Assessing Shared Material Usage in the High Performance Computing (HPC) Education and Training Community

Susan Mehringer
Center for Advanced Computing
Cornell University
Ithaca, New York
shm7@cornell.edu

Kate Cahill
Ohio Supercomputer Center
Columbus, Ohio
kcahill@osc.edu

John-Paul Navarro
University of Chicago
Argonne National Lab
Naperville, Illinois
navarro@anl.gov

Charlie Dey
Texas Advanced Computing Center
Austin, Texas
charlie@tacc.utexas.edu

Scott Lathrop
University of Illinois
Shodor Education Foundation, Inc.
Urbana-Champaign, Illinois
lathrop@illinois.edu

Mary Thomas
San Diego Supercomputing Center
University of California San Diego
La Jolla, California
mpthomas@ucsd.edu

Jeaime H. Powell
Texas Advanced Computing Center
Austin, Texas
jpowell@tacc.utexas.edu

ABSTRACT

This paper shares the results of a survey conducted October-November 2022. The survey's intent was to learn how the community both shares and discovers training and education materials, whether those needs were being met, and if there were interest in improving how materials are shared. The survey resulted in 112 responses primarily from content authors who are, or support, academics. While the majority of respondents considered themselves successful in finding materials, most also encountered barriers, such as finding materials, but not at the needed depth or level. Most respondents were both interested in, and able to, work toward community efforts to improve finding materials, with most citing lack of staff time as a barrier to doing so. Proposed efforts in community engagement to work toward these efforts are discussed.

Keywords

Survey, education, training, community engagement.

1 INTRODUCTION

The use of computing technologies is rapidly expanding in many sectors, necessitating access to high-quality education and training materials to facilitate research computing. The demand for instructional materials, encompassing a wide range of topics related to the development and application of research computing technologies across disciplines, is crucial for both formal classroom settings, informal training, and self-paced learning.

One way to meet this need and keep up with the ever-evolving landscape of HPC educational and training material development is to improve how the community shares and finds materials. In order

to gauge the needs of the HPC Education and Training community with regard to sharing training materials we sought input from stakeholders. To this end, we conducted a survey to explore interest and key factors related to sharing and discovering training and education materials. The results of this survey highlighted the barriers to finding relevant materials and the barriers to sharing materials developed. Overall, we learned that there is a great deal of interest in sharing materials developed more widely and making access to these materials easier for learners.

2 SURVEY BACKGROUND & MOTIVATION

The survey aimed to assess if individuals or organizations have training materials they wish to share or make more widely accessible, and if local communities require more efficient means of locating relevant materials. We were interested in learning how the research computing community shares and finds these materials and if they thought there should be more or better options for doing this. The term "repositories" was used broadly to encompass portals, collections, libraries, and lists of training and education materials and resources for the purposes of this survey; this statement was included in the survey preamble.

3 SURVEY METHODOLOGY

The survey [2] authors developed the survey questions which then were reviewed by a focus group from the HPC education and training community. To ensure accessibility, the survey was conducted using Google Forms and remained open for several weeks in October and November 2022. Invitations to participate were sent out through well-known mailing lists in the HPC support community, such as CaRCC People Network, Campus Champions, Virtual Residents, Coalition for Academic Scientific Computation, and the EDU Special Interest Group on High Performance Computing. This effort resulted in a total of up to 112 responses received for each question. We were targeting professionals in the HPC support community at research computing centers. We shared interim results at the Ninth SC Workshop on Best Practices for HPC Training and Education (BPHTE22) at SC22.

4   SURVEY RESULTS

The survey results based on 112 responses to the questions listed in Appendix A are described in this section. The survey data is available online [2].

4.1   About the Respondents

The survey begins with two questions pertaining to the respondent's role and the communities they support. Both questions allowed multiple selections and were answered by all 112 respondents. Tabulated results, summarized in Table 1 and Table 2, show that 84% are, or support, academics, closely followed by 71% for both the Grad/Post doc and Undergrad communities. 37% of the respondents specifically selected these three options only. 84% are content authors, while 61% curate appropriate materials for their community. 25% see themselves as filling all four roles.

Table 1. Which communities do you support or participate in?

Community	Responses
Academia	94
Grad/Post doc	80
Undergrad	80
Government	25
Pre-college	18
Industry	16
Other	4
Total	**112**

Table 2. When it comes to training and education materials, I consider myself to be:

Respondent's role	Responses
Content author	94
Curator collecting appropriate materials for my community	68
Consumer of materials hosted by other organizations	57
Consumer of materials hosted by my organization	35
Other	6
Total	**112**

4.2   Finding Materials

Question 3, with 111 responses, show only 20% found it difficult to find appropriate training and education materials on specific topics, for themselves. Question 4, "How easy is it for you to find appropriate training and education materials on specific topics, for

group(s) you support?", resulted in only 32% of 111 respondents saying that it was difficult or very difficult.

In question 5, we asked a multiple selection question to learn where respondents look for material, with 111 responses. 94% use search engines, 64% use portals or repositories hosted by other organizations, 45% use a portal or repository hosted locally, and 17% selected Other.

Question 6, "Which portals, repositories, search engines or other resources do you use or find helpful?" resulted in a broad array of both general and specific responses by 93 respondents. In question 7 we asked "How important is, or would be, having easy access to repositories of training and education materials from multiple organizations to your community?"; on a scale of 1 for "not important" to 5 for "very important," the average for 112 responses was 4.3. Figure 1 shows the results of questions 3, 4, and 7, displaying ease of finding materials along with the importance to material access to their community.

Question 8, regarding barriers encountered when searching for materials, answered by 109, asked for all barriers encountered, with results shown in Table 3. 66% said they can find materials, but not at the right depth or level needed. Question 9, answered by 108, asked for the single barrier that it would be most helpful to remove; The top answer was again that they can find materials, but not at the right depth or level needed, with 43%, as shown in Table 4.

4.3   Working Toward Solutions

Question 10 asked if the respondent's organization wants to make it easier to find their materials, question 11 asked for the biggest challenge to sharing, and question 12 asked whether the organization would be willing and able to provide metadata in a standard format. Figure 2, which combines results from questions 10 and 12, shows that most of the respondents fall into the top right area, both interested in, and able to, share materials. Question 11, which called for free response, was answered by 99 people with a broad variety of responses, including time, cost, copyright, and issues raised by the organization.

87 people responded to question 13, a multiple selection question seeking to identify which roadblocks prevent an organization from sharing content information in a standard format. Results in Table 5 show that 75% cited lack of staff time. 7% selected all options, while 30% selected a single reason, lack of staff time.
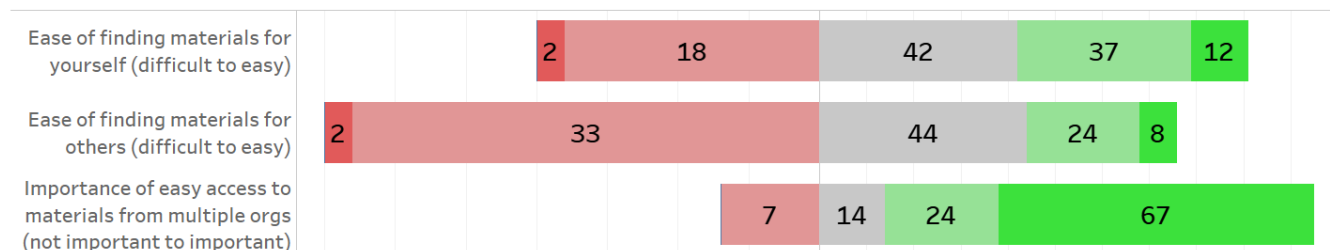

Figure 1. Top: Question 3, finding material for self. Middle: Question 4, finding material for others. Bottom: Question 7, importance of access to materials from multiple organizations.

In question 14, 47 people indicated they are interested in joining a group working on this project by providing contact information. In response to question 15, 56 respondents indicated they are willing to be contacted regarding survey responses. 36 people responded to the closing comments question.

Table 3. What barriers have you encountered when searching for materials?

Barriers encountered	Responses
I can't find materials on the topic I need	35
I can find materials on the topic, but not at the depth or level I need	72
I find too many materials, and I can't effectively sort through them all	44
I am aware of specific appropriate materials, but search engines don't list them in the top results	26
Other	28
Total	**93**

Table 4. Which barrier, if removed, would be most helpful for finding appropriate materials?

Barrier to remove	Responses
I can't find materials on the topic I need	12
I can find materials on the topic, but not at the depth or level I need	46
I find too many materials, and I can't effectively sort through them all	29
I am aware of specific appropriate materials, but search engines don't list them in the top results	10
Other	11
Total	**109**

Table 5. If your organization is not willing and able to provide metadata about your materials in a standard format, what are your roadblocks?

Roadblocks to providing metadata	Responses
Lack of staff time	65
Lack of funding	38
Inadequate staff expertise	28
Our materials aren't in a catalog	37
Other	12
Total	**87**

5   SURVEY ANALYSIS

The 112 survey responses and results show a strong interest and importance (question 7) in the topic of finding and sharing (question 10) education and training materials in the cybertraining community. While we saw strong interest, the results also showed many barriers (questions 8 & 9).

In questions 3, 4, and 7, we found that most of the respondents considered themselves successful in finding appropriate materials for themselves and others, while question 8 shows 109 responses listing barriers encountered when searching for materials, shown in Table 3. Perhaps this indicates that finding materials, while possible by dedicated professionals, could be significantly improved.

Figure 2 displayed two sparse quadrants. In the upper left quadrant, it is unsurprising that only one respondent is both able to provide metadata but uninterested. It is more interesting to see that the bottom right quadrant is also sparse; there are only 4 respondents saying that they want to make finding data easier, but don't have the ability. This shows great potential in the community moving forward with solutions.

Altogether, the results imply that the community sees the potential for improving discovery of materials and many have the interest and ability to contribute to a solution.
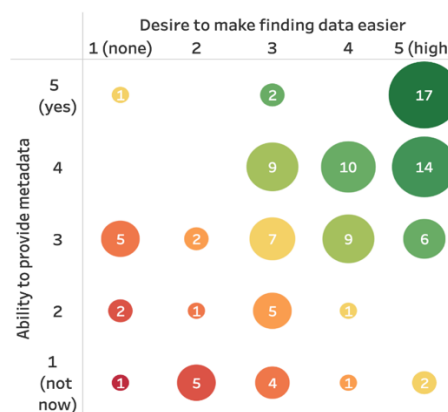

Figure 2. Sharing materials: interest and ability

6   RECOMMENDATIONS & FUTURE WORK

6.1 Community Engagement

6.1.1   Collaborating with HPC Education and Training Communities

We plan to collaborate with educational and training material organizations that focus on high-performance computing (HPC). This will involve our participation in various organizations such as the ACM SIGHPC and the NSF/IEEE-TCPP Curriculum Initiative on Parallel and Distributed Computing, working with institutions such as the NSF ACCESS MATCH program, Cornell Center for Advanced Computing, Kean University, San Diego Supercomputer Center, Texas Advanced Computing Center, Ohio Supercomputer Center, Pittsburgh Supercomputing Center, and others. In addition, we intend to connect with organizations that have received NSF CyberTraining awards to explore opportunities for sharing their

training products and increasing usage within and across disciplines.

6.1.2   Build an HPC Professional Trainer Community

Building a diverse community for HPC education and training begins with building a diverse trainer community. A diverse training community is an important goal because it confronts preconceived stereotypes in learning and education, allowing enhancements to both workplace and community cohesion. Diversity encourages critical thinking while helping students learn to communicate effectively with people of varied backgrounds [3]. The scientific research community is very diverse; therefore, a diverse group of educators is important. The Train-the-Trainer model is very effective in this pursuit [1]. By directly engaging and encouraging the underrepresented minority community in HPC at symposiums such as ADMI (The Association of Computer Science Departments at Minority Institutions), we can support their training and understanding of HPC systems and technologies. Those participants can then take their newly enhanced knowledge to their education and research institutions and train their fellow colleagues. The primary issue to overcome is finding high quality material that has been properly curated, which a federated and decentralized catalog of HPC training material can solve.

6.1.3   Organize Birds-of-a-Feather (BoFs) Meetings

We intend to hold Birds-of-a-Feather (BOFs) to share our findings and to gather more input from relevant communities. We will target key technical meetings such as PEARC23 [5], Supercomputing (SC23) [6], and ISC [4] where large community gatherings occur, allowing us to discuss and work toward solutions to the opportunities for improvement in finding and sharing materials.

6.2 Organize Community Hosted Training Material Services

Our survey showed that training materials across institutions are currently isolated and distributed, and the community recognizes the need for improving discovery and sharing of materials. Our goals include extending the reach of our training materials into underserved communities and identifying gaps in training. The lack of a central platform for sharing training and event services in HPC is a key factor in hindering discovery and advertising of training opportunities.

We plan to work with the HPC Training and Education Communities to identify best practices for sharing training resources. This includes using metadata tagging, adopting publishing mechanisms like GitHub or ReadTheDocs, open sharing of training materials, and collecting and disseminating educational material reviews and ratings. Training gaps can be filled through regular communication between contributors and the community, enhancing local portals by adding training materials shared by others. Shared material contributions would come from organizations that have a history of creating, developing, collecting, and displaying computational science education and training materials, as well as individual developers.

ACKNOWLEDGMENTS

REFERENCES

[1] Elizabeth Bautista and Nitin Sukhija, 2021. Employing directed internship and apprenticeship for fostering HPC training and education. *JOCSE, 12,* 2. *https://doi.org/10.22369/issn.2153-4136/12/2/8*

[2] Katherine Cahill, David Joiner, Scott Lathrop, Susan Mehringer, JP Navarro, and Aaron Weeden, *Final results: National survey on educational and training materials repositories.* Retrieved from https://www.cac.cornell.edu/about/pubs/Survey2022.pdf

[3] Patricia Gurin, Eric Dey, Sylvia Hurtado, and Gerald Gurin. 2002. Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review, 72*, 3. https://doi.org/10.17763/haer.72.3.01151786u134n051

[4] ISC High Performance. n.d. Retrieved from https://www.isc-hpc.com/

[5] PEARC. n.d. Retrieved from https://pearc.acm.org/pearc23/

[6] SC23. 2023. 2023. Retrieved from https://sc23.supercomputing.org/

APPENDIX A: SURVEY INSTRUMENT

The following is the survey as it appeared while it was active in 2022; this section shows both the introductory text and lists the questions in the order they appeared.

A working group has been formed within the ACM SIGHPC Education Chapter to discuss metadata standards for sharing materials across all interested organizations. We are asking the community to complete a short survey to understand the challenges and opportunities for the ACM SIGHPC Education Chapter to consider in promoting metadata standards. We will share what we learn from the survey at the Ninth SC Workshop on Best Practices for HPC Training and Education (BPHTE22) at SC22 and during the ACM SIGHPC Education Chapter working group.

We hope you will join us for the SC22 workshop presentations and discussions, in person or virtually, and we welcome you to join the ACM SIGHPC Education Chapter working group.

For the purposes of this survey, we use the term repositories to broadly include portals, collections, libraries, and lists of training and education materials and resources.

SURVEY PROCEDURES & CONFIDENTIALITY

If you agree to participate in the survey, you will not be required to provide any identifying information, and you will not be required to complete all questions. You will have the option of providing your name and contact information if future contact is desired. Efforts will be made to keep confidential any self-identifying information that you intentionally or inadvertently disclose. Your identity will be held in confidence in reports in which the survey results may be published and/or databases in which results may be stored. We may use aggregated data or anonymous comments from the survey in reports.

1. Which communities do you support or participate in? (check all that apply): *Multiple selections: (a) Pre-College, (b) Undergrad, (c) Grad/Post Doc, (d) Academia, (e) Government, (f) Industry, (g) Other*

2. When it comes to training and education materials, I consider myself to be (check all that apply): *Multiple selections: (a) Consumer of materials hosted by my organization, (b) Consumer of materials hosted by other organizations, (c) Content author, (d) Curator collecting appropriate materials for my community, (e) Other*

3. How easy is it for you to find appropriate training and education materials on specific topics, for yourself? *Likert scale: 1 (very difficult) - 5 (very easy)*

4. How easy is it for you to find appropriate training and education materials on specific topics, for group(s) you support? *Likert scale: 1 (very difficult) - 5 (very easy)*

5. Where do you look for material? *Multiple selections: (a) Portal or repository hosted by my organization, (b) Portals or repositories hosted by other organizations, (c) Search engine, (d) Other*

6. Which portals, repositories, search engines or other resources do you use or find helpful? *Free response text*

7. How important is, or would be, having easy access to repositories of training and education materials from multiple organizations to your community? *Likert scale: 1 (not important) - 5 (very important)*

8. What barriers have you encountered when searching for materials? *Multiple selections: (a) I can't find materials on the topic I need, (b) I can find materials on the topic, but not at the depth or level I need, (c) I find too many materials, and I can't effectively sort through them all, (d) I am aware of specific appropriate materials, but search engines don't list them in the top results, (e) Other*

9. Which barrier, if removed, would be most helpful for finding appropriate materials? *Single selection: (a) I can't find materials on the topic I need, (b) I can find materials on the topic, but not at the depth or level I need, (c) I find too many materials, and I can't effectively sort through them all, (d) I am aware of specific appropriate materials, but search engines don't list them in the top results, (e) Other*

10. Does your organization want to make it easier for the public to find your training and education materials? *Likert scale: 1 (not at all) - 5 (very much)*

11. What do you consider to be the biggest challenge(s) in sharing your materials? *Free response text*

12. Would your organization be willing and able to share your training and education materials in a public catalog by providing metadata about your materials in a standard format? *Likert scale: 1 (not at this time) - 5 (yes, even if it takes a few weeks)*

13. If your organization is not willing and able to provide metadata about your materials in a standard format, what are your roadblocks? *Multiple selections: (a) Lack of staff time, (b) Lack of funding, (c) Inadequate staff expertise, (d) Our materials aren't in a catalog, (e) Other*

14. Are you interested in joining a group working on this project? If so, please provide your contact information or write to hpc.edu.train@gmail.com. *Free response text*

15. Are you willing to be contacted by the survey organizers for follow-up regarding your responses? If so, please provide your contact information. *Free response text*

16. We would be happy to hear any additional comments you have on this topic. *Free response text*