A multipronged approach to building a diverse workforce and cultivating an inclusive professional environment for DOE high-performance computing

Prepared by the ECP Task Force on Broader Engagement

Ashley Barker,* OLCF
Dan Martin,* LBL
Mary Ann Leung, Sustainable Horizons Institute
Lois Curfman McInnes,* ANL
Jini Ramprakash, ALCF
Julia White,* ORNL
Jim Ahrens, LANL
Erik Draeger, LLNL
Shaun Fomby, ORNL
Yasaman Ghadar, ALCF
Rinku Gupta, ANL
Mahantesh Halappanavar, PNNL
Mike Heroux, SNL
Christopher Kelly, BNL
Mark Miller, LLNL
Hai Ah Nam, NERSC
Slaven Peles, ORNL
Damian Rouson, LBL
Dan Turner, SNL
Terry Turton, LANL

Asterisk denotes task force leadership team

Liaisons from the Computing Research Leadership Council (CRLC):
David Brown, LBL
Valerie Taylor, ANL

December 13, 2021

Response to question #5 of the DOE Request for Information on Software Stewardship

Acknowledgment: This work was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative.

Introduction

The U.S. Department of Energy (DOE) is a long-standing leader in research and development of high-performance computing (HPC) in the pursuit of science in the national interest. DOE's investments have pushed the growth of computational science and engineering as an essential driver of scientific and technological progress in conjunction with theory and experiment. Moreover, DOE researchers have developed a wide range of high-impact scientific software, now broadly recognized as key crosscutting technology that connects advances in applied mathematics, computer science, and core disciplines of science and engineering for advanced modeling, simulation, discovery, analysis, and learning. Researchers face unprecedented software challenges, however, due to disruptive changes in high-performance computer architectures and increasing complexity required to address next-generation computational science. In addition, recent community reports have expressed the imperative to firmly embrace the fundamental role of open-source scientific software as a valuable research product and cornerstone of collaboration—and thus to increase direct investment in the software itself, not just as a byproduct of other research [1, 2, 3, 4, 5, 6, 7]. This situation brings with it a mandate to fundamentally change how scientific software is designed, developed, and sustained, while tackling urgent challenges in workforce training and recruitment for the computing sciences [8, 9, 10, 11 12, 13].

Addressing these workforce challenges requires broad community collaboration to change the culture and demographic profile of computational science. Impactful DOE-wide programs [14, 15, 16, 17, 18, 19] and activities in the wider computing community [20, 21, 22] are making headway. Led by staff who specialize in workforce, education, and improving culture with respect to diversity, equity, and inclusion (DEI), laboratory-specific initiatives are also underway to address challenges in workforce and training, capitalizing on each lab's unique perspectives and regional connections to underrepresented populations. The DOE HPC community has a compelling opportunity to address workforce challenges through a complementary lens that focuses on the distinct needs and culture of DOE high-performance computing.

Consequently, in August 2021 the DOE Exascale Computing Project (ECP) established a task force on broader engagement, whose members represent ANL, BNL, LBL, LLNL, LANL, ORNL, PNNL, and SNL, as well as DOE Office of Science Computing Facilities ALCF, NERSC, and OLCF. The task force builds on the special collaborative character of ECP—an aggressive multilab research, development, and deployment project focused on delivery of mission-critical applications, an integrated software stack, and exascale hardware technology advances. In partnership with the ECP and Facilities communities, as well as with lab staff who focus on education and workforce issues, the task force has begun a multipronged initiative to expand the pipeline and workforce for DOE high-performance computing. This response addresses RFI questions for topic #5 (*challenges in building a diverse workforce and maintaining an inclusive professional environment*), while also providing an overview of initial work by the task force.

Response to RFI Questions

1. **What challenges do you face in recruiting and retaining talented professionals to develop software for scientific and high-performance computing?**

With the fundamental opportunity to change the scientific software paradigm come challenges and opportunities with respect to developing the corresponding workforce. The 2014 DOE Advanced Scientific Computing Advisory Committee (ASCAC) Workforce Subcommittee Letter [8] states:

> "All large DOE national laboratories face workforce recruitment and retention challenges in the fields within Computing Sciences that are relevant to their mission. … Future projections indicate an increasing workforce gap and a continued underrepresentation of minorities and females in the workforce unless there is an intervention."

Likewise, the recent NSF report on "Pioneering the Future Advanced Computing Ecosystem: A Strategic Plan" [10] states:

> "Recruitment, retention, and cultivation of a highly capable, adaptive, and agile workforce is also difficult … Partnerships are also an important vehicle for recruitment and hiring."

Thus, there are workforce development challenges throughout the computing sciences at the DOE national laboratories.

2. What additional challenges exist in recruiting and retaining talented professionals from groups historically underrepresented in STEM and/or individuals from underserved communities?

The DOE national laboratories, like many other scientific research organizations, face growing needs and challenges in recruiting and retaining a skilled workforce. Government and academic sectors face fierce competition for talent attracted to lucrative and progressive industrial workplace benefits. Greater attrition rates among a large pool of talent made up of individuals from underrepresented groups and other historically underserved communities further complicate these workforce challenges. In addition, the emergence of scientific software as a field on its own creates new challenges and opportunities for workforce development.

Many successful recruitment models exist; however, they often rely on existing social networks and have largely resulted in a homogeneous workforce. The challenge is not only to develop new approaches to broaden the reach, but also to change longstanding recruitment, onboarding, and retention practices.

3. What challenges exist in maintaining inclusivity and equity in the development of community for scientific and high performance computing software?

Numerous studies have shown that diverse organizations, teams, and communities perform more creatively and effectively—and thus are demonstrably more innovative and productive. However, many of the existing computing sciences workplaces contain largely white male dominated demographic profiles, typically with fairly small percentages of women and even smaller minority populations. These numbers create scenarios where members of an underrepresented group, often are the "only one" of their demographic within their organizational group, e.g., the only woman in the group, the only Black person, etc. Resultant phenomena include isolation, lack of inclusion, imposter syndrome, etc. One of the key challenges is how to foster a sense of belonging to members of underrepresented groups while also respecting cultural frames of reference [23].

4. What successful strategies have you employed to help overcome these challenges?

To overcome the problem of small numbers as a starting state, it is important to balance critical mass with integration, not separation. That is, to enable members of underrepresented groups to develop community with other members of underrepresented groups without isolating them, but rather promoting inclusion [23]. The Sustainable Research Pathways Program [24], along with the Broader Engagement [25] programs at the Society of Industrial and Applied Mathematics (SIAM) Computational Science and Engineering (SIAM-CSE) conferences, are successful implementations of this approach. Working within mainstream organizations, the programs alternatively and iteratively concentrate members from underrepresented groups into small clusters to prepare them and then disperse them into the broader more homogeneous environment. Moreover, individuals and groups can take steps forward within their own spheres of influence to broaden participation of underrepresented groups [26].

5. What opportunities for professional recognition and career advancement exist for those engaged in developing scientific and high-performance computing software?

Developing high-performance scientific software requires the combined contributions of many people, whose skills span applied mathematics, computer science, core disciplines of science and engineering, software engineering, scientific programming, performance analysis, team collaboration, and more. Due to traditional metrics for career advancement focusing on publications rather than software contributions, people who focus on the important work of software development often face challenges in professional recognition and career

advancement. Community organizations [27] are working for change—advancing understanding of the importance of high-quality software in multidisciplinary collaboration and the integrity of computational research, and articulating key issues and needs to stakeholders, agencies, and the broader research community to effect changes in policies, funding, metrics, and reward structure. For example, Research Software Engineering (RSE, https://society-rse.org and https://us-rse.org) has emerged as an increasingly recognizable career track, helping to build community, mentoring, and professional recognition for scientific software specialists. Also, the Better Scientific Software Fellowship Program (https://bssw.io/fellowship), launched in 2018 as an initiative of the IDEAS productivity project (https://ideas-productivity.org) [28], provides recognition and funding to leaders and advocates of high-quality scientific software.

ECP Broader Engagement Initiative

In addition to directly responding to these RFI questions, we provide a brief overview of work underway in the ECP Broader Engagement Initiative. We are establishing a sustainable plan to recruit and retain a diverse HPC workforce by fostering a supportive and inclusive culture within the computing sciences at DOE national laboratories. Specifically, we are:

- Engaging talented people with the potential for strong skills and interest in HPC from underrepresented groups, including Black or African American, Hispanic/Latinx, American Indian, Alaska Native, Native Hawaiian, Pacific Islanders, women, persons with disabilities, first-generation scholars, people from smaller colleges & universities, and others;
- Raising awareness of DOE activities and needs related to scientific applications, software technologies, hardware, and infrastructure; and
- Providing pathways for interactions, including training, internships, and collaborations, while raising awareness of career opportunities.

Our approach recognizes that HPC achievements — including computational and data-enabled science, analytics, learning, and artificial intelligence (AI) — impact virtually all aspects of our world, and that these interdisciplinary fields offer countless opportunities across mathematics, computer science, diverse scientific and engineering disciplines, data science, and AI. We are reaching out to students and faculty who want to get involved in the ECP and Facilities communities, engaging in cutting-edge research and development that provides the foundation for fulfilling careers and broad impacts on society.

Our approach has three complementary thrusts, which leverage and augment existing workforce efforts in DOE national laboratories, computing facilities, and the HPC computational science community.

1: Establishing an HPC Workforce Development and Retention (HPC-WDR) Action Group, whose primary focus is fostering a supportive and inclusive culture in ECP, DOE labs, and communities. Its goal is to promote the workforce pipeline for, and the retention of, a diverse DOE lab HPC workforce. In addition to engaging underrepresented communities and strengthening our workforce pipelines, we also must focus on HPC culture within DOE laboratories. While the HPC community, including DOE labs, have undertaken a variety of initiatives to improve culture with respect to diversity, equity, and inclusion, sharing of information across institutions is limited. The HPC-WDR Action Group will provide a mechanism for better sharing and communicating about these initiatives, best practices, lessons learned, and other DEI experiences, thereby benefiting the HPC community across the DOE lab complex. Key elements of work during phase 1, beginning early 2022, are:

- Leading a new webinar series: *Best Practices for HPC Workplace Development and Retention*;
- Developing a website and curating resources about HPC workforce development and retention; and
- Determining recommendations and strategies for improvement.

2: Building an "Introduction to HPC" Training and Workforce Pipeline Program (HPC-Intro) to provide accessible introductory material on HPC, scalable AI and analytics, thereby addressing gaps in — and expanding the pipeline of — people with foundational HPC skills. Basic HPC is not typically taught at early

stages of students' careers, and the capacity and knowledge of HPC at many institutions are limited. Even so, such topics are prerequisites for advanced opportunities such as the Argonne Training Program for Extreme-Scale Computing (ATPESC) and the DOE Computational Science Graduate Fellowship (CSGF) Program, and ultimately for careers in HPC. The national lab computing complex has expertise, capabilities, and a long history of cross-lab coordination on joint efforts. The HPC-intro training and pipeline program will target advanced undergraduates, students in gap years, and early graduate students. Key elements of work during phase 1, beginning early 2022, are:

- Conducting listening sessions with administrators and faculty at historically black colleges and universities (HBCUs), minority-serving institutions (MSIs), and schools in EPSCOR states, in order to understand the challenges and needs of target schools and students;
- Developing the curriculum and content for a 2-week "Intro to HPC" course and teaching this material in a pilot session in 2022, based at DOE labs; and
- Establishing partnerships with universities to implement the HPC-intro curriculum at their institutions, including a pilot session in 2022.

By intersecting the needs of the workforce with the needs of DOE HPC, we will build self-sustaining and integrated programs to expand the pipeline of professionals joining the HPC workforce.

3: Creating Sustainable Research Pathways for HPC (SRP-HPC); based on the Sustainable Research Pathways Program (a partnership with Sustainable Horizons Institute and Lawrence Berkeley Laboratory) and the Broader Engagement Program at SIAM-CSE Conferences.

We are establishing a multilab cohort of students from underrepresented groups (and faculty working with them), who will work side-by-side with ECP teams and Facilities staff on world-class HPC projects, including:

- Advancing discoveries in chemistry, materials, energy, Earth and space science, data analytics, optimization, AI, national security, and more;
- Building software technologies that power HPC discoveries (open-source programming models, math libraries, data and visualization packages, development tools, and more); and
- Deploying an exascale software stack, including software integration across leadership computing environments and improving software productivity and sustainability to help advance scientific productivity.

Currently work is underway in identifying project hosts (from the ECP and Facilities communities) and potential visiting faculty/students (from US institutions) for a 2022 pilot program.

Conclusion. The work begun by the ECP task force on broader engagement provides vehicles for engaging talented people with the potential for strong skills and interest in the computing sciences to learn about DOE scientific computing activities and needs, while building pathways for further interactions. Consequently, this work helps to address pressing challenges in building the research innovation workforce, while providing creative incentives and reward mechanisms, and increasing partnerships between academia, national laboratories, nonprofits, and industry. This work also will help to build a rich repository of freely available learning materials on scientific software, as well as engagement opportunities to attract and guide talented individuals toward careers in DOE computing sciences.

These preliminary steps are just the starting point of long-term work to change the culture and demographic profile of HPC computational science. We face two urgent challenges: (1) the ECP project will conclude in 2023, and (2) we need to engage the entire DOE HPC community in this initiative, not just people funded through ECP. A critical challenge is identifying a path toward long-term sustainability of work initiated by the task force to broaden participation in DOE HPC, including next phases beyond these initial pilot projects.

References

[1] M. R. Benioff and E. D. Lazowski, PITAC Co-Chairs, Computational science: Ensuring America's competitiveness: President's Information Technology Advisory Committee. http://vis.cs.brown.edu/docs/pdf/Pitac-2005-CSE.pdf, 2005.

[2] W. Gropp, R. Harrison et al., Future directions for NSF advanced computing infrastructure to support U.S. science and engineering in 2017-2020. National Academies Press, 2016. http://www.nap.edu/catalog/21886/.

[3] M. A. Heroux, G. Allen, et al., Computational Science and Engineering Software Sustainability and Productivity (CSESSP) Challenges Workshop Report, Sept 2016. https://www.nitrd.gov/PUBS/CSESSPWorkshopReport.pdf.

[4] S. Hettrick, Research software sustainability. Report on a Knowledge Exchange Workshop, 2016. http://repository.jisc.ac.uk/6332/1/Research_Software_Sustainability_Report_on_KE_Workshop_Feb_2016_FINAL.pdf.

[5] H. Johansen, L. C. McInnes, D. Bernholdt, J. Carver, M. Heroux, R. Hornung, P. Jones, B. Lucas, A. Siegel, and T. Ndousse-Fetter, Software productivity for extreme-scale science, 2014. Report on DOE Workshop, Jan 2014. https://science.osti.gov/-/media/ascr/pdf/research/cs/Exascale-Workshop/SoftwareProductivityWorkshopReport2014.pdf.

[6] D. Keyes, V. Taylor, et al., National Science Foundation Advisory Committee on CyberInfrastructure, Task Force on Software for Science and Engineering, final report, 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_Software.pdf.

[7] J. Hack, K. Riley, R. Gerber, K. Antypas, R. Coffey, E. Dart, T. Straatsma, J. Wells, D. Bard, S. Dosanj, I. Monga, M. Papka, L. Rotman., Crosscut Report: Exascale Requirements Reviews, 2017. https://www.osti.gov/servlets/purl/1417653.

[8] B. Chapman, H. Calandra, S. Crivelli, J. Dongarra, J. Hittinger, S. Lathrop, V. Sarkar, E. Stahlberg, J. Vetter, D. Williams, DOE Advanced Scientific Advisory Committee (ASCAC): Workforce Subcommittee Letter, 2014, https://dx.doi.org/10.2172/1222711.

[9] Diversity demographics of the 17 DOE National Laboratories, https://nationallabs.org/staff/diversity/.

[10] National Science & Technology Council, National Strategic Computing Initiative Update: Pioneering the Future of Computing, 2019, https://www.nitrd.gov/pubs/National-Strategic-Computing-Initiative-Update-2019.pdf

[11] L. Arafune et al., CI Workforce Development Workshop 2020, https://www.rcac.purdue.edu/ciworkforce2020,

[12] National Science & Technology Council, Pioneering the Future Advanced Computing Ecosystem: A Strategic Plan, 2020, https://www.nitrd.gov/pubs/Future-Advanced-Computing-Ecosystem-Strategic-Plan-Nov-2020.pdf

[13] Office of Advanced Cyberinfrastructure, CISE, NSF, Transforming Science Through Cyberinfrastructure: NSF's Blueprint for a National Cyberinfrastructure Ecosystem for Science and Engineering in the 21st Century: Blueprint for Cyberinfrastructure Learning and Workforce Development, Aug 2021, https://www.nsf.gov/cise/oac/vision/blueprint-2019/CI-LWD.pdf

[14] DOE Visiting Faculty Program, seeks to increase the research competitiveness of faculty members and their students at institutions historically underrepresented in the research community.

[15] DOE Minority Serving Institutions Program (MSI), promotes the education and development of the next generation workforce in critical science, engineering, technology, and math (STEM) related disciplines.

[16] DOE Community College Internships Program (CCI), seeks to encourage community college students to enter technical careers relevant to the DOE mission.

[17] DOE Science Undergraduate Laboratory Internships (SULI), encourages undergraduate students to pursue STEM careers by providing research experiences at DOE laboratories.

[18] Mentorship for Environmental Scholars Program, collaborative effort between Pre-College University and the DOE to increase minority awareness and participation in the environmental science disciplines.

[19] GEM Fellowship Program, seeks to recruit high-quality under-represented students looking to pursue degrees in applied science and engineering.

[20] T. Whitney and V. Taylor, Increasing Women and Underrepresented Minorities in Computing: The Landscape and What You Can Do, Computer, vol. 51, no. 10, pp. 24-31, 2018, 10.1109/MC.2018.3971359.

[21] White House Report by the National Science and Technology Council, Best Practices for Diversity and Inclusion in STEM Education and Research: A Guide by and for Federal Agencies , Sept 2021, https://www.whitehouse.gov/wp-content/uploads/2021/09/091621-Best-Practices-for-Diversity-Inclusion-in-STEM.pdf.

[22] National Academies of Sciences, Engineering, and Medicine, E. Hammonds, V. Taylor, and R. Hutton, editors, Transforming Trajectories for Women of Color in Tech. Washington, DC: The National Academies Press, 2021, https://doi.org/10.17226/26345.

[23] M. A. Leung, "Diversity and Inclusion Through Leadership During Challenging Times" in Computing in Science & Engineering, vol. 22, no. 06, pp. 92-96, 2020.

[24] M. A. Leung, S. Crivelli and D. Brown, "Sustainable Research Pathways: Building Connections across Communities to Diversify the National Laboratory Workforce," in Collaborative Network for Engineering and Computing Diversity (CoNECD), Washington D.C., 2019.

[25] M. A. Leung, "Broadening Engagement by Extending the Classroom to Conferences," in Society of Industrial and Applied Mathematics Education Conference, Portland, 2018.

[26] M.A. Leung, D. Rouson and L.C. McInnes, Increasing Productivity by Broadening Participation in Scientific Software Communities, 2020 Collegeville Workshop whitepaper and BSSw.io blog post, Sept 2020,

[27] D. S. Katz, L. C. McInnes, D. E. Bernholdt, A. C. Mayers, N. P. Chue Hong, J. Duckles, S. Gesing, M. A. Heroux, S. Hettrick, R. C. Jimenez, M. Pierce, B. Weaver, and N. Wilkins-Diehr, Community organizations: Changing the culture in which research software is developed and sustained, special issue of IEEE Computing in Science and Engineering (CiSE) on Accelerating Scientific Discovery with Reusable Software, 21 (2019), pp. 8–24, https://dx.doi.org/10.1109/MCSE.2018.2883051.

[28] M. A. Heroux, L. C. McInnes, D. E. Bernholdt, A. Dubey, E. Gonsiorowski, O. Marques, J. D. Moulton, B. Norris, E. M. Raybourn, S. Balay, R. Bartlett, L. Childers, T. Gamblin, P. Grubel, R. Gupta, R. Hartman-Baker, J. Hill, S. Hudson, C. Junghans, A. Klinvex R. Milewicz, M. C. Miller, H. Nam, J. O'Neal, K. Riley, B. Sims, J. Shuler, B. Smith, L. Vernon, G. Watson, J. Willenbring, and P. Wolfenbarger, Advancing Scientific Productivity through Better Scientific Software: Developer Productivity and Software Sustainability Report, ECP Technical Report ECP-U-RPT-2020-0001, Jan 2020, https://exascaleproject.org/better-scientific-productivity-through-better-scientific-software-the-ideas-report.