

Understanding Community Perspectives on HPC Skills and Training Pathways

Weronika Filingier
EPCC, University of Edinburgh
w.filingier@epcc.ed.ac.uk

Jeremy Cohen
Imperial College London
jeremy.cohen@imperial.ac.uk

ABSTRACT

The “Understanding the Skills and Pathways Behind Research Software Training” BoF session run at ISC’23 provided an opportunity to bring together a group of attendees interested in approaches to enhance skills within the Research Software Engineering community. This included looking at options for understanding and developing pathways that practitioners can follow to develop their skills and competencies in a structured manner from beginner to advanced level. Questions discussed included: How can we highlight the existence of different training opportunities and ensure awareness and uptake? What materials already exist and what’s missing? How do we navigate this largely undefined landscape? In short: how does one train to become an RSE?

One of the interactive parts of this session was based around a live, anonymous survey. Participants were asked a number of questions ranging from their role in the training community to how easy they feel it is to find/access training content targeting different skill levels. They were also asked about challenges faced in accessing relevant content, combining it into a coherent pathway, and linking training content from different sources. Other questions focused on discoverability of material and skills that are most commonly overlooked. The number of respondents and responses varied between questions, with 24 to 50 participants engaging and providing 32 to 59 replies.

The goal of this lightning talk is to present findings, within the context of the community wide effort to make the training materials more FAIR - findable, accessible, interoperable and reusable.

KEYWORDS

Research Software Engineering, Training, Community, FAIR, Survey Results

1 INTRODUCTION

At ISC High Performance 2023, representatives from UNIVERSE-HPC [5], The Carpentries [1], CodeRefinery [2] and the HPC Certification Forum [4] organised a BoF session to discuss the skills and learning pathways behind Research Software Training. The goal was to trace current research software learning pathways and to understand the community perspective on difficulties with finding

appropriate training content and creating routes for gaining specialist RSE skills. Beside a series of short presentations and discussions, the session also included an interactive, live and anonymous survey run through Mentimeter. Here we focus on the survey results.

2 METHODOLOGY

The session was attended by approximately 70 people with up to 50 session participants responding interactively to the questions in the room (the questions could only be answered when displayed on the screen during the workshop). There was no pre-registration for the session so we have no details of the profile of audience members or which of the participants in the room chose to respond to which questions. The session attendees were asked 9 questions, and the number of respondents and responses varied between questions, with 24 to 50 participants engaging and providing 32 to 59 replies. Questions were of two types, those requiring only a numeric score and those with the option for open-ended text responses.

3 RESULTS

The first question asked the audience to define their role(s) in the context of training, selecting as many options as appropriate. 47 attendees answered, selecting 94 options, with 16 using training materials as a learning resource, 28 being a trainer, 8 training others, 27 developing training, and 15 attending the session out of general interest. Therefore, more than half the attendees were actively involved in training development and/or delivery. The next two questions (see Fig. 1) asked attendees about the ease of finding/accessing training materials online and within their organisations.

The responses – scored between 1 (easy) and 5 (difficult) – show that many people feel finding beginner level resources within their own institutions is harder than finding equivalent resources online (difficulty level 2.4 vs. 1.5). The difference in the average scores for the difficulty of finding intermediate level training resources is smaller, with a difficulty of 2.9 within their organisation and 2.6 online. Finally, the average scores for the advanced level training are the same at 3.6. Overall, the attendees feel it’s relatively easy to find or access beginner-level training online (1.5), while it is much harder to find or access similar content within their institutions - the score of 2.4 seems high. Intermediate training material is harder to find and access overall. There is no difference in the perception of how accessible advanced level training is - it’s on average relatively hard to find both online and within the attendees’ organisations.

The next question asked the attendees to score how difficult it is (1 - easy, 5 - difficult), in their opinion, to combine existing training materials into coherent pathways. The average score of 3.7 from the 49 respondents confirms that it is relatively hard to combine material from different sources. Another two questions (7 & 8) were focused on challenges associated with finding relevant

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Copyright ©JOCSE, a supported publication of the Shodor Education Foundation Inc.

© 2024 Journal of Computational Science Education
<https://doi.org/10.22369/issn.2153-4136/15/1/9>

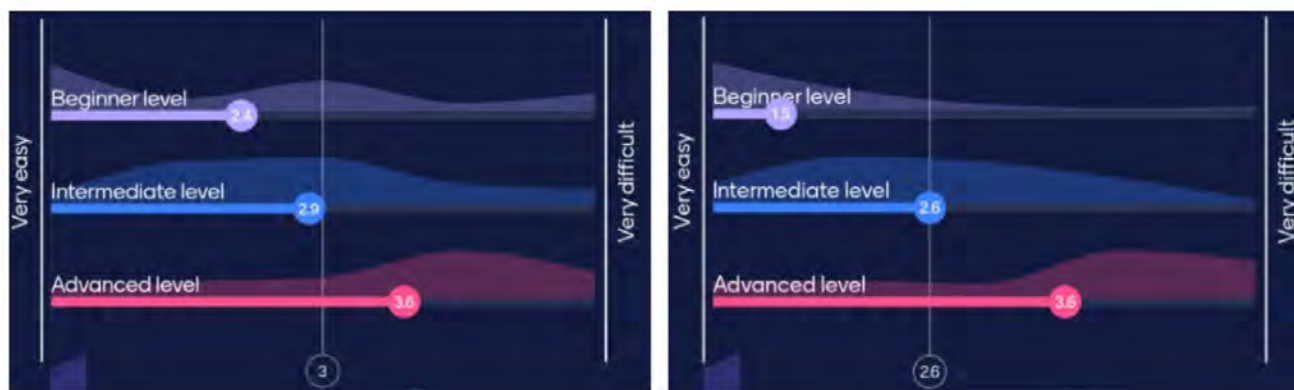


Figure 1: Audience's perception of how easy (1) or difficult (5) it is to find training content at beginner, intermediate and advanced levels within their organisations (left) and online (right). (Slides: Mentimeter.com)

training content and linking it into coherent pathways. 30 respondents provided 38 answers which can be divided into two main categories: accessibility, and appropriateness or relevance of available training materials. The accessibility challenges can be further divided – categories include: too many information sources, the financial cost of accessing content, access to hardware resources and outdated content. Examples from the relevance category include: trivial, too specific or irrelevant examples and difficulties with finding content at the right level and assessing its appropriateness. The main issues with linking content into pathways are centred around accessibility, incompatibility and teaching challenges. Accessibility issues include: broken links, disappearing pages, poor documentation, licensing and materials being either too specific or lacking useful examples. Some of the incompatibilities highlighted were: no consistent notation or terminology, competing standards (e.g. programming languages, research practices), incompatible or different packages/versioning. The teaching issues are related to overheads of producing good quality material not being recognised, time and expertise needed to verify online content, knowledge gaps and content overlaps.

Attendees were also asked about the most frequently overlooked research software skills, with 38 respondents providing 59 answers roughly fitting into categories that include: documentation, project management, communication skills, software engineering skills / best practices (e.g. testing, CI/CD, profiling, debugging), reusability and reproducibility, and personal skills like workload planning, critical thinking and independence. 30 respondents provided 40 answers on the question of how to make training content easier to discover. The main prevailing answer was providing a single point of access, in the form of curated repositories, a training portal or a centralised list, with coherent and consistent tagging. Other ideas included collaborations, organised user forums, raising awareness, dissemination in cooperation with educational institutions and advertising via social media. Additionally, the community should not only create incentives for improving skills, but ensure people can dedicate more time to learning/training.

Finally, attendees were asked about the skill(s) they would like to learn next. There were 35 answers from 20 respondents, mostly

fitting into 6 categories: languages/models, GPU skills, parallel computing, ML/AI, quantum computing and algorithms, and general software skills.

All responses collected during this live survey are available as a digital artifact on Zenodo [3]. The authors hope the collected data will be useful to the whole community.

4 CONCLUSIONS

Overall, there are less training materials covering intermediate and advanced topics and beginner level material, despite its abundance, is not always easy to use and re-use. Linking materials from different sources to create a coherent learning pathway tends to be very difficult. The breadth of skills required by our community is constantly growing, making it impossible for every institution to provide the relevant training. Ensuring that training materials are created in a way that makes them easy to share and re-purpose should take priority over constantly developing new content that may well already exist elsewhere. The responses discussed here provide additional evidence and enforce the need for community-wide effort to make training materials more FAIR - findable, accessible, interoperable and reusable. This material and its associated presentation represent another step in that direction.

ACKNOWLEDGMENTS

The authors thank Anja Gerbes, Maciej Cytowski and Matias Jääskeläinen for presenting at the "Understanding the Skills and Pathways Behind Research Software Training" BoF, and the audience members for their participation in the session. WF and JC acknowledge support via the UNIVERSE-HPC project (EPSRC grant EP/W035731/1).

REFERENCES

- [1] The Carpentries. 2023. <https://carpentries.org/>
- [2] CodeRefinery. 2023. <https://coderefinery.org/>
- [3] Weronika Filingier and Jeremy Cohen. 2023. Challenges in Research Software Training - survey results. <https://doi.org/10.5281/zenodo.8321376>
- [4] HPC Certification Forum. 2023. <https://www.hpc-certification.org/>
- [5] UNIVERSE-HPC. 2023. Understanding and Nurturing an Integrated Vision for Education in RSE and HPC. <https://www.universe-hpc.ac.uk/>