrop. 2009. Challenges and Opportunities in Preparing Students for Petascale Computational Science and Engineering. *Comput. Sci. Eng.* 11, 5 (Sept.–Oct. 2009), 22–27. DOI:https://doi.org/10.1109/MCSE.2009.134

[5] Steven I. Gordon, James Demmel, Lizanne Destefano, and Lorna Rivera. 2015. Extending Access to HPC Skills Through a Blended Online Course. In *XSEDE '15: Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*. July 26, 2015, St. Louis, Missouri. Association for Computing Machinery, New York, NY, USA, 1–5. DOI: https://doi.org/10.1145/2792745.2792760

[6] Jennifer K. Houchins, Jeffrey D. Krause, Scott A. Lathrop, and Robert M. Panoff. 2011. A multi-cultural success story for achieving diversity in multi-core/many-core internships. In *TG '11: Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery*, July 2011. Salt Lake City, UT. ACM. Article 51, 1 page. DOI:https://doi.org/10.1145/2016741.2016796

[7] Jennifer Houchins, Scott Lathrop, Robert Panoff, and Aaron Weeden. 2020. Blue Waters Workforce Development: Delivering National Scale HPC Workforce Development. *Journal of Computational Science Education* 11, 1 (January 2020), 26–28. DOI:https://doi.org/10.22369/issn.2153-4136/11/1/5

[8] Patricia Jacobs, Phillip List, Mobeen Ludin, Aaron Weeden, and Robert M. Panoff. 2014. The Blue Waters Student Internship Program: Promoting Competence and Confidence for Next Generation Researchers in High-Performance Computing. In *2014 Workshop on Education for High Performance Computing*, November 16, 2014, New Orleans, LA. IEEE. 49–55. DOI:https://doi.org/10.1109/EduHPC.2014.6

[9] Scott A. Lathrop, Katharine Cahill, Steven I. Gordon, Jennifer Houchins, Robert M. Panoff, and Aaron Weeden. 2020. Preparing a Computationally Literate Workforce. *Comput. Sci. Eng.* 22, 4 (July–Aug. 2020), 7–16. DOI:https://doi.org/10.1109/MCSE.2020.2994763

[10] Scott Lathrop and Steven I. Gordon. Impact of the Blue Waters Graduate Fellowship Program. Retrieved November 2022 from https://uofi.box.com/s/0z1kirremhfrtw3zm8u8xagvnepnxq5d

[11] Scott Lathrop, Steven I. Gordon, Aaron Weeden, and Lizanne DeStefano. Impact of Blue Waters Education and Training. Retrieved November 2022 from https://uofi.box.com/s/k4gpj704cnv2irt49txue5wahwu5bmqo

[12] Scott Lathrop, Marilyn McClelland, Steven Gordon, D. E. Stevenson, and David Joiner. 2005. Applying verification, validation, and accreditation processes to digital libraries. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '05)*, June 7–11, 2005, Denver, Colorado. IEEE. 382–382, https://doi.org/10.1145/1065385.1065485

[13] Scott Lathrop, Robert M. Panoff, and Aaron Weeden. Impact of the Blue Waters Student Internship Program. Retrieved November 2022 from https://uofi.box.com/s/i65thkjwo80rop95yjavdhm002i75q8k

[14] National Science Foundation. Established Program to Stimulate Competitive Research (EPSCoR). Retrieved November 2022 from https://beta.nsf.gov/funding/initiatives/epscor

[15] PICUP. 2022. PICUP: Partnership for Integration of Computation into Undergraduate Physics. Retrieved November 2022 from https://www.compadre.org/PICUP/

[16] Shodor. 2022. Blue Waters is Wide. Retrieved November 2022 from https://www.youtube.com/watch?v=Dor2_B_OkR8

[17] Shodor. 2022. Blue Waters Petascale Education Program. Retrieved November 2022 from http://shodor.org/petascale/

[18] Shodor. 2022. CSERD. Retrieved November 2022 from http://shodor.org/cserd/

[19] Shodor. 2022. HPC University. Retrieved November 2022 from http://www.hpcuniversity.org/

[20] Shodor. 2022. Interactivate. Retrieved November 2022 from http://shodor.org/interactivate/

[21] Shodor. 2022. Journal of Computational Science Education. Retrieved November 2022 from https://jocse.org/

[22] Shodor. 2022. Old Songs, New Words. Retrieved November 2022 from http://shodor.org/songs/

[23] Shodor. 2022. Petascale Institute Curriculum. Retrieved November 2022 from http://shodor.org/petascale/materials/institute/

[24] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr. 2014. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* 16, 5 (Sept.–Oct. 2014), 62–74. DOI:https://doi.org/10.1109/MCSE.2014.80

[25] University of Illinois Board of Trustees. 2022. Cyberinfrastructure Tutor (CI-Tutor). Retrieved November 2022 from https://researchit.illinois.edu/resource/71

[26] University of Illinois Board of Trustees. 2022. Blue Waters Training Resources. Retrieved November 2022 from https://publish.illinois.edu/bluewaters/blue-waters-training-and-outreach/blue-waters-training-resources/

[27] University of Illinois Board of Trustees. 2022. Blue Waters Webinars. Retrieved November 2022 from https://publish.illinois.edu/bluewaters/blue-waters-training-and-outreach/blue-waters-webinars/

[28] Video Conferencing Infrastructure: Distributed Synchronous Instructional Laboratories. Retrieved November 2022 from https://uofi.app.box.com/s/5cvnbznsv9yyo91f8fxarzxdi75sdydd

[29] Virtual School of Computational Science and Engineering: Lessons Learned. Retrieved November 2022 from https://uofi.app.box.com/s/1ffybmchwlz2uvsrkgacb5bpu9qalfhl

[30] Aaron Weeden. 2021. XSEDE EMPOWER: Engaging Undergraduates in the Work of Advanced Digital Services and Resources. *Journal of Computational Science Education* 12, 2 (February 2021), 22–24. DOI:https://doi.org/10.22369/issn.2153-4136/12/2/5

[31] Aaron Weeden, Colleen Heinemann, Skylar Thompson, Cameron Foss, Mobeen Ludin, and Jennifer K. Houchins. 2019. The Blue Waters Petascale Institute: Longitudinal Impact and Assessment-Driven Development of an Intensive, Hands-on Curriculum for Teaching Applications in HPC. In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* 85 (July 2019), 1–8. DOI:https://doi.org/10.1145/3332186.3337771